Identyfikacja nowych modyfikatorów niekanonicznej biosyntezy toksycznego białka poliglicynowego ze zmutowanego mRNA *FMR1*

Katarzyna Tutak

Rozprawa doktorska

Promotor: prof. dr hab. Krzysztof Sobczak

Promotor pomocniczy: dr inż. Anna Baud

Zakład Ekspresji Genów,

Instytut Biologii Molekularnej i Biotechnologii,

Uniwersytet im. Adama Mickiewicza w Poznaniu

Poznań, 2024

Identification of novel modifiers of noncanonical biosynthesis of toxic polyglycine protein from mutant *FMR1* mRNA

Katarzyna Tutak

PhD thesis

Supervisor: prof. dr hab. Krzysztof Sobczak

Assistant supervisor: Anna Baud, PhD

Department of Gene Expression,

Institute of Molecular Biology and Biotechnology,

Adam Mickiewicz University Poznań

Poznań, 2024

To all the people who supported me on this journey

Content

Articles included in the dissertation 6

Funding 7

Abbreviations 8

STRESZCZENIE 11

ABSTRACT 14

1. INTRODUCTION 16

1.1 Expanded microsatellites in human genome can cause several neurological diseases 16

1.2 Genetic basis and phenotype of fragile X-associated syndromes 17

1.3 Molecular pathomechanisms of CGG repeats expansion in FMR1 19

1.3.1 Secondary structures formed within expanded CGG repeats in mutant FMR1 mRNA 19

1.3.2 R loop formation 20

1.3.3 Sequestration of RNA binding proteins 21

1.3.4 Repeat-associated non-ATG translation 21

1.3.5 RAN translation of FMR1 with CGGexp 22

1.4 Mechanisms and proteins implicated in the regulation of CGG-related RAN translation 23

1.5 FMRpolyG driven pathology in PM-linked conditions 27

1.6 Ribosomal heterogeneity in a nutshell 30

2. AIMS OF THE STUDY 33

3. THE LITERATURE REVIEW: *Partners in crime: Proteins implicated in RNA repeat expansion diseases* 35

4. THE MANUSCRIPT/PREPRINT: *Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions* 36

5. UNPUBLISHED RESULTS: *Characterization of other factors affecting noncanonical polyglycine synthesis from mutant FMR1 mRNA* 37

5.1 Results 39

5.1.1 Alternative preparation of MS2-based pull down samples for MS analysis revealed different proteins bound to FMR1 RNA 39

5.1.2 Silencing of DDX21 and DHX15 downregulated the level of FMRpolyG in two cellular models 40

5.1.3 The effect of ALYREF silencing on FMRpolyG 43

5.1.4 Technical aspects of mass spectrometry sample preparation affected identification of proteins bound to FMR1 RNA 44
5.2 Discussion 48
5.3 Materials and methods 52
5.3.1 MS2-based pull down followed by acetone protein precipitation and SP3 purification 52
5.3.2 Mass spectrometry (MS) 53
5.3.3 MS data analysis 53
5.3.4 Cell culture and transfection 53
5.3.5 Quantitative real-time PCR (RT-qPCR) 54
5.3.6 Gene Ontology 54
5.3.7 Venn diagrams 54
5.4 Supplementary information 55
5.4.1 Tables 55
5.4.2 Supplementary figures 56
5.5 Bibliography 58
6. CONCLUDING REMARKS 71

Articles included in the dissertation

I. Baud A*, Derbis M*, **Tutak K***, Sobczak K. (2022), Partners in crime: proteins implicated in RNA repeat expansion diseases. WIREs RNA. doi: 10.1002/wrna.1709 *these authors contributed equally

II. **Tutak K**, Broniarek I, Zielezinski A, Niewiadomska D, Baud A, Sobczak K. (2024), Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions. Published as a preprint on bioRxiv server under the doi number: doi.org/10.1101/2024.03.27.586952

Funding

The work was supported by the following funding:

I. The Polish National Science Centre grant (2019/35/D/NZ2/02158) to Anna Baud

II. The Polish National Science Centre grant (UMO-2020/38/A/NZ3/00498) to Krzysztof Sobczak

III. Adam Mickiewicz University Foundation Scholarship for best doctoral students awarded to Katarzyna Tutak for academic year 2023/2024

Abbreviations

5'UTR 5' untranslated region
40S eukaryotic small ribosomal subunit 40S
43S PIC 43S preinitiation complex
60S eukaryotic large ribosomal subunit 60S
80S eukaryotic 80S ribosome
ACG near-cognate ACG start codon
ALYREF THO complex subunit 4
ANKZF1 Ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1
ASFMR1 Antisense transcript at the *FMR1* locus
ASFMRpolyA polyalanine-containing RAN protein translated from *ASFMR1*
ASFMRpolyP polyproline-containing RAN protein translated from *ASFMR1*
ASFMRpolyR polyarginine-containing RAN protein translated from *ASFMR1*
C9 ALS-FTD amyotrophic lateral sclerosis and/or frontotemporal dementia
C9orf27 Chromosome 9 Open Reading Frame 27
CAG CAG triplet repeat
CTG CTG triplet repeat
CGGexp expanded CGG repeats
CGIs CpG islands
Ct cycle threshold
Ctrl control
DBA Diamond-Blackfan Anemia
DDR DNA damage response
DDX3X ATP-dependent RNA helicase DDX3X
DGCR8 DiGeorge syndrome critical region 8
DHX9 ATP-dependent RNA helicase A
DHX15 ATP-dependent RNA helicase DHX15
DM1 myotonic dystrophy type 1
DMPK1 Dystrophia myotonica protein kinase 1
DOX doxycycline
DROSHA Drosha ribonuclease type 3
eIF Eukaryotic initiation factor
eIF2α Eukaryotic translation initiation factor 2
eIF4A Eukaryotic initiation factor 4A
eIF4E Eukaryotic initiation factor 4E
FM full mutation of *FMR1* gene characterized by more than 200 CGG repeats
FMR1 fragile X messenger ribonucleoprotein 1 gene/mRNA
FMRP Fragile X messenger ribonucleoprotein 1 protein
FMRpolyA polyalanine-containing RAN protein translated from mutated *FMR1*
FMRpolyG polyglycine-containing RAN protein translated from mutated *FMR1*
FMRpolyR polyarginine-containing RAN protein translated from mutated *FMR1*
FXAND fragile X-associated neuropsychiatric disorders
FXPAC fragile X-premutation-associated conditions
FXPOI fragile X-associated primary ovarian insufficiency
FXS fragile X syndrome
FXTAS fragile X-associated tremor/ataxia syndrome
G4C2 GGGGCC hexanucleotide repeat expansions in C9orf72

GFP Green fluorescent protein
GO gene ontology
GUG near-cognate GUG start
H3.3 Histone protein 3.3
HD Huntington's disease
hnRNP A2/B1 heterogeneous nuclear ribonucleoprotein A2/B1
Hox homeobox genes
HTT huntingtin gene/mRNA
ID uniprot unique protein identifier
IPSCs induced pluripotent stem cells
IRS integrated stress response
IRES internal ribosome entry site
L22e Large ribosomal subunit protein eL22
LAP2β Lamina-associated polypeptide 2 beta
LFQ label free quantification
LTN1 E3 ligase listerin
LUC7L3 Luc7-like protein 3
m7G 5' methyl-7-guanosine cap modification
MBNL1 Muscleblind-like protein 1
Met-tRNA initiator methionyl-tRNA
mRNA messenger RNA
MS mass spectrometry
MS2 bacteriophage ms2 coat protein
MYBBP1A Myb-binding protein 1A
NAF1 H/ACA ribonucleoprotein complex non-core subunit NAF1
NEMF Nuclear export mediated factor
NOP58 Nucleolar protein 58
NXF1 Nuclear RNA export factor 1
NXT1 NTF2-related export factor 1
ns non-significant
ORF open reading frame
PCBP2 Poly(rC)-binding protein 2
PERK PKR-like ER kinase
PM premutation of *FMR1* gene characterized by 55-200 CGG repeats
POI primary ovarian insufficiency
Pol II RNA polymerase II
PRC2 Polycomb repressive complex 2
Pur α Purine-rich binding protein α
RACK1 Small ribosomal subunit protein RACK1
RAN translation repeat-associated non-ATG translation
RBPs RNA binding proteins
REDs repeat expansion disorders
R-loop RNA:DNA hybrid
RPS6 Small ribosomal subunit protein eS6
RPS10 Small ribosomal subunit protein eS10
RPS15 Small ribosomal subunit protein eS15
RPS25 Small ribosomal subunit protein eS26

RPS26 Small ribosomal subunit protein eS26
RPL10A Large ribosomal subunit protein uL1
RPL38 Large ribosomal subunit protein eL38
RQC ribosome-associated quality control pathway
Sam68 Src-Associated substrate during mitosis of 68-kDa
SCA spinocerebellar ataxia
SD standard deviation
SDS sodium dodecyl sulfate
SILAC stable isotope labeling using amino acids in cell culture
siRNA small interfering RNA
SP3 single-pot solid-phase-enhanced sample preparation
SRSF1 Serine/arginine-rich splicing factor 1
SRSF3 Serine/arginine-rich splicing factor 3
SRPK 1 Serine/arginine protein kinase 1
STR short/simple tandem repeat
Suz12 Polycomb protein Suz12 (mouse)
TDP-43 TAR DNA binding protein of 43 kDa
TSR2 Pre-rRNA-processing protein TSR2 homolog
tRNA transfer RNA
uORF upstream open reading frame
UPS ubiquitin proteasome system
WT wild type
YTDC1 YTH domain-containing protein 1

STRESZCZENIE

Zespoły łamliwego chromosomu X obejmują choroby genetyczne wywołane mutacją dynamiczną w genie *FMR1* (ang. *Fragile X messenger ribonucleoprotein 1*), który koduje białko FMRP (ang. *Fragile X messenger ribonucleoprotein 1 protein*). U zdrowego człowieka, w sekwencji rejonu 5' niepodlegającemu translacji (5'UTR) genu *FMR1* znajduje się zwykle 25-35 powtórzeń trójki nukleotydowej CGG. Jednakże, długość tej sekwencji jest wysoce polimorficzna, a powtórzenia CGG mają tendencję do wydłużania, często znacznego. Proces ten nazywany jest ekspansją powtórzeń CGG. Gdy długość sekwencji powtórzeń mieści się między 55 a 200, stan ten nazywany jest premutacją genu *FMR1* i dotyczy tzw. stanów chorobowych związanych z premutacją w chromosomie X (ang. *fragile X-premutation-associated conditions,* FXPAC). Te stany obejmują m.in. neurodegeneracyjną chorobę wieku późnego zwaną zespołem drżenia i ataksji związanym z łamliwym chromosomem X (FXTAS) oraz zespół przedwczesnego wygasania funkcji jajników (FXPOI), prowadzący do przedwczesnej menopauzy. Stan, w którym liczba powtórzeń CGG przekracza 200 określa się mianem pełnej mutacji i stanowi on podłoże genetyczne zespołu łamliwego chromosomu X (FXS). FXS jest chorobą neurorozwojową, będąca najczęstszą przyczyną wrodzonego upośledzenia umysłowego, zwłaszcza u chłopców.

Z molekularnego punktu widzenia, w chorobach z grupy FXPAC, dochodzi do prawidłowej ekspresji białka FMRP, pomimo mutacji w genie *FMR1*, podczas gdy w FXS rejon promotorowy, zawierający ekspansję CGG, ulega hipermetylacji, prowadząc do wyciszenia genu i braku ekspresji FMRP. Uważa się, że rozwój chorób z grupy FXPAC jest efektem trzech niezależnych, wzajemnie przenikających się patomechanizmów molekularnych. Pierwszy z nich związany jest z toksycznością RNA zawierającego wydłużony ciąg powtórzeń CGG. Taki toksyczny RNA, w obrębie powtórzeń CGG tworzy strukturę drugorzędową typu spinka do włosów (ang. *hairpin*), na której mogą być sekwestrowane białka, co prowadzi do upośledzenia ich prawidłowych funkcji w komórce. Drugi mechanizm dotyczy kotranskrypcyjnego powstawania hybrydowych struktur typu pętle R (hybryda DNA:RNA), które prowadzą do akumulacji pęknięć w DNA, stwarzając zagrożenie dla utrzymania integralności genomu komórki. Trzecim mechanizmem jest niekanoniczna biosynteza białka

nazwana zależną od powtórzeń translacją niewymagającą kodonu inicjatorowego ATG (ang. ***r****epeat-****a****ssociated ****n****on-ATG translation;* RAN), w skrócie translacją RAN. Proces ten prowadzi do powstawania toksycznych białek zawierających ciągi monoaminokwasowe kodowane przez sekwencję powtórzeń CGG, które mają tendencję do agregacji. Białka te składają się z powtórzonego ciągu aminokwasów jednego rodzaju, np. glicyny (kodon GGC), tworząc tzw. białka poliglicynowe (FMRpolyG). Białko FMRpolyG gromadzi się w postaci złogów wewnątrzjądrowych w komórkach pacjentów, stanowiąc istotny element patogenezy FXPAC, poprzez zwiększenie śmiertelności komórek.

Ponieważ dokładny mechanizm translacji RAN nie jest jeszcze w pełni zrozumiały, celem niniejszej pracy doktorskiej było poszukiwanie nowych modyfikatorów tego procesu. W ramach projektu zastosowano system znakowania cząsteczek RNA zawierających 99 powtórzeń CGG w 5'UTR *FMR1* oraz identyfikację białek, które związały się z badanym transkryptem w komórkach, za pomocą spektrometrii mas. W wyniku przeprowadzonych badań zidentyfikowano ponad 60 białek, które w warunkach natywnych wiążą się do rejonu 5'UTR RNA *FMR1* zawierającego powtórzenia CGG. Analiza ontologii genów wykazała, że większość zidentyfikowanych białek należy do klasy białek wiążących RNA oraz takich, które uczestniczą w procesach związanych z biogenezą rybosomu, translacją oraz procesowaniem cząsteczki mRNA. Niektóre z zidentyfikowanych białek znajdują potwierdzenie w literaturze na temat interaktomu mRNA *FMR1*, a zastosowana w tej pracy technologia pozwoliła na identyfikację szeregu nowych białek podlegających interakcji z badanym RNA.

Spośród zidentyfikowanych białek wybrano dziesięciu kandydatów i przetestowano ich zdolność do regulacji procesu translacji RAN, wykorzystując technikę wyciszania genów z zastosowaniem krótkich interferujących RNA. Wyciszenie kilku z tych białek obniżyło poziom toksycznego białka FMRpolyG. Na przykład, wyciszenie białka rybosomalnego eS26 małej podjednostki 40S rybosomu (ang. *small ribosomal subunit protein eS26;* RPS26), które ze względu na lokalizację w pobliżu kanału mRNA kontaktuje się podczas translacji lub skaningu z tymi RNA, spowodowało zmniejszenie ilości białka poliglicynowego w kilku niezależnych modelach

komórkowych. Dodatkowo, niedobór dwóch helikaz RNA, DHX15 (ang. *ATP-dependent RNA helicase DHX15*) oraz DDX21 (ang*. Nucleolar RNA helicase 2*), a także czynnika transportującego ALYREF (ang. *THO complex subunit 4*) negatywnie wpłynął na biosyntezę białka FMRpolyG. Co istotne, wyciszenie tych białek nie miało wpływu na poziom białka FMRP, wskazując zdolności do specyficznej regulacji otwartej ramki odczytu białka FMRpolyG. Ponadto, wyciszenie RPS26 doprowadziło do ograniczenia tworzenia się złogów białkowych, powodując częściowe zniesienie toksyczności białek poliglicynowych w modelu komórkowym. Dodatkowo, ilościowa analiza proteomu komórek linii HEK293 po wyciszeniu RPS26 ujawniła, że tylko niewielka liczba białek jest wrażliwa na niedobór RPS26. Analiza transkryptów kodujących te białka wykazała wzbogacenie nukleotydów guaninowych i cytozynowych w rejonie 5' UTR, co sugeruje podobieństwo biochemicznie tych transkryptów do mRNA *FMR1*.

W celu lepszego zrozumienia mechanizmu translacji RAN, zweryfikowano także funkcję czynnika TSR2 – białka opiekuńczego RPS26 (ang. *Pre-rRNA-processing protein TSR2 homolog*). Wykazano, że TSR2 pozytywnie wpływa na proces biosyntezy FMRpolyG. Ponadto, wykazano, że inne białko małej podjednostki rybosomu, RPS25 (ang. *Small ribosomal subunit protein eS25*) również reguluje proces translacji RAN białka FMRpolyG**.**

Podsumowując, przeprowadzone badanie przesiewowe w oparciu o analizę proteomiczną, pozwoliło zidentyfikować pulę białek oddziałujących ze zmutowanym *FMR1* zawierającym ciąg powtórzeń CGG, co stanowi cenne źródło wiedzy na temat biologii tej cząsteczki RNA**.** Głównym osiągnięciem pracy doktorskiej jest identyfikacja pięciu nowych modulatorów translacji RAN oraz propozycja koncepcji, zgodnie z którą skład małej podjednostki rybosomu odgrywa istotną rolę w regulacji niekanonicznej syntezy białka poliglicynowego – czynnika patogenetycznego w zespołach chorobowych związanych z premutacją genu *FMR1.*

ABSTRACT

Fragile X-premutation-associated conditions (FXPAC) are genetic diseases caused by dynamic mutations of the fragile X messenger ribonucleoprotein 1 gene (*FMR1*) located on the X chromosome encoding fragile X messenger ribonucleoprotein 1 protein (FMRP). The gene usually contains 25–35 CGG repeats in the 5'-untranslated region (5'UTR). However, these triplet repeats are highly polymorphic and tend to expand. The premutation (PM) state of CGG expansion (CGGexp) corresponds to 55–200 repeats and is associated with multiple FXPAC such as a late onset neurodegenerative disease called fragile X-associated tremor/ataxia syndrome (FXTAS) and fragile X-associated primary ovarian insufficiency (FXPOI). On the contrary, when CGGexp exceeds 200, it is called full mutation and underlies the pathogenesis of neurodevelopmental disease named fragile-X syndrome (FXS), the most common form of inherited intellectual disability.

In case of FXPAC, FMRP protein is produced despite the presence of CGGexp in 5'UTR of *FMR1*, whereas, in FXS, methylation of *FMR1* promoter leads to the gene silencing. It is postulated that the interplay between three major molecular pathomechanisms drives FXPAC. At first, RNA with CGGexp is toxic and forms a secondary hairpin structure that sequesters RNA binding proteins into RNA foci leading to their functional depletion. Second, co-transcriptional formation of R loops, a DNA:RNA hybrids which triggers DNA breakage and compromises genomic stability. Third mechanism involves repeat-associated non-ATG initiated (RAN) translation, which leads to the production of toxic and aggregation-prone proteins called RAN proteins, which contain long tract of a repeated monoamino acid, that can be either polyglycine, polyalanine or polyarginine depending on the open reading frame, however polyglycine-containing proteins (FMRpolyG) predominate. Toxic FMRpolyG aggregates and forms intranuclear inclusions in patient's brain, a hallmark of FXTAS pathology.

