ADAM MICKIEWICZ
UNIVERSITY
POZNAŃ

The role of *cis*-regulatory elements in mutant mRNA of *FMR1* gene containing expanded CGG repeats in R-loop formation and regulation of noncanonical translation of pathogenic protein

Daria Brygida Niewiadomska

Doctoral thesis carried out in the Laboratory of Gene Therapy
at the Department of Gene Expression
in the Institute of Molecular Biology and Biotechnology
at Adam Mickiewicz University
supervised by **Prof. dr hab. Krzysztof Sobczak**

Poznań, 2023

UNIWERSYTET
IM. ADAMA MICKIEWICZA
W POZNANIU

Rola elementów regulatorowych *cis* w zmutowanym mRNA *FMR1* zawierającym ekspansję powtórzeń CGG w tworzeniu struktur typu R-loop i regulacji niekanonicznej translacji patogennego białka

Daria Brygida Niewiadomska

Praca doktorska wykonana w Laboratorium Terapii Genowej, w Zakładzie Ekspresji Genów, w Instytucie Biologii Molekularnej i Biotechnologii
na Uniwersytecie im. Adama Mickiewicza w Poznaniu
pod opieką **Prof. dra hab. Krzysztofa Sobczaka**

Poznań, 2023

PODZIĘKOWANIA

W tym miejscu pragnę podziękować wszystkim osobom, bez których ukończenie tej pracy nie byłoby możliwe.

Prof. Krzysztofowi Sobczakowi, za ogromne wsparcie na każdym etapie doktoratu, za dzielenie się swoją wiedzą i doświadczeniem, za wspaniałą opiekę mentorską. Za zaufanie którym mnie obdarzył oraz za nieskończone pokłady cierpliwości i życzliwości. A także za bycie po prostu świetnym Szefem, który rozumie, że małe dzieci często chorują.

Mojemu Mężowi, za bycie moją opoką. Za bezgraniczną wyrozumiałość, cierpliwość i empatię. Za nieustające wspieranie i podnoszenie mnie na duchu gdy wątpiłam w siebie i w sens doktoratu. A także za to, że jest cudownym Partnerem i Tatą, przy którym nie ma rzeczy niemożliwych.

Moim Rodzicom, za ukształtowanie mojego charakteru, który doprowadził mnie do tego miejsca, a także za nieocenioną pomoc w opiece nad Leosiem podczas spisywania poniższej pracy.

Koleżankom i Kolegom z Laboratorium Terapii Genowej, tym obecnym i byłym, za niezwykle przyjazną atmosferę w pracy, za wszystkie naukowe i te mniej naukowe dyskusje, za dzielenie się swoim doświadczeniem i pomysłami. Dziękuję za wszystkie przyjacielskie relacje, które wykraczają poza granice laboratorium.

Wszystkim **Pracownikom Instytutu Biologii Molekularnej i Biotechnologii** za tworzenie tak inspirującego i rozwijającego miejsca pracy.

Pani dr Karolinie Cerbin, Pani Iwonie Kanonik-Jędrzejak oraz **Pani Magdalenie Walczak** za ich nieocenioną pomoc w załatwianiu wszystkich formalności.

Bożej Opatrzności.

Patrykowi i Leosiowi

FUNDING

The work was supported by the following funding:

1. Passport to the future – Interdisciplinary doctoral studies at the Faculty of Biology, Adam Mickiewicz University POWR.03.02.00-00-I022/16 (financing of conferences, a one-month internship in Renate Hukema Lab at the University Medical Center Rotterdam – Erasmus MS, Department of Clinical Genetics; and incentive scholarship)
2. Faculty of Biology's "Dean's grant" GDWB-627-18/19 to Daria Niewiadomska
3. The Foundation for Polish Science grant (POIR.04.04.00-00-5C0C/17-00) to Krzysztof Sobczak
4. The Polish National Science Centre grant (UMO-2020/38/A/NZ3/00498) to Krzysztof Sobczak

ACADEMIC ACHIEVEMENTS

During the implementation of the doctoral thesis, I was a co-author of the following publications:

Articles related to the subject of the dissertation:

Derbis M, Kul E, **Niewiadomska D**, Sekrecki M, Piasecka A, Taylor K, Hukema RK, Stork O, Sobczak K. "*Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring expanded CGG repeats*", ***Nature Communications***, 12, 1265 (2021), DOI: 10.1038/s41467-021-21021-w.

Articles related to topics other than dissertation:

Konieczny P, Mukherjee S, Stepniak-Konieczna E, Taylor K, **Niewiadomska D**, Piasecka A, Walczak A, Baud A, Dohno C, Nakatani K, Sobczak K. "*Cyclic mismatch binding ligands interact with disease-associated CGG trinucleotide repeats in RNA and suppress their translation*", ***Nucleic Acids Research***, 49 (16), 9479–9495 (2021), DOI: 10.1093/nar/gkab669.

Kajdasz A, **Niewiadomska D**, Sekrecki M, Sobczak K. "*Distribution of alternative untranslated regions within the mRNA of the CELF1 splicing factor affects its expression*", ***Scientific Reports*** 12, 190 (2022), DOI: 10.1038/s41598-021-03901-9.

ABBREVIATIONS

40S	eukaryotic small ribosomal subunit 40S
5-azadC	5-aza-2'-deoxycytidine
5'UTR	5' untranslated region
60S	eukaryotic large ribosomal subunit 60S
80S	eukaryotic 80S ribosome
aa	amino acid
ACG (+0)	ACG near-cognate codon in +0 frame
ACG (+1)	ACG near-cognate codon in +1 frame
ASFMR1	AntiSense transcript at the *FMR1* locus
ASFMRpolyA *ASFMR1*	polyalanine-containing RAN protein translated from
ASFMRpolyP *ASFMR1*	polyproline-containing RAN protein translated from
ASFMRpolyR *ASFMR1*	polyarginine-containing RAN protein translated from
CELF1	CUGBP Elav-Like Family Member 1
CGG_{dut}	CGG knock-in mouse model from the Willemsen lab in the Netherlands, also called "Dutch" mouse
CGGexp	expansion of CGG repeats/ expanded CGG repeats
CGG_{nih}	CGG knock-in mouse model from the National Institutes of Health, also called "NIH" mouse
CGIs	CpG islands; cytosine and guanine nucleotides with the "p" representing the linking phosphate
CRISPR-Cas9	Clustered Regularly Interspaced Palindromic Repeats-CRISPR associated protein 9
DDR	DNA damage response
DGCR8	DiGeorge syndrome critical region 8
DNMTs	DNA methyltransferases
DRIP-seq	DNA/RNA immunoprecipitation followed by high-throughput sequencing
DROSHA	Drosha ribonuclease type 3
eIF4A	eukaryotic initiation factor 4A
eIF4E	eukaryotic initiation factor 4E

eIF	eukaryotic initiation factor
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
FMRpolyG	in this work – RAN translation product containing the *FMR1* N-terminal sequence of 12 aa followed by polyglycine stretch (composed either by 16 or 85 polyglycine aa) and the C-terminal sequence (40 aa) containing *FMR1* ex1 region in-frame with Nluc-FLAG protein
FRAXA	fra(X)(q27.3) A; folate-sensitive fragile site at the *FMR1*
FRDA	friedreich ataxia
FXAND	fragile X-associated neuropsychiatric disorders
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
GUG (+1)	GUG near-cognate codon in +1 frame
H2AFX	H2A histone family member X
H3K4me3 protein	tri-methylation at the 4th lysine residue of the histone H3
H3K9me2 protein	di-methylation at the 9th lysine residue of the histone H3
hnRNP A2/B1	heterogeneous nuclear ribonucleoprotein A2/B1
ID	intellectual disability
IRES	internal ribosome entry site
KI	knock-in mouse model
LAP2β	Lamina-associated polypeptide 2 beta
LNA	locked nucleic acid
m7G	5’ methyl-7-guanosine cap modification
Met-tRNA	initiator methionyl-tRNA
mRNA	messenger RNA

MSH2	MutS Homolog 2; crucial protein in the human MutSα mismatch repair complex
MMR	mismatch repair pathway
NES	nuclear export sequence
NLS	nuclear localization sequence
nt	nucleotide
ORF	open reading frame
PIC	preinitiation complex
PM	premutation of *FMR1* gene characterized by 55-200 CGG repeats
Pol II	RNA polymerase II
polyA signal	polyadenylation signal
pri-miRNA	primary microRNA
Pur α	purine-rich binding protein α
RAN translation	repeat-associated non-AUG initiated translation
RBP	RNA binding protein
rCGGexp	mRNA containing expanded CGG repeats
RED	repeat expansion disorder
Ribosomal P-site	peptidyl, second binding site for tRNA in the ribosome
R-loop	RNA:DNA hybrid
Sam68	Src-Associated substrate during mitosis of 68-kDa
ssDNA	single stranded DNA
STR	short/simple tandem repeat
TIS	translation initiation site
tRNA	transfer RNA
TSS	transcription start site
UFM	unmethylated full mutation
uORF	upstream open reading frame
UPS	ubiquitin proteasome system
Xq27.3	band 27.3 on the long arm of the X chromosome
γH2AX	the phosphorylated form of H2AX

ABSTRACT

The expansion of short tandem repeats located in either coding or non-coding regions of different genes underlies the pathogenesis of diverse human neurological diseases. The expansion of an unstable CGG repeat sequence within the 5' untranslated region (5'UTR) of *FMR1* has been implicated in the pathogenesis of multiple fragile X-linked syndromes.

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset neurodegenerative disorder caused by the expansion of 55-200 CGG repeat, named as premutation of *FMR1*. The main symptoms of FXTAS are intention tremor, ataxia, and dementia. At the molecular level the disease is caused by the toxic *FMR1* mRNA that folds into a thermodynamically stable secondary structure at the region of excessively expanded CGG repeats (rCGGexp). Due to the sequestration by rCGGexp of many RNA-binding proteins the metabolism of RNA is highly disturbed and toxic inclusions containing this RNA are formed. The toxic mRNA with expanded CGG repeats is the template for non-canonical translation which results in the synthesis of toxic protein containing polyglycine tract (FMRpolyG) from the same mRNA from which the natural product of the *FMR1*, FMRP protein, is synthesized. Due to the strong aggregation properties, the protein is known to create intranuclear aggregates that lead to the disturbance of neurons and their death. Finally, the co-transcriptionally formed RNA:DNA hybrids, called R-loop structures, in the region of CGGexp are considered as another FXTAS pathomechanism leading to alterations in the transcription and driving DNA damage *via* cellular stress.

On the contrary, the fragile X syndrome (FXS) is associated with the expansion of more than 200 CGGs within the 5'UTR of the *FMR1* gene (name as full mutation) and is a neurodevelopmental disease, the most common form of inherited intellectual disability. The FXS patients are characterized by full mutation of *FMR1* which usually leads to the epigenetic silencing of the *FMR1* gene and consequently loss of FMRP protein. Although the silencing of *FMR1* is a complex process it has been shown that it is, at least partially, dependent on R-loops formation within CGGexp region.

The first part of the project aimed to establish the role of R-loops in the pathogenesis of FXTAS and FXS disorders. After confirming that R-loops are formed within 5'UTR of the *FMR1* in the premutation range of CGG repeats the transcription efficiency regulated by the presence of R-loops was verified both *in vitro* and *in cellula*. Then, the contribution of short chemically modified antisense oligonucleotides (ASO-CCG) directly targeting

CGGexp, involved in R-loop formation, on this structure stability and therefore on the *FMR1* transcription efficiency was tested. Finally, since *FMR1* full mutation in FXS leads to the silencing of *FMR1* transcription the long treatment with ASO-CCG was utilized to verify whether the reactivation of the *FMR1* transcription leading to the FMRP translation is possible. Results obtained in this part confirmed that R-loops forming within *FMR1* 5'UTR in FXTAS conditions have a negative effect on the *FMR1* transcription which can be partially abolished by ASO-CCG. However, according to FXS, ASO-CCG treatment did not reactivate the *FMR1* transcription from FXS cells which were characterized by full *FMR1* silencing. On the contrary, ASO-CCG treatment of FXS cells which possessed partially active *FMR1* locus resulted in the increased transcription rate of *FMR1* and its enhanced mRNA pool in the cytoplasm. Nevertheless, elevated *FMR1* mRNA level did not translate into increased FMRP level.

The second part of the project was concerning the *cis*-regulatory elements within *FMR1* 5'UTR and their involvement in the regulation of initiation of toxic FMRpolyG synthesis from near-cognate ACG or GUG start codons. In line with that, among others, the effect of different nucleotide sequence context in the vicinity of one of the near-cognate start codons on the FMRpolyG translation was established. Also, the effect of stable secondary RNA structure formed by the sequence located downstream of ACG near-cognate start codon on the translation initiation, and how different size of CGG repeats would affect the initiation of FMRpolyG synthesis were validated. Obtained data showed that both sequence context as well as stable secondary structure within mRNA have enormous effect on the initiation of FMRpolyG translation suggesting that this process is potentially regulated by many *cis*- and *trans*-factors targeting various regions/elements within the *FMR1* mRNA sequence.

Key words: *FMR1*, CGG repeat expansion, R-loop, fragile X syndromes, noncanonical translation

STRESZCZENIE

Ekspansja krótkich powtórzeń tandemowych zlokalizowanych w części kodującej lub niekodującej genu leży u podstaw patogenezy wielu chorób neurologicznych występujących u ludzi. Ekspansja niestabilnych powtórzeń CGG w regionie 5' niepodlegającym translacji (5'UTR) genu *FMR1* w zależności od wielkości ekspansji została powiązana z patogenezą wielu chorób związanych z łamliwym chromosomem X.

Zespół drżenia i ataksji związany z łamliwym chromosomem X (FXTAS) jest chorobą neurodegeneracyjną wieku późnego spowodowaną przez ograniczoną ekspansję (55-200) powtórzeń CGG, zwaną premutacją. Do głównych objawów FXTAS należą drżenie zamiarowe, ataksja chodu i demencja. Na poziomie molekularnym choroba jest spowodowana występowaniem toksycznej cząsteczki mRNA *FMR1*, która tworzy termodynamicznie stabilną strukturę drugorzędową w regionie nadmiernie wydłużonych powtórzeń CGG (rCGGexp). Zmutowana cząsteczka rCGGexp uczestniczy w trzech patogennych procesach. Po pierwsze cząsteczka ta sekwestruje wiele białek wiążących się z RNA, co prowadzi do tworzenia patogennych inkluzji zawierających zmutowane RNA, w wyniku czego metabolizm setek innych cząsteczek RNA jest istotnie zakłócony. Po drugie toksyczna cząsteczka rCGGexp stanowi matrycę dla translacji inicjowanej z niekanonicznego kodonu start, która skutkuje syntezą toksycznego białka zawierającego trakt poliglicynowy (FMRpolyG) z tej samej cząsteczki mRNA, z której jest syntetyzowany naturalny produkt genu *FMR1* – białko FMRP. Ze względu na silne właściwości agregujące, powstające białko tworzy wewnątrzjądrowe agregaty, które prowadzą do zaburzeń funkcji neuronów i ich obumierania. Po trzecie podczas transkrypcji *FMR1,* w obrębie nadmiernie wydłużonych powtórzeń CGG, tworzone są hybrydy RNA:DNA nazywane strukturami typu R-loop. Tworzenie tych struktur prowadzi do zaburzeń transkrypcji i indukuje uszkodzenia DNA wywołujące stan stresu komórkowego.

W przeciwieństwie do FXTAS, klasyczny zespół łamliwego chromosomu X (FXS) jest związany z ekspansją powyżej 200 powtórzeń CGG, zwaną pełną mutacją, i jest chorobą neurorozwojową stanowiącą najbardziej powszechną formę wrodzonej niepełnosprawności intelektualnej. Pełna mutacja prowadzi przeważnie do epigenetycznego wyciszenia genu *FMR1* i w konsekwencji do braku syntezy białka FMRP w komórkach pacjentów z FXS. Pomimo że wyciszenie *FMR1* jest procesem

złożonym wykazano, że przynajmniej częściowo proces ten zależy od tworzenia struktur typu R-loop w obrębie nadmiernie wydłużonych powtórzeń CGG.

Celem pierwszej części projektu było ustalenie roli struktur typu R-loop w procesie patogenezy obu chorób – FXTAS i FXS. Po potwierdzeniu tworzenia struktur typu R-loop w regionie 5'UTR genu *FMR1* z premutacją powtórzeń CGG wykazano wpływ tych struktur na efektywność transkrypcji *FMR1* zarówno w warunkach *in vitro,* jak i *in cellula.* Następnie zbadano wpływ krótkich, chemicznie modyfikowanych antysensowych oligonukleotydów wiążących się bezpośrednio z nadmiernie wydłużonymi powtórzeniami CGG (ASO-CCG) na stabilność tworzonych struktur typu R-loop i w konsekwencji na efektywność transkrypcji *FMR1*. Jako że pełna mutacja genu *FMR1* prowadzi do wyciszenia jego transkrypcji, zostało przeprowadzone długotrwałe traktowanie komórek wyprowadzonych od pacjentów z FXS cząsteczkami ASO-CCG w celu zbadania, czy jest możliwa reaktywacja transkrypcji *FMR1,* w wyniku której będzie syntetyzowane białko FMRP. Wyniki uzyskane w tej części pracy potwierdziły, że struktury R-loop tworzone w obrębie regionu 5'UTR genu *FMR1* w warunkach FXTAS mają negatywny wpływ na transkrypcję *FMR1,* co może być częściowo osłabione poprzez zastosowanie ASO-CCG. Jednakże w odniesieniu do warunków FXS, zastosowanie ASO-CCG nie prowadziło do reaktywacji transkrypcji *FMR1* w komórkach wyprowadzonych od pacjenta FXS, które charakteryzowały się całkowitym wyciszeniem *FMR1*. Natomiast traktowanie cząsteczkami ASO-CCG komórek pacjenta z FXS, które posiadały częściowo aktywne locus *FMR1,* prowadziło do wzmożenia transkrypcji *FMR1* oraz zwiększenia puli mRNA *FMR1* w cytoplazmie, co jednak nie spowodowało zwiększenia poziomu białka FMRP w tych komórkach.

Druga część projektu dotyczyła roli elementów regulatorowych *cis* zlokalizowanych w regionie 5'UTR *FMR1* w regulacji translacji toksycznego białka FMRpolyG inicjowanej z kodonów ACG lub GUG. Zbadano między innymi wpływ kontekstu sekwencji nukleotydowej w pobliżu jednego z kodonów start na translację białka FMRpolyG. Dodatkowo określono wpływ stabilnej struktury drugorzędowej RNA tworzonej przez sekwencje zlokalizowane poniżej kodonu start ACG oraz różnej długości powtórzeń CGG na efektywność translacji FMRpolyG. Uzyskane wyniki wykazały, że zarówno kontekst sekwencyjny, jak i stabilna struktura drugorzędowa tworzona w obrębie mRNA *FMR1* mają ogromny wpływ na inicjację translacji białka FMRpolyG, co sugeruje, że

proces ten może być również regulowany przez wiele czynników *trans* wiążących się z regionami *cis* w obrębie sekwencji mRNA *FMR1*.

Słowa kluczowe: *FMR1*, ekspansja powtórzeń CGG, struktura typu R-loop, łamliwy chromosom X, translacja niekanoniczna

CONTENTS

1. INTRODUCTION

Microsatellite repeats, interchangeably known as simple or short tandem repeats (**STRs**), constitute short (1-8 nt) repeat units that are contiguously repeated. STRs are distributed throughout the human genome and account for roughly 3% of the entire genetic information[1]. Therefore, microsatellite repeats, highly polymorphic in terms of the number of repeats in individual locus, were found in hundreds of genes. Expansion of STRs located in either coding or non-coding regions of different genes underlies the pathogenesis of diverse human neurological diseases known as repeat expansion disorders (**REDs**). This class of diseases contains over 50 inherited neurological disorders including neurodevelopmental, neuromuscular and neurodegenerative genetic conditions[2,3]. One characteristic feature of REDs is genetic anticipation. This term describes the phenomenon in which, from generation to generation, the severity of the disease increases and the age of onset decreases. This mechanism is also positively correlated with the number of repeats. It is worth mentioning that the threshold at which the repeat expansions become symptomatic varies between diseases and depends on the expansion type and its localization within the gene. An interesting example of the importance of repeat size are CGG repeats within the fragile X messenger ribonucleoprotein 1 gene (***FMR1)*** encoding for Fragile X messenger ribonucleoprotein 1 (**FMRP**) which is crucial for the translation of dendritic mRNAs in response to synaptic activation.