Mechanistic insights into RAN translation remain elusive, therefore we sought to identify novel RAN translation modifiers, which constitutes the main part of the PhD thesis. We applied an in cellulo RNA tagging system combined with mass spectrometry (MS) based protein identification and discovered more than 60 proteins that bind to

5’UTR of mutant *FMR1* mRNA containing CGGexp. Gene ontology analysis performed on identified proteins revealed that majority of them represent RNA binding properties, are involved in ribosome biogenesis, translation or mRNA processing. Some of identified proteins overlapped with already identified interactors of *FMR1*, however our dataset contains newly identified factors.

Among identified proteins, we selected ten candidates and we verified their RAN translation regulatory properties using small interfering RNA. As a result, we identified few proteins, which depletion affected the level of FMRpolyG. For instance, depletion of small ribosomal subunit protein eS26 (RPS26), a component of 40S subunit which contacts mRNA sequence during translation, significantly impeded RAN translation in multiple tested models. In addition, insufficiency of two RNA helicases, ATP-dependent RNA helicase DHX15 (DHX15) and Nucleolar RNA helicase 2 (DDX21) as well as THO complex subunit 4 (ALYREF) negatively affected biosynthesis of FMRpolyG. Additionally, silencing of RPS26, DDX21 and ALYREF did not affect the level of FMRP indicating specificity of regulation towards FMRpolyG frame. Importantly, we showed that depletion of RPS26 decreases the amount of aggregates formed by FMRpolyG and alleviated their toxicity in cellular model. In addition, using quantitative MS approach, we found that the number of proteins produced by RPS26-sensitive translation is limited and that the 5’UTRs of the mRNAs encoding these proteins are rich in guanosine and cytosine, similar to *FMR1* mRNA.

In order to gain mechanistic insights into RPS26-sensitive translation regulation, we verified the function of RPS26 chaperone, Pre-rRNA-processing protein TSR2 homolog (TSR2). We demonstrated that TSR2 positively regulated the production of FMRpolyG. Finally, we verified that another component of the 40S subunit, Small ribosomal subunit protein eS25 (RPS25) also regulates CGG-related RAN translation.

To sum up, performed MS-based screening provided a unique and valuable source of information about *FMR1* interacting proteins in cellulo, which may be implicated in the biology of this molecule. The main achievement of this study is the identification of five novel RAN translation modifiers and the proposal of a concept suggesting that the composition of the 40S subunit plays a pivotal role in regulating noncanonical CGG-related RAN translation in FXPAC.

1. INTRODUCTION

1.1 Expanded microsatellites in human genome can cause several neurological diseases

Short tandem repeats (STRs) also known as microsatellites are repeated tracts of three to seven base pairs such as CTG, CGG of GGGGCC (1). More than 1.5 million STRs have been identified in coding and non-coding regions and constitute about 3% of human genome (2, 3). When microsatellites are expanded to a critical, pathological number, they contribute to the development of almost 50, neurodevelopmental, neurodegenerative and neuromuscular diseases named repeats-expansion related disorders (REDs) (1, 4, 5). For example, expanded CTG repeats in 3' untranslated region (UTR) of *dystrophia myotonica protein kinase* gene (*DMPK*) underlies the myotonic dystrophy type 1 (DM1) (6). Pathological number of CAG repeats present in the exon 1 of *huntingtin* (*HTT*) gene contribute to the development of Huntington's disease (HD) (7). Expansion of GGGGCC (G4C2) repeats in intronic region of *c9orf72* gene underlies amyotrophic lateral sclerosis and/or frontotemporal dementia (C9 ALS/FTD) (8). Other diseases such as several spinocerebellar ataxias (SCA) and fragile X syndrome (FXS) also belong to REDs (5).

The STRs are unstable, prone to expand and to a lesser extend contract. It is thought that microsatellites cause replication stalling and are hotspots of chromosomal double strand breaks (4, 9). Repeats instability occurs predominantly in germline tissue, often leading to inheritance of longer repeats tract in progeny (10). Such phenomenon is known as genetic anticipation and inherited trait manifests in more severe disease phenotype beginning earlier in the offspring's life (11). The repeats are also unstable in somatic cells causing tissue-specific somatic mosaicism (10, 12).

Repeats expansion above a certain threshold can lead to the gene loss or gain of function phenomenon (13). The gene loss of function manifests in gene silencing. For instance, loss of fragile X messenger ribonucleoprotein 1 protein (FMRP) drives pathology in FXS (14). The gene gain of function correlates with acquiring new deleterious features by an RNA or a protein triggering disease development (13). For example, mutated Huntingtin in HD serves as an example of toxic protein gain of

function and in case of RNA gain of function, the great example is toxic RNA with expanded CUG repeats in DM1, which sequesters several cellular proteins leading to their functional depletion (15).

1.2 Genetic basis and phenotype of fragile X-associated syndromes

Fragile X messenger ribonucleoprotein 1 gene (*FMR1*) in 5'UTR contains CGG repeats that are prone to breakage and were linked to so-called fragile site on the X chromosome in the late 1960s (16). The CGG STR in this locus is conserved in mammals, and the average size of CGG repeats in humans varies mostly between 25 and 30 (17) (Figure 1.1). In 1991 the CGG expansion (CGGexp) in the *FMR1* was independently identified by four research groups and described as the causative mutation of neurodevelopmental disease known as FXS (18–21). When CGGexp exceeds 200, it induces hypermethylation of the neighboring *FMR1* promoter, leading to the transcriptional silencing and consequently lack of the primary gene product, FMRP (14). FMRP is an RNA binding protein with known implications in synaptic plasticity (22) and its insufficiency is believed to cause autism and other forms of cognitive and intellectual disability in FXS patients (23).

The premutation (PM) of *FMR1* gene corresponds to 55-200 CGG repeats expansion in 5'UTR of *FMR1* and underlies fragile X-premutation-associated conditions (FXPAC) (24). FXPAC is an umbrella term, which covers three distinct medical conditions associated with PM: fragile X-associated tremor/ ataxia syndrome (FXTAS) (25), fragile-X-associated primary ovarian insufficiency (FXPOI) (26) and fragile X-associated neuropsychiatric disorders (FXAND) (27). The estimated prevalence of PM is 1 in 150–300 women and 1 in 400–850 men. However, due to incomplete penetrance of the mutation, approximately 1 in 5,000–10,000 men will develop some form of FXPAC in their fifties or later, while random X inactivation in woman PM carrier, greatly reduces the risk of disease manifestation (28–30). On molecular level, in opposite to FXS, FMRP expression is not affected or slightly decreased in FXPAC, however the *FMR1* mRNA is elevated (31–33).

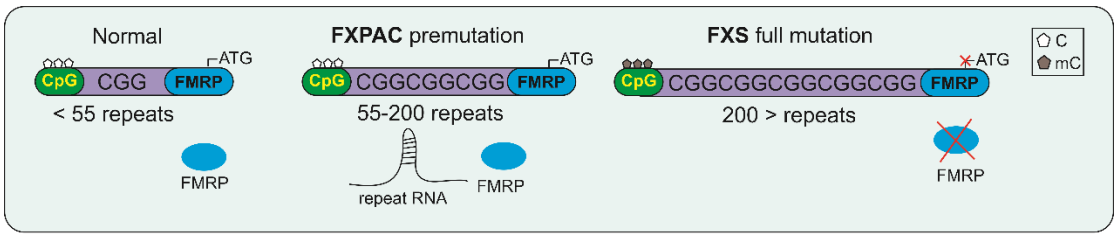


Figure 1.1. CGG repeats instability in *FMR1* locus and its molecular implications. Schematic representation of CGG repeats expansion in 5'UTR of *FMR1* and its molecular consequences in FXS and FXPAC; CpG, refers to CpG island in *FMR1* promoter, mC indicates methylation of cytosine.

Fragile X-associated tremor/ ataxia syndrome (FXTAS)

In early 2000s, FXTAS was described as a late onset, incurable neurodegenerative disease that affects mostly men (30, 34, 35). The main clinical features of FXTAS include intention tremor, cerebellar ataxia, neuropathy, balance problems, parkinsonian features, dementia as well as cognitive decline, memory problems and executive function deficits (28, 34). The pathological hallmark of the disease are ubiquitin-positive intranuclear or perinuclear inclusions in nerve cells (neurons and glia) (36–39) and neurodegeneration resulting in mild brain atrophy and white matter lesions (35, 40). FXTAS progresses with age and so far no effective cause-oriented treatment is available (28).

Fragile-X-associated primary ovarian insufficiency (FXPOI)

Primary ovarian insufficiency (POI) is described as premature cessation of ovary function leading to early menopause before the age of 40 (41). It is estimated that female PM carriers are 20% more likely to develop POI in their lifetime and suffer from infertility problems (42). Similarly to FXTAS, FMRpolyG and ubiquitin positive inclusions were found in ovaries of FXPOI patients as well as in mice model which suggests one of the disease pathomechanism (43, 44). In addition, dysfunctional hypothalamic–pituitary–gonadal-axis may contribute to FXPOI development (43). Similarly to FXTAS, only symptomatic treatment is available for FXPOI patients.

Fragile X-associated neuropsychiatric disorders (FXAND)

The neuropsychiatric disorders such as anxiety, depression, social defects, chronic fatigue, sleep disturbances but also obsessive compulsive behavior are the most common problems of psychological and/or psychiatric nature reported by almost 50% of PM patients. PM carriers who suffer from those symptoms often do not meet the criteria for FXTAS and/or FXPOI, thereby demand different treatment. Such high occurrence of neuropsychiatric symptoms led to a proposal of separate, novel term - FXAND in order to enhance recognition and promote effective research in that area (27).

1.3 Molecular pathomechanisms of CGG repeats expansion in FMR1

1.3.1 Secondary structures formed within expanded CGG repeats in mutant FMR1 mRNA

High guanosine (G) and cytosine (C) content of *FMR1* RNA determinates the structural features of this molecule. Complementary binding between C and G bases is recognized as the strongest pairing between nucleotides and leads to formation of thermodynamically stable secondary structures, either hairpins and/or G-quadruplexes (structures formed by the stacking of several planar guanine quadruplets) within expanded CGG repeats in *FMR1* mRNA (45–47).

Structured, mutant GC-rich *FMR1* RNA can cause several, deleterious molecular consequences for the cell (Figure 1.2). First, high GC content triggers R loop formation, a DNA:RNA hybrid formed during transcription (48). Second, sequestration of RNA binding proteins on *FMR1* RNA containing CGGexp disrupts their functionality (49, 50). Third, mutant *FMR1* serves as template for noncanonical FMRpolyG protein synthesis *via* mechanism named repeat-associated non-ATG (RAN) translation (40). It is considered that the interplay between three main molecular pathomechanisms related to toxic RNA and protein gain of function phenomenon is involved in FXPAC development and progression (Figure 1.2) (51, 52).

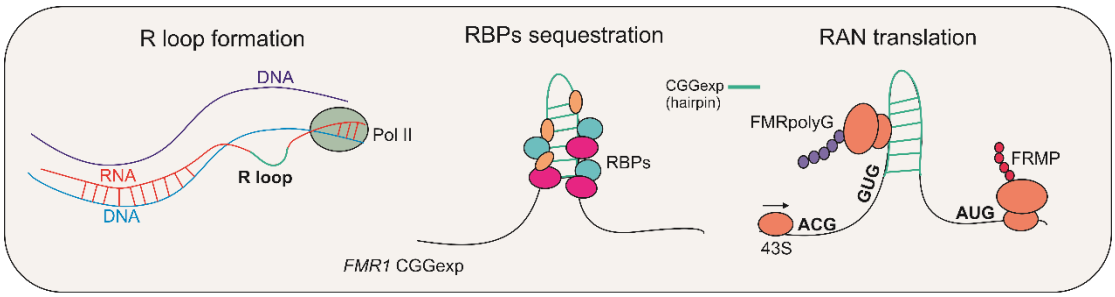


Figure 1.2. Three major molecular pathomechanisms contributing to FXPAC. Synergy between R-loop formation, RNA binding proteins sequestration and FMRpolyG biosynthesis drive FXPAC pathogenesis; Pol II – polymerase II, RBPs – RNA biding proteins, AUG – canonical start codon, ACG and GUG – noncanonical, near-cognate RAN translation start codons, 43S – preinitiation complex.

1.3.2 ***R loop formation***

R loops are three stranded DNA:RNA hybrids formed behind elongating Polymerase II (Pol II), when the binding between nascent RNA and the template DNA (RNA:DNA) is more thermodynamically advantageous than binding between template single-stranded DNA (DNA:DNA) (53). Such situation occurs frequently on GC rich DNA template (54) and high GC content of *FMR1* gene promotes the Polymerase II (Pol II) co-transcriptional formation of R loops (48, 55). They can be formed in CpG islands (CGIs) of *FMR1* promoter region and/or within expanded CGG repeats (48, 53, 56). One hypothesis is that R loops formed in CGIs-containing promoters can relax chromatin and promote transcription, while R loops formed over CGGexp cause Pol II stalling and result in decreased transcription efficiency (53, 57, 58). With the increase of CGG repeats number, the R loop formation occurs more frequently causing DNA breakage strongly correlated with transcriptional-replication machineries collision. Accumulation of these events can be lethal to a cell by activating DNA damage response (DDR) and threatening genomic stability (48, 55, 58, 59).

Off note, in addition to impeded transcription and activated DDR, stalled Pol II can recruit Polycomb repressive complex 2 (PRC2). Methylation of histone H3 at lysine 27 by PRC2 promotes heterochromatinization of *FMR1* locus giving possible explanation for stable gene silencing in FXS (53, 60, 61).

Altogether, that activated DDR caused by R loops in *FMR1* locus can lead to cellular death constituting one of the pathomechanism underlying PM-linked disorders (48, 55).

1.3.3 Sequestration of RNA binding proteins

Toxic RNA with expanded repeats organizes within a cell in nuclear clusters known as RNA foci, which are stabilized by RNA-RNA, RNA-protein and protein-protein interactions (62, 63). Within foci, RNA containing expanded repeats sequester cellular RNA binding proteins (RBPs) disabling their functions and dysregulating many cellular processes (49, 50, 64). Depending on repeat length and sequence, multivalent base-pairing and occupancy of sequestered RBPs, RNA foci formation can promote phase transition leading to RNA gelation observed in many REDs (65–67).

Multiple RBPs were found to be sequestered on toxic *FMR1* RNA with CGGexp. These include Microprocessor complex subunit DiGeorge Syndrome Critical Region 8 (DGCR8), Ribonuclease 3 (DROSHA) (50), Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1) (68) and Transcriptional activator protein Pur-alpha (Pur-α) (69), Src-associated in mitosis 68 KDa protein (SAM68) (49) and TAR DNA binding protein of 43 kDa (TDP-43) (70). Microprocessor complex disruption results in the deregulation of microRNA maturation and SAM68 sequestration causes aberrant splicing in FXTAS models, respectively (49, 50). Importantly, overexpression of DGCR8, Pur-α and hnRNP A2/B1 in *Drosophila melanogaster* FXTAS model reduced neurodegeneration providing a proof of concept for RNA gain of function mediated toxicity (50, 68, 69).

1.3.4 Repeat-associated non-ATG translation

In 2011, Zu and colleagues described for the first time the repeat-associated non-ATG (RAN) translation, a noncanonical protein synthesis triggered by CAG repeats expansion (linked to SCA8) and CTG repeats expansion (underlying DM1). This process was shown to be initiated in multiple reading frames without canonical AUG start codon (71). Later, RAN translation induced by other repeats: G4C2 repeats in *C9orf72* gene, CAG repeats in *HTT* and CGG repeats in *FMR1* underlying ALS/FTD, HD and FXTAS,FXPOI, respectively, was reported (38, 72–74). RAN translation derived proteins (for simplicity named here RAN proteins) usually contain homopolymeric or dipeptide sequence comprised of one or two repeated amino acids. For instance, CGGexp in

FMR1 repeats encode mostly FMRpolyG, protein with long polyglycine stretch (38), and G4C2 repeats encode dipeptides such as polyglycine and polyalanine polyGA (72). RAN proteins are toxic, prone to aggregate and their accumulation in patient's tissues is believed to contribute to the development and progression of all mentioned REDs (38, 71–75).

1.3.5 RAN translation of FMR1 with CGGexp

The open reading frame (ORF) for FMRP resides downstream to CGG hairpin of *FMR1* mRNA (Figure 1.3). FMRP translation initiates at canonical AUG codon, whereas upstream open reading frames (uORFs) for RAN proteins begin at less-favored, near-cognate start codons such as ACG or GUG (75, 76) (Figure 3). RAN translation of *FMR1* mRNA with CGGexp results in noncanonical biosynthesis of polyglycine (FMRpolyG), polyalanine (FMRpolyA) and polyarginine (FMRpolyR) containing proteins depending on the uORF. FMRpolyG predominates, whereas FMRpolyA and FMRpolyR are less abundant in cell (77, 78). RAN translation for FMRpolyG and FMRpolyR begins upstream to CGG repeats starting from ACG and GUG codons, while FMRpolyA synthesis is initiated within CGG repeats. In addition to homopolymeric tract of repeated single amino acid, RAN proteins contain amino acids encoded upstream to CGG tract as well as peptides encoded in first and partially second exon of FMRP (38, 76). Moreover, formation of hybrids of these three RAN proteins is possible as a product of frameshifting on CGGexp sequence (79). Additionally, RAN translation also occurs on antisense *ASFMR1* transcript containing CCG repeats resulting in a synthesis of polyproline (ASFMRPolyP), polyarginine (ASFMRpolyR), and polyalanine (ASFMRpolyA) proteins (80). In fact, PM of *FMR1* gene results in production of many RAN proteins, while FMRpolyG is the most abundant and the most studied one, yet biological relevance of other RAN products is not clarified and remains questionable in the field. Regardless of uORFs of RAN proteins, majority of studies indicate that FMRP translation is independent from RAN translation event. However, Rodriquez and colleagues (2020) showed that antisense oligonucleotides treatment targeting RAN start sites impeded RAN translation and enhanced FMRP synthesis in human neurons demonstrating the native regulatory function of uORFs over FMRP frame (81).

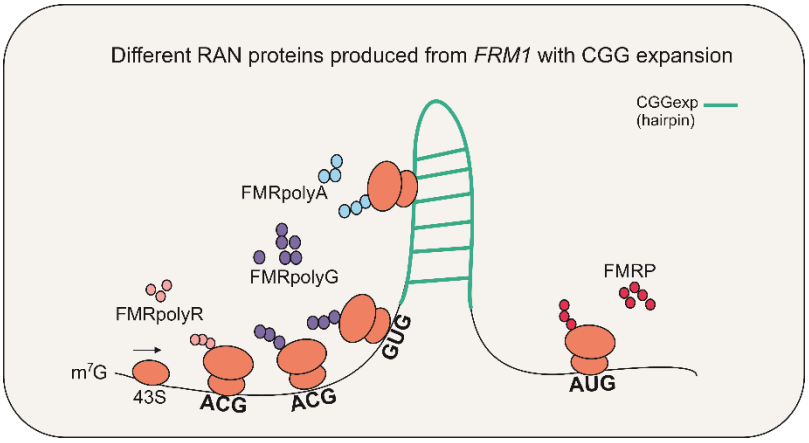


Figure 1.3. Schematic representation of canonical FMRP and noncanonical (RAN) translation of RAN proteins encoded in toxic *FMR1* RNA with CGGexp. m^7G indicates 5'methyl-7-guanosine cap, 43S – preinitiation complex. AUG – canonical start codon, ACG and GUG – noncanonical, near-cognate RAN translation start codons.

1.4 Mechanisms and proteins implicated in the regulation of CGG-related RAN translation

Despite the important role RAN translation in multiple neurodegenerative diseases, our understanding of mechanism of this process remains incomplete. In fact, although RAN translation is 30-40% less efficient than traditional protein biosynthesis, it adapts canonical mechanism of eucaryotic initiator factors (eIF4A, eIF4E and others) recruitment to 5' methyl-7-guanosine cap (5' m^7G) followed by subsequent 43S Preinitiation complex (43S PIC) scanning (76). On the other hand, in different genetic and cellular contexts RAN translation can occur in cap-independent manner utilizing tRNA initiators of translation other than methionyl-tRNA (78, 82–84). However, it is puzzling why PIC chooses near-cognate codons for translation initiation. One possible explanation refers to steric hindrance of structured CGGexp which abolishes PIC scanning and causes its stalling resulting in skewed start codon fidelity and forced translation initiation in less-favored non-AUG codons (Figure 1.4) (76). This hypothesis is supported by the increase of RAN translation rate correlated with the increase of repeats length, and the more frequent initiation at ACG codon which resides upstream to hairpin than the initiation at GUG located within hairpin (76). In addition, ribosomal stalling on structured RNA can cause translational frameshifting which results in the production of aggregation-prone chimeric proteins such as hybrid of polyarginine-polyglycine (79)

CGGexp secondary structure can be resolved by RNA helicases which positively affects RAN translation (Figure 1.4). Linsalata and colleagues (2019) showed that DEAD-box helicase DDX3X which possibly unwinds hairpin, facilitates RAN translation, as the helicase depletion resulted in the significant decline of RAN products (Figure 1.4) (85). Such results might be explained by a mechanism where DDX3X resolves hairpin paving the way for scanning 43S PIC, hence without DDX3X, near-cognate codons are hardly accessible, especially GUG codon within hairpin. In addition, Tseng and coworkers (2021) provided the evidence that another helicase, DEAH-box RNA helicase DHX36 plays a role in CGGexp G-quadruplexes and/or hairpin relaxation supporting the idea that physical hindrance formed by repeated sequences is pivotal for RAN translation initiation and efficiency (86). Altogether, one can say that RAN translation initiation is governed by fine tuning of noncanonical codons selection by 43S PIC. More complexity adds the fact that RAN translation occurs also on shorter, physiological number of repeats questioning the hypothesis that the pathological expansion of the repeats drives RAN translation (76, 81). This opens up a possibility that RAN translation may play a physiological regulatory role *in vivo* (81).

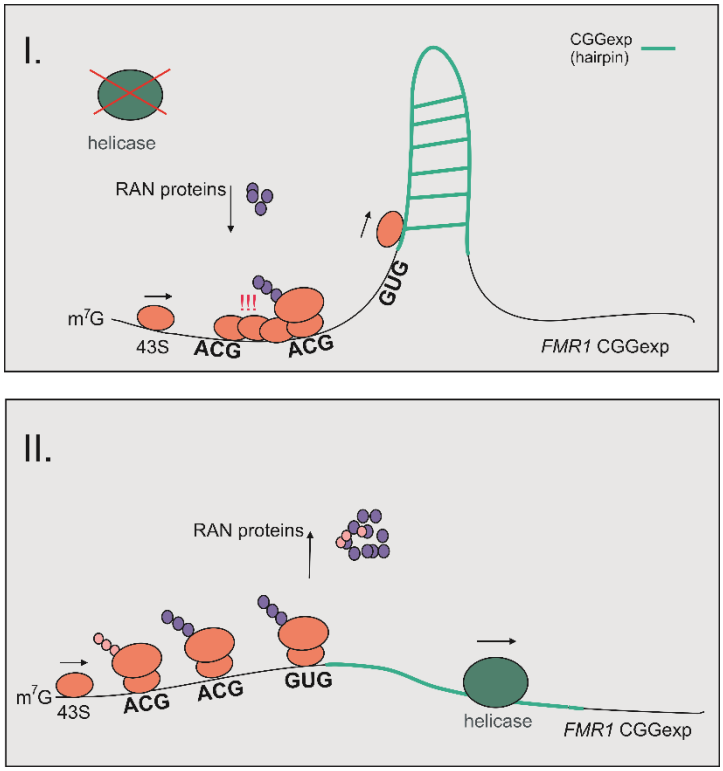


Figure 1.4. Schematic representation of two mechanisms implicated in RAN translation. I. 43S stalling while scanning (marked as !!!) caused by steric hindrance such as hairpin, induces the selection of near-cognate codons. **II.** Unwinding the RNA secondary structures by RNA helicases facilities faster scanning and more frequent selection of near-cognate codons which enhances RAN translation. m^7G indicates 5'methyl-7-guanosine cap, 43S – preinitiation complex. AUG – canonical start codon, ACG and GUG – noncanonical, near-cognate RAN translation start codons.

Expanded repeats and their secondary structures cause aberrations in translation initiation, yet the same structures may impede and abort translational elongation. This hypothesis was very recently explored pointing out that components of ribosome-associated quality control (RQC) pathway play important role in the regulation of RAN translation elongation (87). RQC pathway protects the cell from misfolded or incompletely generated polypeptides by ubiquitination and subsequent proteasomal degradation, ribosome recycling, degradation of template RNA and other processes (88). Briefly, RQC pathways consists of two sequential steps, first sensing of stalled ribosome and disassembly of 80S ribosomal subunits and second formation of RQC complex *via* initial recruitment of nuclear export mediated factor (NEMF) to 60S subunit which binds and stabilizes binding by the E3 ligase listerin (LTN1). At this stage, NEMF synthesizes carboxy-terminal alanine and threonine tails to expose lysine

residues that are inside the ribosomal exit tunnel, to be ubiquitylated by LTN1. Next, ubiquitinated, nascent peptides are released by ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1 (ANKZF1) and condemned to proteasomal degradation (89). Tseng and colleagues demonstrated that depletion of RQC complex components, NEMF, LTN1 and ANKZF1 enhanced RAN protein synthesis (87). This indicates that elongating ribosome stalls on expanded structured repeats, activates RQC pathways, which in return leads to the degradation of nascent polypeptide chain and halting RAN translation. Hence, activated RQC prevents accumulation of RAN misfolded proteins by ubiquitination and subsequent proteasomal degradation (87).

RAN protein are toxic for the cell and trigger elevated integrated stress response (ISR) (83, 90). Green and colleagues (2017) demonstrated that the presence of misfolded RAN proteins leads to the activation of endoplasmic reticulum-resident kinase (PERK), and subsequent phosphorylation of the α subunit of eukaryotic initiation factor-2 (eIF2α) at serine 51, which is the core event of an ISR (90). Phosphorylated eIF2α evokes cascade of events resulting in the inhibition of global, canonical protein synthesis but enhances RAN translation by alteration in start codon stringency and initiation kinetics. At the end, elevated ISR stimulates the formation of phosphorylated-eIF2α-dependent stress granules causing neuronal death. Hence, authors proposed a deleterious feedforward loop, where RAN proteins activate ISR, leading to global translation shut down and selective enhancement of RAN translation (90).

In addition to more direct regulatory process such as start codon fidelity, RAN translation can be modulated in an indirect way, be retention of transcript with CGGexp in the nucleus. Malik and colleagues (2021) identified direct interaction between serine/arginine splicing factor 1 (SRSF1) and *FMR1* mRNA (91). They also showed that impeded activity of serine/arginine protein kinase 1 (SRPK1), which normally phosphorylates SRSF1, results in reduction of SRSF1 nuclear import, which in return halts *FMR1* transcripts in the nucleus. It was also demonstrated that pharmacological inhibition of SRPK1 prevented stress-induced enhancement of RAN translation. Hence, SRPK1 modulates translocation of toxic RNA preventing it from becoming a template for RAN translation in cytoplasm (91).

In conclusion, several regulatory pathways contribute to RAN translation regulation starting with transport, unwinding of structured RNA, start codon fidelity, RQC and stress response, however continuation of mechanistic studies underpinning noncanonical protein synthesis may broaden our understanding of this process and thus open up novel therapeutic perspectives.

Please note that other proteins implicated in RAN translation regulation not presented here are reviewed in "*Partners in crime: Proteins implicated in RNA repeat expansion diseases*", Baud A et al., 2022, that constitutes the first part of PhD thesis (please find chapters entitled *RAN Translation, Immune response to RNA^{exp} and products of RAN translation, Targeting RBPs as potential therapeutic strategies*).

1.5 ***FMRpolyG driven pathology in PM-linked conditions***

In FXTAS patients, FMRpolyG and ubiquitin-positive intranuclear inclusions are predominantly present throughout the brain in neurons, astroglia and Purkinje cells of cerebellum, however they can be also found in other tissues (Figure 1.5) (36–38, 75, 92–94). Except FMRpolyG and to a lesser extend FMRpolyA (38, 75, 77), the exact composition of intranuclear inclusions is not fully determined and varies in different studies, however, a pool of proteins interacting with FMRpolyG and/or present in inclusions was identified by few research groups (39, 67, 75, 93). These include Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1), Musceblind like 1 (MBNL1), Pur-α, Lamin A, Nucleolin, several heat shock proteins and other mitochondrial, cytoskeleton, proteasome and exosomal proteins. The exact sequence of inclusion formation is unresolved, however, one hypothesis is that toxic RNA acts as a scaffold for multistep RNA-RBPs complex assembly (50, 68, 95). For instance, it was shown that hnRNP A2/B1 directly interacts with CGGexp and that the CUGBP1 protein interacts with toxic RNA *via* hnRNP A2/B1 (68). Similarly, initial sequestration of DGCR8 and DORSHA to CGGexp is required for subsequent SAM68 recruitment (50). In addition, Asamitsu and colleagues (2021) have shown the direct interaction between pathogenic CGG repeat-derived RNA G-quadruplexes (CGG-G4RNA) and FMRpolyG (67) *via* the polyglycine region of FMRpolyG, promoting liquid-to-solid transition and aggregation of FMRpolyG. This suggested that FMRpolyG forms a

complex with RBPs and exosomal proteins *via* toxic RNAs (67). On the contrary, Hoem and colleagues (2019) showed that FMRpolyG itself, without co-expression of the CGGexp RNA hairpin can form aggregates and induce toxicity in several cell lines (96).

RAN translation occurs in cytoplasm, yet FMRpolyG-positive aggregated are found in nucleus. How these aggregates are transported from the cytoplasm to the nucleus is unclear. One could assume that aggregates are mobile and travel along microtubules, however this hypothesis was denied as no colocalization between aggregates and microtubule organizing center was observed (96). In addition, soluble, diffused fraction of FMRpolyG in cytoplasm and nucleus stays mobile until it attaches to aggregate, and becomes immobile, indicating possible snow-ball-like dynamic of aggregate formation (96). Altogether, the complex nature of FMRpolyG aggregates require further investigation.

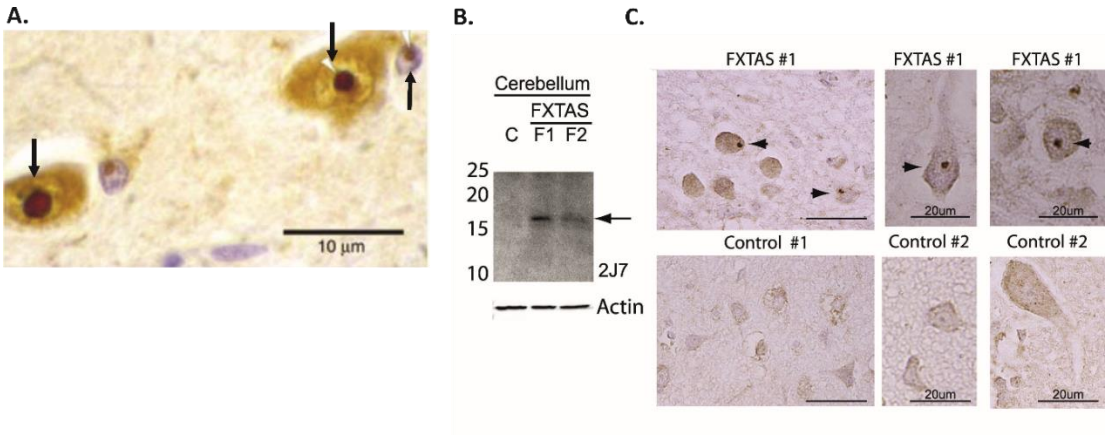


Figure 1.5. FMRpolyG detected in FXTAS patient's tissues. A. Intranuclear inclusions stained with anti-ubiquitin antibodies detected in FXTAS patient's neurons (indicated by arrows) **B.** Western blot with 2J7 anti-FMRpolyG antibody of cerebellar lysates from two FXTAS patients and an age-matched control. Arrow indicate FMRpolyG of expected size. **C.** Images of 2J7 immunostaining from frontal cortex and hippocampus of control and FXTAS brain. Arrows indicate FMRpolyG-positive inclusions. Adapted from Greco C et al., 2002 (A) and Todd P et al., 2013 (B & C).

Regardless of well documented presence of intranuclear inclusions in FXTAS cells, the mechanism explaining their contribution to the neurodegeneration remains uncovered. One possible explanation relates to disruption of the normal architecture of lamin A/C within the nucleus of cell expressing toxic RNA (97), which was further

supported by the discovery of interaction between FMRpolyG and Lamina-associated polypeptide 2 beta (LAP2β), which disorganizes nuclear lamina architecture in neurons derived from FXTAS patient's induced pluripotent stem cells (IPSCs) (75). Overexpression of LAP2β rescues neurodegeneration induced by expression of FMRpolyG indicating that destabilized lamina architecture contributes to FXTAS pathogenesis (75). In addition to lamina disruption, it was demonstrated that FMRpolyG can be sorted to exosomes, which are further propagated from cell-to-cell, negatively affecting synaptic plasticity and thus eliciting neuronal dysfunction (67). Other hypothesis regarding FMRpolyG toxicity refers to protein quality control pathways. Ubiquitin and proteasomal proteins were detected in intranuclear inclusions (93) and FMRpolyG is a stable protein that is primarily degraded *via* the ubiquitin proteasome system (UPS) (96). However, it has been shown that RAN translation and accumulation of RAN proteins leads to impairment of the UPS (98). Hence, these results suggest a mechanism in which RAN translation of FMRpolyG leads to the inability of the UPS to remove RAN proteins from the cell (96, 98).

In FXPOI, FMRpolyG-positive inclusions were found in ovary tissues and granulosa cells derived from PM women carriers and their toxicity was directly confirmed in FXPOI mice models showing significant fertility impairment in the presence of FMRpolyG (43, 44, 99). In fact, Shelly and colleagues (2021) showed that sole expression of toxic RNA was sufficient enough to impair responses to hormonal stimulation which lead to ovarian dysfunction, but expression of both RNA with CGGexp and FMRpolyG lead to premature cessation of breeding (44).

FMRpolyG-positive inclusions for many years constituted the hallmark of the disease, although, recently it was shown that the presence of RAN proteins-positive aggregates was not sufficient to induce clear behavioral disease phenotype in FXTAS mice model (100). In addition, the native FMRpolyG is hardly detectable in human tissue using MS and its quantification was possible after extensive enrichment from the analyzed patient's material (39, 78). This could be explained by the limitations of FMRpolyG detection as it was demonstrated that for successful identification a minimum 60 to 70 CGG repeats are required (75) and many patients with PM display less than 70 repeats (29). Technical limitations also play a role as RAN proteins are

small and proteolytic digestion results in very few unique peptides (39, 78). On the contrary, Sellier and colleagues showed that in FXTAS mice the sole expression of CGGexp RNA was not pathogenic, whereas FMRpolyG was required to induce locomotor deficiency (75). These findings opened a debate, if and to what extent RAN translation contributes to FXTAS pathology and what it more toxic, RNA containing CGGexp or deleterious RAN protein, or is it an interplay between those two entities that drive pathogenesis? (101) Given that FMRpolyG was initially detected by immunostaining (36, 38, 75) and other methods seem to fail or underdeliver RAN proteins identification, it is necessary to develop a reliable methodology to detect and quantify FMRpolyG to answer this question in future (39, 78).

1.6 Ribosomal heterogeneity in a nutshell

Central dogma of gene expression underlies our understanding how encoded genetic information is deciphered, leading to the biosynthesis of a final gene product. For decades, it was believed that transcribed RNAs are translated by well-defined ribosomal machinery, however, with the development of new technologies, the last two decades brought up new perspectives to this subject. It was shown that composition of the ribosome may vary and this structural rearrangements result in the translation of different pools of mRNA shifting our understanding about gene expression towards ribosome-mediated control of gene regulation. Such phenomenon named ribosomal heterogeneity or specialized ribosomes was initially described in invertebrate model organisms, although, today we know that this process is conserved among different organisms and occurs also in mammalian cells (102, 103). It is important to highlight that majority of ribosomal proteins are necessary for proper ribosomal functionality but some ribosomal proteins may be disassembled from the ribosome without harming translational efficiency, yet determining what RNAs can or cannot be translated (104–106). For instance, ribosomes containing small ribosomal subunit protein eS25 (RPS25) or Large ribosomal subunit protein uL1 (RPL10A/uL1) preferentially translate transcripts encoding proteins implicated in development, cell cycle or vitamin B12 metabolism (104).