1.1 Dynamic mutations in fragile X messenger ribonucleoprotein 1 gene

The *FMR1* gene mapping on Xq27.3 loci of chromosome X consists of 17 exons spanning ~ 38 kb of genomic DNA and is inherited as a X-linked dominant trait. The expansion of an unstable CGG repeat sequence, known as dynamic mutation, within the 5' untranslated region (**5'UTR**) of *FMR1* has been implicated in the pathogenesis of the fragile X-linked syndromes. Term "fragile" is derived from the folate-sensitive fragile site at the FRAXA locus on the terminal end of the long arm of the X chromosome (Xq27.3). The fragile site was described first time for fragile X syndrome (FXS) by Lubs in 1969 and termed as "marker chromosome with secondary constriction"[4].

In the general population, the number of CGG repeats within the *FMR1* 5'UTR is highly variable, however, the majority of *FMR1* alleles have 29-30 repeats[5,6]. The expansions of CGG repeats outside of the normal range fall into two distinct categories: the premutation

(**PM**) with 55-200 repeats, and the full mutation (**FM**) with expansions greater than 200[7,8]. However, some studies distinguish also another category known as the "gray zone" (~45-54 CGG repeats) which consists of alleles that are not short enough to be classified as normal alleles, and are not long enough to belong to the PM category. Nonetheless, gray zone alleles were stated to be prone to repeat size instability upon transmission[8,9,10], and indeed it has been confirmed that they expand to full mutation within typically 2-3 generations[6,11]. Therefore, according to the American College of Medical Genetics and Genomics recommendations four main allelic forms of the *FMR1* gene can be distinguished: normal alleles, intermediate (gray zone), premutation, and full mutation[12]. Nevertheless, the boundaries between different categories are becoming more blurred with the increase in empirical data and research. Importantly, the CGG repeat sequence is usually interrupted by 1-3 non-CGG triplets, mostly AGG, in healthy individuals. However, loss of the interruptions within the expanded CGG repeats was connected with more severe disease phenotype and greater repeat instability.

The *FMR1* gene encodes the fragile X messenger ribonucleoprotein 1 (FMRP) which is a classic RNA binding protein that is necessary for normal brain development and synaptic plasticity[13]. It has been revealed that *FMR1* is broadly expressed in most tissues. The localization of the FMRP is mainly cytoplasmic, however, due to the presence of nuclear localization and nuclear export sequences (NLS, and NES, respectively) the protein can shuttle between the nucleus and cytoplasm. This protein is an important mRNA transport protein but also translational repressor. Emerging studies revealed that FMRP binds to target mRNAs encoding proteins involved in synaptic plasticity and inhibits their translation[14,15]. Thus, FMRP is an essential regulator in the synapses and the dendrites development, and the loss of this protein has been linked with the fragile X syndrome (FXS).

1.1.1. Premutation driven fragile X-associated disorders and fragile X syndrome

The premutation carriers are at risk of developing multiple fragile X-associated disorders including fragile X-associated tremor/ataxia syndrome (FXTAS)[16,17], fragile X-associated primary ovarian insufficiency (FXPOI)[18], and fragile X-associated neuropsychiatric disorders (FXAND)[19,20]. On the other hand, the CGG expansion to more than 200 is related to the FXS development.

Although FXTAS and FXS are developed as the consequence of CGG expansions (**CGGexp**) within the same locus the molecular mechanisms by which these neurological conditions arise are largely unrelated. The FXS base on the loss-of-function of the *FMR1* gene while in the premutation-associated diseases, there are RNA and protein gain-of-function pathomechanisms.

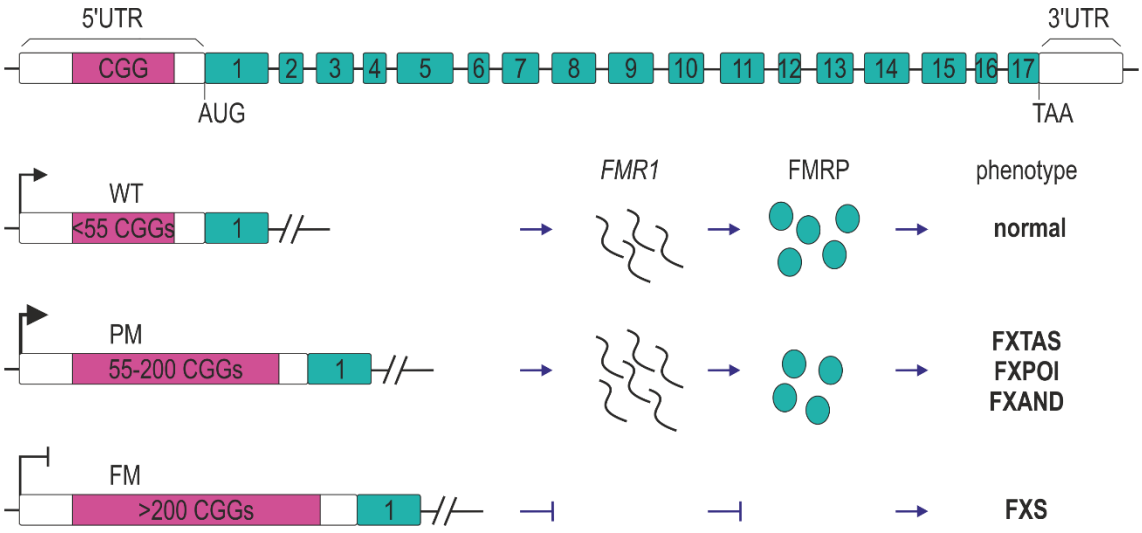


Figure 1. Scheme of the fragile X messenger ribonucleoprotein 1 gene (FMR1) structure and its various allelic forms implicated in human diseases. *Based on the CGG expansion the following classes of alleles are shown, with their transcriptional activity indicated by the arrow: normal (WT), premutation (PM) with an increased transcription (bold arrow), and slight decrease of translation associated with the fragile X-associated disorders, and full mutation (FM) leading to silencing of FMR1 transcript and fragile X messenger ribonucleoprotein 1 (FMRP) and consequently to fragile X syndrome (FXS). Based on the figure from*[21].

1.1.1.1. Fragile X-associated tremor/ataxia syndrome.

Fragile X-associated tremor/ataxia syndrome (**FXTAS**) is an X-linked, late-onset inherited neurodegenerative disorder caused by the expansion of 55-200 CGG repeats (premutation; **PM**) in the 5' UTR of the *FMR1* gene encoding FMRP. The main clinical features of FXTAS are intention tremor and cerebellar gait ataxia. However, these symptoms usually occur with several comorbidities such as short-term memory loss, cognitive decline, parkinsonism, dementia and autism-spectrum phenotypes[17]. On the molecular level, the FXTAS is characterized by the presence of ubiquitin-positive intranuclear aggregates which constitute one of the hallmarks of FXTAS neuropathology[22].

FXTAS usually affects males over 50 years of age while females are less commonly affected and present less severe symptoms, probably due to the protective effect of the second X-chromosome. The prevalence of PM carriers in the general population varies significantly between distinct studies and populations with estimates of 1:151[23], 1:178[24], 1:201[25], and 1:209[26] for females, and 1:400[24,26], and 1:468[23] for males. Nevertheless, according to the literature and data collected by the National Fragile X Foundation, considering premutation carriers, due to incomplete penetrance it is estimated that about 40% of males older than 50 years and 8 – 16% of women older than 40 years develop FXTAS[27,28].

FXTAS patients present an increased level of *FMR1* mRNA which correlates positively with the increasing number of CGG repeats[29,30,31,32,33]. Interestingly, the elevated *FMR1* mRNA level is inconsistent with the level of FMRP which is unchanged or slightly reduced in PM carriers. Of note, the role of marginally reduced FMRP level in PM driven disorders was initially rejected. However, currently, it is suggested that the decline in FMRP level observed in PM carriers might explain some neurodevelopmental problems observed in children.

It has to be mentioned that the increase in the number of CGG repeats is also correlated with the gradual impairment of *FMR1* translation efficiency and reduced level of FMRP[34,30,35,36]. The current state of knowledge claims that the reduced translational efficiency of FMRP is partially compensated through the increased transcriptional activity of *FMR1*[33]. It was established, that the increased *FMR1* mRNA does not result from the elevated message stability since the rate of decay of the *FMR1* mRNA was not altered in the PM carriers after actinomycin D treatment which is an inhibitor of RNA Polymerase II (**Pol II**)[33]. Tassone and colleagues[37] presented that the growth in the level of *FMR1* mRNA in PM carriers follows primarily from the increased transcription rate. This effect can be also supported by the phenomenon characteristic for expanded alleles which present a shift in the usage of transcription start sites (**TSSs**) to upstream sites as the CGG repeats length increase[38,39].

1.1.1.2. Fragile X syndrome

Fragile X syndrome (**FXS**) is the most common form of inherited intellectual disability (ID) and the leading form of the monogenic cause of autism. As was already mentioned, the FXS is a disease associated with the expansion of CGG repeats (>200) within the

5'UTR of the *FMR1* gene. The patients are characterized by full mutation (**FM**) of *FMR1* which is usually accompanied by hypermethylation of the promoter region of the *FMR1* gene, including CpG dinucleotides (CpG islands) and CGG repeats, and the epigenetic silencing of the *FMR1*[40]. Consequently, the loss of FMRP is observed.

The prevalence of FXS, based on data presented by the National Fragile X Foundation, is estimated to be 1 in 7,000 males and 1 in 11,000 females. The behavioral phenotype of FXS is characterized by hyperactivity, emotional problems including anxiety, and difficulties in maintaining eye contact. The cognitive deficit can range from mild learning disabilities to severe mental retardation[41]. From the molecular point of view FXS results mainly from the full mutation of *FMR1* leading to the gene loss-of-function. Nevertheless, it should be mentioned that 1-2% of FXS cases are developed as a consequence of other *FMR1* mutations leading to the *FMR1* loss-of-function[42].

The genetic instability of FM alleles in early embryogenesis leads to somatic mosaicism meaning that cells of the same individual carry different sizes of CGG repeats within *FMR1* which can be characterized by different status of methylation[43,44]. Interestingly, the phenotype can be modulated *via* this premutation-full mutation mosaicism. Due to the presence of this phenomenon in some cells the FMRP is translated thus mosaicism-positive patients are usually characterized by milder intellectual disability[45]. On the other hand, it has been proven that the severity of mental retardation is positively correlated with the degree of cytosine methylation within the *FMR1* 5'UTR[46,47,40,48]. In line with that, patients harboring FM with partially or completely unmethylated alleles (unmethylated full mutation; UFM) characterized by less severe symptoms have been described. These data also support the statement that expanded CGG repeats alone are unable to repress transcription of the *FMR1* and that DNA methylation is required. Indeed, it has been presented that in the FXS patients containing unmethylated full mutation alleles the increased level of *FMR1* mRNA resulting from still active transcription was observed [40,48,33]. Therefore the suggestion was made that the loss of FMRP observed in FXS patients is not always a consequence of the silenced transcription[30]. Finally, the correlation between the severity of the disease and the number of interruptions, which usually constitute the single substitution within the CGG unit resulting in an AGG or TGG triplet, has been confirmed. Consequently, even a moderate size of pure, uninterrupted CGG repeat tract can have a more deleterious effect than longer, but containing interruptions, tract of CGGs.

Therefore, due to mentioned issues, the mosaic background reflecting different *FMR1* and FMRP expression may add complexity to the FXS phenotype and widen the spectrum of altered mechanisms in the cell.

1.2. *Molecular basis of premutation-driven fragile X-associated disorders*

1.2.1. Antisense transcript at the *FMR1* locus

Emerging studies revealed that antisense transcription is a common feature of genes that are transcribed from different loci of repeat expansion diseases[49,50,51,52]. In 2007 Ladd and colleagues[53] confirmed that next to the regular sense transcription of *FMR1*, an antisense transcription occurs. The **A**nti**S**ense transcript at the ***FMR1*** locus (***ASFMR1***) is overlapping the region of CGG repeats of the *FMR1* gene in the antisense orientation. Antisense *FMR4* and *FMR6*[54,55] are also produced, albeit they do not contain repeat elements. Interestingly, these long noncoding RNAs present similarly to *FMR1* mRNA increased expression in PM carriers and are not detected in FXS patients with full mutation suggesting that they may contribute to the pathomechanisms of fragile X-associated disorders[56]. It has been proven that *ASFMR1* is transported to the cytoplasm. As similar mechanisms have been reported for other diseases the protein products of *ASFMR1* may be involved in the pathogenesis of fragile X-linked syndromes. Indeed, it has been confirmed that ubiquitinated neuronal inclusions in FXTAS patients are positive for anti-ASFMRpolyP and anti-ASFMRpolyA staining[57]- the protein products of *ASFMR1*.

1.2.2. RNA gain-of-function

The expansion of CGG repeats in the PM range within *FMR1* mRNA is directly correlated with RNA gain-of-function toxicity. In this model the excessively expanded CGG repeats within *FMR1* form the thermodynamically stable secondary structure which sequesters many RNA binding proteins (**RBPs**) impairing their physiological functions involved in the processes such as alternative splicing, regulation of transcription, microRNA biogenesis, mRNA maturation, transport, stability, and translation. It is important to highlight that the toxicity in PM carriers arises due to the limited expansion of CGG repeats not the augmented level of *FMR1* expression, as overexpression of *FMR1* with normal size of CGG repeats did not trigger neuronal death[58].

1.2.2.1. RNA foci

According to the RNA gain-of-function mechanism it has been presented that the CGGexp containing mRNA (**rCGGexp**) and the sequestered proteins co-aggregate and form intranuclear inclusions (RNA *foci*). To date, due to mass spectrometric analyses combined with immunohistochemical analyses, many RBPs have been confirmed to bind to hairpin structures formed by rCGGexp. Since the composition of RNA *foci* differs between distinct studies, mainly due to the technical limitations and troubles connected with the proteins' isolation from inclusions, no dominant protein was found. Nevertheless, the list of proteins found in the RNA *foci* includes but is not limited to heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP A2/B1), CUGBP Elav-Like Family Member 1 (CELF1)[59], purine-rich binding protein α (Pur α)[60], Src-Associated substrate during mitosis of 68-kDa (Sam68)[61], DiGeorge syndrome critical region 8 (DGCR8), and its partner – Drosha ribonuclease type 3 (DROSHA)[62].

Recently, two splicing factors hnRNP A2/B1 and CELF1 were reported to bind to expanded rCGG repeats[59]. However, as it was presented by Sofola and colleagues[59] the hnRNP A2/B1 is involved in the direct interaction with rCGGexp and recruits CELF1 in *trans*, potentially sequestering this protein from their cellular function. Nevertheless, the overexpression of both proteins led to the suppression of CGG-mediated toxicity in the *Drosophila* model which suggests that these proteins play a role in the neuropathology of FXTAS[59,60]. Similarly to hnRNP A2/B1, it was shown that Pur α binds directly to rCGGexp and the overexpression of Pur α in *Drosophila* led to the suppression of neurodegeneration phenotype[60].

The DGCR8 and DROSHA[62] are major components of a microprocessor, which is involved in the processing of primary precursors of microRNA (pri-miRNA). The sequestration on rCGGexp leads to the impairment of the activity of this enzymatic complex. As a consequence, the levels of mature microRNAs are decreased in neuronal cells which directly results in the neurodegeneration[62]. Sam68 is a nuclear RNA-binding protein involved in alternative splicing regulation[63,64] which has also been reported to be sequestered on CGG repeats in FXTAS[61]. Therefore, the Sam68-dependent splicing is impaired in FXTAS. Interestingly, the neuronal cell death induced by the expression of CGG repeats in cultured mouse cortical neurons was rescued by the overexpression of DGCR8 but not Sam68. This result suggests that the loss of DGCR8 function plays a

crucial role in the neuropathology observed in FXTAS and that Sam68 does not bind CGG repeats in a direct manner[61,62].

Importantly, the *FMR1* mRNA, but not the FMRP protein, was found in the inclusions[65,66]. This observation together with the fact that the number of inclusions increases as the CGG repeats expand is in line with the statement that the RNA gain-of-function mechanism plays an important role in FXTAS pathogenesis. Of note, Sellier and co-workers[67] presented that depending on the surroundings of CGG repeats the efficiency of RNA *foci* formation is variable. The RNA Fluorescence in situ hybridization (FISH) is a technique utilizing fluorescently labeled nucleic acid probes to detect RNA within cells. The RNA FISH against CGG repeats indicated that rCGGexp embedded in the 5'UTR of *FMR1* formed fewer RNA *foci* than expanded CGG repeats without the surrounding *FMR1* sequence. The results from RT-PCR performed on nuclear and cytoplasmic fractions established that the majority of the rCGGexp embedded in the *FMR1* 5'UTR were exported to the cytoplasm (and potentially underwent the translation), while the rCGGexp without *FMR1* context largely retained in the nucleus.

In conclusion, both, the RBPs sequestration and the formation of intranuclear inclusions lead to the altered RNA processing and constitute one of the major pathomechanisms in FXTAS and other PM-driven conditions.

1.2.3. Repeat associated non-AUG initiated (RAN) translation

Although one of the consequences of RNA toxicity in PM carriers is the formation of RNA *foci* it is unlikely that the RNA gain-of-function alone would be sufficient to create such large (2-5 μm) ubiquitin-positive intranuclear inclusions in the brains of FXTAS patients. What is more, these inclusions contain also proteins that are not involved in the rCGGexp binding and are reminiscent of the aggregates found in protein-mediated neurodegenerative disorders[22,68]. Therefore, recently, an additional mechanism has been proposed to highlight the pathology of FXTAS which is the CGGexp-associated non-AUG initiated (**RAN**) translation[69,70]. The mechanism of RAN translation is based on the evidence that trinucleotide repeats can be translated into protein even if they do not reside in an AUG-initiated open reading frame (**ORF**). According to the *FMR1*, the CGG repeats are embedded within the *FMR1* 5'UTR 69 nt upstream of the AUG start codon of the FMRP ORF, the main protein product of the *FMR1* mRNA. RAN translation of CGG repeats can occur in all possible ORFs of a transcript generating multiple protein products

from a single repeat tract. Importantly, proteins synthesized *via* RAN translation have toxic properties as they are prone to aggregate and create nuclear or cytoplasmic inclusions which may sequester other proteins and abolish their function.

This non-canonical mode of translation, however, has something in common with canonical translation. Similarly, as in AUG-initiated translation, the initiation of RAN translation occurs *via* a cap-dependent scanning model. Importantly, the RAN translation initiates at near-cognate start codons, which differs from the AUG codon by one nucleotide, and are located upstream or within expanded repeats[70,71,72]. Taking into consideration the evolutionary pressure to use the AUG start codons it would be expected that initiation from non-AUG codons results from the aberrations performed by the translation machinery. On the other hand, there is a great number of proteins playing important functions that are translated solely from the non-AUG start codons[73,74]. Additionally, emerging studies presented the potential regulatory role of proteins translated from near-cognate start codons which usually constitute upstream ORFs (**uORFs**). However, to not impair the downstream main ORF, these codons are usually embedded within the weak Kozak sequence context which is a sequence of nucleotides surrounding the start codon strongly influencing the codon utilization by the ribosome (***see Methods 1.2.4.1 "RNA sequences and structures affect start codon utilization"***).

1.2.3.1. RAN translation of expanded CGG repeats within FMR1

Even though the RAN translation of *FMR1* was first described a decade ago[69] the precise mechanism of action is not fully understood. Recently, it has been shown that RAN translation of *FMR1* mRNA may occur from both sense (CGG) and antisense (CCG) transcripts leading to the synthesis of homopolymeric cytotoxic proteins in potentially six different ORFs[75]. From the sense strand of *FMR1* mRNA the polyglycine- (GGC frame; called **FMRpolyG**), polyalanine- (GCG; **FMRpolyA**), and polyarginine-containing proteins (CGG; **FMRpolyR**) can be translated, while from the antisense strand the polyproline- (CCG; **ASFMRpolyP**), polyalanine- (GCC; **ASFMRpolyA**), and polyarginine- (CGC; **ASFMRpolyR**) containing proteins[67,69,75].

The open reading frame of FMRP initiated at the AUG codon is stated as a +0 frame, therefore (+1) reading frame with respect to the downstream initiation site of FMRP is producing FMRpolyG, while (+2) reading frame is producing FMRpolyA. Importantly, the FMRpolyR, which is in-frame with the AUG start codon for FMRP, can generate

FMRP protein with N-terminal extension (***Figure 26a***). Noteworthy, nearly all the RAN proteins translated from both sense (CGG) and the antisense (CCG) expanded repeats within *FMR1* 5'UTR have been reported to colocalize with ubiquitinated neuronal inclusions[76,67,75,77,57] pointing out that products of RAN translation contribute to the neuropathology of FXTAS. However, RAN proteins are not equally produced and abundant within aggregates. Thus, ASFMRpolyP- and ASFMRpolyA-positive staining was detected only in some studied aggregates[57] while ASFMRpolyR has not yet been detected in patients' samples. On the contrary, the FMRpolyG was found to be the most abundant RAN product within protein aggregates in FXTAS cases[75]. Interestingly, the (+2) reading frame encoding FMRpolyA is translated only at ~30% of the FMRpolyG efficiency[78], hence FMRpolyA is hard to detect[67]. Initiation in this reading frame is believed to occur at the CGG repeats itself since the introduction of stop codon upstream the CGG repeats did not affect its translation[70,69]. The ACG (+0) near-cognate start codon located 57 nt upstream of the CGGs was confirmed to be the translation initiation site (**TIS**) of FMRpolyR, however, the protein was easily detectable only in the reporter system when no CGG repeats were present[70]. In line with that, the FMRpolyR translation in (+0) reading frame has been reported to be highly reduced by the CGG repeats showing the direct negative correlation between the efficiency of FMRpolyR *in vitro* translation and the number of CGG repeats[70]. Moreover, the efficiency of FMRpolyR translation constitutes less than 5% of FMRpolyG translation efficiency with 25 CGGs and less than 1% with 100 CGGs[70,69]. Therefore, to my knowledge, along with ASFMRpolyR, FMRpolyR has so far not been detected in the pathological samples from FXTAS cases. Hence, although the RAN translation from *ASFMR1* has already been confirmed, the RAN products from sense transcript, especially FMRpolyG, are still much better explored.

1.2.3.1.1. RAN translation of FMRpolyG

RAN translation of FMRpolyG may be initiated at ACG (+1) or GUG (+1) near-cognate start codons located 32 nt and 8 nt upstream CGG repeats, respectively[67,69,70]. Nonetheless, the ACG (+1) near-cognate start codon has been proven to be the major TIS for FMRpolyG[67]. The initiation of canonical translation depends on the base-pairing between the anticodon of initiator Met-tRNA and the AUG start codon, although the RAN translation of *FMR1* initiates at ACG (+1) near-cognate start codon it has been proven that this codon is also decoded by the initiator Met-tRNA despite imperfect match[67].

Additionally, CGGexp-associated RAN translation of *FMR1*, similar to canonical translation, depends on eukaryotic translation initiation factor 4E (eIF4E), the protein which binds cap structure and 5' methyl-7-guanosine cap (**m7G**) recognition. The mechanism requires also the eukaryotic initiation factor 4A (eIF4A), an RNA helicase that is crucial for both cap binding and scanning of the small ribosomal subunit along the mRNA. Therefore, the initiation of CGG-induced RAN translation of *FMR1* is similar to the AUG-induced canonical translation, albeit only 30 – 40% as efficient[79].

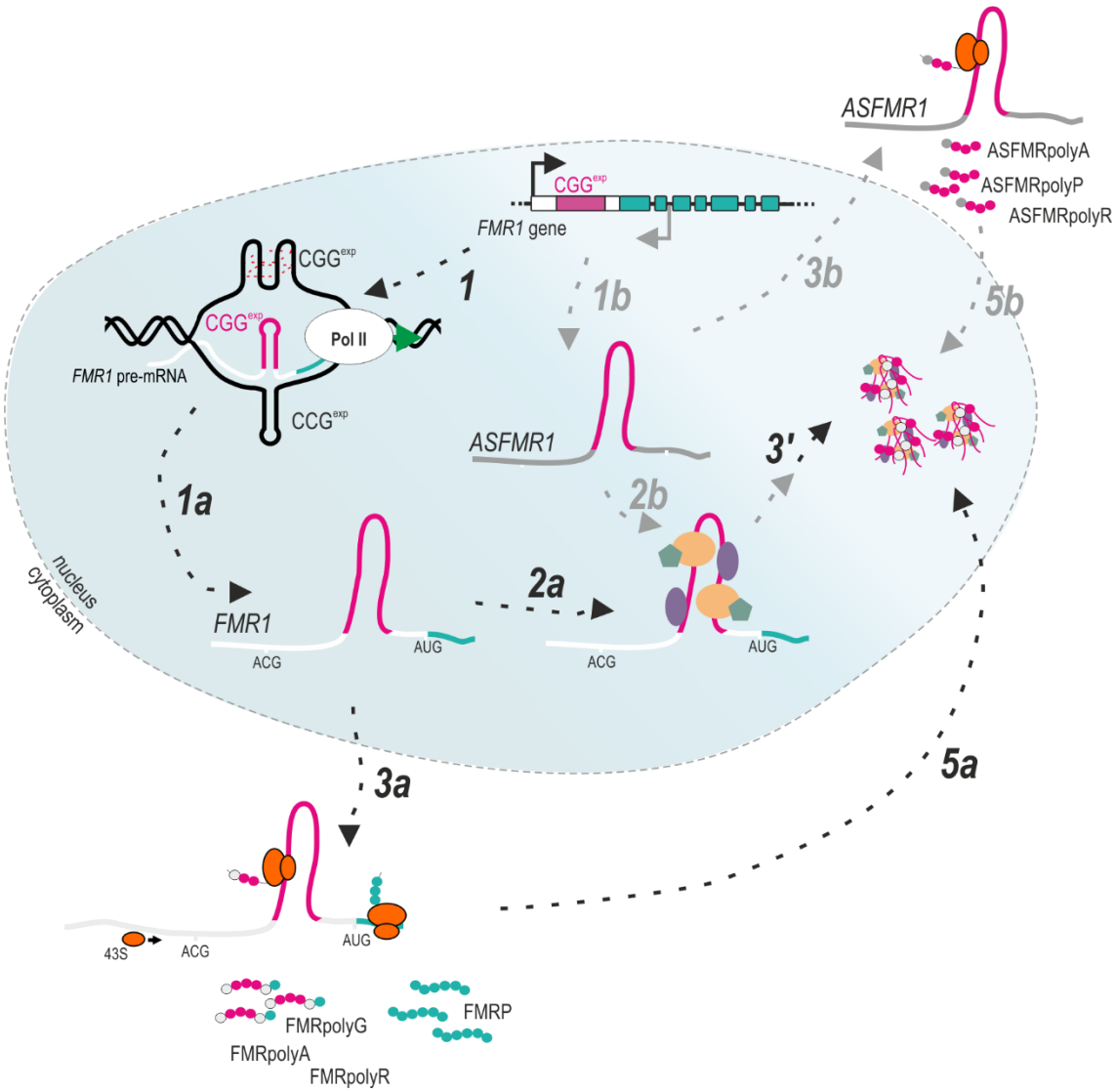


Figure 2. Molecular basis of premutation-driven Fragile X-associated disorders. *Mutant FMR1 containing expanded CGG repeats (CGGexp) in 5'UTR is transcribed in both sense (black arrow) and antisense (grey arrow) direction. Some rounds of transcription result in* ***(1)*** *R-loop formation which can lead to more open chromatin structure or blockage of incoming RNA Polymerase (Pol II), genomic instability, etc.;* ***(1a)*** *The majority of transcribed CGGexp in FMR1 mRNA, and* ***(1b)*** *CCGexp in ASFMR1 transcript form a stable secondary structure that is a potential cause of main pathologic consequences:* ***(2)*** *interaction of RNA secondary structure (containing either* ***(2a)***

*CGG or **(2b)** CCG repeats) with proteins is causing its partial retention and accumulation within nucleus and sequestration of RNA-binding proteins; **(3')** in the consequence the sequestered RBPs by rCGGexp form RNA foci; **(3a)** when FMR1 mRNA and **(3b)** ASFMR1 are transported to cytoplasm stable secondary structure facilitate RAN translation initiation; **(5a, 5b)** Toxic RAN proteins aggregate within the nucleus with other proteins.*

1.2.3.1.2. FMRpolyG-driven toxicity in premutation conditions

The RAN translation of the *FMR1* expanded CGG repeats leads to the accumulation of the FMRpolyG within the aggregates forming ubiquitin-positive intranuclear or perinuclear aggregates. Noteworthy, the RAN translation has been confirmed to occur also at mRNA containing a normal range of CGG repeats[80] and it has been proven that FMRpolyG is prone to length-dependent aggregation. However, the increase in the number of CGG repeats was also correlated with the gradual impairment of *FMR1* translation efficiency of major ORF and reduced level of FMRP[34,30,35,36]. On the other hand, the *FMR1* mRNAs with expanded repeats are translated at higher levels in reporter systems than those with normal lengths of CGG repeats[70].

Interestingly, the threshold of the number of CGG repeats required for FMRpolyG detection by western blot in cells vary between distinct studies. It has been revealed that depending on the size of the fused tag, the threshold of FMRpolyG detection is different. For example, fusion of the CGG repeats to GFP coding sequence in the (+1) reading frame allows for FMRpolyG detection even with short repeats (~30 CGGs) while fusion to small tag, like FLAG-tag, results in the FMRpolyG detection only if the number of CGG repeat is greater than ~50[69,67]. On a technical note, it has been demonstrated that fusion with FLAG resulted in a positive correlation between the FMRpolyG level and number in the CGG repeats while the level of FMRpolyG-GFP fusion protein was negatively correlated with the increasing number of CGG repeats[67].

The FMRpolyG aggregates were found in FXTAS fly and mouse models as well as in the post-mortem brain samples, including the frontal cortex, cerebellum, and hippocampus, of FXTAS patients[69]. Although the toxicity of FMRpolyG produced as the result of RAN translation was confirmed, the involvement of FMRpolyG in the FXTAS neurodegeneration remains elusive.

The formation of intranuclear aggregates results in the disruption of nucleus circularity. Lamina-associated polypeptide 2 beta (LAP2β) is the isoform of LAP2 protein that

interacts with lamin proteins which creates an envelope of the inner nuclear membrane[81]. It has been proven that the C-terminus of FMRpolyG binds the LAP2β protein and disrupts the nuclear lamina architecture leading to neuronal cell death[67]. Interestingly, the *LMNA* mRNA, of lamin A/C protein, was reported to be upregulated in the FXTAS-derived tissues[82] suggesting there is a feedback loop that the cell activates to recompensate the reduced activity of lamin A/C protein.

Ubiquitin-positive aggregates were found in the nuclei of neurons, astrocytes[83], and Purkinje cells[77] of PM carriers. However, non-central nervous system organs positive for the intranuclear aggregates were also reported[84]. The composition of isolated inclusions was established in a few studies and revealed that over 200 proteins were shown to be enriched within aggregates when compared to the composition of the whole nuclei in FXTAS patients[85]. It has been shown that aggregates are composed of proteins including molecular chaperones, stress response proteins, and components of proteasome.

To establish the role of FMRpolyG aggregates formation in the neurodegeneration in FXTAS patients scientists took advantage of various animal models. The first developed mouse models of FXTAS were the Dutch and NIH premutation CGG-repeat knock-in (KI) mouse models (named **CGG_{dut}** and **CGG_{nih}**, respectively)[86,87]. Despite the differences in the cloning strategies used to make these mouse lines (the presence of UAA stop codon upstream of the CGG repeats in the CGG_{nih}) it was shown that ubiquitin-positive intranuclear inclusions were found in both models, but were more common in neurons and astrocytes in the CGG_{dut} KI model[69,88,86]. Noteworthy, the placement of 5' leader sequence from CGG_{dut} mouse upstream 30 CGG repeats led to the FMRpolyG detection by western blot whereas the protein was not detected from CGG_{nih} mouse sequence[69] suggesting that the presence of native stop codon within murine *Fmr1,* at least partially, inhibit the RAN translation of FMRpolyG. On the other hand, as already mentioned, both mice models were characterized by the presence of ubiquitin-positive aggregates suggesting that RAN translation occurs in both models. On the other hand, in accordance with the assumption that RAN translation is not supported in CGG_{nih} mouse the difference in the severity of phenotype was observed. Behaviorally, the memory impairment was confirmed in both models[89,90], however, the CGG_{dut} KI mouse showed increased anxiety[91] which could be explained by the additive toxicity effect, next to RNA toxicity, induced by FMRpolyG.

Nonetheless, the mice models developed by Sellier and colleagues[67] revealed that: *(i)* translation of the CGG repeats in the (+1) glycine frame requires the presence of an upstream *FMR1* sequence, *(ii)* the polyglycine region is responsible for aggregation, and *(iii)* polyglycine repeats together with C-terminus of FMRpolyG are driving the toxicity *in vivo* through binding to the LAP2β and disrupting the nuclear lamina architecture. Nonetheless, the main conclusion from this study was that mice expressing FMRpolyG developed reduced longevity while mice expressing the rCGGexp without FMRpolyG protein (only polyglycine stretch without the flanking sequence of *FMR1* 5'UTR) were indistinguishable from control mice. On the contrary, Jin and colleagues[92] reported that overexpression of expanded CGG repeats without the native *FMR1* sequence context was sufficient to drive the neurodegeneration and a rough eye phenotype in *Drosophila* in a repeat-length dependent manner[92]. Similar observations were made by Castro and co-workers[93] who used a doxycycline-inducible mouse model and presented that 90 CGG repeats expressed outside of the context of the *FMR1* gene reproduced the FXTAS-like behavioral phenotype. Oh and colleagues[94] in turn, suggested that FMRpolyG drives the toxicity in FXTAS as a consequence of ubiquitin proteasome system (UPS) impairment. The statement that FMRpolyG alters the quality control pathways is supported by the fact that the components of the proteasome including ubiquitin and heat shock proteins were found in the intranuclear aggregates in FXTAS[68]. Also, the revealed protein composition suggests that formed inclusions are the consequence of protein aggregation due to the exceeded threshold of proteasomal degradation[85].

Additionally, the direct interaction between rCGGexp and FMRpolyG has been shown to promote FMRpolyG aggregation, and more specifically the liquid-to-solid transition, and therefore lead to neuronal dysfunction[95]. On the other hand, the ability of FMRpolyG to induce the formation of aggregates without the presence of rCGGexp has also been presented[96]. Todd and co-workers presented that only ~40% of RNA *foci* co-localized with the FMRpolyG-positive inclusions in the *Drosophila* model[69]. Therefore, it has been proposed that RAN translation of rCGGexp occurs at a low level, hence it is hard to define to what extent it contributes to the FXTAS pathology[85,97]. In conclusion, whether CGGexp-containing RNA *foci* and FMRpolyG aggregates constitute the same inclusion individual remains an open question and is far from being answered.

Taken together, these data confirm that the FMRpolyG plays a role in expanded CGG repeat-associated toxicity in FXTAS, however, due to inconsistent data, whether it is a crucial factor or requires the assistance of toxic RNA to drive neurodegeneration remains unanswered. Thus, the role of RNA gain-of-function mechanism and RAN translation in the toxicity in FXTAS is still a topic of many debates in the field. Presented above data suggest that, among others, there is a potential bias in the experiments based on the artificial models with expanded CGG repeats and that depending on the conditions including diverse expression levels of *FMR1* transgenes, size of fused tag, *FMR1* sequence context, and cell type and/or generally molecular background, the RAN translation may be strongly regulated *via cis* and *trans* factors influencing the efficiency of FMRpolyG synthesis and the downstream toxic effects.

1.2.4. Factors regulating FMRpolyG RAN translation initiation

The initiation of translation in eukaryotes is a complex process in which initiator tRNA together with ribosomal subunits, the 40S, and 60S, are assembled *via* eukaryotic initiation factors (**eIFs**) into the 80S ribosome on initiation codon. The whole process begins *via* the scanning mechanism which was introduced by Marylin Kozak[98,99]. The preinitiation complex (**PIC**) consisting of a 40S ribosomal subunit loaded with initiator Met-tRNA and eIFs enters the 5' end of mRNA *via* the recognition of 5' methyl-7-guanosine (m7G) cap structure and proceeds with scanning of the 5'UTR in the 5'-to-3' direction. Then, when the start codon is recognized initiator factors dissociate and the 60S ribosomal subunit joins to form an 80S ribosome competent for polypeptide elongation.

Similarly, as the translation elongation, the 43S scanning does not proceed through the mRNA with the same, equal dynamics. The kinetics of the scanning ribosome can be altered by many factors including the presence of a stable secondary RNA structure which has to be unwinded to allow further scanning of the downstream sequence by the 43S PIC or the translocation of the 80S ribosome during elongation. According to the *FMR1* 5'UTR sequence, among others, the rCGGexp as well as normal length rCGG can form hairpin RNA secondary structures[100], while the rCGGexp were also confirmed to be involved in the G4-RNA-quadruplex formation[101]. Because of the foregoing, the dynamic of the ribosome, both at initiation and elongation steps, can be disturbed. Indeed, it has been shown in yeast that the presence of stable hairpin structure (ΔG=–52.1 kcal/mol at 30°C) reduced the speed of scanning by ~30% in comparison to the 5'UTR region without

hairpin structure[102]. As a consequence, arising from the presence of more stable secondary RNA structures, different deleterious events can occur including ribosome dissociation.

Next to the cap-dependent mechanism the initiation of *FMR1* translation *via* IRES-mediated mechanism has been recently reported[103,104,105]. Pyrimidine-rich and well conserved among different species element found to function as Internal Ribosome Entry Site (**IRES**) is located 95 nt upstream of CGG repeats within *FMR1* 5'UTR[105,104]. This region contains two UUUC sequences and a CUUC sequence, each separated by one or two purines. The IRES-mediated, cap-independent mode of translation may use an internal ribosome entry site to directly recruit ribosomes to the *FMR1* mRNA. Nevertheless, IRES-mediated initiation should be still sensitive to translation impairment since the IRES sequence in *FMR1* was confirmed to be located upstream of CGG repeats thus the structural obstacle formed by expanded CGG repeats can still slow down the recruited ribosome. In conclusion, the initiation of translation at *FMR1* mRNA may arise from both cap-dependent and cap-independent mechanisms providing another level of complexity in translation regulation.

From the mechanistic point of view, start codon recognition depends on codon-anticodon interaction that is formed when the initiator Met-tRNA is positioned in the ribosomal P-site. In the canonical mode of translation, the Met-tRNA interacts with the AUG start codon, however, probably due to the reduced control of the base pairing in the P-site the initiator tRNA is allowed to recognize also other, non-AUG, start codons. In addition, the likelihood of translation initiation is regulated by the surrounding *cis*-regulatory factors.

1.2.4.1. RNA sequences and structures affect start codon utilization

In the 1980s Marylin Kozak performed many landmark experiments providing the very first suggestions about how sequences and secondary structures in the vicinity of the start codon influence the translation initiation[106,107,108]. She established that the particular nucleotides surrounding the AUG start codon strongly affect the efficiency of initiation across various vertebrate mRNAs[106,108,109,110]. These demanding nucleotides are now commonly referred to as the "Kozak sequence". The Kozak sequence context in vertebrates is GCCRCC**ATG**G where R is a purine (A or G)[109]. Positions -3 and +4 (where +1 refers to A in AUG) are considered the most important in regulating the efficiency of the translation initiation process due to the stabilizing interactions with

PIC[111,112]. Importantly, these surrounding nucleotides have a larger influence on the recognition of non-AUG start codons than on the AUG start codons.

Upstream ORFs (uORFs) are mRNA sequences defined by either AUG or near-cognate start codons in the 5′ UTR which are in-frame or out-of-frame with the main, downstream AUG-initiated ORF. The translation of uORFs usually results in the synthesis of short polypeptides which are believed to have an inhibitory effect on the translation of downstream ORF[113,114,115]. The inhibition of the downstream main ORF by uORF is usually mild because uORFs mostly begin from the near-cognate start codons or the AUG start codons embedded in the poor Kozak sequence context. Also, the possible re-initiation of the ribosome after translation of short uORF may support this effect[116]. Taken together, translation initiation may occur on mRNAs at more than one start codon. Of interest, recent studies revealed that non-AUG translation initiation sites are more abundant than AUG TISs[117,118] suggesting that protein products of uORFs may regulate the translation of downstream AUG-initiated ORFs. Nevertheless, some polypeptides translated from uORFs have been reported to be functional[119,120]. In the light of the foregoing, as the translational efficiency o *FMR1* and the levels of FMRP are decreased in FXTAS this phenomenon may result from the fact that RAN translation act as uORF to repress downstream FMRP synthesis. However, the decreased level of FMRP may also result directly from the disturbed ribosome scanning through the expanded CGG repeats.

According to the scanning model of translation, it has been shown that mRNA structures located either upstream or downstream of the start codons can influence the efficiency of initiation by the influence on the PIC movement. Therefore, the initiation from the near-cognate start codon or the AUG start codon located in the poor context can be increased when a stable secondary structure is placed downstream from the start codon[121,122]. It is worth mentioning, that the distance between the start codon and the secondary structure is important since it directly results in the positioning of stalled PIC on the mRNA. Indeed, it has been shown, based on the known size of ribosomes, that ~14 nt is a distance allowing for start codon positioning in the P-site of the 43S ribosome[121]. Additionally, the presence of stable secondary RNA structures has been suggested to be crucial for RAN translation regulation since the inhibition of the RNA helicase eIF4A which is involved in the ribosome scanning abolished the RAN translation of expanded CGG repeats[70]. The involvement of secondary structure in the modulation of RAN translation

has been also confirmed for another RED caused by the expansion of GGGGCC repeats (G4C2), within the *C9orf72* gene, forming in RNA very stable G-quadruplex (G4) structures[71,123,124]. Recent genome-wide analysis, in agreement with mentioned results, revealed that secondary structures located downstream functional near-cognate start codons are a common feature of mRNAs[118]. The structures formed by expanded CGG repeats are predicted to be extremely stable therefore they may form a physical obstacle for scanning PIC. Thus, due to the stalled 43S ribosomes and increased dwell time at the particular site of mRNA, it is possible that the efficiency of translation initiation, even at near-cognate start codon, will be boosted. This hypothesis is in line with the observations of the increased FMRpolyG translation correlated with the increased number of CGG repeats[70].