Ribosomal heterogeneity has the significant consequences for body patterning and organism development overall (107, 108). The great example is Ribosomal Protein 38L (RPL38), which selectively facilitates translation of *homeobox* (*Hox*) mRNAs – crucial for formation of body plan (109, 110). Kondrashov and colleagues (2011) demonstrated that mutations in RPL38 led to abnormal body patterning in mice (109). Of similar importance is a role of RPL10A in the production of mesoderm lineage (111). Additionally, it was demonstrated that several ribosomal proteins are differentially expressed throughout the organism, for instance, ribosomal protein L22e family members in *Drosophila* (112). These RPLs can be alternatively spliced and differentially expressed in the organism. While one paralog is ubiquitous, the other has tissue-specific expression (112). Moreover, mutations of ribosomal proteins encoding genes result in the development of tissue-specific, sever congenital diseases called ribosomopathies (113). Well known example is Diamond-Blackfan Anemia (DBA), which is linked to mutations in gene encoding Small ribosomal subunit protein eS10 (RPS10) and Small ribosomal subunit protein eS26 (RPS26) (114, 115). It was shown that depletion of RPS26 from the cell abolishes erythropoiesis which may drive DBA (116). Ribosomal composition may also affect translation of RAN proteins from expanded STRs. For example, Yamada and colleagues (2019) demonstrated the role of RPS25 in RAN translation regulation. They showed that RPS25-depleted ribosomes less efficiently translate RAN peptides derived from G4C2 and CAG repeats alleviating RAN-translation related toxicity in several C9 ALS/FTD models (117). Importantly authors showed that RPS25 is non-essential for ribosome functionality as its depletion from translational machinery did not affect global translation efficiency (117).

In addition to the role of ribosomal heterogeneity in the development and disease, it was demonstrated that ribosomal composition changes upon environmental stimuli. For instance, in yeast, high-salt and high-pH stress induces the release of Rps26 from mature ribosomes by its chaperone pre-rRNA-processing protein Tsr2, enabling the translation of mRNAs engaged in stress response pathways (105, 106, 118). In eukaryotic cells, RPS26 is also involved in stress responses, although on the contrary to the yeast model, RPS26 remains associated to the ribosome under energy stress (119). It was demonstrated that cells with mutated C-terminus of RPS26

were more resistant to glucose starvation, than the wild type cells due to modulation of energy metabolism mediated by changes in the translatome induced by RPS26-dependent mechanism (119). Remarkably, C-terminal domain of RPS26 was shown to specifically interact with nucleotides upstream to AUG codon near E-site of actively translating ribosome (118, 119), most likely contributing to the selection of transcripts undergoing effective translation, however the exact motif facilitating the interaction between RPS26 and RNA is not determined.

To sum up, biological relevance of specialized ribosomes is robust and extensive studies are actively conducted to better understand this phenomenon and its implications.

2. AIMS OF THE STUDY

Fragile X premutation-associated conditions (FXPAC) are linked to the pathogenic expansion of 55-200 CGG repeats in 5'UTR of *FMR1* gene. One of the pathomechanisms underlying FXPAC is noncanonical protein synthesis named repeat-associated non-ATG (RAN) translation of CGG repeats which leads to the production of toxic RNA proteins containing a tract of single, repeated amino acid, for instance, the polyglycine-containing protein named FMRpolyG. FMRpolyG is prone to aggregate and form intranuclear inclusions in the patient's tissues, which is believed to be one of the factors driving disease progression. Although RAN translation was discovered more than a decade ago, the exact mechanism and factors involved in CGG-related RAN translation regulation are yet to be uncovered, hence further studies concerning noncanonical protein synthesis in FXPAC are required. Therefore, **the main aim of the study was to identify novel modifiers of CGG-related RAN translation.**

The specific goals of the study were to:

1. Identify proteins interacting with the 5' UTR of *FMR1* containing expanded CGG repeats in cellulo, using mass spectrometry-based screening.
2. Validate the potential regulatory properties of selected candidates in modulating RAN translation by using targeted siRNAs for gene silencing.
3. Investigate the mechanism underlying RAN translation regulation facilitated by the factors that have been verified in the previous step.

The PhD thesis is structured into **three parts**. **The first part** comprises a comprehensive literature review that synthesizes current knowledge on RNA binding proteins implicated in disease-associated repeat expansion. This review was authored by Baud A, Derbis M, Tutak K & Sobczak K and was published in *Wiley Interdisciplinary Reviews: RNA* in 2022 (Chapter 3). The second and third parts of the dissertation present the results pertaining to the primary objective of the study. **Part two** is the manuscript titled "*Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*" authored by Tutak K et al., published as a preprint on bioRxiv server (Chapter 4). **Part three** titled "*Characterization of other factors affecting noncanonical polyglycine synthesis from*

mutant FMR1 mRNA" constitutes unpublished results which detail the identification and characterization of additional proteins identified through MS-based screening (Chapter 5). Furthermore, it discusses the verification of their regulatory properties in regulating RAN translation.

3. THE LITERATURE REVIEW: *Partners in crime: Proteins implicated in RNA repeat expansion diseases*

Baud A*, Derbis M*, **Tutak K***, Sobczak K. (2022), Partners in crime: proteins implicated in RNA repeat expansion diseases. WIREs RNA. doi: 10.1002/wrna.1709 *these authors contributed equally

Open access:

https://wires.onlinelibrary.wiley.com/doi/epdf/10.1002/wrna.1709

Katarzyna Tutak,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią mojej rozprawy doktorskiej, polegał na uczestniczeniu w konceptualizacji pracy, przygotowaniu rozdziałów zatytułowanych *RAN translation*, *Targeting RBPs as potential therapeutic strategies,* opracowaniu tabeli, wprowadzeniu elementów, o które postulowali recenzenci oraz edycji i korekcie finalnej wersji tekstu.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy

Katarzyna Tutak

Podpis współautora

Podpis promotora

dr. Anna Baud,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią rozprawy doktorskiej Katarzyny Tutak, polegał na uczestniczeniu w konceptualizacji pracy, przygotowaniu rozdziałów zatytułowanych *Introduction*, *Transciription and splicing of RNA^{exp}*, *Sequestration of proteins on RNA^{exp} molecules*, *Immune Response to RNA^{exp} and Products of RAN translation*, wprowadzeniu elementów, o które postulowali recenzenci, edycji i korekcie finalnej wersji tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy

Anna Baud

Podpis współautora

dr. Magdalena Derbis,
Instytut Chemii Bioorganicznej Polskiej Akademii Nauk

14.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią rozprawy doktorskiej Katarzyny Tutak, polegał na uczestniczeniu w konceptualizacji pracy, przygotowaniu rozdziałów zatytułowanych *Toxic RNA^{exp} Molecules, Nucleocytoplasmic transport, Phase Separation*, przygotowaniu rycin, wprowadzeniu elementów, o które postulowali recenzenci oraz edycji i korekcie finalnej wersji tekstu.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy

Magdalena Derbis

Podpis współautora

Prof. dr hab. Krzysztof Sobczak,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią rozprawy doktorskiej Katarzyny Tutak, polegał na uczestniczeniu w konceptualizacji pracy, edycji i korekcie finalnej wersji tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy

Podpis współautora

4. THE MANUSCRIPT/PREPRINT: *Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*

Tutak K, Broniarek I, Zielezinski A, Niewiadomska D, Baud A, Sobczak K. (2024), Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions. Published as a preprint on bioRxiv server under the doi number: doi.org/10.1101/2024.03.27.586952

Open access:

https://www.biorxiv.org/content/10.1101/2024.03.27.586952v2

Katarzyna Tutak,
Zakład Ekspresji Genów
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią mojej rozprawy doktorskiej, mój udział polegał na uczestniczeniu w konceptualizacji pracy, wygenerowaniu modelu L-99xCGG, zaprojektowaniu i przeprowadzeniu większości eksperymentów (z wyjątkiem tych, które są wyszczególnione w oświadczeniach innych współautorów), wykonaniu analiz statystycznych, przygotowaniu rycin oraz suplementów, interpretacji wyników, a także napisaniu oraz edycji tekstu.

*Wymienieni autorzy są autorami korespondencyjnymi

Katarzyna Tutak

Podpis współautora

K. Sobczak

Podpis promotora

dr. Izabela Broniarek,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24,Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na wygenerowaniu i charakteryzacji modeli badawczych S-95xCGG i S-16xCGG.

*Wymienieni autorzy są autorami korespondencyjnymi

Izabela Broniarek

Podpis współautora

dr hab. Andrzej Zieleziński,
Zakład Biologii Obliczeniowej
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24,Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na przeprowadzeniu analizy bioinformatycznej transkryptów kodujących białka zidentyfikowane w eksperymencie SILAC-MS.

*Wymienieni autorzy są autorami korespondencyjnymi

Andrzej Zieleziński

Podpis współautora

dr. Daria Niewiadomska,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24,Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na przygotowaniu konstruktu genetycznego z mutacją −4G<A.

*Wymienieni autorzy są autorami korespondencyjnymi

Daria Niewiadomska

Podpis współautora

dr. Anna Baud,
Zakład Ekspresji Genów
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24,Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na uczestniczeniu w konceptualizacji badań, nadzorowaniu prac, przeprowadzeniu analizy in vitro, wykonaniu testu na agregację białek poliglicynowych po wyciszeniu białka RPS26, pomocy w przygotowaniu próbek do analizy za pomocą spektrometrii mas (MS), analizie surowych danych MS, napisaniu i korekcie tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy są autorami korespondencyjnymi

Anna Baud

.....

Podpis współautora

Prof. dr hab. Krzysztof Sobczak,
Zakład Ekspresji Genów
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24,Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na uczestniczeniu w konceptualizacji badań, nadzorowaniu prac, napisaniu i edycji tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy są autorami korespondencyjnymi

Podpis współautora

5. UNPUBLISHED RESULTS: *Characterization of other factors affecting noncanonical polyglycine synthesis from mutant FMR1 mRNA*

Given that sample preparation for MS analysis affects detectability of peptides leading to differences in the identification and the quantification of proteins (120), we repeated the biological experiment of MS2-based pull down and applied different MS sample preparation protocols including acetone precipitation of the proteins followed by single-pot solid-phase-enhanced sample preparation (SP3) procedure (described in Methodology section in details). Briefly, SP3 procedure includes immobilization of either proteins or peptides on magnetics beads followed by rinsing steps with organic solvent such as acetonitrile (121). Such intervention improves removal of common contaminants such as sodium dodecyl sulfate (SDS), which presence may negatively affect MS analysis. Results described in preprint (Tutak et al.), were obtained using another methodology, where, proteins captured on investigated bait RNA (5'UTR of *FMR1* with expanded CGG repeats) were digested on the streptavidin beads without the elution step. In another preparative approach, proteins were eluted from the beads, acetone precipitated, purified using SP3 approach and trypsin-digested prior to MS analysis (Figure 5.1). Results related to precipitation and SP3 purification variant of MS sample preparation are described below.

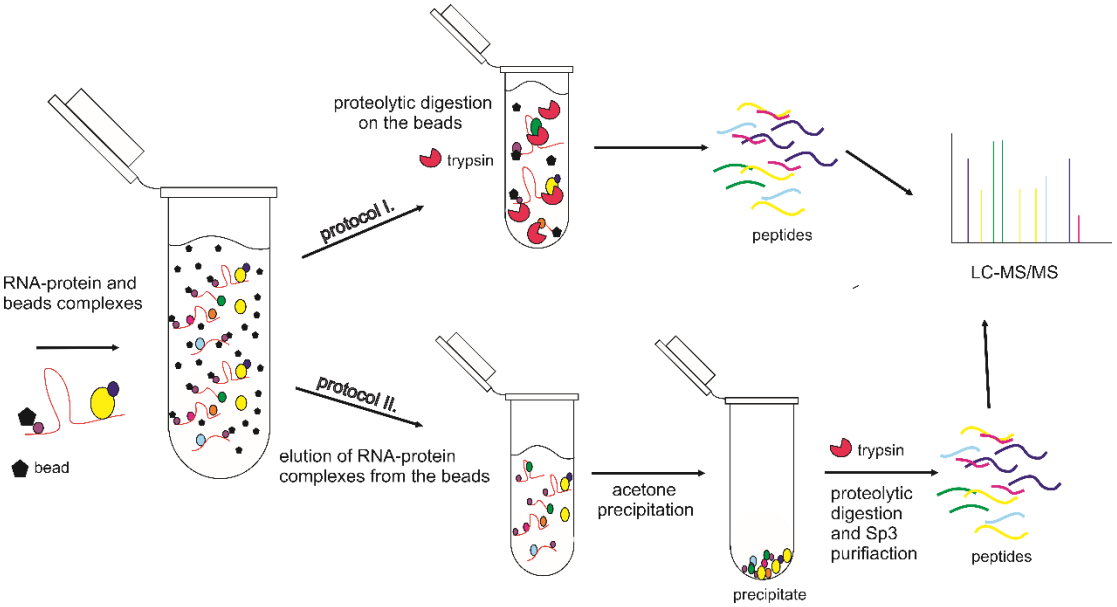
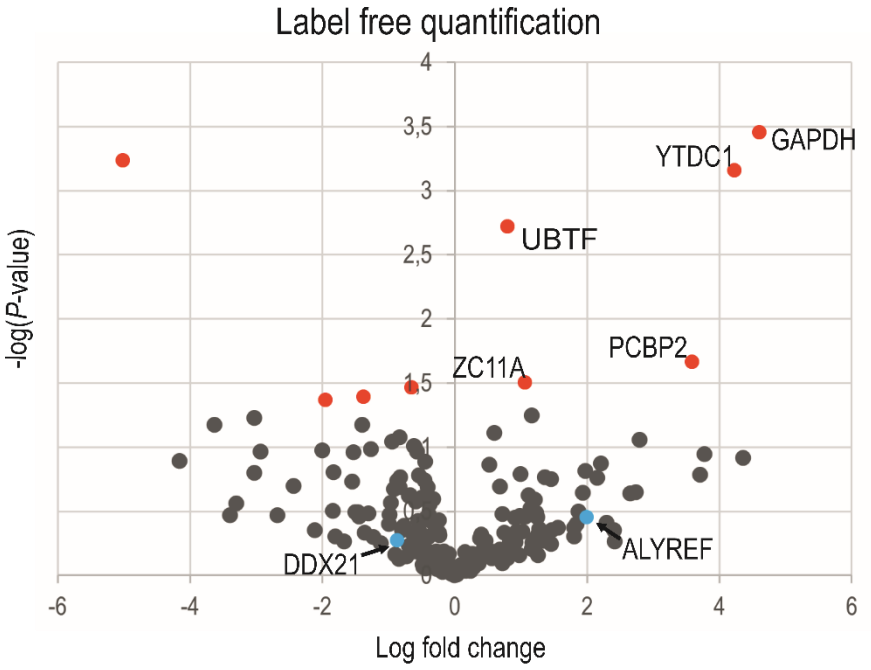


Figure 5.1. Schematic representation of two preparative approaches applied for MS sample preparation. Protocol I includes direct, on the beads digestion, protocol II includes elution of RNA-protein complexes from the beads, acetone precipitation followed by proteolytic digestion and SP3 purification. Digested peptides were analyzed by mass spectrometry.

5.1 Results

5.1.1 Alternative preparation of MS2-based pull down samples for MS analysis revealed different proteins bound to FMR1 RNA

After the RNA-protein pull down followed by MS, the bioinformatic analysis of MS RAW data was performed in an identical way as the data set described in Tutak et al., using MaxQuant software and protein identification based on Uniprot database. This analysis revealed 155 proteins common between *FMR1 RNA* replicates (proteins identified in all samples are listed in the Table 5.1). Further, we performed label free quantification (LFQ) to identify proteins enriched in *FMR1 RNA* in comparison to *GC-rich RNA.* LFQ analysis revealed five proteins significantly enriched in *FMR1 RNA* (Figure 5.2, Table 5.2).



FD	-log(*P*-value)	ID	Protein name
4.60	3.46	P04406-2	Glyceraldehyde-3-phosphate dehydrogenase GAPDH
4.22	3.16	Q96MU7-2	YTH domain-containing protein 1 YTDC1
0.80	2.72	P17480-2	Nucleolar transcription factor 1 UBTF
3.58	1.66	Q15366-7	Poly(rC)-binding protein 2 PCBP2
1.06	1.51	O75152	Zinc finger CCCH domain-containing protein 11A ZC11A

Figure 5.2. Label free quantification of proteins enriched in *FMR1 RNA* samples. The volcano plot representing proteins captured during MS-based screening showing the magnitude of enrichment (Log2 fold change) and the statistical significance (−log *P*-value); red dots on the right site indicate proteins significantly enriched ($P < 0.05$) on *FMR1 RNA* compared to *GC-rich RNA.* Blue dots represent DDX21 & ALYREF analyzed in the context of potential RAN translation

regulatory properties. Proteins significantly enriched in *FMR1 RNA* in comparison to *GC-rich RNA* are listed below the graph; FD - log2 fold change, ID - unique protein identifier from Uniprot database.

Given that proteins identified as not significantly enriched in *FMR1 RNA* samples may be also relevant to *FMR1* RNA biology, we decided to verify RAN translation regulatory properties of Nucleolar RNA helicase 2 (DDX21) and Aly/REF Export Factor (ALYREF). We speculated that DDX21, similar to many other identified helicases may play a role in the unwinding of CGGexp hairpin structure, or thermodynamically stable RNA structure in general, and therefore modulating RAN translation efficiency (76, 85, 86). ALYREF is a component of the TRanscription EXport (TREX) protein complex that is thought to couple mRNA transcription, processing and nuclear export (122, 123). This protein represents high affinity towards 5' and 3' end of RNAs in vivo (124) and facilitate nuclear export *via* NXF1 - NTF2-related export factor 1 (NXT1) pathway (122). Components of NXF1-NXT1 pathway were described to represent RAN translation modulatory features (91, 125). Additionally, ALYREF was already identified as protein binding to mutant *FMR1* in similar screening performed by other research group (126). Moreover, both DDX21 and ALYREF were reported to be involved in guarding the genomic stability either by prevention (127) or unwinding the R loops (128) – structures involved in FXPAC development (48). Altogether, it seemed reasonable to verify whether DDX21 and ALYREF modulate CGG-related RAN translation.

5.1.2 Silencing of DDX21 and DHX15 downregulated the level of FMRpolyG in two cellular models

In order to verify RAN translation regulatory properties of DDX21, we evaluated the level of FMRpolyG derived from transient transfection system upon silencing DDX21 (verified by RT-qPCR) in COS7 cell (Figure 5.3A). DDX21 depletion downregulated FMRpolyG by approximately 70%, however the level of exogenous *FMR1-GFP* mRNA was also reduced (Figure 5.3A). Despite the fact, that results were not statistically significant (perhaps due to high standard deviation (SD) in control samples), we pursued investigation of the DDX21 role in cell lines stably expressing FMRpolyG derived from expanded (95) and normal (16) CGG repeats (*S-95xCGG* & *S-16xCGG*

models, described in Tutak et al.). In addition, we included into the experiment ATP-dependent RNA helicase DHX15, which was initially identified in MS-based screening approach described in Tutak et al. The prior identification of DHX15 and verification of its RAN translation regulatory properties, encouraged us to test whether DHX15 depletion induces RAN translation suppression also in cell lines stably expressing FMRpolyG. At first, we confirmed the efficient DDX21 and DHX15 silencing by RT-qPCR (Figure 5.3A&B, lower panel) and observed that knockdown of these helicases induced lowering of FMRpolyG level regardless of CGG repeats content, as RAN translation was impeded in both *S-95xCGG* and *S-16xCGG* models (Figure 5.3B). Moreover, we checked the level of FMRP after depletion of both helicases. In *S-95xCGG* model, DHX15 depletion induced slight increase of FMRP (Figure 5.3B). On the contrary, in *S-16xCGG* model, we did not observe such effect (Figure 5.3B). Silencing DDX21 in both models, did not impact the level of FMRP (Figure 5.3B). Further, we evaluated whether observed reduction of FMRpolyG was due to change in its mRNA level. RT-qPCR analysis revealed no changes in transgenic *FMR1-GFP* mRNA level in both cell lines (Figure 5.3C).