These data suggest that stable structures within mRNAs may play an important and underestimated role in the regulation of translation initiation, therefore significantly regulating the diversity and quantity of the cell proteome.

1.2.5. Premutation-driven R-loops within the *FMR1* locus

All of the mentioned models describing FXTAS pathogenesis – sequestration of RNBs on the toxic rCGGexp, generation of antisense rCCGexp transcripts, and RAN translation – are exclusively based on the post-transcriptional mechanisms. However, another potential pathomechanism of FXTAS, concerning the co-transcriptional level, has been suggested.

R-loops are nucleic acid structures which are RNA:DNA hybrids forming co-transcriptionally when nascent RNA hybridizes to the DNA template strand behind the elongating RNA polymerase II, with a simultaneous displacement of the non-template single-stranded DNA[125]. In general, R-loop formation is promoted when there is a thermodynamic advantage of the binding between the nascent RNA and the DNA template strand over the corresponding DNA:DNA duplex[126]. Hence, the main hotspots for R-loops formation constitute the GC-rich sequences, regions with elevated GC skew (an enrichment of guanine over cytosine on the non-template strand), and the abundance of G-clusters[127]. The other factors promoting R-loops formation include breaks in the non-template DNA strand[128], negative supercoiling that facilitates DNA unwinding, and

non-canonical DNA structures forming on the displaced DNA strand including G-quadruplexes (G4s)[129,130,131].

1.2.5.1. Regulatory role of R-loops

During the last decade, scientists began to study R-loops more intensively and they found that R-loops are not only 'by-products' formed during transcription but these structures are involved in the regulation of hundreds of genes and processes. Interestingly, depending on their genomic location they may play a beneficial or deleterious function. In general, they can be divided into two types: physiological, when their presence is crucial for the biological process, and pathological when their presence leads to genomic instability *via* many detrimental mechanisms. Regulatory R-loops are involved in chromatin structure and gene regulation, both by activating and silencing gene expression. Genome-wide studies have shown that R-loops are enriched over the loci with decreased DNA methylation and increased chromatin accessibility determined by DNase hypersensitivity[132]. The chromatin signature of sequences involved in the R-loop formation resembles those associated with transcription at promoters and the transcription start sites[130,129]. Indeed, it has been presented that in humans R-loops are enriched over the promoter regions containing CpG islands (**CGIs**)[131,129] and that they are involved in the protection of these regions from DNA methylation. Two main mechanisms explaining this process have been proposed. The first one assumes that R-loops can inhibit the binding of DNA methyltransferases (**DNMTs**) and thus protect promoters from methylation[129,133]. The second one proposes that R-loops may recruit to the promoter regions either the protective H3K4me3 mark or the DNA demethylation complex[129]. In agreement with these data, R-loops can induce chromatin decondensation, however, in parallel, they can also be involved in the heterochromatin assembly[125,134] and chromatin compaction[135]. At the same time, R-loops forming over the G-rich pause sites downstream of the polyadenylation signal in human genes are essential for the transcription termination of RNA Pol II[136]. Besides, it has been shown that the formation of R-loops on plasmids containing CTG/CAG repeats in *E.coli* and mini-gene constructs in human cells was associated with the promotion of repeat instability, leading to expansions, pointing towards their potential role in disease[125,137].

1.2.5.2. R-loops formed within 5'-part of FMR1 locus

R-loops have been implicated in the development of many diseases, especially those containing microsatellite repeat expansions, like triplet repeat expansion diseases (e.g. Huntington's disease or myotonic dystrophy). As one of many hot spots for R-loop formation is elevated GC skew and the abundance of G-clusters[127] it seems that the CGGexp in the 5'-part of *FMR1* gene meets the assumptions of R-loop forming sequence. Although the first premise of R-loop formation on the *FMR1* gene was published in 2014 only since very recently R-loops are considered one of the putative FXTAS pathomechanisms.

It has been presented *via in vitro* transcription that CGG trinucleotide repeats alone are able to form R-loops[138]. Till now, it has been confirmed that R-loops can be formed at the endogenous human *FMR1* locus[126,139,125], however, the precise localization and the borders of R-loops differ in different studies. The *FMR1* promoter is very CpG rich and has many features of the CpG-island promoter which makes it also a putative R-loop forming sequence. Importantly, the RNA:DNA hybrids formed over expanded triplet repeats may be different from R-loops formed over CpG islands-containing promoters. R-loops over expanded repeats may form a structural block, directly interfering with Pol II transcription initiation and/or elongation and decreasing the transcription efficiency while the R-loops formed over CpG islands-containing promoters may promote transcription[125,136,140].

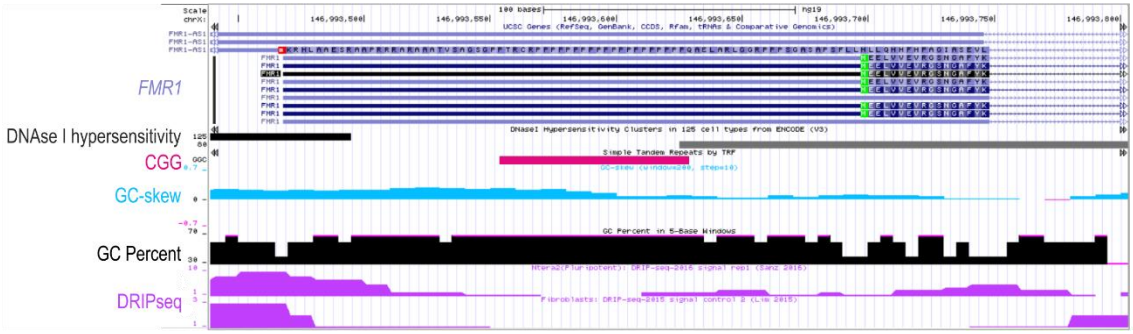


Figure 3. The genomic characterization of the 5'-part of the FMR1 gene. *The R-loop characteristic features within FMR1 5'leader sequence are presented in the UCSC Genome Browser (www.genome.ucsc.edu) containing data from the RloopDB (www.rloop.bii.a-star.edu.sg) and example DRIP-seq results obtained in the following studies*[141,142]*. Since the R-loop formation on the CGG repeats was confirmed, the observed 'gap' in the DRIP signals covering the CGGexp, may result from the fact that CGG repeats, similarly to other repeated sequences, are extremely hard to sequence, so the lack of signal rather comes from the technical limitations of the method than from the absence of R-loop structure in this region. In addition, it*

should be kept in mind during data analysis that the DRIP technique is also limited in the context of resolution.

Loomis and colleagues[126] suggested that R-loop forming at the CGGexp may result in a more open chromatin structure and thus contribute to the increase of *FMR1* mRNA in FXTAS patients, however, no direct evidence for this suggestion has been provided. Nevertheless, it can not be excluded that the more open chromatin state would promote transcription by increased access to transcription factors. On the other hand, as a co-transcriptional process, R-loop formation can be increased by enhanced transcription[126]. Consequently, the excessive R-loop formation can activate the DNA damage response (DDR) and result in DNA breakage. Of note, the accumulation of γH2AX, a histone variant related to DNA damage repair was present in the inclusions of FXTAS patient neurons[68,143]. From yet another side the CGG expansion can result in the formation of longer R-loops that can be prone to fold into more complex structures resulting in the structural blockage directly interfering with the Pol II during transcription.

As was already mentioned, the well-known hotspots for R-loop formation are GC skew, high GC percent, and the presence of CGIs. The CGGexp region of *FMR1*, on which R-loop formation was confirmed, contains many features of R-loop forming sequences, however, there are more R-loop prone regions within the 5'-part of *FMR1* that may potentially be involved in R-loop formation than CGG repeats itself. Based on the data from R-loopDB, UCSC Genome Browser, and available DRIP-seq data (DNA/RNA immunoprecipitation followed by high-throughput sequencing) one can observe that GC skew, CpG, and open chromatin regions determined by the DNase I hypersensitivity are in agreement with DRIP-seq results which clearly indicates that CGG repeats are not the only hot spots for R-loop formation within normal variant of *FMR1*.

Results from the genome-wide mapping studies presented that R-loops are abundant at promoters of RNA Pol II-transcribed genes[129,131,144,132]. At the CpG-containing promoters, R-loop may facilitate transcription *via* the protection of underlying DNA from methylation[129,133]. This mechanism is in agreement with the fact that DNA methyltransferases poorly bind to RNA:DNA hybrids[133]. On the other hand, it was shown that R-loops may constitute the promoters for Pol II transcribed genes by themselves[145]. Till now it is unknown how to distinguish the physiological from pathological R-loops, however, it has been emphasized how crucial it is to keep the perfect R-loop balance

because both failure in R-loop formation and failure in R-loop removal can affect a biological process similarly.

1.3. Molecular basis of fragile X syndrome

The full mutation of CGGexp, above 200 repeats, results in DNA hypermethylation and is thought to lead to heterochromatin formation at the *FMR1* promoter region[146,147,148] and subsequent *FMR1* loss-of-function resulting from transcriptional silencing of the *FMR1* gene[40,149,150]. Nevertheless, the process of *FMR1* silencing is a consequence of rather complex epigenetic modifications[149].

In FXS the cytosines within the region of approximately 1 kb upstream to CGG repeats, including the *FMR1* promoter, are methylated[40,151]. Importantly, the alleles containing a normal range of CGG repeats are also methylated in the *FMR1* promoter region, however, not in close proximity to the CGG repeat tract which constitutes the "boundary" between methylated and non-methylated DNA in healthy individuals. This boundary is lost in the FXS alleles thus the methylation of cytosines is spread through the *FMR1* gene. In addition to altered methylation status, excessive studies revealed that the hypermethylation of *FMR1* is associated with the local histones H3 and H4 deacetylation, reduced methylation of lysine 4 (K4), and increased methylation of lysine 9 (K9) on histone H3[152,153]. Together, these epigenetic modifications support that heterochromatin formation leads to *FMR1* silencing.

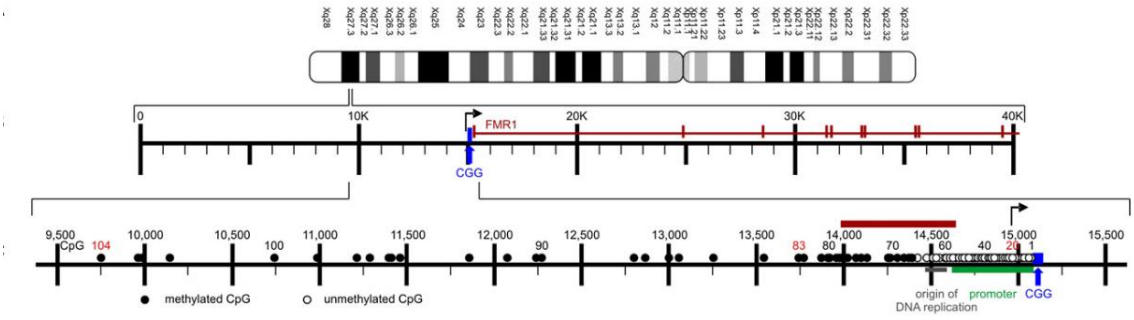


Figure 4. Methylation boundary in the mouse Fmr1 upstream region. *Taken from*[151]*. Black dots – methylated CpG, white dots – unmethylated CpG.*

Although the chromatin modifications important for *FMR1* silencing are already established, the timing and sequence of events leading to these epigenetic changes remain

elusive. Till now, the mechanism responsible for *FMR1* silencing in FXS is still not fully understood, however, a few hypotheses have been proposed[151,154,155,156].

1.3.1. R-loops in FXS

Colak and colleagues[146] performed studies in which they presented that: *(i)* the methylation of *FMR1* promoter in FXS occurs only in the presence of *FMR1* transcript; *(ii)* the lack of *FMR1* silencing in normal and premutation carriers is a consequence of lack of binding of *FMR1* pre-mRNA fragment to the promoter; *(iii)* there is a direct binding between *FMR1* pre-mRNA and coding DNA strand of *FMR1* gene. Based on these data it has been suggested that the methylation of *FMR1* promoter in FXS patients is correlated with the R-loop formation. The involvement of CGG-R-loop in the *FMR1* methylation was also suggested in other studies[125,157].

The precise mechanism explaining how R-loops formed within *FMR1* alleles with full mutation lead to heterochromatinization is not known yet, however, there are suggestions that R-loops, similarly to CGG hairpins[158], may recruit specific chromatin modifiers and thus lead to the *FMR1* silencing. Of note, a similar mechanism has been already confirmed for another RED. Friedreich ataxia (**FRDA**) is an autosomal recessive neurodegenerative disorder caused by the expansion of GAA repeats (100-1,500) within the first intron of the *FXN* gene[159]. Similarly as in the FXS condition the expanded GAA repeats in FRDA trigger the epigenetic silencing of the gene and lead to a deficiency of encoded protein. According to suggestions about the role of R-loops in the silencing processes, it has been shown that stable R-loops recruit the G9a histone methyltransferase, crucial for di-methylation of H3K9 (H3K9me2), to expanded GAA repeats within the *FXN* gene in Friedreich ataxia cells[125].

On the contrary, the most recent study published by Lee and co-workers[160] presents that CpG demethylation and induced R-loop formation within the *FMR1* locus lead to CGG repeat contraction and therefore result in the restoration of FMRP level. The authors performed the demethylation of the *FMR1* promoter which as expected resulted in the transcription of the *FMR1* locus. According to the conclusions drawn by the authors the demethylation induced *de novo* transcription of *FMR1* and led to the formation of Pol II-mediated R-loops. Due to the expanded CGG repeats the formed R-loops were suggested to be aberrantly stabilized and triggered DNA damage signals. As a consequence the MSH2/MMR DNA repair pathway was activated and the CGG repeat contraction

occurred. Interestingly, the contraction was not observed at the normal range of CGG repeats and although the mechanism behind this phenomenon is not known, the authors proposed a hypothesis explaining the “contraction threshold”. About that, it has been suggested that longer CGG repeats are involved in the complex secondary structure formation in both RNA and DNA and therefore, such structures formed in the non-template ssDNA could affect positively R-loop stability and recruit additional modifiers leading to CGG contraction.

So far, there is no cure for FXS. Excessive studies were performed verifying various potential therapeutic approaches and molecules[161,162,163,164,165,160], however most obtained results constitute the proof of concept thus, still more data about the mechanisms behind the FXS etiology is required to develop efficient and targeted therapy.

2. OBJECTIVES OF THE WORK

The fragile X tremor/ataxia syndrome (FXTAS) and fragile X syndrome (FXS) are human neurodegenerative and neurodevelopmental disorders, respectively, caused by the expansion of trinucleotide CGG repeats (CGGexp) in the 5'UTR of fragile X messenger ribonucleoprotein 1 gene (*FMR1*). Although both conditions arise from the mutation within the same locus, the pathomechanisms driving the development of the diseases are completely different. While FXTAS results from the RNA-, and protein-gain-of function mechanisms the FXS is characterized by RNA-, and protein-loss-of function. Interestingly, the R-loops were suggested to play a role in the pathology of both conditions, however, in FXTAS they were correlated with the decreased transcription efficiency and induction of cellular stress while in FXS they were probably involved in the silencing of the *FMR1* gene.

The presence of toxic mutant rCGGexp and the nature of the 5'UTR sequence of *FMR1* may result in a broad range of disruptions in a cell at both transcriptional and translational levels. Although these processes are completely different, regarding the assumptions of the central dogma of molecular biology, they are interrelated by toxic RNA molecule. Due to the complexity of raised issues, the project has been divided into two main parts:

- The aim of the first part of my project focused on the role of R-loops in the pathogenesis of FXTAS and FXS disorders. Firstly, I wanted **to confirm the ability of expanded CGG repeats (premutation range) located in the context of *FMR1* 5'UTR to form R-loops during *in vitro* transcription and therefore to correlate the formation of R-loops with the transcription efficiency both *in vitro* and *in cellula***. Since it has been previously shown that R-loops in 5′-parts of other genes may reduce transcription efficiency I wanted **to determine the effect of short chemically modified antisense oligonucleotides (ASOs, herein ASO-CCG) directly targeting rCGGexp involved in R-loop formation on the structure stability and therefore on the *FMR1* transcription efficiency**. As *FMR1* full mutation in FXS leads to the silencing of *FMR1* transcription I wanted **to establish whether binding of ASO-CCG to the CGG repeats could reactivate transcription *via* affecting R-loop stability,** resulting in the FMRP biosynthesis and therefore act as a potential therapeutic for FXS disease.

- The aim of the second part of my project was concerning the *cis*-regulatory elements within *FMR1* 5'UTR and their involvement in the regulation of FMRpolyG synthesis. Probably the main factor which drives the efficiency of RAN translation is the formation of ribosome 43S preinitiation complex (43S PIC) on the near-cognate ACG (+1) start codon present in 5'UTR of *FMR1*. This phenomenon depends on factors that may affect the speed/kinetics of ribosome scanning and/or ribosome pausing. To better understand the mechanisms of FMRpolyG RAN translation initiation efficiency it was assumed **to verify: *(1)* the effect of different nucleotide sequence context in the vicinity of near-cognate ACG (+1) start codon on the RAN translation, *(2)* the effect of stable secondary RNA structure formed by the sequence located downstream of ACG (+1) codon on the RAN translation initiation, and *(3)* how different size of rCGG repeats would affect efficiency of FMRpolyG biosynthesis.**

3. MATERIALS AND METHODS

3.1 Genetic constructs and cloning

3.1.1. Constructs used in the project which were already available in the laboratory

- 5′(CGGexp)-GFP(+1)[67] (Addgene #63091) construct was a kind gift from N. Charlet-Berguerand. Briefly, 5′(CGGexp)-GFP(+1) contains the 5′UTR of the *FMR1* gene with 99 CGG repeats and is fused to the eGFP sequence. Both proteins, polyglycine (FMRpolyG) and GFP are expressed as a fusion protein (FMRpolyG-GFP).
- pXPG-CMV-Fluc construct was a kind gift from Michał Sekrecki. Briefly, pXPG-CMV-Fluc contains ORF for Firefly luciferase under the control of CMV promoter.
- ACG-99xCGG Complete STOP and ACG-16xCGG Complete STOP constructs were previously cloned in our lab. These plasmids contain a 5'UTR sequence of the *FMR1* gene with 99 and 16 CGG repeats, respectively, without any tag. These constructs due to the native 3' end of the *FMR1* sequence were used to amplify insert for my constructs.

3.1.2. Constructs prepared for the dissertation

The majority of cloning procedures were prepared based on the In-Fusion Cloning technique (TaKaRa) which is an adaptation of the method described by Ochman et al., in 1988[166]. Briefly, the modified protocol of this technique assumes the mutagenesis by the inverse PCR using primers (containing 15 nt overhangs) that overlap each other at their 5' ends. The whole vector is then amplified and PCR product is added directly (if a homogenous product is produced), or *via* agarose gel-out purification, to the HiFi reaction mix (NEBuilder) where the enzyme removes nucleotides from 3' ends and allows complementary base pairs to join and anneal (re-circularize at the site of the 15 nt overlap). Two-fragment DNA assembly was performed either based on the InFusion system or by standard ligation procedure. 2-5 µl of the reaction mix were then used to transform competent cells.

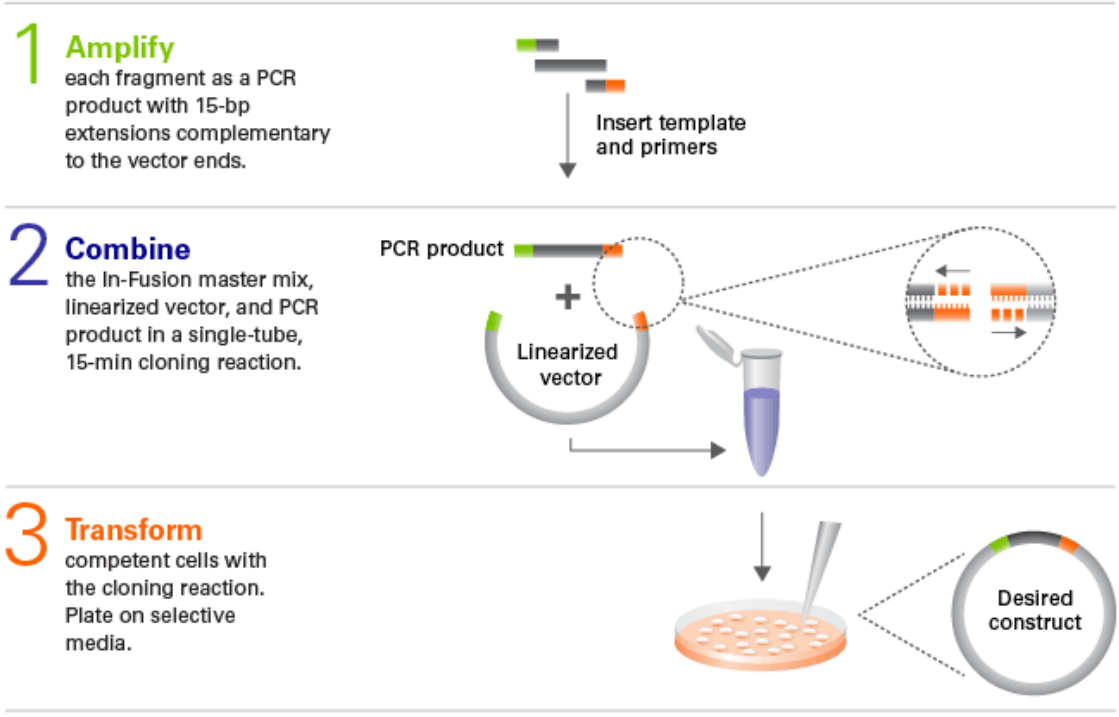


Figure 5. In-Fusion cloning protocol. *Taken from* [*https://www.takarabio.com/learning-centers/cloning/in-fusion-cloning-general-information/in-fusion-cloning-overview*](https://www.takarabio.com/learning-centers/cloning/in-fusion-cloning-general-information/in-fusion-cloning-overview)

3.1.2.1. Alternative cloning approach

Due to the extremely high content of GC pairs within some regions of the *FMR1* 5'UTR sequence (%GC content: 82-87), a few mutants were cloned by inverse PCR with primers without overhangs (to minimize the Tm of primers). PCR products amplified by such primers (linearized vector) were then either phosphorylated by T4 Polynucleotide Kinase (#EK0031; Thermo Fisher Scientific) according to manufacturer's instructions followed by self-circularization during ligation (at 4°C, overnight) by T4 DNA Ligase (#EL0011; Thermo Fisher Scientific) according to manufacturer's instructions or were amplified by phosphorylated primers and after purification directly used for self-circularization by ligation. In both cases, 2-5 µl of ligation mix were used to transform competent cells.

3.1.2.2. Description of the initial constructs

For sequence and structural context studies of *FMR1* 5'UTR, the new model based on NanoLuciferase (Nluc) reporter system has been developed. As the backbone the pNL1.1.CMV Vector (#N1091; Promega) has been chosen, which is a CMV-driven NanoLuciferase vector. The basic construct has cloned *FMR1* 5'UTR and Nluc in-frame with FLAG-tag. Each designed mutation has been cloned in parallel in two open reading frames generating either FMRpolyG-Nluc-FLAG (+1 open reading frame) or FMRP-

Nluc-FLAG (+0 open reading frame, allowing also for FMRpolyR-Nluc-FLAG detection (***see Figure 26b***). Depending on the studied aspect constructs had either 16 or 85 CGG repeats within *FMR1* 5'UTR. A detailed description of particular cloning steps is written below.

3.1.3. Cloning procedures

3.1.3.1. The backbone cloning

3.1.3.1.1. FLAG-tag addition to C-terminus of NanoLuciferase

To generate CMV-Nluc-FLAG construct the FLAG tag sequence (5' – GATTACAAGGATGACGACGATAAG – 3') was added to the C-terminus of Nluc sequence during inverse PCR (plasmid has been opened and the FLAG-tag sequence has been added during one reaction) by the following primers – Nluc_FLAG_F and Nluc_FLAG_R, containing 15 nt homologous overhangs. The sequences of primers are presented in ***Table 17***.

The following construct have been cloned in this step – CMV-Nluc-FLAG.

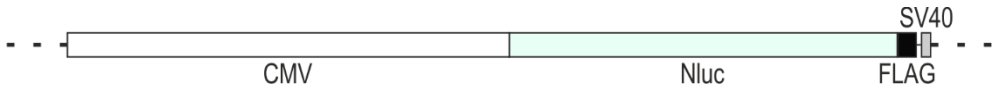


Figure 6. Scheme of cloned construct – CMV-Nluc-FLAG. *CMV corresponds to the CMV promoter sequence, Nluc – the sequence coding for Nanoluciferase enzyme fused with FLAG tag, and SV40 – transcription termination sequence.*

3.1.3.1.2. PGK-Firefly Luciferase-SV40 polyA signal cloning

As the internal reference the open reading frame for Firefly luciferase (Fluc) has been cloned into CMV-Nluc-FLAG plasmid. The sequence of PGK promoter, Fluc and SV40 polyA signal from pmirGLO vector (#E1330; Promega) has been amplified by primers F_add_firefly and R_add_firefly with 15 nt long overhangs. Simultaneously the CMV-Nluc-FLAG backbone was opened by inverse PCR with the following primers: F_open_pNL1.1 and R_open_pNL1.1. Amplified insert (PGK-Fluc-SV40) was cloned into the backbone (CMV-Nluc-FLAG) using the homology between 15 nt long overhangs. The sequences of primers are listed in ***Table 17***.

The following construct have been cloned in this step – CMV-Nluc-FLAG-PGK-Fluc.

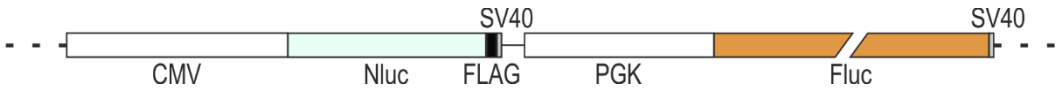


Figure 7. Scheme of cloned construct – CMV-Nluc-FLAG-PGK-Fluc. *CMV corresponds to the CMV promoter sequence, Nluc – the sequence coding for Nanoluciferase enzyme fused with FLAG tag, SV40 – transcription termination sequence, PGK – the PGK promoter sequence, and Fluc – the sequence coding for Firefly luciferase.*

Unfortunately, probably because the luminescence generated by the Nluc is approximately 100x "brighter" than those generated by the Fluc the firefly did not behave as an internal control. Because titration of plasmid did not resolve this issue I decided to resign from Fluc as internal control localized on the same plasmid as Nluc, since the expression of Fluc could be potentially altered by the Nluc. Therefore, the western blots were normalized to housekeeping proteins while Nano-Glo Dual-Luciferase Reporter Assay was performed on cells co-transfected with 20 ng of Nluc- containing plasmids and 100 ng of pXPG-CMV-Fluc plasmid, from which higher amount of Fluc was produced and the amount could be regulated by application of different ratio of plasmids used in co-transfection.

3.1.3.1.3. The FMR1 5'UTR sequence cloning

To ensure the transcription start site at the beginning of *FMR1* 5'UTR, the sequence of 109 nt between the CMV promoter and the *FMR1* sequence had to be removed. The deletion was performed by inverse PCR during the opening of plasmid CMV-Nluc-FLAG-PGK-Fluc with the following primers: F_open_pNLv2 and R_open_pNLv2. In parallel, the *FMR1* 5'UTR containing 16 CGG repeats has been amplified from ACG-16xCGG Complete STOP plasmid by F_5UTR_FMRP and R_5UTR_FMRP primers (+0 frame) and F_5UTR_FMRpolyG and R_5UTR_FMRpolyG inserting one extra nucleotide (+G) to perform frameshift for FMRpolyG (+1 frame). Both PCR products – linearized vector (CMV-Nluc-FLAG-PGK-Fluc) and insert (*FMR1* 5'UTR), contained 15 nt long overhangs. The sequences of primers are presented in ***Table 17***.

The following constructs have been cloned in this step – 16FMRP-Nluc-FLAG and 16FMRpolyG-Nluc-FLAG. The rest of the construct names will be omitted in the next chapters.

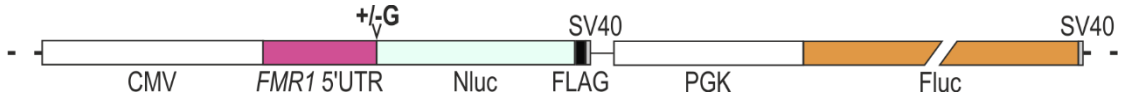


Figure 8. Scheme of cloned constructs – 16FMRP-Nluc-FLAG (-G) and 16FMRpolyG-Nluc-FLAG (+G). *The description of construct's elements is the same as in* ***Figure 7****, however, the FMR1 5'UTR sequence with the additional G nucleotide (+/-G) changing the open reading frame was marked. 16 – corresponds to the number of CGG repeats.*

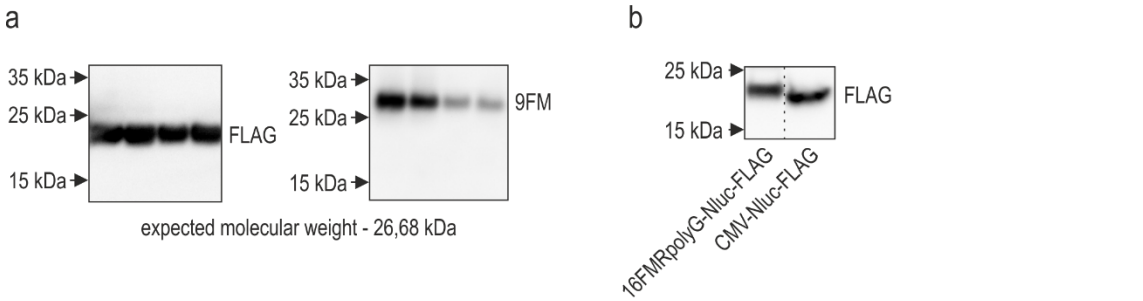
3.1.3.1.4. GUG (+1) near-cognate start codon mutation

The *FMR1* 5'UTR contains two well-known near-cognate start codons which are involved in FMRpolyG translation initiation, the ACG (+1) and GUG (+1) (***see Figure 26a***). To ensure that FMRpolyG translation will initiate from a single, particular start codon (ACG (+1)), the GUG (+1) codon was mutated to GUU by NEB_mutGUG_GUU_F_5P and NEB_mutGUG_GUU_R_5P primers without homologous overhangs. Amplified PCR products were run on agarose gel, purified from agarose, and used for self-circularization during ligation (T4 DNA Ligase) followed by competent cells transformation. The sequences of primers are presented in ***Table 17***.

The following constructs have been cloned in this step – 16FMRP-Nluc-FLAG_GTT and 16FMRpolyG-Nluc-FLAG_GTT

3.1.3.1.5. Additional mutations performed on constructs in the FMRpolyG frame

Proteins translated from 16FMRpolyG-Nluc-FLAG_GTT construct detected by anti-FLAG antibody (#F1804; Sigma) were inconsistent with the expected molecular weight (***see Figure 9a***), however, if anti-FMRpolyG antibody (9FM, recognizes the C-terminal part of FMRpolyG[76]; #MABN1788; Sigma) was used, the detected protein has the expected molecular weight. Although, the ATG start codon of Nluc remained unchanged the additional product was bigger than Nluc-FLAG (20,28 kDa) suggesting that translation was initiated upstream of the ATG start codon of Nluc (***Figure 9b***).

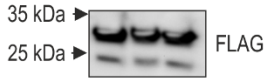


c

FMRP-/FMRpolyG-Nluc-FLAG_GTT

CTCAGTCAGGCGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGGCCGCCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGG
CGGCGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTTCGGCAGCG(CGG)CTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGG
GGCGGGCTCCCGGCGCTAGCAGGGCTGAAGAGAAGATGGAGGAGCTGGTGGTGGAAGTGCGGGGCTCCAATGGCGCTTTCTACAAGG

d



e

FMRP-/FMRpolyG-Nluc-FLAG_GTT_A/G

CTCAGTCAGGCGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGGCCGCCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGG
CGGCGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTTCGGCAGCG(CGG)CTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGG
GGCGGGCTCCCGGCGCTAGCAGGGCTGAAGAGAAGATGGAGGAGCTGGTGGTGGAAGTGCGGGGCTCCAAAAGCGCTTTCTACAAGG

	Kozak Similarity Score
GTG	≥0,7 <0,8
ACG	≥0,7 <0,8
CTG	≥0,6 <0,7
CTG	≥0,6 <0,7

Figure 9. Analysis of the FMR1 5'UTR sequence in the context of additional translation initiation sites. a) *The difference in molecular weight of protein product of 16FMRpolyG-Nluc-FLAG_GTT construct detected by anti-FLAG antibody (left), and anti-FMRpolyG antibody (right* [76]*);* ***b)*** *Western blot presenting difference in molecular weight between protein products of 16FMRpolyG-Nluc-FLAG_GTT and CMV-Nluc-FLAG constructs;* ***c)*** *The localization of ATG codon within FMR1 ex1 region, selected for a mutation to AAA (grey);* ***d)*** *Western blot presenting protein products of 16FMRpolyG-Nluc-FLAG_GTT_A/G construct detected by an anti-FLAG antibody;* ***e)*** *The localization of near-cognate start codons (grey) within FMR1 5'UTR which are possible donors for extra proteins translated in-frame with Nluc-FLAG. The corresponding Kozak Similarity Score is written; In* ***c*** *and* ***e*** *the ACG (+0), ACG (+1), and mutated GUG (+1) to GUU codon are bolded (red), CGG repeats are shown in brackets (pink). The insertion of the G nucleotide to generate frame shift to the FMRpolyG frame is bolded (the last nucleotide in the sequence; pink); The sequence of FMR1 ex1 is marked in green.*

Due to these discrepancies, the sequence of 16FMRpolyG-Nluc-FLAG_GTT has been analyzed in the context of other codons that could be donors for extra proteins translated in-frame with Nluc-FLAG. It turned out that within the *FMR1* 5'UTR sequence, exactly within the *FMR1* ex1 region there was an ATG codon, which was in-frame with Nluc-FLAG in 16FMRpolyG-Nluc-FLAG_GTT construct (***Figure 9c***). Therefore, this codon

was mutated to AAA and simultaneously the ATG start codon of Nluc was mutated to GGG by the primers – polyG_2ATG_AAA_GGG_F and polyG_2ATG_AAA_GGG_R. The sequences of used primers are presented in ***Table 17***.

The following constructs have been cloned in this step – 16FMRpolyG-Nluc-FLAG_GTT_A/G.

Performed mutations unfortunately did not result in homogenous protein synthesis (detected by anti-FLAG antibody) from 16FMRpolyG-Nluc-FLAG_GTT_A/G construct (***Figure 9d***). Hence, the sequence was analyzed in the context of potential near-cognate start codons that could be donors for the translation of extra proteins. The analysis was performed with the usage of available online tools – Open Reading Frame Finder (https://www.ncbi.nlm.nih.gov/orffinder/) and TIS predictor (https://www.tispredictor.com/tis#). The result of the analysis is presented in the ***Figure 9e***, however, only translation initiation sites for putative proteins translated in-frame with Nluc-FLAG are marked. Based on the molecular weight of translated protein it can be assumed that the last predicted near-cognate start CUG codon (8 nt upstream *FMR1* ex1 sequence) is a donor of additional protein visible on western blot. However, due to the presence of other codons that could change significantly the FMRpolyG level if they would be mutated (see [70]), and the fact that the sequence of *FMR1* 5'UTR was desired to be as close to native as possible, I decided to not mutate this codon. Because of that, to ensure reliable interpretation of generated results all mutants were further analyzed by both western blot and Nano-Glo Dual-Luciferase Reporter Assay.

3.1.3.2. Mutations of FMR1 5'UTR

3.1.3.2.1. Mutations of ACG (+1) Kozak sequence context

For better readability, the name of the constructs will be formed according to the model – 16FMRP/FMRpolyG-**name of introduced mutation**. Albeit, please keep in mind that all mutants, in both frames, possess GUG (+1) mutation and there are two additional mutations exclusively for the FMRpolyG frame.

All mutants considering the region of ACG (+1) Kozak context were obtained by the use of inverse PCR with primers containing 15 nt overhangs. PCR products were run on the agarose gel, appropriate bands were cut, DNA was purified and used in a reaction with HiFi NEBuilder mix according to the manufacturer's instructions. 2-5 µl of reaction mix

were used for competent cells transformation. The name of primers used for the mutagenesis of particular mutants are presented below. The sequences of primers are presented in ***Table 17***.

Table 1. List of primers used for Kozak ACG (+1) sequence context mutagenesis

Construct name	Forward primer name	Reverse primer name
16FMRP/FMRpolyG-Kozak1	F_Kozak1	R_Kozak1
16FMRP/FMRpolyG-Kozak2	F_Kozak2	R_Kozak2
16FMRP/FMRpolyG-Kozak3	F_Kozak3	R_Kozak3
16FMRP/FMRpolyG-Kozak4	F_Kozak4	R_Kozak4
16FMRP/FMRpolyG-Kozak4b	Kozak-4G-A_F	Kozak-4G-A_R
16FMRP/FMRpolyG-Kozak5	F_Kozak5	R_Kozak5

Described mutations were performed for both open reading frames thus the following constructs were generated in this step: 16FMRP/FMRpolyG-Kozak1, 16FMRP/FMRpolyG-Kozak2, 16FMRP/FMRpolyG-Kozak3, 16FMRP/FMRpolyG-Kozak4, 16FMRP/FMRpolyG-Kozak4b, and 16FMRP/FMRpolyG-Kozak5.

3.1.3.2.2. Mutations of ACG (+1) near-cognate start codon

The following mutations of ACG (+1) codon were introduced by the primers with 15 nt long overhangs based on the InFusion cloning – ACG→CTG, ACG→GTG, and ACG→AAA. Mutation ACG→ATG was obtained by the inverse PCR with primers without overhangs. Before ligation, the linearized plasmid (16FMRP/FMRpolyG-ACG→ATG) was phosphorylated by the T4 Polynucleotide Kinase (#EK0031, Thermo Fisher Scientific) according to the manufacturer's instruction.

The names of primers used for mutagenesis are presented below. The sequences of primers are listed in ***Table 17***.

Table 2. List of primers used for ACG (+1) near-cognate start codon mutagenesis

Construct name	Forward primer name	Reverse primer name
16FMRP/FMRpolyG-ACG→AAA	F_mutACG_AAA	R_mutACG_AAA
16FMRP/FMRpolyG-ACG→ATG	F_ACG-ATG_NEB	R_ACG-ATG_NEB
16FMRP/FMRpolyG-ACG→CTG	F_KmutACG_CTG	R_KmutACG_CTG
16FMRP/FMRpolyG-ACG→GTG	F_KmutACG_GTG	R_KmutACG_GTG

Described mutations were performed for both open reading frames thus the following constructs were generated in this step: 16FMRP/FMRpolyG-ACG→AAA,

16FMRP/FMRpolyG-ACG→ATG, 16FMRP/FMRpolyG-ACG→CTG, and 16FMRP/FMRpolyG-ACG→GTG.

3.1.3.2.3. Constructs containing randomly introduced extra ACG codon in +1 frame

One point mutations were performed to generate two mutants with **r**andomly introduced **ACG** (+1) codon (rACG1 and rACG2). Both constructs were generated by inverse PCR based on InFusion cloning. The names of primers used for the mutagenesis of particular mutants are presented below. The sequences of primers are presented in ***Table 17***.

Table 3. List of primers used for rACG1 and rACG2 cloning

Construct name	Forward primer name	Reverse primer name
16FMRP/FMRpolyG-rACG1	F_rACG1	R_rACG1
16FMRP/FMRpolyG-rACG2	F_rACG2	R_rACG2

Described mutations were performed for both open reading frames thus the following constructs were generated in this step: 16FMRP/FMRpolyG-rACG1, 16FMRP/FMRpolyG-rACG2.

3.1.3.2.4. Constructs with additional hairpin forming sequence

To analyze the structural effect on the translation initiation efficiency of FMRpolyG the sequence predicted to form a stable hairpin structure was cloned at different distances downstream ACG (+1) codon. The 42 nt long sequence was selected[167] and the following modifications have been designed.

Original hairpin sequence

GCCTAGGCCGGAGCGCCCA**GAT**CTGGGCGCTCCGGCCTAGGC

Mutation of STOP codon (green)

TAG→AAG; Mutation of CTA binding with TAG was also performed to ensure unchanged hairpin stability

GCCTTGGCCGGAGCGCCCA**GAT**CTGGGCGCTCCGGCCAAGGC

Mutation of CUG codon in +1 frame (yellow)

CTG→CCG

GCCTTGGCCGGAGCGCCCG**GAT**CCGGGCGCTCCGGCCAAGGC

The stability of the predicted RNA hairpin structure calculated by RNAfold WebServer as the Gibbs free energy was ΔG=–46.1 kcal/mol.