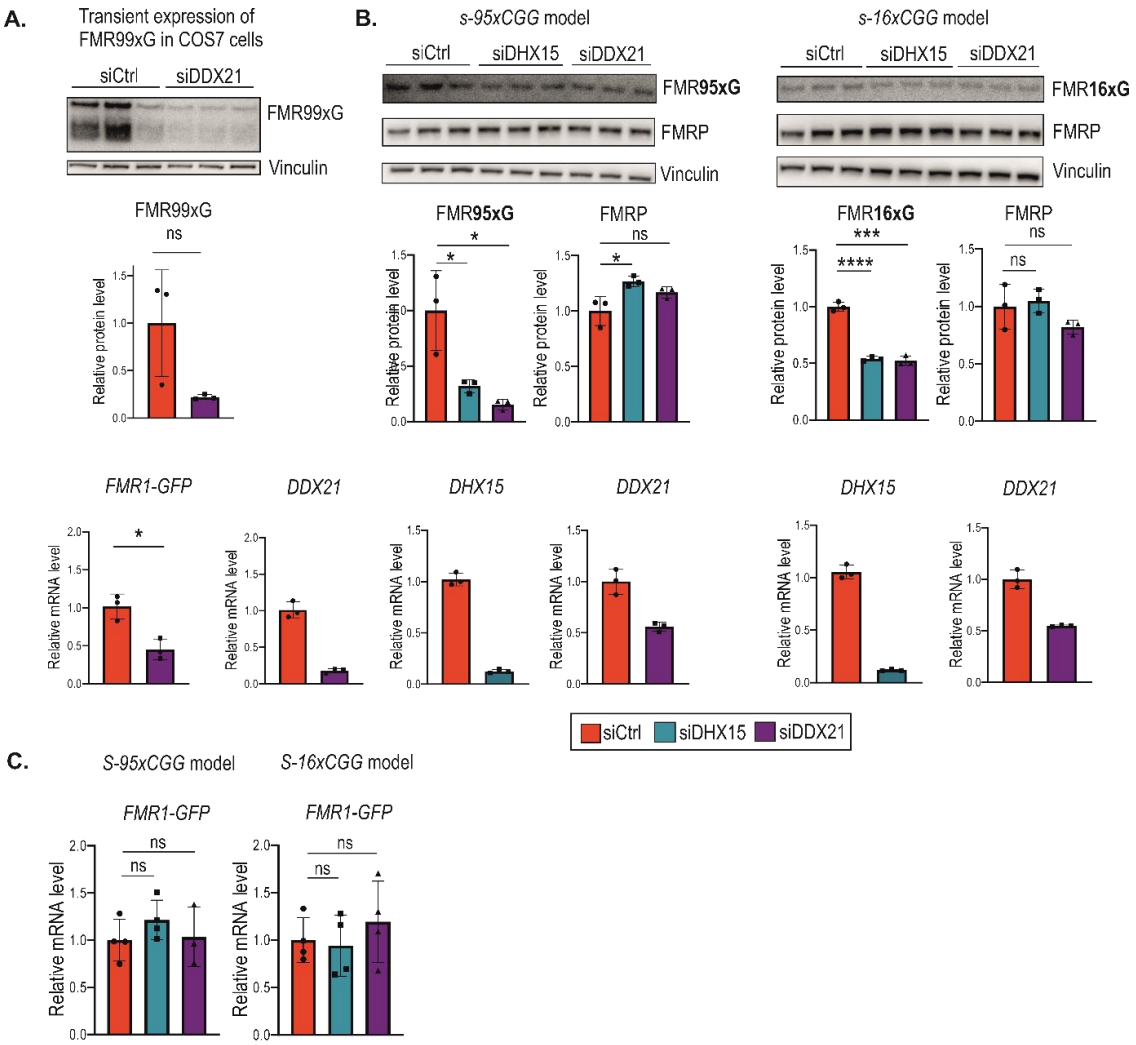


Figure 5.3. Depletion of RNA helicases, DDX21 and DHX15 downregulated FMRpolyG level in cellular models. A. Westen blot analysis of transiently expressed FMR99xG normalized to Vinculin and RT-qPCR analysis of *FMR1-GFP* and *DDX21* normalized to *GAPDH* upon DDX21 silencing in COS7 cells. **B.** Western blot analysis of FMRP and FMRpolyG normalized to Vinculin (upper panel) and RT-qPCR knockdown verification of *DDX21* and *DHX15* normalized to *GAPDH* (lower panel) in stable *S-95xCGG* and *S-16xCGG* models upon DHX15 and DDX21 siRNA treatment. **C.** RT-qPCR analysis of transgenic *FMR1-GFP* normalized to *GAPDH* upon siDDX21 or siDHX15 treatment in *S-95xCGG* and *S-16xCGG* cell lines. The graphs present means from N = 3 (A, B; western blots, C) or N = 4 biologically independent samples (B; RT-qPCR) with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, non-significant. (The full size images are placed in supplementary figures section).

5.1.3 The effect of ALYREF silencing on FMRpolyG

Furthermore, we evaluated the level of FMRpolyG derived from transient transfection system upon silencing ALYREF in COS7 cells. Obtained results suggested negative effect of ALYREF insufficiency on FMRpolyG, however the results were not significant (perhaps due to high SD in control samples) (Figure 5.4A). We verified if the level of *FMR1-GFP* transcript was affected by ALYREF depletion and RT-qPCR analysis revealed the increase of transgenic mRNA (Figure 5.4A). Next, we tested the effect on ALYREF silencing on FMRpolyG production in *S-95xCGG* and *S-16xCGG* models. Despite the fact that in *S-95xCGG* model, results of western blot analysis of FMR95xG were not statistically significant (perhaps due to high SD in control sample and the weak band signal hindering the quantification), similarly to transient overexpression experiment, ALRYEF insufficiency seemed to evoke the downregulation of FMRpolyG level (Figure 5.4B). In case of *S-16xCGG* model, silencing of ALYREF did not change FMR16xG level. In both models, ALYREF depletion had no effect on FMRP level (Figure 5.4B). It is important to highlight, that presented results concerning ALYREF CGG-related RAN translation regulatory properties are preliminary and drawing confident conclusion require repetition of experiments.

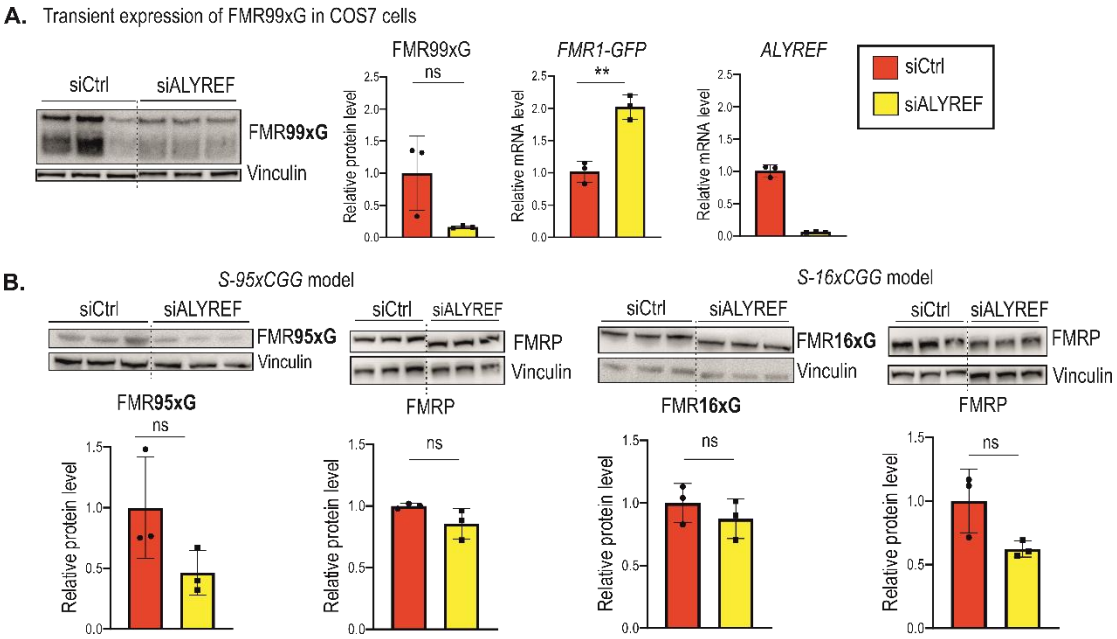


Figure 5.4. The effect of ALYREF silencing on FMRpolyG in cellular models. A. Westen blot analysis of transiently expressed FMR99xG normalized to Vinculin and RT-qPCR analysis of

transgenic *FMR1-GFP* and *ALYREF* mRNA normalized to *GAPDH* upon ALYREF silencing in COS7 cells. **B.** Western blot analysis of FMRP and FMRpolyG normalized to Vinculin derived from expanded and normal CGG repeats in stable *S-95xCGG* and *S-16xCGG* models, respectively, after ALYREF depletion. Note that presented blots were cropped (the full size images are placed in supplementary figures section). The graphs present means from N = 3 biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: **, $P < 0.01$; ns, non-significant.

5.1.4 Technical aspects of mass spectrometry sample preparation affected identification of proteins bound to FMR1 RNA

In order to characterize pool of proteins bound to *FMR1 RNA* derived from the samples treated with acetone precipitation and SP3 purification, we performed a comparative analysis between technical replicates. Only proteins identified by unique peptides were considered. Analysis revealed 155 common proteins among analyzed samples (Figure 5.5A, Table 5.3). Gene ontology (GO) analysis indicated that majority of identified proteins represent RNA binding properties, are a constituent of the ribosome, participate in translation or are involved in other processes related to RNA metabolism (Figure 5.5A, Table 5.4). Given that depending on MS sample preparation identified proteins may vary (120), we compared proteins identified in two different MS sample preparation approaches, either direct, on-beads digestion or precipitation with acetone and SP3 purification. To exclude possibility that observed differences may stem from bioinformatic analysis, we applied the same bioinformatic pipeline (described in Tutak et al.) to analyze both raw data sets. Number of proteins identified in both approaches was similar, 254 and 213, for direct digestion or acetone precipitation approach respectively (while considering proteins identified in minimum 2 technical replicates), and 67 proteins (including RPS26) were overlapping between data sets (Figure 5.5B, Table 5.5). Enriched GO terms for 67 proteins were similar to the ones revealed for 155 proteins common among *FMR1 RNA* replicates (Figure 5.5A&B), however some terms were exclusive for those 67 proteins, such as rRNA metabolic process or Synapse (Figure 5.5B, Table 5.6).

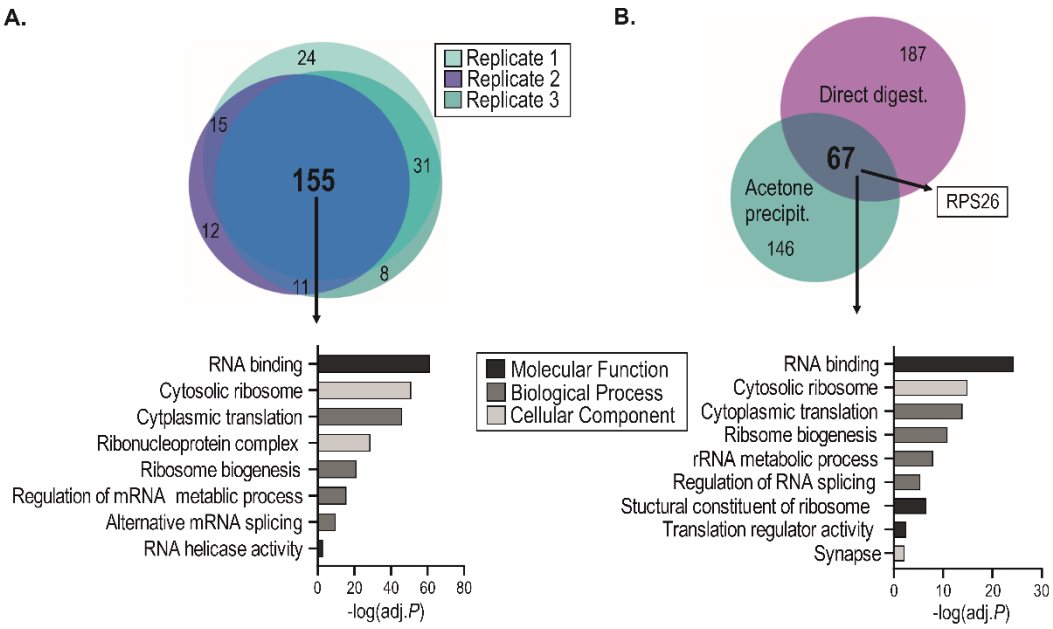


Figure 5.5. Comparison between proteins interacting with *FMR1 RNA* identified in samples prepared with two different workflows for MS analysis. A. Venn diagram represents proteins common between three replicates of *FMR1 RNA* samples treated with acetone and SP3 purification. Lower graph represents GO analysis performed on 155 common proteins between triplicates. **B.** Venn diagram represents proteins identified in at least 2 replicates of *FMR1 RNA* samples treated by two different MS sample preparation workflows (either Direct digestion or Acetone precipitation*). Lower graph represents GO analysis performed on 67 common proteins among two sample groups. Selected, significantly enriched GO terms are presented and full lists are presented in Table 5.4 and 5.6, respectively. *Please note that the higher number of proteins analyzed in this data set (B – Acetone precipitation) is different than the one analyzed in A due to less stringent selection criteria.

Table 5.5 67 proteins identified in *FMR1 RNA* samples overlapping between two MS analysis

Proteins identified in at least two technical replicates in two, separate MS data sets. Bolded proteins indicate known interactors of *FMR1* described by other research groups.

67 common proteins in two MS data sets		
Uniprot ID	**Entry Name**	**Protein names**
Q13085	ACACA	Acetyl-CoA carboxylase 1
P11498	PYC	Pyruvate carboxylase, mitochondrial
P05165	PCCA	Propionyl-CoA carboxylase alpha chain
Q96RQ3	MCCA	Methylcrotonoyl-CoA carboxylase subunit alpha
O00763-3	ACACB	Acetyl-CoA carboxylase 2
P49327	FAS	Fatty acid synthase
P05166	PCCB	Propionyl-CoA carboxylase beta chain
P0DMV9	HS71B	Heat shock 70 kDa protein 1B
P26641	EF1G	Elongation factor 1-gamma
P08238	HS90B	Heat shock protein HSP 90-beta
Q9BVP2-2	GNL3	Guanine nucleotide-binding protein-like 3
O60841	IF2P	Eukaryotic translation initiation factor 5B
P39023	RL3	Large ribosomal subunit protein uL3
P04075	ALDOA	Fructose-bisphosphate aldolase A
P09874	PARP1	Poly [ADP-ribose] polymerase 1-terminus
P63261	ACTG	Actin
O75367	H2AY	Core histone macro-H2A.1
P09651-3	**HNRNPA1**	**Heterogeneous nuclear ribonucleoprotein A1**
P13639	EF2	Elongation factor 2
Q9UQ35	**SRRM2**	**Serine/arginine repetitive matrix protein 2**
P27635	RL10	Large ribosomal subunit protein uL16
Q9BZE4-3	GTPB4	GTP-binding protein 4
P62241	RS8	Small ribosomal subunit protein eS8
P63241	IF5A1	Eukaryotic translation initiation factor 5A-1
P19338	**NUCL**	**Nucleolin**
P26599-1	PTBP1	Polypyrimidine tract-binding protein 1
Q96SB4-4	**SRPK1**	**SRSF protein kinase 1**
Q99880	H2B1L	Histone H2B type 1-L
O95232	LC7L3	Luc7-like protein 3
P07437	TBB5	Tubulin beta chain
P07900	HS90A	Heat shock protein HSP 90-alpha
P11387	TOP1	DNA topoisomerase 1
P22626-2	**HNRNPA2B1**	**Heterogeneous nuclear ribonucleoproteins A2/B1**
P30414	NKTR	NK-tumor recognition protein
P51532-5	SMCA4	Transcription activator BRG1
P52272-2	**HNRPM**	**Heterogeneous nuclear ribonucleoprotein M**
P61978-3	**HNRPK**	**Heterogeneous nuclear ribonucleoprotein K**
P62244	RS15A	Small ribosomal subunit protein uS15
P62750	RL23A	Large ribosomal subunit protein uL23

P62753	RS6	Small ribosomal subunit protein eS6
P62805	H4	Histone H4
P62847-2	RS24	Small ribosomal subunit protein eS24
Q00839-2	**HNRPU**	**Heterogeneous nuclear ribonucleoprotein U**
Q07065	CKAP4	Cytoskeleton-associated protein 4
Q14241	ELOA1	Elongin-A
Q14839	CHD4	Chromodomain-helicase-DNA-binding protein 4
Q1ED39	KNOP1	Lysine-rich nucleolar protein 1
Q8NC51-4	SERB1	SERPINE1 mRNA-binding protein 1
Q9BQG0	MBB1A	Myb-binding protein 1A
Q9NX58	LYAR	Cell growth-regulating nucleolar protein
Q6FI13	H2A2A	Histone H2A type 2-A
P38919	IF4A3	Eukaryotic initiation factor 4A-III
P62854	RS26	Small ribosomal subunit protein eS26
Q99459	CDC5L	Cell division cycle 5-like protein
Q9Y2X3	NOP58	Nucleolar protein 58
P22087	FBRL	rRNA 2'-O-methyltransferase fibrillarin
P23396	RS3	Small ribosomal subunit protein uS3
P38432	COIL	Coilin
P81605	DCD	Dermcidin
Q13263-2	TIF1B	Transcription intermediary factor 1-beta
Q96T37-4	RBM15	RNA-binding protein 15
P52597	**HNRPF**	**Heterogeneous nuclear ribonucleoprotein F**
P63173	RL38	Large ribosomal subunit protein eL38
P46777	RL5	Large ribosomal subunit protein uL18
P49411	EFTU	Elongation factor Tu
P62263	RS14	Small ribosomal subunit protein uS11
P62829	RL23	Large ribosomal subunit protein uL14

5.2 Discussion

The main aim of the PhD thesis was to identify novel RAN translation modifiers specific for mutant *FMR1* harboring expanded CGG repeats. We adapted RNA-tagging technique, which allowed in cellulo RNA-protein interaction capture taking into account RNA secondary structures as well as localization of studied molecule in different cellular compartments. During experimental procedure we fixed interactions between proteins and RNAs with formaldehyde crosslinking in account for identification of weak and transient interactions, however this step may have introduced a bias for secondary interactions. Given that downstream MS sample preparation affects proteins identification (120), we applied MS2-based pull down followed by two distinct MS sample preparation protocols to minimize potential losses due to technical limitations. First technical variant relied on direct on-beads proteolytic digestion which limits potential losses of material, yet may leave more contaminants such a SDS in the sample that hinders further MS analysis. In second technical variant, eluted proteins were precipitated with acetone and digested peptides were treated with SP3 purification procedure to remove common contaminants, although precipitation step might have influenced the solubility of proteins affecting their further identification. These factors need to be taken into consideration while interpreting the obtained data.