All constructs were generated by inverse PCR based on InFusion cloning. The hairpin-forming sequence was cloned downstream ACG (+1) codon at the following places: 2 nt-, 6 nt-, 14 nt-, and 20 nt-downstream ACG (+1). The names of primers used for the mutagenesis of particular mutants are presented below. The sequences of primers are presented in ***Table 17***.

Table 4. List of primers used for cloning of plasmids containing structure-forming-sequence

Construct name	Forward primer name	Reverse primer name
16FMRP/FMRpolyG-Hp2nt	F_hairpin2nt	R_hairpin2nt
16FMRP/FMRpolyG-Hp6nt	F_hairpin6nt	R_hairpin6nt
16FMRP/FMRpolyG-Hp14nt	F_hairpin14nt	R_hairpin14nt
16FMRP/FMRpolyG-Hp20nt	F_hairpin20nt	R_hairpin20nt

Described mutations were performed for both open reading frames thus the following constructs were generated in this step: 16FMRP/FMRpolyG-Hp2nt, 16FMRP/FMRpolyG-Hp6nt, 16FMRP/FMRpolyG-Hp14nt, and 16FMRP/FMRpolyG-Hp20nt.

3.1.3.2.5. Constructs with increased distance between ACG (+1) near-cognate start codon and CGG repeats

The mutants containing additional non-structure-forming sequences (CAlinker) between ACG (+1) and 16 CGG repeats were generated. The linker was 18 nt long and was cloned 14 nucleotides downstream ACG (+1). The sequence 5' CACACACACACACACACA 3' was predicted to not form any secondary RNA structure (ΔG=–0 kcal/mol).

Constructs were generated by inverse PCR based on InFusion cloning. The names of primers used for mutagenesis are presented below. The sequences of primers are presented in ***Table 17***.

Table 5. List of primers used for cloning of constructs containing an additional non-structure-forming sequence

Construct name	Forward primer name	Reverse primer name
16FMRP/FMRpolyG-CAlinker	F_CAn14nt	R_CAn14nt

Described mutation was performed for both open reading frames thus the following constructs were generated in this step: 16FMRP/FMRpolyG-CAlinker.

3.1.3.2.6. Mutation of ACG (+0) near-cognate start codon - translation initiation site for FMRpolyR

The mutation of the ACG (+0) codon was introduced by the inverse PCR with phosphorylated primers without overhangs. The names of primers used for mutagenesis are presented below. The sequences of primers are listed in ***Table 17***.

Table 6. List of primers used for ACG (+0) near-cognate start codon mutagenesis

Construct name	Forward primer name	Reverse primer name
16FMRP/FMRpolyG-ACG (+0)AAA	ACG(+)AAA_NEB_F	ACG(+)AAA_NEB_R_5PS

Described mutations were performed for both open reading frames thus the following constructs were generated in this step: 16FMRP/FMRpolyG-ACG (+0)AAA.

3.1.3.2.7. Cloning of FMRP/FMRpolyG-Nluc-FLAG constructs with long CGG repeats

As the source of long CGG repeats the 5′(CGGexp)-GFP(+1)[67] plasmid was used. Due to the lack of an appropriate restriction site that could be used to digest 5′(CGGexp)-GFP(+1) plasmid and cut out CGG repeats and to ligate them into target plasmids – 16FMRpolyG/FMRP-Nluc-FLAG, new restriction site, *Nru*I, has been cloned directly upstream CGG repeats on both plasmids. For this purpose, phosphorylated primers without overhangs were used. During a single PCR reaction the plasmids were opened and mutated. PCR products were run on the agarose gel and appropriate bands were cut out. Purified DNA was then used for ligation followed by transformation into competent cells.

The names of primers used for mutagenesis of particular plasmids are presented below. The sequences of primers are presented in ***Table 17***.

Table 7. List of primers used for *Nru*I restriction site insertion

Template plasmid DNA	Forward primer name	Reverse primer name
5′(CGGexp)-GFP(+1)	*Nru*I_GFP_F	*Nru*I_GFP_R_5PS
16FMRpolyG/FMRP-Nluc-FLAG	*Nru*I_Nluc_F	*Nru*I_Nluc_R_5PS

The number of CGG repeats in obtained colonies was checked by colony PCR with primers – 5'UTR_F and 5'UTR_R (***see Methods 3.3. "Colony PCR – screening for the***

number of CGG repeats"). Selected colonies were used to inoculate the LB medium, bacterial cultures were grown overnight at 30°C followed by centrifugation for 5 min at 14,000 rpm and plasmid isolation. Introduced mutation and the number of CGG repeats were verified by the Sanger sequencing. The lower number of CGG repeats in obtained constructs – 85 – in comparison to donor plasmid (5′(CGGexp)-GFP(+1) – 99 CGG repeats) results from the repeat instability during the bacterial culture growth.

The following constructs were generated in this step: 5'(85CGG)-GFP(+1)*Nru*I (donor of CGG repeats) and 16FMRpolyG/FMRP-Nluc-FLAG-*Nru*I (target backbone).

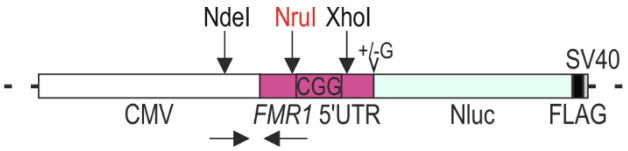


Figure 10. Scheme of FMRP/FMRpolyG-Nluc-FLAG-NruI plasmids. *The NruI restriction site was introduced upstream CGG repeats. NruI and XhoI restriction sites were used to prepare the backbone of FMRP/FMRpolyG-Nluc-FLAG and to cut out CGG repeats from 5'(CGGexp)-GFP(+1) plasmid. The localization of primers used for amplification of FMR1 5'UTR with NdeI and NruI restriction sites is shown. Appropriate mutated regions of FMR1 5'UTR were cloned between NdeI and NruI restriction sites in further steps.*

In the second step the 5'(85CGG)-GFP(+1)*Nru*I plasmid was digested with *Nru*I and *Xho*I (recognizes restriction site downstream CGG repeats) to obtain insert containing 85 CGG repeats that could be ligated to destination 16FMRpolyG/FMRP-Nluc-FLAG-*Nru*I vectors already digested with the same restriction enzymes. However, 16FMRpolyG/FMRP-Nluc-FLAG-*Nru*I vectors have two restriction sites recognized by *Xho*I, one downstream CGG repeats (in the same position as in 5'(85CGG)-GFP(+1)*Nru*I plasmid) and the other one downstream Firefly luciferase sequence. Hence, the titration of the *Xho*I enzyme and digestion time optimization was required to establish conditions in which *Xho*I will digest 16FMRpolyG/FMRP-Nluc-FLAG-*Nru*I vectors at a single site, downstream CGG repeats (***Figure 11b***). After optimization the following conditions have been chosen – digestion of 1 µg of plasmid DNA with 20 U of *Nru*I and 1 U of *Xho*I for 1 h at 37°C followed by enzyme deactivation at 65°C for 20 min. During digestion, the plasmid has been also dephosphorylated by Quick CIP (#M0525S; NEB) according to the manufacturer's instruction. Properly digested 16FMRpolyG/FMRP-Nluc-FLAG-*Nru*I vectors (without CGG repeats) and insert DNA containing 85 CGG repeats were then used in ligation followed by competent cells transformation.

The following constructs were generated in this step: 85FMRpolyG/FMRP-Nluc-FLAG.

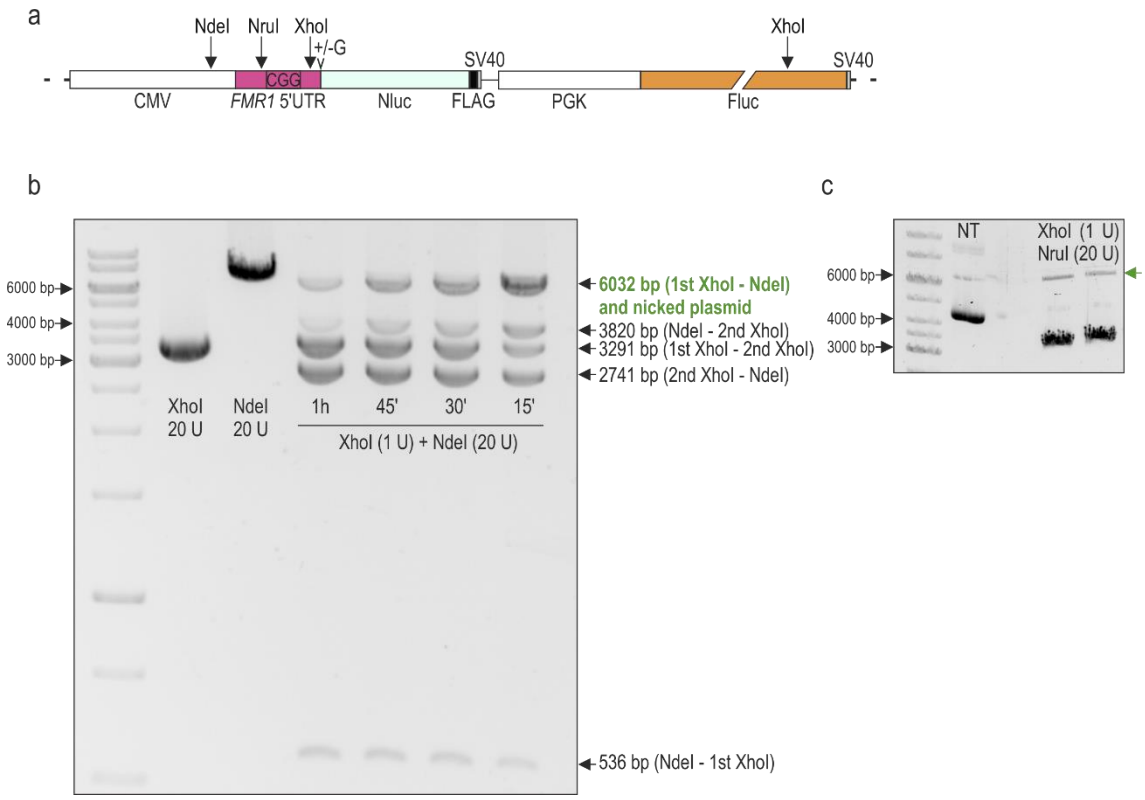


Figure 11. Optimization of plasmid 16FMRP/FMRpolyG-Nluc-FLAG-NruI digestion. a) *Schematic of restriction sites localization within 16FMRP/FMRpolyG-Nluc-FLAG-NruI plasmids;* ***b)*** *Restriction digestion analysis of 16FMRP-Nluc-FLAG-NruI plasmid was performed with XhoI and NdeI enzymes for better visualization of resulting fragments. To verify the efficiency of digestion by used enzymes 1 µg of plasmid DNA was digested by 20 U of XhoI or 20 U of NdeI for 1 h at 37°C followed by enzyme deactivation at 65°C for 20 min. To optimize digestion conditions 1µg of plasmid DNA was digested by 20 U of NdeI and 1 U of XhoI for 1 h, 45 min, 30 min, or 15 min. Each reaction was stopped by enzyme deactivation as above;* ***c)*** *16FMRP/FMRpolyG-Nluc-FLAG-NruI constructs were digested with 20 U of NruI and 1 U of XhoI enzyme for 1 h at 37°C followed by enzyme deactivation to generate backbone for 85 CGG repeats ligation.*

To study the effect of a hairpin structure formed by 85 CGG repeats on the FMRpolyG translation initiation the following mutants were planned: 85FMRpolyG-Kozak3, 85FMRpolyG-Kozak5, and 85FMRpolyG-CAlinker. 85FMRP-ACG (+0)→AAA mutant was also designed.

To avoid PCR amplification through CGG repeats the 5'UTR regions containing desired mutations (constructs already prepared with 16 CGG repeats) were used as templates for standard PCR. The Forward primer contained the *Nde*I restriction site which was natively present within the CMV promoter of all constructs. The Reverse primer was introducing the *Nru*I restriction site, thus after amplification PCR product was purified by Clean-Up

kit (#021-250; A&A Biotechnology) and digested with *Nde*I and *Nru*I enzymes for 1 hour at 37°C followed by enzyme deactivation at 65°C for 20 min. Simultaneously, 85FMRpolyG/FMRP-Nluc-FLAG-*Nru*I plasmids were digested with the same enzymes for 3 hours at 37°C followed by enzyme deactivation at 65°C for 20 min. Digested insert DNA fragments were again purified by Clean-Up while digested target backbone was purified after gel electrophoresis from agarose. Both DNA fragments were used in ligation followed by competent cells transformation. Plates with bacteria were incubated overnight at 30°C.

The same pair of primers was used to amplify different mutants of *FMR1* 5'UTR. Primers were compatible with both open reading frames and their sequences are presented in ***Table 17***.

Table 8. List of constructs used as templates for mutation of 85FMRP/FMRpolyG-Nluc-FLAG plasmids

Template plasmid DNA	Forward primer name	Reverse primer name
16FMRpolyG-Nluc-FLAG-Kozak3	5'UTR_F_Nluc_*Nde*I_v2	5'UTR_R_Nluc_*Nru*I_v2
16FMRpolyG-Nluc-FLAG-Kozak5		
16FMRpolyG-Nluc-FLAG-CAlinker		
16FMRP-Nluc-FLAG-ACG (+0)→AAA		

3.2. Polymerase Chain Reaction

Standard Polymerase Chain Reaction (PCR) of all non-problematic sequences was performed with the use of GoTaq G2 Flexi DNA polymerase (#M7845, Promega) according to the manufacturer's instruction.

3.3. Colony PCR – screening for the number of CGG repeats

I optimized conditions for colony PCR enabling screening concerning the number of CGG repeats after cloning before plasmid isolation and verification by Sanger sequencing.

After transformation agar plates with bacteria were incubated overnight at 30°C. Then, single colonies were picked and inoculated in 15 µl of LB medium. A mixture of bacteria

and LB medium was later used as a template for PCR reactions, and after verification of the number of CGG repeats as bacterial inoculum.

Conditions of colony PCR were optimized for GoTaq G2 Flexi DNA polymerase (#M7845, Promega) and F_5'UTR and R_5'UTR primers.

Table 9. Protocol for PCR mixture preparation with GoTaq G2 Flexi DNA polymerase

Reagent	Amount per 10 µl of reaction	Final concentration
H_2O	0.25 µl	
5x Green GoTaq Flexi Reaction Buffer	2 µl	1x
10 mM dNTP	0.2 µl	0.2 mM
10 µM Forward primer	1 µl	1 µM
10 µM Reverse primer	1 µl	1 µM
GoTaq DNA polymerase (5 U/µl)	0.05 µl	0.25 U
Template *	2 µl	
DMSO	0.5 µl	5%
5 M Betaine	3 µl	1.5 M

Template – 2 µl of LB medium inoculated by a single bacterial colony.

Table 10. PCR reaction programme for colony PCR

Temperature	Time	No. of cycles
98 °C	3 min	1
98 °C	30 sec	35
60 °C	30 sec	35
72 °C	35 sec	35
72 °C	2 min	1
4 °C	∞	1

3.4. Mutagenesis of GC rich sequences – optimization of PCR conditions

For cloning purposes two polymerase kits were used – CloneAmp HiFi PCR Premix (#639298, Takara Bio) and Phusion High Fidelity Polymerase (#F530L, Thermo Fisher Scientific).

Table 11. Protocol for PCR mixture preparation with CloneAmp HiFi PCR Premix

Reagent	Amount per 25 µl of reaction	Final concentration
CloneAmp HiFi PCR Premix	12.5 µl	
10 µM Forward primer	0.5 µl	0.2 µM
10 µM Reverse primer	0.5 µl	0.2 µM
DNA template 5 ng/µl	1 µl	0.2 ng/µl

5 M Betaine	9.25 µl	1.85 M
H_2O	1.25 µl	

Table 12. PCR reaction programme for PCR with CloneAmp HiFi PCR Premix

Temperature	Time	No. of cycles
98 °C	3 min	1
98 °C	10 sec	35
55/68 °C*	15 sec	35
72 °C	45 sec	35
72 °C	2 min	1
4 °C	∞	1

* Annealing temperature was equal either 55°C or 68°C, depending on the used primers pair

Table 13. Protocol for PCR mixture preparation with Phusion High Fidelity Polymerase

Reagent	**Amount per 20 µl of reaction**	**Final concentration**
H_2O	5.8 µl	
5 M Betaine	6 µl	1.5 M
5x Phusion GC Buffer	4 µl	1x
10 mM dNTP	0.4 µl	0.2 mM
10 µM Forward primer	1 µl	0.5 µl
10 µM Reverse primer	1 µl	0.5 µl
DNA template 5 ng/µl	1 µl	0.25 ng/µl
DMSO	0.6 µl	3%
Phusion DNA Polymerase 2 U/µl	0.2 µl	0.4 U

Table 14. 2-step PCR reaction programme for PCR with Phusion High Fidelity Polymerase

Temperature	Time	No. of cycles
98 °C	3 min	1
98 °C	10 sec	35
72 °C	1 min 50 sec	35
72 °C	2 min	1
4 °C	∞	1

3.5. Bacterial transformation procedure

DH5alfa competent cells (#18263012, Thermo Fisher Scientific) were used for the transformation of constructs without *FMR1* 5'UTR sequence or those containing *FMR1* 5'UTR with 16 CGG repeats.

Standard heat shock transformation procedure

Bacterial cells were thawed on ice for 5 – 10 min and DNA of interest was added. Cells were incubated on ice for the next 30 min and after that treated at a high temperature of 42°C for 1 min followed by 5 min incubation on ice. Then, 1 mL of LB medium was added and bacteria were incubated for 1 h at 37°C (shaking at 350 rpm). After incubation, bacteria were centrifuged for 2 min at 5000 rpm, supernatant was discarded and bacterial pellets were resuspended in 50 – 100 µl of LB medium. Therefore, bacteria were spread onto plates with a solid medium containing the appropriate antibiotics. Plates were incubated overnight at 37°C.

For the transformation procedure of constructs containing long CGG repeats (85 or 99 CGG repeats) chemically competent NEB Stable Competent *E. coli* cells (#C3040H, NEB) were used. These cells are suitable for high-efficiency transformation and isolation of plasmid clones containing repeat elements and unstable inserts. The transformation was performed according to the manufacturer's instruction followed by overnight incubation at 30°C. All bacterial cultures, to maintain CGG repeats were also incubated at 30°C.

3.6. Sanger sequencing

Since the 5'UTR region of *FMR1* is characterized by a high percentage of GC pairs and sequencing through the CGG repeats is very challenging the special protocol for my problematic templates has been established in the AMU Molecular Biology Techniques Laboratory (http://wptbm.amu.edu.pl/faq-en/). All Sanger sequencing performed to verify whether appropriate mutation occurred or to verify the number of CGG repeats in particular clones were done using the ProblemSeq protocol with a note – **GC rich template; TM kit; program seq_50**. Samples were always prepared as ready for sequencing, prepared in PCR tubes got from Sequencing Lab, according to the recipe – 150 ng of template DNA and 5 µl of 10 µM primer. The total volume of the mix did not exceed 8 µl. For the 5'UTR *FMR1* region with CGG repeats these two primers were used: DN_new117_F and DN_new79_R (sequences are presented in ***Table 17***).

3.7. Antisense oligonucleotides and siRNA

Antisense oligonucleotides (ASOs) were synthesized and HPLC purified by Kaneka Eurogentec. The sequences of used ASOs are presented in ***Table 15.***

Steric blockers

ASOs targeting CGG repeats (ASO-CCG) were 9-nucleotide-long and were composed of 8 LNA units and a 2′-O-Me unit at 3′ end. All LNA positions were phosphorothioated. The sequence of ASO–CCG-Cy3 contained additional Cy3 modification at the 5′ end. The ASOs targeting flanking regions of CGG repeats (ASO1 and ASO3) were 20-nucleotide-long and were exclusively composed of 2'- methoxyethyl (2'MOE) units. All positions were phosphorothioated.

Gapmers

ASOs inducing RNase H – dependent mRNA degradation were 15-nucleotide-long and contained 3 LNA-modified nucleotides at both 5' and 3' ends and 9 DNA nucleotides as the central core. All positions were phosphorthioated.

Table 15. List of Antisense oligonucleotides (ASOs) used in the project

Name	5' – 3' sequence
ASO-CCG	CCGCCGCCG
ASO-CCG-Cy3	Cy3 – CCGCCGCCG
ASO-Ctrl	TGAACATAA
ASO-1	TGCCAGGGGGCGTGCGGCAG
ASO-3	GGCGGCTGGGCCTCGAGCGC
ASO-Scr	GCCGGACGCCACGCTCGCGC
gap-Ctrl	GTGACTAAGGTGCTA
gap-CCG	CCGCCGCCGCCGCCG
gap-FMR1	CTTCAGCCCTGCTAG

Table 16. List of siRNA duplexes used in the project

Name	5' – 3' sequence	Company
RNaseH1_s	p-CCGGAAGUUUCAGAAGGGCAUGAAAdTdT	Future Synthesis
RNaseH1_as	p- UUUCAUGCCCUUCUGAAACUUCCGGdTdT	
RNaseH2_s	p- CCACUGGGCUUAUACAGUAUGCAUUdTdT	Future Synthesis
RNaseH2_as	p- AAUGCAUACUGUAUAAGCCCAGUGGdTdT	
siCTRL_s	p-UAAGGCUAUGAAGAGAUACdTdT	Future Synthesis
siCTRL_as	p-GUAUCUCUUCAUAGCCUUAdTdT	
siDHX9	ON-TARGETplus siRNA, Human DHX9 (1660) Individual (J-009950-06-0002)	Dharmacon

dT – deoxythymidine; p – phosphate

3.8. Oligonucleotides

Table 17. List of oligonucleotides used in the project

Name	Sequence

Nluc_FLAG_F	TTACAAGGATGACGACGATAAGTAATTCTAGAGTCGGGGCGGC
Nluc_FLAG_R	TCGTCATCCTTGTAATCTCCCGCCAGAATGCGTTCGCAC
F_add_firefly	GCAGCGCTCTTCCGCGGGTAGGGGAGGCGCTTT
R_add_firefly	TCAGTGAGCGAGGAAAACTTGTTTATTGCAGCTTATAATG
F_open_pNL1.1	TTCCTCGCTCACTGACTCGC
R_open_pNL1.1	GCGGAAGAGCGCTGCCGG
F_open_pNLv2	ATGGTCTTCACACTCGAAGATTTCG
R_open_pNLv2	ACGGTTCACTAAACGAGC
F_5UTR_FMRP	CGTTTAGTGAACCGTCTCAGTCAGGCGCTCAGC
R_5UTR_FMRP	GAGTGTGAAGACCATCTTGTAGAAAGCGCCATTGGAGC
F_5UTR_FMRpolyG	CGTTTAGTGAACCGTCTCAGTCAGGCGCTCAGCT
R_5UTR_FMRpolyG	GAGTGTGAAGACCATCCTTGTAGAAAGCGCCATTGGA
NEB_mutGUG_GUU_F_5P	CAGGGGGCGTTCGGCAGCGCG
NEB_mutGUG_GUU_R_5P	GCAGCGGCGCCTCCGTCAC
polyG_2ATG_AAA_GGG_F	AGCGCTTTCTACAAGGGGGGTCTTCACACTCGAAGATTTCGTTG
polyG_2ATG_AAA_GGG_R	CTTGTAGAAAGCGCTTTTGGAGCCCCGCACTTCCAC
F_Kozak1	CGGCGGCGTTGACGGAGGCGCCGCTGC
R_Kozak1	CCGTCAACGCCGCCGCCCGCGCTCG
F_Kozak2	CGGTGACGAAGGCGCCGCTGCCAGGGG
R_Kozak2	GCGCCTTCGTCACCGCCGCCGCCCG
F_Kozak3	CGGCGTTGACGAAGGCGCCGCTGCCAGGGG
R_Kozak3	CCTTCGTCAACGCCGCCGCCCGCGCTCG
F_Kozak4	GCGGCGGCCGTGACGGAGGCGCCGCTGC
R_Kozak4	CGTCACGGCCGCCGCCCGCGCTCGC
Kozak-4G-A_F	GCGGCGGCAGTGACGGAGGCGCCGCTGC
Kozak-4G-A_R	CGTCACTGCCGCCGCCCGCGCTCGC
F_Kozak5	CGGCGGCCGCGACGGAGGCGCCGCTGCC
R_Kozak5	CCGTCGCGGCCGCCGCCCGCGCTCGC
F_mutACG_AAA	GCGGTGAAAGAGGCGCCGCTGCCAGGG
R_mutACG_AAA	CGCCTCTTTCACCGCCGCCGCCCGCG
F_ACG-ATG_NEB	GCGGCGGTGATGGAGGCGCCG
R_ACG-ATG_NEB	CGCCCGCGCTCGCCGTCG
F_KmutACG_CTG	GGCGGTGCTGGAGGCGCCGCTGCCAGGG
R_KmutACG_CTG	GCCTCCAGCACCGCCGCCGCCCGCGC
F_KmutACG_GTG	GGCGGTGGTGGAGGCGCCGCTGCCAGGG
R_KmutACG_GTG	GCCTCCACCACCGCCGCCGCCCGCGC
F_rACG1	CTGAGCGGACGGCGGGCCGACGGCGAG
R_rACG1	CCGCCGTCCGCTCAGAGGCGGCCCTC
F_rACG2	CGCGGGCGACGGCGGTGACGGAGGCGC
R_rACG2	CCGCCGTCGCCCGCGCTCGCCGTCG
F_hairpin2nt	CGCCCGGATCCGGGCGCTCCGGCCAAGGCGGCGCCGCTGCCAGGGGG
R_hairpin2nt	GCCCGGATCCGGGCGCTCCGGCCAAGGCTCCGTCACCGCCGCCGCC
F_hairpin6nt	CGCCCGGATCCGGGCGCTCCGGCCAAGGCCCGCTGCCAGGGGGCGTTCG

R_hairpin6nt	GCCCGGATCCGGGCGCTCCGGCCAAGGCCGCCTCCGTCACCGCCGC
F_hairpin14nt	CGCCCGGATCCGGGCGCTCCGGCCAAGGCAGGGGGCGTTCGGCAGCG
R_hairpin14nt	GCCCGGATCCGGGCGCTCCGGCCAAGGCGGCAGCGGCGCCTCCGTCACC
F_hairpin20nt	CGCCCGGATCCGGGCGCTCCGGCCAAGGCCGTTCGGCAGCGCGGCGG
R_hairpin20nt	GCCCGGATCCGGGCGCTCCGGCCAAGGCCCCCCTGGCAGCGGCGCC
F_CAn14nt	ACACACACACACACACAAGGGGGCGTTCGGCAGCG
R_CAn14nt	TGTGTGTGTGTGTGTGGGCAGCGGCGCCTCCGTCACC
ACG(+)AAA_NEB_F	CGGCGGGCCGAAAGCGAGCGCGG
ACG(+)AAA_NEB_R_5PS	CCCGCTCAGAGGCGGCCC
*Nru*I_GFP_F	CAGGGGGCGTTCGCGAGCGCGGCGG
*Nru*I_GFP_R_5PS	GCAGCGGCGCCTCCGTCA
*Nru*I_Nluc_F	GGGGCGTTCGCGAGCGCGGCGG
*Nru*I_Nluc_R_5PS	CTGGCAGCGGCGCCTCCG
5'UTR_F_Nluc_*Nde*I_v2	CATCAAGTGTATCATATGCCAAGTCC
5'UTR_R_Nluc_*Nru*I_v2	CCGCGCTCGCGAACG
hFMR1_F	ATCCCAACAAACCTGCCACA
hFMR1_R	ATGTGCTCGCTTTGAGGTGA
hGAPDH_F	GAGTCAACGGATTTGGTCGT
hGAPDH_R	TTGATTTTGGAGGGATCTCG
MALAT_F	GACGGAGGTTGAGATGAAGC
MALAT_R	ATTCGGGGCTCTGTAGTCCT
F_FMR1_Nluc_mRNA	GCAGGGCTGAAGAGAAGATG
R_FMR1_Nluc_mRNA	TGGATCGGAGTTACGGACAC
FMR1_pre-mRNA_proximal_F	AGAAGATGGAGGAGCTGGTG
FMR1_pre-mRNA_proximal_R	CCTGAAAAGCACTCAAACTGGA
FMR1_pre-mRNA_distal_F	TGTGTCCCCATTGTAAGCAA
FMR1_pre-mRNA_distal_R	CTCAACGGGAGATAAGCAG
RNASEH1_F	CACAGAGGATGAGGCCTG
RNASEH1_R	CAGTGGCTCACGGAGTC
RNaseH2A_F	CTGGGCGTCGATGAGG
RNaseH2A_R	CCGCTCGCTCTCCAATAG
DHX9_F	TTGGCAGTACACGGTATGGA
DHX9_R	ATAGCCTCCACCAACACCTG
DN_new117_F	CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG
DN_new79_R	ACACCCCGAGATTCTGAAACAAACTGGACACACCTC

3.9. Cell culture and transfection

All experiments with FMRP/FMRpolyG-Nluc-FLAG constructs were performed in the HEK-293 cell line while R-loop studies were performed in FXTAS- and FXS-patients-derived fibroblasts.

CGGnorm/- (1), C6; CGGnorm/- (2), C0603; CGGnorm/CGGexp, FX11-02; CGGexp/CGGexp, WC26 were previously described[168,82]as well as FX08-01 and FX13-01 which were described[168,169]. 1044-07 cells were a kind gift from Paul J. Hagerman.

Table 18. List of patient-derived fibroblasts

Name	Internal name	Phenotype	number of CGG repeats
CGGnorm/- (1)	C6	control	20
CGGnorm/- (2)	C0603	control	31
CGGnorm/CGGexp	FX11-02	FXTAS	20, 79
CGGexp/CGGexp	WC26	FXTAS	60, 90
1044-07	1044-07	FXTAS	97
FX08-01	FX08-01	FXS	>435
FX13-01	FX13-01	FXS	>435

HEK-293 cells were grown in a high glucose DMEM medium supplemented with L-glutamine (#L0104; Biowest), 10% fetal bovine serum (#S181H; Biowest), and 1% antibiotic/antimycotic (#A5955; Sigma) at 37°C in 5% CO2.

All fibroblasts were grown in MEM medium (#L0416; Biowest) supplemented with 15% fetal bovine serum (#S181H; Biowest), 1% MEM nonessential amino acids (#M7145; Sigma), and 1% antibiotic/antimycotic (#A5955; Sigma) at 37°C in 5% CO2.

For R-loops analyses *in cellula* FXTAS – derived fibroblasts were plated on a 12-well plate and transfected at ~80% confluency with 15 nM siRNAs. After 24 h transfection with 200 nM ASOs was performed. Cells were harvested after an additional 48 h.

For transfections with genetic constructs and antisense oligonucleotides (ASOs), HEK-293 cells were plated on the appropriate cell culture vessels. Cells were transfected 3 h from plating with genetic constructs at ~80% confluency. After the next 3 h, cells were transfected with 200 nM ASOs and harvested after another 48 h.

For transfection with FMRP/FMRpolyG-Nluc-FLAG genetic constructs for western blot analysis HEK-293 cells were plated on 48-well plates and transfected with 100 ng of Nluc-containing plasmids. Cells were harvested either 24 or 48 h post-transfection.

As a positive control of *FMR1* transcription reactivation FXS – derived fibroblasts were seeded on a 6-well plate and grown for 7 days in MEM medium supplemented with 1 μM 5-Aza-2′-Deoxycytidine (5-azadC; #A3656; Sigma). The medium was replaced with a

fresh one every day. After treatment cells were cultured for additional 1 to 30 days (according to performed experiment) in the medium without 5-azadC. Simultaneously the same cells were grown for two weeks in the presence of ASO-Ctrl/ASO-CCG which were delivered to cells *via* two transfections (on the 1st and 8th day). The detailed descriptions of these experiments are written under appropriate figures in the Result section.

In all experiments, ASOs were denatured before transfection for 30 s at 95°C and chilled on ice. All transfections were performed with the use of Lipofectamine 3000 (#L3000015, Thermo Fisher Scientific) according to the manufacturer's instructions.

3.10. RNA isolation and reverse transcription

The isolation of total RNA from cells was performed using TRI Reagent (#AM9738, Thermo Fisher Scientific) and Total RNA Zol-Out™ D kit (#043-100; A&A Biotechnology). During purification on columns, the RNA was digested by the DNase. Finally, 300 – 500 ng of the total RNA was used for reverse transcription (RT) with GoScript Reverse Transcriptase (#A5004; Promega) and random primers (#A2801; Promega) according to the manufacturer's protocol.

3.11. RT-PCR

All RT-PCR reactions were performed using GoTaq G2 Flexi DNA polymerase (#M7845, Promega) according to the manufacturer's instruction. PCR products were separated on the agarose gel (1 – 2%) with ethidium bromide (0.5 μg/ml) and visualized by G:BOX (Syngene). The intensity of DNA bands was measured by GeneTools software (Syngene).

3.12. RT-qPCR

RT-qPCR reactions were performed with the use of Maxima SYBR Green/ROX qPCR Master Mix (#K0223, Thermo Fisher Scientific) according to the manufacturer's instructions in QuantStudio 7 Flex Real-Time PCR System (Thermo Fisher Scientific). Ct values were normalized against GAPDH. Fold differences in expression level were calculated according to the $2^{-\Delta\Delta Ct}$ method[170]. The sequences of used primers are listed in ***Table 17***.

3.13. Nano-Glo Dual-Luciferase Reporter Assay

HEK-293 cells were plated on 96-well plates and co-transfected after 24 h with 20 ng of FMRP/FMRpolyG-Nluc-FLAG plasmids and 100 ng of pXPG-CMV-Fluc construct using Lipofectamine 3000 (#L3000015, Thermo Fisher Scientific). Cells were lysed

either 24 h or 48 h post-transfection in 100 µl of RIPA buffer (#R0278, Sigma) and incubated for 30 min on ice. 20 µl of lysate from each sample were transferred on a black well plate (#137101; Thermo Fisher Scientific) equilibrated to room temperature and proceeded with Nano-Glo Dual-Luciferase Reporter Assay (#N1620, Promega). Nluc was normalized to Fluc to control for variation in transfection efficiencies.

3.14. In vitro transcription

Visualization of R-loops in the 5'-end of *FMR1* was performed using two approaches based on *in vitro* transcription performed on either circular or linearized DNA.

The first approach – *In vitro* transcription using the non-digested 5′(CGGexp)-GFP(+1) genetic construct (~500 ng of circular plasmid). The reactions were performed at 37°C for 2 h and then were treated with 1 µl of RNase A (12.5 ng/µl) and 1 µl of RNase H (10 U; #AM2293; Thermo Fisher Scientific) or an equal amount of 50% glycerol in untreated samples for an additional 30 min at 37°C. Reactions were stopped and nucleic acids were extracted with phenol-chloroform and precipitated with 96% ice-cold ethanol and 3 mM sodium acetate. Pelleted nucleic acids were resuspended in the water with 6 x BLUE DNA Loading buffer (#AG16; Blirt) and analyzed on 1% agarose gels with ethidium bromide (0.5 µg/ml) run in 1 x Tris-Borate-EDTA buffer at 70 V for 2 h.

The second approach – *In vitro* transcription performed on the 5′(CGGexp)-GFP(+1) genetic construct (~500 ng) digested with *Avr*II enzyme recognizing restriction site downstream putative R-loop forming sequence and CGG repeats (allow for transcription of whole exon 1 of *FMR1*). The reactions were performed at 37°C for either 20 min or 90 min in the presence of 1 µl of RNase H (10 U) or 50% glycerol in the untreated samples. Reactions were stopped by the addition of 6×BLUE DNA Loading buffer which contained 60 mM EDTA and analyzed on 1% agarose gels with or without ethidium bromide (0.5 µg/ml) run in 1 x Tris-Borate-EDTA buffer at 70 V for 2 h.

The standard *in vitro* transcription reactions were performed based on the second approach on the 5′(CGGexp)-GFP(+1) construct linearized by *Avr*II enzyme. The construct contained a T7 polymerase promoter upstream to 5′UTR of *FMR1*. Reactions were performed with ~500 ng of template DNA in 10 µl of a mixture containing 1× transcription buffer (#P118B; Promega), 10 mM DTT, 10 U of RNasin (#N2615; Promega), rNTPs (0.5 mM each), 4 U of T7 polymerase (#P207B; Promega) and (if required) 2.5 µM 9-nucleotide-long ASO-ctrl, ASO-CCG or ASO-CCG-Cy3. In the case

of RNase H treatment, 1 µl of RNase H (10 U) was added to the mixture of proper samples at the beginning of the reaction. Samples without RNase H were treated with an equal volume of 50% glycerol to provide the same density of the mixture. The further steps were the same as described above in the second approach procedure.

To confirm the interaction of ASO-CCG with R-loop structures and sense strand of DNA template with expanded CGG repeats *in vitro* transcription was performed as described above (in the presence of ASO-CCG or fluorescently labeled ASO-CCG-Cy3), albeit the following modifications have been performed. Samples were digested with either 2 U of DNase TURBO (#AM1907; Thermo Fisher Scientific) or 1 µl of RNase A (0.3 µg/µl and 5 µg/µl) for 30 min on ice. Control samples were treated with 50% glycerol. Reactions were stopped as described above and analyzed on 1% agarose gels run in 1 x Tris-Borate-EDTA buffer at 70 V for 3 h. To visualize CGG-containing nucleic acids the agarose gels were scanned using Amersham Typhoon RGB Biomolecular Imager and the fluorescent signal coming from ASO-CCGCy3 was detected using a Cy3 filter (without EtBr). Then gels were stained with ethidium bromide (0.5 µg/ml) for 20 min and scanned again to visualize DNA templates and all RNA products (with EtBr; using Cy3 filter).

Intensities of the nucleic acids signals were measured and quantitated with Multi Gauge 3.0 software (Fujifilm) and ImageQuant TL 8.1.0.0 (Cytiva). Each signal (for gels stained with EtBr) was normalized to the intensity of the signal coming from a genetic construct in the same lane. This procedure was applied to compensate for random variations in the sample signal intensities due to gel loading errors.

3.15. Cytoplasm/nucleus fractionation

The nucleocytoplasmic fractionation was performed based on the protocol[171] with some modifications. Briefly, fibroblasts (1044-07 and FX13-01) were seeded on a 100-mm plate and if needed transfected at ~80% of confluency with 9-nucleotide-long ASOs at 200 nM final concentration. Fibroblasts were harvested 48 h post-transfection, washed in ice-cold 1× PBS, and centrifuged at 500 g for 5 min. Cell pellets were resuspended by gentle pipetting in 1 mL of ice-cold HLB buffer (Hypotonic Lysis Buffer: 10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3 mM MgCl2, 0.3% (v/v) NP-40, 10% (v/v) glycerol), complemented with RNasin (#N2515, Promega). Cells were lysed on ice for 30 min and after that briefly vortexed. Then, cells were centrifuged at 2000 g for 5 min at 4 °C. After centrifugation the supernatant containing the cytoplasmic fraction was transferred to a

new tube and kept on ice. 5 M NaCl solution was added to the cytoplasmic fraction to adjust the NaCl concentration to 140 mM. The nuclei pellet was washed 4 with HLB, pipetting, and centrifuging at 500 g for 2 min at 4 °C. The total RNA was isolated followed by cDNA synthesis and RT-qPCR analysis. Primers used for RT-qPCR are listed in the ***Table 17***.

3.16. Western blot

Fibroblasts extracts were prepared in lysis RIPA buffer (pH = 8, 150 mM NaCl, 1% NP-40) supplemented with 1× protease inhibitor (#78439, Thermo Fisher Scientific) and 1 mM PMSF. Lysates were vortexed, and frozen overnight at −80 °C. Protein extracts, after being thawed, were centrifuged at 10,000 g for 10 min at 4 °C, and protein concentration was measured by Pierce BCA Protein Assay Kit (#23225, Thermo Fisher Scientific). In total, 20–30 µg of protein was heat-denatured for 10 min at 70 °C with the addition of Bolt LDS buffer (#B0008, Thermo Fisher Scientific). Electrophoresis was performed in Bolt 4–12% Bis-Tris Plus gel (Thermo Fisher Scientific) in Bolt MES SDS Running Buffer (#B0002, Thermo Fisher Scientific). Proteins were transferred to PVDF transfer membrane (#GE10600021, GE Healthcare) for 1 h, at 100 V in 1xLeammli buffer with 20% methanol

The following modifications were introduced for HEK-293 cells expressing FMRP/FMRpolyG-Nluc-FLAG constructs. Cells were lysed in RIPA buffer (#R0278, Sigma) with 1× protease inhibitor and were sonicated for 10 cycles (10 s ON/10 s OFF) using Bioruptor Plus (Diagenode). Instead of protein concentration measurements, an equal volume of protein extracts (10 µl of lysate from a 48-well plate) was loaded on the polyacrylamide gel. Heat-denaturation was performed for 5 min at 95 °C. Proteins were transferred to the PVDF transfer membrane in Bolt Transfer Buffer (#BT00061, Thermo Fisher Scientific) supplemented with 20% methanol.