As the result, we obtained two data sets containing *FMR1* binding proteins (Tutak et al.; Figure 1 and Figure 4A, Supplementary Table 3). Number of proteins identified in both approaches was similar and 67 proteins (Table 5.5) were overlapping between two data sets (Figure 5.5B). The differences among identified proteins may be explained by distinct downstream MS sample preparation protocols. Unsurprisingly, regardless of investigated data set, GO terms were mostly common to all analyses including following GO terms: RNA binding properties, ribosomal constituent or translation regulation (Tutak et al.; Figure 5.5A&B). However, in data containing 67 common proteins some GO terms were exclusive for this group, such as rRNA processing or Synapse. This result is especially interesting because *FMR1* encodes FMRP, a protein with well documented role in synaptic plasticity (129). Remarkably, among identified proteins some of them were already identified as *FMR1* interactors (such as serine/arginine rich proteins, for instance, SRPK1 (91)) or were

found in intranuclear inclusions in FXTAS patient's tissues (such as Nucleolin or multiple Heterogeneous nuclear ribonucleoproteins, hnRNPs and others (39, 67, 75, 93, 126, 130)). (Forementioned proteins were bolded in the Table 5.5.) Hence, this implies that our data set containing 67 interactors of *FMR1 RNA* may be a particularly valuable resource of information about mutant *FMR1* biology.

LFQ analysis elucidated proteins enriched in *FMR1 RNA* samples in comparison to *GC-rich RNA* control samples (Figure 5.2, Table 5.1). However, it is worth mentioning that precipitation steps might have incurred loss of some proteins present in samples, thus negatively affecting further quantification and skewing the data interpretation. Additionally, LFQ data should be analyzed with assumption that protein binding to two RNA molecules (*FMR1 RNA* and *GC-rich RNA*) may represent key regulatory functions specific for the one or both of the RNAs. Therefore, proteins enriched in *GC-rich RNA* samples should not be discarded by default from further analysis in the context of *FMR1* regulation. Among five proteins significantly enriched in *FMR1 RNA* samples (Figure 5.2), Poly(rC)-binding protein 2 (PCBP2) and YTH domain-containing protein 1 (YTDC1) seem to be particularly interesting in the context of potential regulation of *FMR1* mRNA. YTDC1 is a regulator of alternative splicing that specifically recognizes and binds N6-methyladenosine (m6A)-containing RNAs (131). Moreover, YTDC1 binds to Serine/arginine-rich splicing factor 3 (SRSF3) and Nuclear RNA export factor 1 (NXF1) facilitating nuclear export of RNAs harboring m6A modifications (132). Given that our preliminary experiments indicate that *FMR1* contains m6A modification (data unpublished), YTDC1 may have so far not described roles in *FMR1* processing and metabolism. PCBP2 is known to interact with structured RNAs such as internal ribosome entry site (IRES) elements (133) and play an important role in modulating cellular response to viral infection (134, 135). Binding to IRES sequences may suggest potential RAN translation regulator properties, especially that known RAN translation modifier, RPS25 binds to IRES as well (117, 136–138). To our surprise, GAPDH was identified as the most enriched protein on *FMR1 RNA*. This may be explained as a technical artifact or stem from the lack of GAPDH identification in *GC-rich RNA* samples. However, this enzyme is characterized as non-canonical RNA binding protein (139), which suggests its potential role in *FMR1* metabolism. Nevertheless, other

proteins identified in LFQ analyses (Tutak et al.; Figure 1, Supplementary table 5) regardless of statistical significance are worth further investigating in the context of RAN translation regulation of *FMR1* metabolism in general.

We verified RAN translation regulatory properties of two RNA helicases; DHX15 and DDX21. Out results are in line with current knowledge concerning roles of these enzymes in RAN translation regulation (76, 85, 86). Observed RAN translation regulation facilitated by DHX15 and DDX21 seemed not to be dependent on CGG content as both types of FMRpolyG (with normal and long glycine stretches) were sensitive to DXH15 and DDX21 depletion (Figure 5.3B). Similar regulation was observed for DDX3X, which was shown to bind to *FMR1* regardless of CGG repeats content (85). On the contrary to studied helicases, ALYREF depletion had an effect only on shorter form of FMRpolyG (Figure 5.4B).

In contrast to stable cell lines, silencing DDX21 in COS7 cells transiently expressing FMRpolyG, downregulated the level of *FMR1-GFP* (Figure 5.3A&C). Given documented role of DDX21 in resolving the R loop structure and therefore promoting transcription (128, 140), it is possible that DDX21 depletion hindered transcription from multiple copies of plasmid DNA due to insufficient R loop unwinding. Additionally, level of *DDX21* mRNA in stable cell lines indicated that silencing procedure was approximately 50% less efficient in comparison to COS7 cells. (Figure 5.3A&B). Perhaps, partial expression of DDX21 in *S-95/16xCGG* models was sufficient enough to resolve potential R loops and not affect the transcription, although this is a speculation. Surprisingly, silencing ALYREF resulted in doubling the amount of *FMR1-GFP* mRNA while downregulating the FMRpolyG (Figure 5.4A). This might be explained by transcriptional activator role of ALYREF in addition to its role in nuclear-cytoplasmic transport (141–143). Hence, the role of ALYREF in *FMR1* transcription regulation or other RNA metabolism processes requires further investigation in future.

Importantly, DDX21 and ALYREF depletion did not affect FMRP level (Figure 5.3B, 5.4B), on the contrary to DHX15 insufficiency, which induced the increase of FMRP level in *S-95xCGG* model (Figure 5.3B). This suggests that studied proteins represent some specificity towards FMRpolyG over FMRP frame. It should be noted that in humans, FMRpolyG and FMRP are translated from the same mRNA, while in

artificial models these two proteins are not synthetized from the same transcript, complicating the data interpretation.

Overall, DHX15, DDX21 and ALYREF were not previously described in the context of RAN translation regulation and we provide the evidence indicating their roles in this process. Nevertheless, conducted experiments verifying DXH15, DDX21 and ALYREF RAN translation properties were preliminary and did not cover mechanistic aspects, hence further studies to reveal the nature of discovered regulation are required. Some aspects are worth addressing in the future. For instance, verifying RNA-protein interaction (ideally by an orthogonal technique) and addressing whether CGG repeats are necessary for this interaction, or checking if overexpression of selected proteins will evoke anticipated upregulation of FMRpolyG biosynthesis. Additionally, testing multiple siRNAs targeting selected proteins by validating the level of target protein expression by western blotting may be desired to maximize the reproducibility and minimize possibility of technical artifacts affecting results interpretation.

To sum up, MS sample preparation protocol including acetone precipitation and SP3 purification steps revealed multiple proteins bound to *FMR1 RNA* including DDX21 and ALYREF. Both of the proteins were tested in the context of potential RAN translation modulation and our results indicated their involvement in regulating FMRpolyG level, thus pointing novel potential therapeutic targets for FXPAC.

Finally, 67 proteins identified in both types of MS2-based pull down samples treated with different downstream protocols are particularly valuable source of information about biology of mutant *FMR1* mRNA and our dataset likely contains numerous RAN translation related proteins yet to be uncovered.

5.3 Materials and methods

The methodology of all experiments was presented in Tutak et al., except for the procedures described below.

5.3.1 MS2-based pull down followed by acetone protein precipitation and SP3 purification

Detailed experimental protocol for MS2-based pull down is described in section titled "Mass spectrometry-based proteins screening" (Tutak et al.). The protocol applied here differs from the one described in the manuscript in terms of sample preparation for MS analysis. Differences begin after pull down procedure followed by washing steps and are presented as follow; samples were transferred to the fresh Protein Lobind tubes (Eppendorf) and RNA-protein complexes were eluted from the beads by heating the samples for 10 min at 95°C in 100 µl of elution buffer (0.1% sodium dodecyl sulphate, 0.05 mM biotin in phosphate buffered saline). Next, 400 µl of ice-cold acetone was added to the sample. Samples were mixed and incubated for 60 min in 4°C followed by 10 min centrifugation at 13,000 g. Supernatant was discarded, and the protein pellet was air dried. Subsequently, protein pellets were processed in Mass Spectrometry Laboratory, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a Street, 02-106 Warsaw, Poland. Samples were suspended in dissolution buffer (100 mM NH4HCO3). First, the cysteines were reduced by 1 hour incubation with 20 mM tris(2-carboxyethyl)phosphine (TCEP) at 60°C followed by 10 min incubation at a RT with 50 mM methyl methanethiosulfonate (MMTS). Digestion was performed at 37°C overnight with 1 µg of trypsin (Promega). After digestion, peptides were vacuum-dried, re-suspended in 10 µl of extraction buffer (0.1% TFA 2% acetonitrile) with sonication, and processed using single-pot solid-phase-enhanced sample preparation (SP3). Magnetic beads mix were prepared by combining equal parts of Sera-Mag Carboxyl hydrophilic and hydrophobic particles (09-981-121 and 09-981-123, GE Healthcare). The bead mix was washed three times with MS-grade water and re-suspended in a working concentration of 10 µg/µl. The bead mix was then added to the samples then suspended in 100% acetonitrile, this step was repeated 3 times. Pure peptides ware eluted from the beads by using 2% acetonitrile in MS-grade water. Using a magnet, the peptides solution was separated from beads. Peptide

mixture was dried in SpeedVac and re-suspended in 80 µl extraction buffer (0.1% TFA 2% acetonitrile) with sonication.

5.3.2 Mass spectrometry (MS)

Samples were analyses using LC-MS system composed of Evosep One (Evosep Biosystems) coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) via Flex nanoESI ion source (Thermo Fisher Scientific). 20 µl of each sample was loaded onto disposable Evotips C18 trap columns (Evosep Biosystems, Odense, Denmark) according to the manufacturer protocol. Chromatography was carried out at a flow rate 220 nl/min using the 88 min preformed gradient on EV1106 analytical column (Dr Maisch C18 AQ, 1.9 µm beads, 150 µm ID, 15 cm long, Evosep Biosystems). Data was acquired in positive mode with a data-dependent method using the following parameters. MS1 resolution was set at 60 000. For MS2, resolution was set at 15 000 and top 40 precursors within an isolation window of 1.6 m/z considered for MS/MS analysis.

5.3.3 MS data analysis

MS data was analyzed as described in Tutak et al. Briefly, raw data obtained from the LC-MS/MS runs were analyzed in MaxQuant v2.0.3.0 (144) software. UniProtKB database for reviewed human canonical and isoform proteins of May 2023 was used. For comparative analysis between samples only protein identified by unique peptides were included. Label-free quantification (LFQ) using Perseus software v2.0.3.0 (145) after filtering for “reverse”, “contaminant” and “only identified by site” proteins. The LFQ intensity was logarithmized (log2[x]), and imputation of missing values was performed with a normal distribution (width = 0.3; shift = 1.8). Proteomes were compared using t-test statistics with a permutation-based FDR of 5% and *P*-values < 0.05 were considered to be statistically significant.

5.3.4 Cell culture and transfection

COS7 and S-95/16xCGG cells were maintained as described in Tutak et al. All siRNAs were purchased from Thermo Fisher Scientific, catalog number indicated in brackets: siDHX15 (#s4029), siDDX21 (#s17564), siALYREF (#s19854) and siCtrl (#4390847). For the delivery of siRNAs with the final concentration 25 nM in culture, reverse transfection protocol was applied using jetPRIME® reagent (Polyplus) according to

manufacturer's instruction. In case of COS7, cells were harvested 48h post siRNA silencing and 24h post transient plasmids expression. *S-95/16xCGG* cells were harvested 48h post siRNA silencing and 72h after doxycycline treatment.

5.3.5 Quantitative real-time PCR (RT-qPCR)

RNA isolation and reverse transcription were performed as described in Tutak et al. Quantitative real-time RT-PCRs were performed in a QuantStudio 7 Flex System (Thermo Fisher Scientific) using Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific) with 5 ng of cDNA in each reaction. RT-qPCRs were run at 58°C annealing temperature. To amplify *DHX15*, *DDX21* and *ALYREF* following primer pairs were used:

DHX15:

Forward: 5' CGCAGATGAGGCCAAGATGA 3', Reverse: 5' CGTTCTAAATGTGCCACCTGC 3'

DDX21:

Forward: 5' GGACCCAAAGGGCAGCAGTT 3', Reverse: 5' AACGACTGGGCATCCTGCCT 3'

ALYREF:

Forward: 5' TGCCACCTCTGTTTACGCTC 3', Reverse 5' TCTGGTCGCAGCTTAGGAAC 3'

Ct values were normalized to *GAPDH* mRNA level. Fold differences in expression level were calculated according to the $2^{-\Delta\Delta Ct}$ method (146).

5.3.6 Gene Ontology

Gene ontology (GO) analysis performed was performed using an online software - g:Profiler (147) with g:SCS algorithm, where at least 95% of matches above threshold are statistically significant. As reference proteome we used total human proteome.

5.3.7 Venn diagrams

Venn diagrams were prepared with BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams (148).

5.4 Supplementary information

5.4.1 Tables

All tables (except Table 5.5) are attached to the PhD dissertation on CD in a form of the excel file titled TABLES – UNPUBLISHED RESULTS (Chapter 5) with following pages:

Table 5.1 MaxOuant output including identified proteins in MS2-screenining followed by acetone precipitation and SP3 purification

Table 5.2 Perseus output with label free quantification analysis

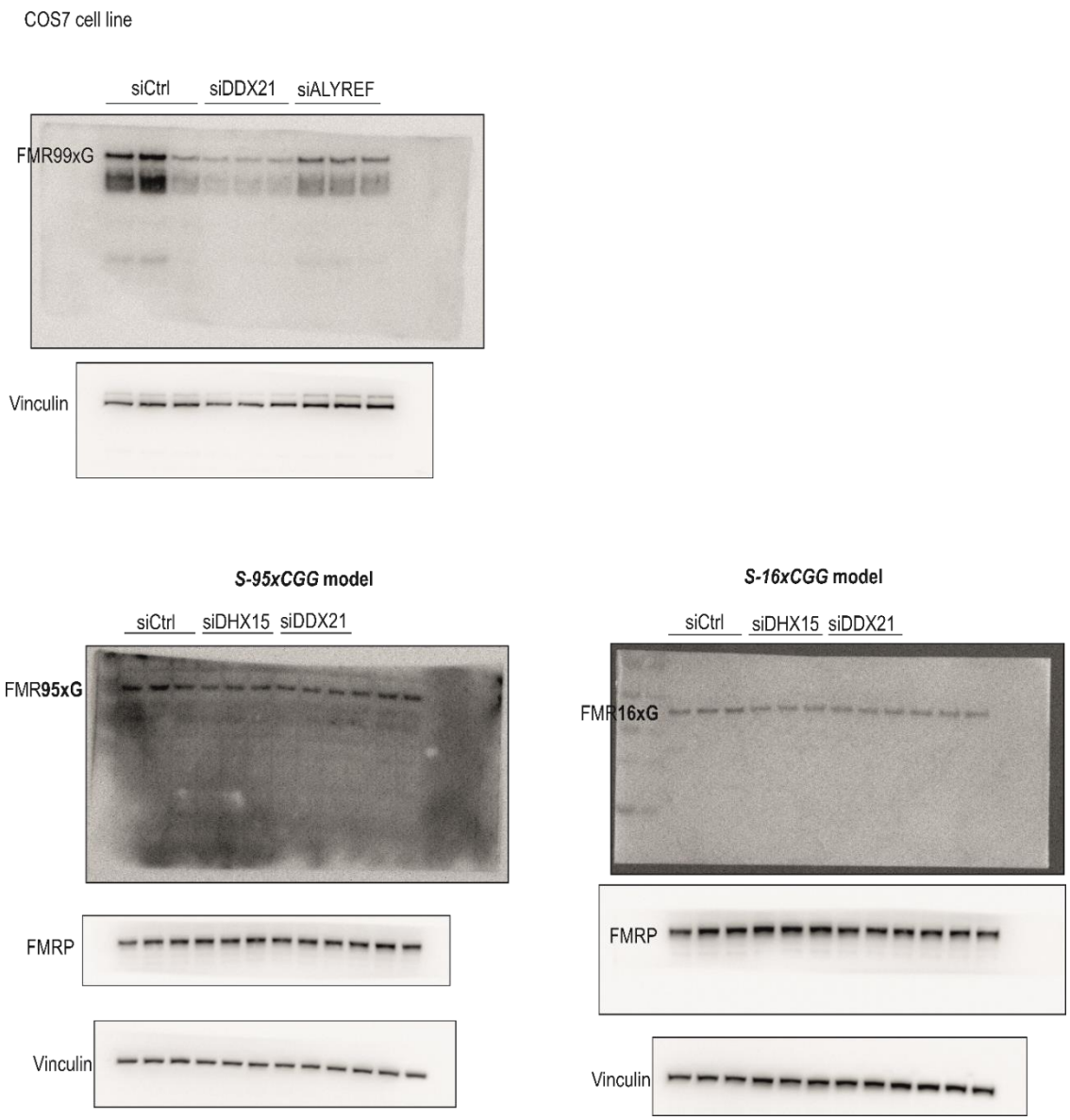
Table 5.3 155 proteins identified in three technical replicates of *FMR1 RNA*

Table 5.4 Gene ontology analysis performed on 155 proteins common for three replicates of *FMR1 RNA*

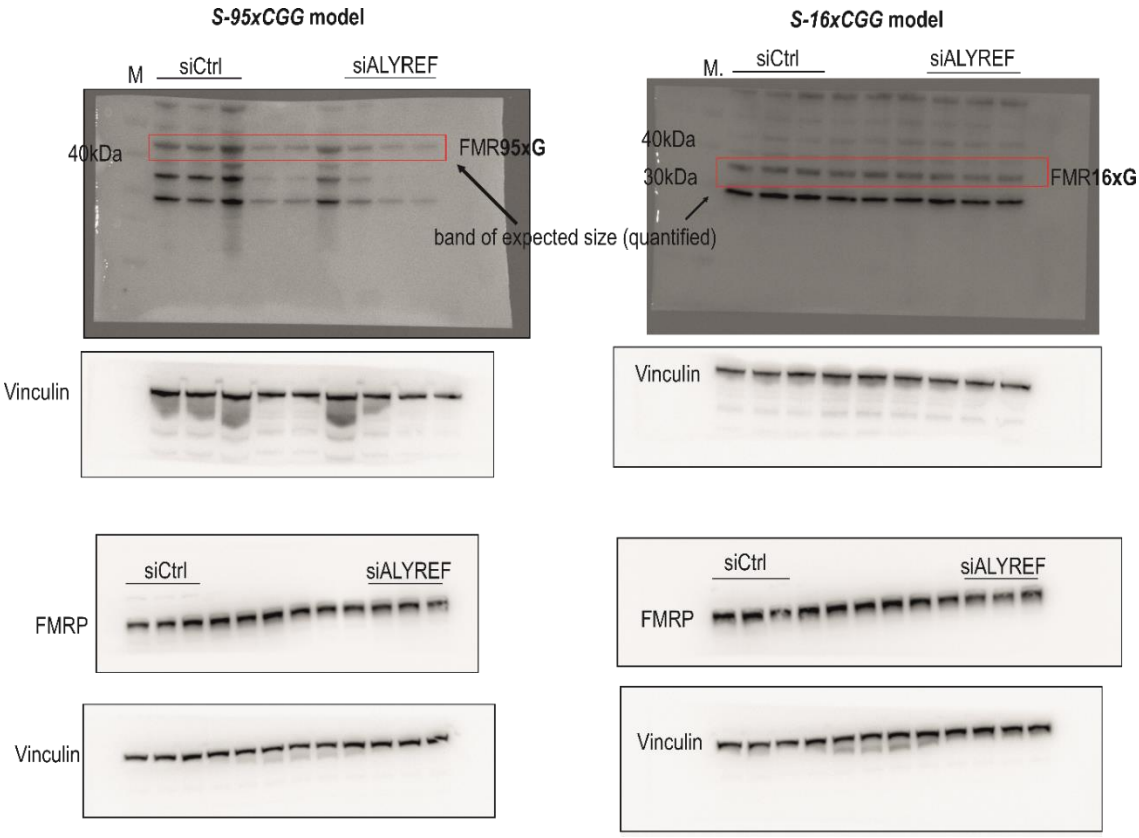
Table 5.5 67 proteins identified in *FMR1 RNA* samples overlapping between two MS analysis (printed in the dissertation)

Table 5.6 Gene ontology analysis performed on 67 proteins bound to *FMR1 RNA*

5.4.2 Supplementary figures



Supplementary Figure 5.1. Full size western blot images corresponding to Figures 5.3 & 5.4.



Supplementary Figure 5.2. Full size western blot images corresponding to Figure 5.4.

5.5 Bibliography

1. Gatchel,J.R. and Zoghbi,H.Y. (2005) Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles. *Nature Reviews Genetics*, **6**, 743–755.

2. Fotsing,S.F., Margoliash,J., Wang,C., Saini,S., Yanicky,R., Shleizer-Burko,S., Goren,A. and Gymrek,M. (2019) The impact of short tandem repeat variation on gene expression. *Nat Genet*, **51**, 1652–1659.

3. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., Fitzhugh,W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

4. Khristich,A.N. and Mirkin,S.M. (2020) On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J Biol Chem*, **295**, 4134–4170.

5. Paulson,H. (2018) Repeat expansion diseases. *Handb Clin Neurol*, **147**, 105-123.

6. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.P., Hudson,T., *et al.* (1992) Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.

7. Harper,P.S. (1999) Huntington's disease: a clinical, genetic and molecular model for polyglutamine repeat disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **354**, 957-961.

8. Renton,A.E., Majounie,E., Waite,A., Simón-Sánchez,J., Rollinson,S., Gibbs,J.R., Schymick,J.C., Laaksovirta,H., van Swieten,J.C., Myllykangas,L., *et al.* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, **72**, 257–268.

9. Gadgil,R., Barthelemy,J., Lewis,T. and Leffak,M. (2017) Replication stalling and DNA microsatellite instability. *Biophys Chem*, **225**, 38–48.

10. Fortune,M.T., Vassilopoulos,C., Coolbaugh,M.I., Siciliano,M.J. and Monckton,D.G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum Mol Genet*, **9**, 439–445.

11. Mirkin,S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.

12. Swami,M., Hendricks,A.E., Gillis,T., Massood,T., Mysore,J., Myers,R.H. and Wheeler,V.C. (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet*, **18**, 3039–3047.

13. Wenstrom,K.D. (2002) Fragile X and other trinucleotide repeat diseases. *Obstet Gynecol Clin North Am*, **29**, 367–388.

14. Sutcliffe,J.S., Nelson,D.L., Zhang,F., Pieretti,M., Caskey,C.T., Saxe,D. and Warren,S.T. (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet*, **1**, 397–400.

15. Fardaei,M., Larkin,K., Brook,J.D. and Hamshere,M.G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res*, **29**, 2766–2771.

16. Lubs,H.A. (1969) A marker X chromosome. *Am J Hum Genet*, **21**, 231-244.

17. Eichler,E.E., Kunst,C.B., Lugenbeel,K.A., Ryder,O.A., Davison,D., Warren,S.T. and Nelson,D.L. (1995) Evolution of the cryptic FMR1 CGG repeat. *Nat Genet*, **11**, 301–308.

18. Kremer,E.J., Pritchard,M., Lynch,M., Yu,S., Holman,K., Baker,E., Warren,S.T., Schlessinger,D., Sutherland,G.R. and Richards,R.I. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science*, **252**, 1711–1714.

19. Fu,Y.H., Kuhl,D.P.A., Pizzuti,A., Pieretti,M., Sutcliffe,J.S., Richards,S., Verkert,A.J.M.H., Holden,J.J.A., Fenwick,R.G., Warren,S.T., *et al.* (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.

20. Verkerk,A.J.M.H., Pieretti,M., Sutcliffe,J.S., Fu,Y.H., Kuhl,D.P.A., Pizzuti,A., Reiner,O., Richards,S., Victoria,M.F., Zhang,F., *et al.* (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–914.

21. Oberlé,I., Rousseau,F., Heitz,D., Kretz,C., Devys,D., Hanauer,A., Boué,J., Bertheas,M.F. and Mandel,J.L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, **252**, 1097–1102.

22. Sidorov,M.S., Auerbach,B.D. and Bear,M.F. (2013) Fragile X mental retardation protein and synaptic plasticity. *Mol Brain*, **6**, 15.

23. Hagerman,R.J., Berry-Kravis,E., Hazlett,H.C., Bailey,D.B., Moine,H., Kooy,R.F., Tassone,F., Gantois,I., Sonenberg,N., Mandel,J.L., *et al.* (2017) Fragile X syndrome. *Nat Rev Dis Primers*, **3,** 17065.

24. Johnson,K., Herring,J. and Richstein,J. (2020) Fragile X Premutation Associated Conditions (FXPAC). *Front Pediatr*, **8**, 266.

25. Hagerman,P. (2013) Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms. *Acta Neuropathol*, **126**, 1–19.

26. Buijsen,R.A.M., Visser,J.A., Kramer,P., Severijnen,E.A.W.F.M., Gearing,M., Charlet-Berguerand,N., Sherman,S.L., Berman,R.F., Willemsen,R. and Hukema,R.K. (2016) c. *Human Reproduction*, **31**, 158–168.

27. Hagerman,R.J., Protic,D., Rajaratnam,A., Salcedo-Arellano,M.J., Aydin,E.Y. and Schneider,A. (2018) Fragile X-Associated Neuropsychiatric Disorders (FXAND). *Front Psychiatry*, **9**, 564.

28. Hagerman,R.J. and Hagerman,P. (2016) Fragile X-associated tremor/ataxia syndrome-features, mechanisms and management. *Nat Rev Neurol*, **12**, 403–412.

29. Tassone,F., Iong,K.P., Tong,T.H., Lo,J., Gane,L.W., Berry-Kravis,E., Nguyen,D., Mu,L.Y., Laffin,J., Bailey,D.B., *et al.* (2012) FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Med*, **4**, 100.

30. Jacquemont,S., Hagerman,R.J., Leehey,M.A., Hall,D.A., Levine,R.A., Brunberg,J.A., Zhang,L., Jardini,T., Gane,L.W., Harris,S.W., *et al.* (2004) Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA*, **291**, 460–469.

31. Tassone,F., Hagerman,R.J., Loesch,D.Z., Lachiewicz,A., Taylor,A.K. and Hagerman,P.J. (2000) Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *Am J Med Genet*, **94**, 232–236.

32. Kenneson,A., Zhang,F., Hagedorn,C.H. and Warren,S.T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum Mol Genet*, **10**, 1449–1454.

33. Tassone,F., Beilina,A., Carosi,C., Albertosi,S., Bagni,C., Li,L., Glover,K., Bentley,D. and Hagerman,P.J. (2007) Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA*, **13**, 555–562.

34. Jacquemont,S., Hagerman,R.J., Leehey,M., Grigsby,J., Zhang,L., Brunberg,J.A., Greco,C., Des Portes,V., Jardini,T., Levine,R., *et al.* (2003) Fragile X Premutation Tremor/Ataxia Syndrome: Molecular, Clinical, and Neuroimaging Correlates. *Am J Hum Genet*, **72**, 869-878.

35. Hagerman,R.J., Leehey,M., Heinrichs,W., Tassone,F., Wilson,R., Hills,J., Grigsby,J., Gage,B. and Hagerman,P.J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*, **57**, 127–130.

36. Greco,C.M., Hagerman,R.J., Tassone,F., Chudley,A.E., Del Bigio,M.R., Jacquemont,S., Leehey,M. and Hagerman,P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, **125**, 1760–1771.

37. Greco,C.M., Berman,R.F., Martin,R.M., Tassone,F., Schwartz,P.H., Chang,A., Trapp,B.D., Iwahashi,C., Brunberg,J., Grigsby,J., *et al.* (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain*, **129**, 243–255.

38. Todd,P.K., Oh,S.Y., Krans,A., He,F., Sellier,C., Frazer,M., Renoux,A.J., Chen,K. chun, Scaglione,K.M., Basrur,V., *et al.* (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*, **78**, 440–455.

39. Ma,L., Herren,A.W., Espinal,G., Randol,J., McLaughlin,B., Martinez-Cerdeño,V., Pessah,I.N., Hagerman,R.J. and Hagerman,P.J. (2019) Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome. *Acta Neuropathol Commun*, **7**, 143.

40. Glineburg,M.R., Todd,P.K., Charlet-Berguerand,N. and Sellier,C. (2018) Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. *Brain Res*, **1693**, 43-54.

41. Rebar,R.W. (2009) Premature ovarian failure. *Obstetrics and Gynecology*, **113**, 1355–1363.

42. Stephanie L. Sherman (2002) Premature ovarian failure in the fragile X syndrome. *Am. J. Med. Genet.*, **97**, 189–194.

43. Buijsen,R.A.M., Visser,J.A., Kramer,P., Severijnen,E.A.W.F.M., Gearing,M., Charlet-Berguerand,N., Sherman,S.L., Berman,R.F., Willemsen,R. and Hukema,R.K. (2016) Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. *Human Reproduction*, **31**, 158–168.

44. Shelly,K.E., Candelaria,N.R., Li,Z., Allen,E.G., Jin,P. and Nelson,D.L. (2021) Ectopic expression of CGG-repeats alters ovarian response to gonadotropins and leads to infertility in a murine FMR1 premutation model. *Hum Mol Genet*, **30**, 923-938.

45. Sobczak,K., de Mezer,M., Michlewski,G., Krol,J. and Krzyzosiak,W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res*, **31**, 5469-5482.

46. Krzyzosiak,W.J., Sobczak,K., Wojciechowska,M., Fiszer,A., Mykowska,A. and Kozlowski,P. (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res*, **40**, 11–26.

47. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A*, **91**, 4950.

48. Loomis,E.W., Sanz,L.A., Chédin,F. and Hagerman,P.J. (2014) Transcription-Associated R-Loop Formation across the Human FMR1 CGG-Repeat Region. *PLoS Genet*, **10**, e1004294.

49. Sellier,C., Rau,F., Liu,Y., Tassone,F., Hukema,R.K., Gattoni,R., Schneider,A., Richard,S., Willemsen,R., Elliott,D.J., *et al.* (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO Journal*, **29**, 1248–1261.

50. Sellier,C., Freyermuth,F., Tabet,R., Tran,T., He,F., Ruffenach,F., Alunni,V., Moine,H., Thibault,C., Page,A., *et al.* (2013) Sequestration of DROSHA and DGCR8 by expanded CGG RNA Repeats Alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep*, **3**, 869–880.

51. Malik,I., Kelley,C.P., Wang,E.T. and Todd,P.K. (2021) Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature Reviews Molecular Cell Biology, **22**,* 589–607.

52. Tassone,F., Protic,D., Allen,E.G., Archibald,A.D., Baud,A., Elvassore,N., Gabis,L. V, Grudzien,S.J., Hall,D.A., Hessl,D., *et al.* (2023) Insight and Recommendations for Fragile X-Premutation-Associated Conditions from the Fifth International Conference on FMR1 Premutation. Cells, **12**, 2330.

53. Groh,M., Lufino,M.M.P., Wade-Martins,R. and Gromak,N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet*, **10**, e1004318.

54. Boque-Sastre,R., Soler,M. and Guil,S. (2017) Detection and Characterization of R Loop Structures. *Methods Mol Biol*, **1543**, 231–242.

55. Abu Diab,M., Mor-Shaked,H., Cohen,E., Cohen-Hadad,Y., Ram,O., Epsztejn-Litman,S. and Eiges,R. (2018) The G-rich Repeats in FMR1 and C9orf72 Loci Are Hotspots for Local Unpairing of DNA. *Genetics*, **210**, 1239–1252.

56. Ginno,P.A., Lott,P.L., Christensen,H.C., Korf,I. and Chédin,F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*, **45**, 814–825.

57. Belotserkovskii,B.P., Shin,J.H.S. and Hanawalt,P.C. (2017) Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res*, **45**, 6589–6599.

58. Crossley,M.P., Bocek,M. and Cimprich,K.A. (2019) R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell*, **73**, 398–411.

59. Gan,W., Guan,Z., Liu,J., Gui,T., Shen,K., Manley,J.L. and Li,X. (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev*, **25**, 2041–2056.

60. Kumari,D. and Usdin,K. (2014) Polycomb group complexes are recruited to reactivated FMR1 alleles in Fragile X syndrome in response to FMR1 transcription. *Hum Mol Genet*, **23**, 6575–6583.

61. Colak,D., Zaninovic,N., Cohen,M.S., Rosenwaks,Z., Yang,W.Y., Gerhardt,J., Disney,M.D. and Jaffrey,S.R. (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science (1979)*, **343**, 1002–1005.

62. Wojciechowska,M. and Krzyzosiak,W.J. (2011) Cellular toxicity of expanded RNA repeats: Focus on RNA foci. *Hum Mol Genet*, **20**, 3811–3821.

63. Tassone,F., Iwahashi,C. and Hagerman,P.J. (2004) FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol*, **1**, 103–105.

64. Fardaei,M., Larkin,K., Brook,J.D. and Hamshere,M.G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res*, **29**, 2766–2771.

65. Jain,A. and Vale,R.D. (2017) RNA phase transitions in repeat expansion disorders. *Nature*, **546**, 243–247.

66. Lin,Y., Protter,D.S.W., Rosen,M.K. and Parker,R. (2015) Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell*, **60**, 208–219.

67. Asamitsu,S., Yabuki,Y., Ikenoshita,S., Kawakubo,K., Kawasaki,M., Usuki,S., Nakayama,Y., Adachi,K., Kugoh,H., Ishii,K., *et al.* (2021) CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Sci Adv*, **7**, eabd9440.

68. Sofola,O.A., Jin,P., Qin,Y., Duan,R., Liu,H., de Haro,M., Nelson,D.L. and Botas,J. (2007) RNA-Binding Proteins hnRNP A2/B1 and CUGBP1 Suppress Fragile X CGG Premutation Repeat-Induced Neurodegeneration in a Drosophila Model of FXTAS. *Neuron*, **55**, 565–571.

69. Jin,P., Duan,R., Qurashi,A., Qin,Y., Tian,D., Rosser,T.C., Liu,H., Feng,Y. and Warren,S.T. (2007) Pur α Binds to rCGG Repeats and Modulates Repeat-Mediated Neurodegeneration in a Drosophila Model of Fragile X Tremor/Ataxia Syndrome. *Neuron*, **55**, 556–564.

70. He,F., Krans,A., Freibaum,B.D., Paul Taylor,J. and Todd,P.K. (2014) TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. *Hum Mol Genet*, **23**, 5036–5051.

71. Zu,T., Gibbens,B., Doty,N.S., Gomes-Pereira,M., Huguet,A., Stone,M.D., Margolis,J., Peterson,M., Markowski,T.W., Ingram,M.A.C., *et al.* (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A*, **108**, 260–265.

72. Mori,K., Weng,S.M., Arzberger,T., May,S., Rentzsch,K., Kremmer,E., Schmid,B., Kretzschmar,H.A., Cruts,M., Van Broeckhoven,C., *et al.* (2013) The C9orf72

GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science*, **339**, 1335–1338.

73. Bañez-Coronel,M., Ayhan,F., Tarabochia,A.D., Zu,T., Perez,B.A., Tusi,S.K., Pletnikova,O., Borchelt,D.R., Ross,C.A., Margolis,R.L., *et al.* (2015) RAN Translation in Huntington Disease. *Neuron*, **88**, 667–677.

74. Ash,P.E.A., Bieniek,K.F., Gendron,T.F., Caulfield,T., Lin,W.L., DeJesus-Hernandez,M., Van Blitterswijk,M.M., Jansen-West,K., Paul,J.W., Rademakers,R., *et al.* (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, **77**, 639–646.

75. Sellier,C., Buijsen,R.A.M., He,F., Natla,S., Jung,L., Tropel,P., Gaucherot,A., Jacobs,H., Meziane,H., Vincent,A., *et al.* (2017) Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron*, **93**, 331–347.

76. Kearse,M.G., Green,K.M., Krans,A., Rodriguez,C.M., Linsalata,A.E., Goldstrohm,A.C. and Todd,P.K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell*, **62**, 314–322.

77. Krans,A., Skariah,G., Zhang,Y., Bayly,B. and Todd,P.K. (2019) Neuropathology of RAN translation proteins in fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun*, **7**, 152.

78. Zhang,Y., Glineburg,M.R., Basrur,V., Conlon,K., Wright,S.E., Krans,A., Hall,D.A. and Todd,P.K. (2022) Mechanistic convergence across initiation sites for RAN translation in fragile X associated tremor ataxia syndrome. *Hum Mol Genet*, **31**, 2317–2332.

79. Wright,S.E., Rodriguez,C.M., Monroe,J., Xing,J., Krans,A., Flores,B.N., Barsur,V., Ivanova,M.I., Koutmou,K.S., Barmada,S.J., *et al.* (2022) CGG repeats trigger translational frameshifts that generate aggregation-prone chimeric proteins. *Nucleic Acids Res*, **50**, 8674–8689.

80. Krans,A., Kearse,M.G. and Todd,P.K. (2016) Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Ann Neurol*, **80**, 871–881.

81. Rodriguez,C.M., Wright,S.E., Kearse,M.G., Haenfler,J.M., Flores,B.N., Liu,Y., Ifrim,M.F., Glineburg,M.R., Krans,A., Jafar-Nejad,P., *et al.* (2020) A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat Neurosci*, **23**, 386–397.

82. Green,K.M., Linsalata,A.E. and Todd,P.K. (2016) RAN translation—What makes it run? *Brain Res*, **1647**, 30–42.

83. Cheng,W., Wang,S., Mestre,A.A., Fu,C., Makarem,A., Xian,F., Hayes,L.R., Lopez-Gonzalez,R., Drenner,K., Jiang,J., *et al.* (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat Commun*, **9**, 51.

84. Gao,F.B., Richter,J.D., and Cleveland,D.W. (2017) Rethinking Unconventional Translation in Neurodegeneration. *Cell* , **171**, 994–1000.

85. Linsalata,A.E., He,F., Malik,A.M., Glineburg,M.R., Green,K.M., Natla,S., Flores,B.N., Krans,A., Archbold,H.C., Fedak,S.J., *et al.* (2019) DDX 3X and specific initiation factors modulate FMR 1 repeat-associated non-AUG-initiated translation. *EMBO Rep*, **20,** e47498.

86. Tseng,Y.J., Sandwith,S.N., Green,K.M., Chambers,A.E., Krans,A., Raimer,H.M., Sharlow,M.E., Reisinger,M.A., Richardson,A.E., Routh,E.D., *et al.* (2021) The RNA helicase DHX36–G4R1 modulates C9orf72 GGGGCC hexanucleotide repeat–associated translation. *Journal of Biological Chemistry*, **297**, 100914.

87. Tseng,Y.-J., Krans,A., Malik,I., Deng,X., Yildirim,E., Ovunc,S., Tank,E.M.H., Jansen-West,K., Kaufhold,R., Gomez,N.B., *et al.* (2024) Ribosomal quality control factors inhibit repeat-associated non-AUG translation from GC-rich repeats. *Nucleic Acids Res*, gkae137.

88. Brandman,O. and Hegde,R.S. (2016) Ribosome-associated protein quality control. *Nat Struct Mol Biol*, **23**, 7–15.

89. Joazeiro,C.A.P. (2019) Mechanisms and functions of ribosome-associated protein quality control. *Nat Rev Mol Cell Biol*, **20**, 368–383.

90. Green,K.M., Glineburg,M.R., Kearse,M.G., Flores,B.N., Linsalata,A.E., Fedak,S.J., Goldstrohm,A.C., Barmada,S.J. and Todd,P.K. (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat Commun*, **8,** 2005.

91. Malik,I., Tseng,Y., Wright,S.E., Zheng,K., Ramaiyer,P., Green,K.M. and Todd,P.K. (2021) SRSF protein kinase 1 modulates RAN translation and suppresses CGG repeat toxicity. *EMBO Mol Med*, **13,** e14163.

92. Ariza,J., Rogers,H., Monterrubio,A., Reyes-Miranda,A., Hagerman,P.J. and Martínez-Cerdeño,V. (2016) A Majority of FXTAS Cases Present with Intranuclear Inclusions Within Purkinje Cells. *Cerebellum*, **15**, 546–551.

93. Iwahashi,C.K., Yasui,D.H., An,H.J., Greco,C.M., Tassone,F., Nannen,K., Babineau,B., Lebrilla,C.B., Hagerman,R.J. and Hagerman,P.J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**, 256–271.

94. Buijsen,R.A.M., Sellier,C., Severijnen,L.A.W.F.M., Oulad-Abdelghani,M., Verhagen,R.F.M., Berman,R.F., Charlet-Berguerand,N., Willemsen,R. and Hukema,R.K. (2014) FMRpolyG-positive inclusions in CNS and non-CNS organs of

a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun*, **2**, 162.

95. Cid-Samper,F., Gelabert-Baldrich,M., Lang,B., Lorenzo-Gotor,N., Ponti,R.D., Severijnen,L.A.W.F.M., Bolognesi,B., Gelpi,E., Hukema,R.K., Botta-Orfila,T., *et al.* (2018) An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep*, **25**, 3422-3434.

96. Hoem,G., Larsen,K.B., Øvervatn,A., Brech,A., Lamark,T., Sjøttem,E. and Johansen,T. (2019) The FMRpolyGlycine protein mediates aggregate formation and toxicity independent of the CGG mRNA hairpin in a cellular model for FXTAS. *Front Genet*, **10,** 249.

97. Arocena,D.G., Iwahashi,C.K., Won,N., Beilina,A., Ludwig,A.L., Tassone,F., Schwartz,P.H. and Hagerman,P.J. (2005) Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. *Hum Mol Genet*, **14**, 3661–3671.

98. Oh,S.Y., He,F., Krans,A., Frazer,M., Taylor Paul,J., Paulson,H.L. and Todd,P.K. (2015) RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. *Hum Mol Genet*, **24**, 4317–4326.

99. Friedman-Gohas,M., Elizur,S.E., Dratviman-Storobinsky,O., Aizer,A., Haas,J., Raanani,H., Orvieto,R. and Cohen,Y. (2020) FMRpolyG accumulates in FMR1 premutation granulosa cells. *J Ovarian Res*, **13**, 22.

100. Haify,S.N., Mankoe,R.S.D., Boumeester,V., van der Toorn,E.C., Verhagen,R.F.M., Willemsen,R., Hukema,R.K. and Bosman,L.W.J. (2020) Lack of a Clear Behavioral Phenotype in an Inducible FXTAS Mouse Model Despite the Presence of Neuronal FMRpolyG-Positive Aggregates. *Front Mol Biosci*, **7,** 599101.

101. Boivin,M., Willemsen,R., Hukema,R.K. and Sellier,C. (2018) Potential pathogenic mechanisms underlying Fragile X Tremor Ataxia Syndrome: RAN translation and/or RNA gain-of-function? *Eur J Med Genet*, **61**, 674–679.

102. Genuth,N.R. and Barna,M. (2018) The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol Cell*, **71**, 364-374.

103. Xue,S. and Barna,M. (2012) Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol*, **13**, 355-369.

104. Shi,Z., Fujii,K., Kovary,K.M., Genuth,N.R., Röst,H.L., Teruel,M.N. and Barna,M. (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol Cell*, **67**, 71-83.

105. Ferretti,M.B., Ghalei,H., Ward,E.A., Potts,E.L. and Karbstein,K. (2017) Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements. *Nat Struct Mol Biol*, **24**, 700–707.

106. Yang,Y.M. and Karbstein,K. (2022) The chaperone Tsr2 regulates Rps26 release and reincorporation from mature ribosomes to enable a reversible, ribosome-mediated response to stress. *Sci Adv*, **8**, 8.

107. Shi,Z. and Barna,M. (2015) Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. *Annu Rev Cell Dev Biol*, **31**, 31–54.

108. Norris,K., Hopes,T. and Aspden,J.L. (2021) Ribosome heterogeneity and specialization in development. *Wiley Interdiscip Rev RNA*, **12,** e1644.

109. Kondrashov,N., Pusic,A., Stumpf,C.R., Shimizu,K., Hsieh,A.C., Xue,S., Ishijima,J., Shiroishi,T. and Barna,M. (2011) Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell*, **145**, 383–397.

110. Xue,S., Tian,S., Fujii,K., Kladwang,W., Das,R. and Barna,M. (2014) RNA regulons in Hox 5′ UTRs confer ribosome specificity to gene regulation. *Nature,* **517**, 33–38.

111. Genuth,N.R., Shi,Z., Kunimoto,K., Hung,V., Xu,A.F., Kerr,C.H., Tiu,G.C., Oses-Prieto,J.A., Salomon-Shulman,R.E.A., Axelrod,J.D., *et al.* (2022) A stem cell roadmap of ribosome heterogeneity reveals a function for RPL10A in mesoderm production. *Nature Commun,* **13**, 5491.

112. Kearse,M.G., Chen,A.S. and Ware,V.C. (2011) Expression of ribosomal protein L22e family members in Drosophila melanogaster: rpL22-like is differentially expressed and alternatively spliced. *Nucleic Acids Res*, **39**, 2701–2716.

113. Mills,E.W. and Green,R. (2017) Ribosomopathies: There's strength in numbers. *Science*, **358**, 6363.

114. Doherty,L., Sheen,M.R., Vlachos,A., Choesmel,V., O'Donohue,M.F., Clinton,C., Schneider,H.E., Sieff,C.A., Newburger,P.E., Ball,S.E., *et al.* (2010) Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *Am J Hum Genet*, **86**, 222–228.

115. Li,J., Su,Y., Chen,L., Lin,Y. and Ru,K. (2023) Identification of novel mutations in patients with Diamond-Blackfan anemia and literature review of RPS10 and RPS26 mutations. *Int J Lab Hematol*, **45**, 766–773.

116. Piantanida,N., La Vecchia,M., Sculco,M., Talmon,M., Palattella,G., Kurita,R., Nakamura,Y., Ronchi,A.E., Dianzani,I., Ellis,S.R., *et al.* (2022) Deficiency of ribosomal protein S26, which is mutated in a subset of patients with Diamond Blackfan anemia, impairs erythroid differentiation. *Front Genet*, **13**, 3573.

117. Yamada,S.B., Gendron,T.F., Niccoli,T., Genuth,N.R., Grosely,R., Shi,Y., Glaria,I., Kramer,N.J., Nakayama,L., Fang,S., *et al.* (2019) RPS25 is required for efficient RAN

translation of C9orf72 and other neurodegenerative disease-associated nucleotide repeats. *Nat Neurosci*, **22**, 1383–1388.

118. Pisarev,A. V., Kolupaeva,V.G., Yusupov,M.M., Hellen,C.U.T. and Pestova,T. V. (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO Journal*, **27**, 1609–1621.

119. Havkin-Solomon,T., Fraticelli,D., Bahat,A., Hayat,D., Reuven,N., Shaul,Y. and Dikstein,R. (2023) Translation regulation of specific mRNAs by RPS26 C-terminal RNA-binding tail integrates energy metabolism and AMPK-mTOR signaling. *Nucleic Acids Res*, **51**, 4415–4428.

120. Varnavides,G., Madern,M., Anrather,D., Hartl,N., Reiter,W. and Hartl,M. (2022) In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. *J Proteome Res*, **21**, 2397–2411.

121. Hughes,C.S., Moggridge,S., Müller,T., Sorensen,P.H., Morin,G.B. and Krijgsveld,J. (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc*, **14**, 68–85.

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

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

124. Shi,M., Zhang,H., Wu,X., He,Z., Wang,L., Yin,S., Tian,B., Li,G. and Cheng,H. (2017) ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo. *Nucleic Acids Res*, **45**, 9640–9653.

125. Hautbergue,G.M., Castelli,L.M., Ferraiuolo,L., Sanchez-Martinez,A., Cooper-Knock,J., Higginbottom,A., Lin,Y.H., Bauer,C.S., Dodd,J.E., Myszczynska,M.A., *et al.* (2017) SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nat Commun*, **8,** 16063.

126. Rosario,R., Stewart,H.L., Choudhury,N.R., Michlewski,G., Charlet-Berguerand,N. and Anderson,R.A. (2022) Evidence for a fragile X messenger ribonucleoprotein 1 (FMR1) mRNA gain-of-function toxicity mechanism contributing to the pathogenesis of fragile X-associated premature ovarian insufficiency. *FASEB J*, **36,** e22612.

127. Argaud,D., Boulanger,M.C., Chignon,A., Mkannez,G. and Mathieu,P. (2019) Enhancer-mediated enrichment of interacting JMJD3-DDX21 to ENPP2 locus prevents R-loop formation and promotes transcription. *Nucleic Acids Res*, **47**, 8424–8438.

128. Song,C., Hotz-Wagenblatt,A., Voit,R. and Grummt,I. (2017) SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev*, **31**, 1370–1381.

129. Sidorov,M.S., Auerbach,B.D. and Bear,M.F. (2013) Fragile X mental retardation protein and synaptic plasticity. *Mol Brain*, **6**, 15.

130. Jin,P., Duan,R., Qurashi,A., Qin,Y., Tian,D., Rosser,T.C., Liu,H., Feng,Y. and Warren,S.T. (2007) Pur α Binds to rCGG Repeats and Modulates Repeat-Mediated Neurodegeneration in a Drosophila Model of Fragile X Tremor/Ataxia Syndrome. *Neuron*, **55**, 556–564.

131. Xiao,W., Adhikari,S., Dahal,U., Chen,Y.S., Hao,Y.J., Sun,B.F., Sun,H.Y., Li,A., Ping,X.L., Lai,W.Y., *et al.* (2016) Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell*, **61**, 507–519.

132. Roundtree,I.A., Luo,G.Z., Zhang,Z., Wang,X., Zhou,T., Cui,Y., Sha,J., Huang,X., Guerrero,L., Xie,P., *et al.* (2017) YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife*, **6,** e31311.

133. Walter,B.L., Parsley,T.B., Ehrenfeld,E. and Semler,B.L. (2002) Distinct poly(rC) binding protein KH domain determinants for poliovirus translation initiation and viral RNA replication. *J Virol*, **76**, 12008–12022.

134. Gu,H., Yang,J., Zhang,J., Song,Y., Zhang,Y., Xu,P., Zhu,Y., Wang,L., Zhang,P., Li,L., *et al.* (2022) PCBP2 maintains antiviral signaling homeostasis by regulating cGAS enzymatic activity via antagonizing its condensation. *Nat Commun*, **13,** 1564.

135. You,F., Sun,H., Zhou,X., Sun,W., Liang,S., Zhai,Z. and Jiang,Z. (2009) PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4. *Nat Immunol*, **10**, 1300–1308.

136. Nishiyama,T., Yamamoto,H., Uchiumi,T. and Nakashima,N. (2007) Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. *Nucleic Acids Res*, **35**, 1514–1521.

137. Fuchs,G., Petrov,A.N., Marceau,C.D., Popov,L.M., Chen,J., OLeary,S.E., Wang,R., Carette,J.E., Sarnow,P. and Puglisi,J.D. (2015) Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proc Natl Acad Sci U S A*, **112**, 319–325.

138. Hertz,M.I., Landry,D.M., Willis,A.E., Luo,G. and Thompson,S.R. (2013) Ribosomal Protein S25 Dependency Reveals a Common Mechanism for Diverse Internal Ribosome Entry Sites and Ribosome Shunting. *Mol Cell Biol*, **33**, 1016–1026.

139. Garcin,E.D. (2019) GAPDH as a model non-canonical AU-rich RNA binding protein. *Semin Cell Dev Biol*, **86**, 162–173.

140. Argaud,D., Boulanger,M.C., Chignon,A., Mkannez,G. and Mathieu,P. (2019) Enhancer-mediated enrichment of interacting JMJD3-DDX21 to ENPP2 locus

prevents R-loop formation and promotes transcription. *Nucleic Acids Res*, **47**, 8424–8438.

141. Osinalde,N., Olea,M., Mitxelena,J., Aloria,K., Rodriguez,J.A., Fullaondo,A., Arizmendi,J.M. and Zubiaga,A.M. (2013) The Nuclear Protein ALY Binds to and Modulates the Activity of Transcription Factor E2F2. *Mol Cell Proteomics*, **12**, 1087-98.

142. Nagy,Z., Seneviratne,J.A., Kanikevich,M., Chang,W., Mayoh,C., Venkat,P., Du,Y., Jiang,C., Salib,A., Koach,J., *et al.* (2021) An ALYREF-MYCN coactivator complex drives neuroblastoma tumorigenesis through effects on USP3 and MYCN stability. *Nat Commun*, **12**, 1881.

143. Klec,C., Knutsen,E., Schwarzenbacher,D., Jonas,K., Pasculli,B., Heitzer,E., Rinner,B., Krajina,K., Prinz,F., Gottschalk,B., *et al.* (2022) ALYREF, a novel factor involved in breast carcinogenesis, acts through transcriptional and post-transcriptional mechanisms selectively regulating the short NEAT1 isoform. *Cell Mol Life Sci*, **79**, 391.

144. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, **26**, 1367–1372.

145. Tyanova,S., Temu,T., Sinitcyn,P., Carlson,A., Hein,M.Y., Geiger,T., Mann,M. and Cox,J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, **13**, 731–740.

146. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.

147. Raudvere,U., Kolberg,L., Kuzmin,I., Arak,T., Adler,P., Peterson,H. and Vilo,J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*, **47**, 191–198.

148. Hulsen,T., de Vlieg,J. and Alkema,W. (2008) BioVenn - A web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, **9**, 488.

6. CONCLUDING REMARKS

The primary aim of the PhD dissertation was to identify novel factors of CGG repeat expansion-related RAN translation. This goal was addressed by applying proteomic screening using two different sample preparation protocols prior MS analysis. Such approach resulted in the **identification of 67 proteins binding natively to toxic *FMR1* mRNA** with expanded CGG repeats, which were present in both MS analyses. These identified proteins serve as a valuable resource for understanding the interactome of mutant *FMR1* and provide a basis for future studies underlying the biology of this molecule. Verification of RAN translation regulatory properties based on siRNA silencing of ten selected candidates resulted in the **characterization of four novel RAN translation modifiers: RPS26, DHX15, DDX21, and ALYREF**. In addition, the fifth one, **RPS25 was shown to modulate CGG-related RAN translation**, although it was not initially identified in MS screening. Notably, depleting the majority of the aforementioned factors did not affect the mRNA encoding FMRpolyG and FMRP levels, indicating a specific regulation of translation associated with the FMRpolyG open reading frame.

The newly discovered role of RPS25 and RPS26 in regulating RAN translation presents an intriguing **perspective that 40S subunit composition affects the noncanonical protein synthesis of FMRpolyG.** This view sheds light on a previously unexplored mechanistic aspect of CGG repeat-related RAN translation. Furthermore, it has been shown that the chaperon protein TSR2, associated with RPS26, was involved in RAN translation regulation, possibly through its involvement in mediating the incorporation of RPS26 into the assembling 40S subunit. Given the current knowledge about specialized ribosomes responsible for translating specific pools of mRNAs, and previous reports indicating RPS26-dependent translation, a global analysis was conducted to investigate proteins sensitive to RPS26 insufficiency, revealing a **relatively small fraction of the total proteome altered by RPS26 depletion**. Moreover, a bioinformatic study examining the features of mRNA encoding affected proteins uncovered that **they posses GC-rich 5'UTRs, similar to the one observed in *FMR1*.**

Finally, it was demonstrated that **RPS26 depletion reduced toxicity caused by RAN translation in cellular model.** Given that current therapeutic strategies for fragile-X associated conditions primarily target symptoms and lack specific treatments, understanding the roles of RPS26, RPS25 DDX21, DHX15, and ALYREF in regulating RAN translation provides a foundation for future research that could yield targeted therapies aimed at alleviating symptoms and impeding the disease progression.