For standard western blot procedure membranes were blocked in 5% nonfat dry milk in PBS with 0.1% Tween 20 (PBS-T) overnight at 4 °C or at least for 1 h at RT. Incubation with the following antibodies: rabbit anti-FMRP antibody (#ab17722, Abcam) 1:1000, mouse anti-FMRpolyG antibody (9FM; recognizes C-terminal part of FMRpolyG[76]; #MABN1788, Sigma) 1:1000, was performed in 5% nonfat dry milk in TBS-T overnight at 4 °C. Mouse anti-GAPDH HRP-conjugated antibody (#sc-47724, Santa Cruz Biotechnology) 1:10,000, rabbit anti-Fluc antibody (#PA5-32209, Thermo Fisher

Scientific) 1:5000, mouse anti-vinculin HRP-conjugated antibody (#sc-73614 HRP, SantaCruz Biotechnology) 1:5000, rabbit anti-alpha-tubulin antibody (#ab52866, Abcam) 1:10,000, and mouse anti-FLAG HRP-conjugated antibody (#A8592, Sigma) 1:20,000 were incubated in 5% nonfat dry milk in TBS-T for 1 h at RT. Membranes were washed in TBS-T and incubated with horseradish peroxidase-conjugated secondary antibodies (not applicable for HRP-conjugated primary antibodies): anti-rabbit (#A9169, Sigma) 1:20,000 or anti-mouse (#A9044, Sigma) 1:20,000 for 1 h and washed with TBS.

Membranes prepared for western blot experiments performed on SNAP ID Protein Detection System (Merck Millipore) were blocked with 0.125% nonfat dry milk in TBS with 0.1% Tween 20 (TBS-T) for 20 min. Incubation with antibodies was performed in 0.125% nonfat dry milk in TBS-T in the following conditions: rabbit anti-FMRP antibody (#ab17722, Abcam) 1:500 for 1 h 10 min, mouse anti-GAPDH antibody (#sc-47724, Santa Cruz Biotechnology) 1:10,000 for 15 min, rabbit anti-alpha-tubulin antibody (#ab52866, Abcam) 1:10,000 for 15 min. Membranes were washed in TBS-T and incubated with horseradish peroxidase-conjugated secondary antibodies: anti-rabbit (#A9169, Sigma) 1:20,000 or anti-mouse (#A9044, Sigma) 1:20,000 for 15 min and washed with TBS.

Antibody–antigen complexes were visualized by enhanced chemiluminescence (ECL) using Immobilon Forte Western HRP substrate (#WBLUF0500, Sigma) and detected with G:BoxSystem (Syngene). Intensities of the protein signals were measured and quantitated with ImageQuant TL 8.1.0.0 (Cytiva). For detection of the signal from different antibodies, the membrane was cropped or washed with stripping buffer (1.5% glycine, 0.1% SDS, 1% Tween 20, pH 2.2).

3.17. RNA secondary structure predictions

Predictions of secondary structures formed by single-stranded RNA sequences were performed with the use of RNAfold WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) which predicts RNA structures based on the minimum free energy and base pair probabilities.

3.18. Statistics and reproducibility

All data presented in this thesis were processed and analyzed with the use of Microsoft Excel. Statistical analysis was done using an unpaired two-tailed Student's t-test. Tests that resulted in $p < 0.05$ have been reported to be statistically significant. The symbols; *,

, *, **** represent values of $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p<0.0001$, respectively. All data were analyzed using Prism software version 8 (GraphPad). Error bars represent standard deviation (SD). All *in cellula* and *in vitro* experiments were repeated at least three times with similar results.

4. RESULTS

Due to the differentiated aims of the study, the results will be divided into two main parts. The first one will concern the formation of R-loop structures within *FMR1* 5'UTR, in the premutation conditions, and their role in the *FMR1* transcription regulation. Also, the putative therapeutical potential of antisense oligonucleotides targeting CGG repeats involved in R-loop structures, in both FXTAS and FXS conditions will be presented (Subchapter 4.1). The second part of this chapter will focus on the *cis*-regulatory elements within *FMR1* 5'UTR and their influence on the FMRpolyG RAN translation efficiency (Subchapter 4.2).

4.1. R-LOOP FORMED OVER EXPANDED CGG REPEATS WITHIN *FMR1* 5'UTR IS DRUGGABLE TARGET FOR ANTISENSE OLIGONUCLEOTIDES IN FXTAS BUT ONLY PARTIALLY IN FXS

Fragile-X-linked syndromes are caused by the expansion of CGG trinucleotide repeats within 5'UTR of the *FMR1* gene. One of the gaining importance pathomechanism involved in FXTAS etiology is the formation of R-loops in the *FMR1* 5' leader sequence containing expanded CGG repeats in the premutation (PM) range of 55-200. Importantly, it has been suggested that R-loops formed also over expanded CGG repeats in full mutation (FM) conditions (over 200 CGG repeats) are involved in the silencing of the *FMR1* gene[125] leading directly to RNA- and protein-loss-of-function in FXS.

Nowadays, it is well known that R-loops, these untypical RNA:DNA hybrid structures, are widely spread through the whole genome, however, their role in particular locus is often unknown. This is the case with the *FMR1* gene which although possesses many features of the R-loop forming sequence and it has been confirmed that CGG repeats are indeed involved in R-loop formation[126,125], very little is known about their function.

One conserved feature of R-loops is their transient nature, however, when their removal is disturbed the presence of R-loops in gene bodies or promoters can interfere with transcription leading to polymerase stalling[172,173]. Hence, the failure in the R-loops removal is directly linked with the genomic stress[136,174]. It is in agreement with the studies presenting that deficiency in enzymes dissolving R-loop structures like RNase H1, senataxin, DExH-Box Helicase 9 (DHX9, also known as NDH II and RHA), or DEAD-box helicase 5 (DDX5) induces global R-loop accumulation and increases genomic instability[175,173,176].

Although the pathomechanisms driven by R-loops in PM and FM carriers are different, R-loops appear to be the underestimated targets for novel therapeutic approaches. The main goal of this part of my doctoral project was to: *(i)* develop an *in vitro* model for studying R-loops formed by the sequence containing CGG repeats within *FMR1* 5'UTR, *(ii)* examine their role in the context of *FMR1* transcription regulation, and *(iii)* establish whether antisense oligonucleotides targeting CGG repeats may modulate the stability/resolve R-loops formed within *FMR1* 5'UTR with PM and FM.

4.1.1. *In vitro* study of R-loops formation in 5'-part of *FMR1*

4.1.1.1. In vitro R-loop formation assay

Described in the literature protocols for *in vitro* R-loop detection base on the *in vitro* transcription of plasmid DNA containing R-loop forming sequence under the T7 promoter followed by organic extraction and electrophoresis in agarose gel[127,177].

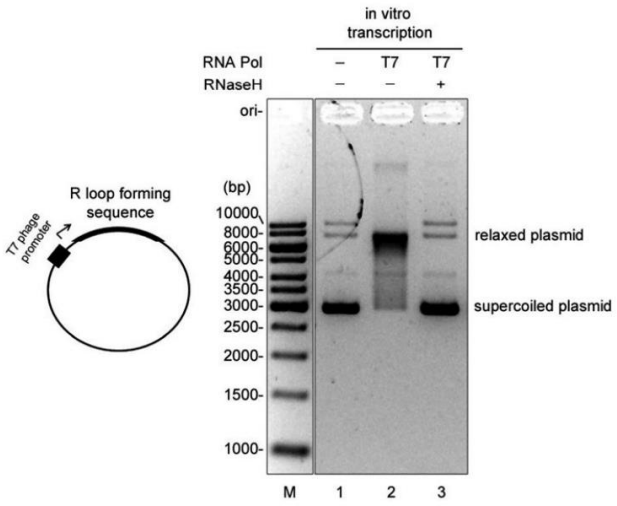


Figure 12. In vitro R-loop formation assay. *In vitro transcription was carried out on the plasmid DNA containing the putative R-loop forming sequence under the T7 promoter. The reaction mix was then run on the 1% agarose gel and stained with ethidium bromide. A shift of the bands present in lane 2 demonstrates the stable interaction between template DNA and the transcribed RNA. To confirm the presence of RNA:DNA hybrids in the shifted band the reaction was performed in the presence of the bacterial recombinant RNase H (lane 3). Lane 1 constitutes a control where a mock incubation without T7 polymerase was analyzed. M, 1 Kb DNA ladder. Figure from ,,Promoter Associated RNA. Chapter 13 – Detection and Characterization of R Loop Structures." (Raquel Boque-Sastre, Marta Soler, and Sonia Guil, 2017, p. 236 – 237)*[127].

Based on available protocols I wanted to confirm the formation of R-loops in the *FMR1* 5′UTR during *in vitro* transcription of plasmid DNA template. Thus, the plasmid containing 5′UTR of *FMR1* with ~100 CGGs (5′(CGGexp)-GFP(+1)) under T7

polymerase promoter was used as a template for *in vitro* transcription (***see Methods 3.14. "The first approach"***). To confirm the presence of R-loop structures the samples after transcription were treated with RNase H. To avoid potential problems with visualization and false-positive interactions between nucleic acids, RNase A, which specifically degrades single-stranded RNA at C and U residues, was added after transcription to digest excess RNA.

Obtained results (***Figure 13***) confirmed that within the 5'UTR of *FMR1* RNase H-sensitive R-loops are formed (the R-loop smear and the bands' mobility shift between RNase H – treated and non-treated lanes). The optimization of RNase A concentration was performed in the range between 5 to 50 ng/µl, and the 12.5 ng/µl concentration was chosen as the most optimal.

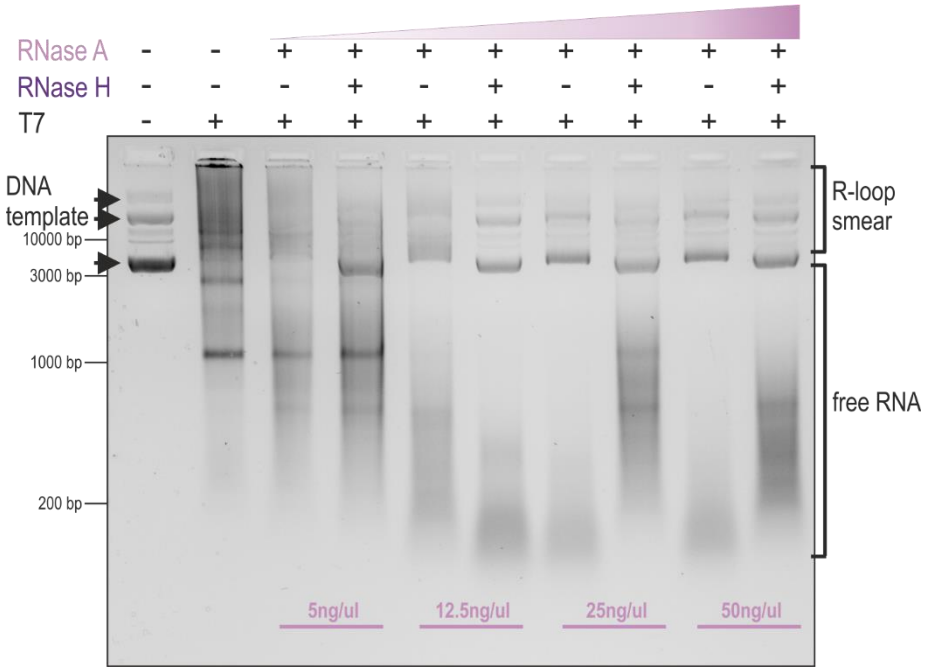


Figure 13. Detection of R-loops formed within FMR1 5'UTR during in vitro transcription. *In vitro transcription was performed on the 5'(CGGexp)-GFP(+1) plasmid DNA. The reactions were performed at 37°C for 2 h and then were treated with 1 µl of RNase A (5 ng/µl; 12.5 ng/µl; 25 ng/µl or 50 ng/µl) and 10 U of RNase H (+) or equal amount of 50% glycerol in untreated samples (-) for an additional 30 min at 37°C. Reactions were stopped and nucleic acids were extracted with phenol-chloroform and precipitated with 96% ice-cold ethanol and 3 mM sodium acetate. Pelleted nucleic acids were resuspended in the water with 6 x BLUE DNA Loading buffer and analyzed on 1% agarose gels with ethidium bromide. The R-loop smear represents the interaction between template DNA and the transcribed RNA which is removed after RNase H treatment.*

4.1.1.2. Development of a new assay for in vitro R-loop detection

Although I successfully confirmed the formation of R-loops *in vitro* on 5′(CGGexp)-GFP(+1) plasmid DNA I was skeptical about the reproducibility and accuracy of this method. The usage of the circular template during *in vitro* transcription could result in false-positive results since R-loops could be formed within other (out of interest) regions of the vector backbone. Besides, using the circular template leads to that transcription occurs continuously without termination and therefore an extremely long transcript is produced. This situation has nothing in common with native transcription which occurs in the cells. Lastly, a step using the organic extraction to stop the reaction and further precipitate nucleic acids introduces the discrepancy between samples. Due to all mentioned reasons, I decided to investigate whether another approach could be used to detect and monitor R-loops formation *in vitro*. To achieve this goal I performed *in vitro* transcription on the 5′(CGGexp)-GFP(+1) construct linearized by *Avr*II restriction enzyme. This enzyme recognizes and digests the sequence at the end of the *FMR1* 5′UTR and thus ensures that observed R-loops are formed within the region of interest and the termination of transcription occurs. Additionally, reactions were stopped by the addition of 60 mM EDTA and then directly loaded on the agarose gel without the organic extraction step. A detailed description of both approaches is present in the Methods section.

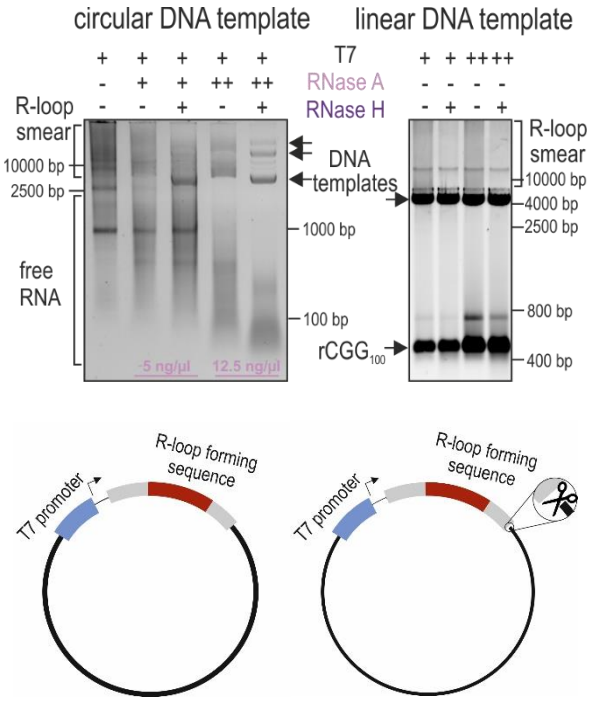


Figure 14. Visualization of R-loops in 5'-part of FMR1 by two in vitro approaches. *(Left) In vitro transcription was performed using the non-digested 5'(CGGexp)-GFP(+1) genetic construct (circular plasmid). The reactions were performed as in* ***Figure 13****; (Right) In vitro transcription was performed on the 5'(CGGexp)-GFP(+1) genetic construct digested with AvrII enzyme recognizing restriction site downstream putative R-loop forming sequence and CGGexp (allow for transcription of whole exon 1 of FMR1). The reactions were performed using different amounts of T7 RNA polymerase ("+" or "++") at 37°C for 20 min in the presence of 10 U of RNase H (+) or 50% glycerol in the untreated samples (-). The synthesized RNA – rCGG$_{100}$ – containing the whole region of FMR1 5'UTR with 99 CGG repeats, is marked. A significant decrease of R-loop smear signal intensity upon RNase H addition can be observed. The schemes of DNA templates are presented below the appropriate gels.*

It turned out that this approach could be successfully used to detect and analyze R-loops forming *in vitro* and therefore to quantify the efficiency of *FMR1* transcription measured as the amount of synthesized RNA (**rCGG$_{100}$**) which contained 99 CGG repeats flanked by 132 nt upstream, and 69 nt downstream of a full native sequence of *FMR1* 5'UTR, and 51 nt of FMRP coding sequence (***Figure 14, right panel***). The usage of digested plasmid as a template for *in vitro* transcription abolished the risk of observing R-loops formed on the random sequence from the vector backbone as well as the variations in the sample signals intensities due to sample extraction biases. Thus, all *in vitro* transcriptions were performed on the properly digested plasmid. However, for efficient and specific R-loops digestion the concentration of RNase H had to be optimized. Hence, *in vitro* transcription was performed in the presence of antisense oligodeoxynucleotides designed to induce the degradation of target RNA *via* RNase H (**Gapmers**) and different amounts of RNase H (***Figure 15a***). The RNase H should cleave the RNA moiety within RNA:DNA heteroduplexes, thus 60 nt shorter rCGG$_{100}$ RNA should be detected after efficient digestion. Indeed, the usage of 10 U of RNase H per reaction was established as the most optimal (1 U/µl) since this concentration resulted in the rCGG$_{100}$ band shift migration (marked as an asterisk).

However, due to the divergent structural character of *FMR1* 5'UTR, the chosen concentration was validated using two gapmers targeting different regions within *FMR1* 5'UTR (***Figure 15b***). Both experiments confirmed that independently from the localization of RNA:DNA hybrid, the concentration of RNase H equal 1 U/µl was efficient for RNA degradation within RNA:DNA heteroduplexes (rCGG$_{100}$ band shift migration marked as an asterisk). Therefore, all further *in vitro* transcriptions were performed on linearised 5′(CGGexp)-GFP(+1) plasmid DNA and treated with an established amount of RNase H.

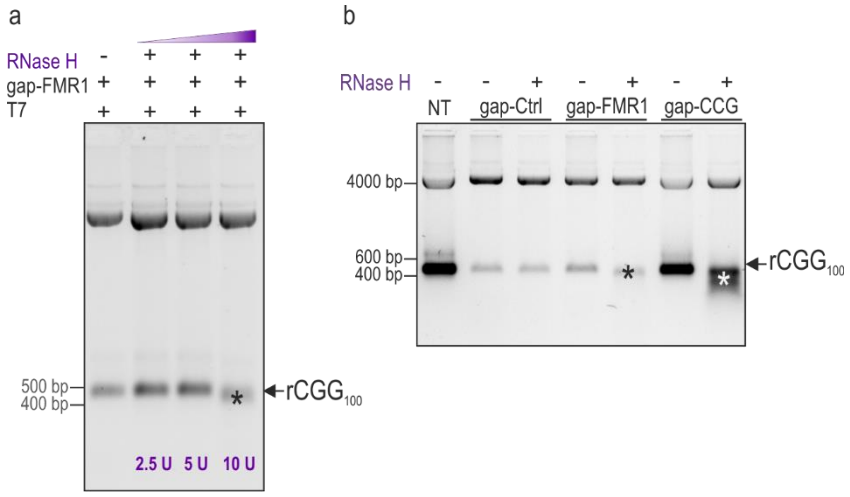


Figure 15. Optimization of RNase H concentration. *In vitro transcription was performed on the 5'(CGGexp)-GFP(+1) genetic construct linearized by AvrII;* ***a)*** *The reactions were performed at 37°C for 2 h in the presence of different amounts of RNase H (marked in purple, +) or 50% glycerol in the untreated samples (-), and 1μM Gapmer targeting region within 5'UTR of FMR1 (gap-FMR1). The product of RNase H digestion, shorter by 60 nt, is marked by an asterisk;* ***b)*** *The reactions were performed at 37°C for 90 min in the presence of 1μM of appropriate Gapmer (gap-Ctrl, gap-FMR1 or gap-CCG-targeting CGG repeats) followed by digestion with 10 U of RNase H at 37°C for 30 min. Reactions were stopped by the addition of buffer containing 60 mM EDTA and analyzed on 1% agarose gel. The products of RNase H digestion are marked by asterisks.*

4.1.1.3. R-loops formed within FMR1 5'UTR in FXTAS are disturbed by ASO-CCG and influence the FMR1 transcription efficiency

Since the system for R-loops detection *in vitro* has been established I wanted to analyze the effect of R-loops formation within *FMR1* 5'UTR on the *FMR1* transcription efficiency. To achieve this goal the *in vitro* transcription was performed in the presence of 50% glycerol (-) or RNase H (+) (***Figure 16, left panel***). Obtained results showed that the synthesis of transcripts containing ~100 CGG repeats ($rCGG_{100}$) significantly increased in the presence of RNase H. This data confirm that R-loops formed within *FMR1* 5'UTR negatively regulate *FMR1* transcription potentially due to structural block formation which can directly interfere with T7 transcription elongation leading to blockage of the further rounds of transcription. As expected, the reduction in the signal coming from the R-loop smear in lanes treated with RNase H is also visible.

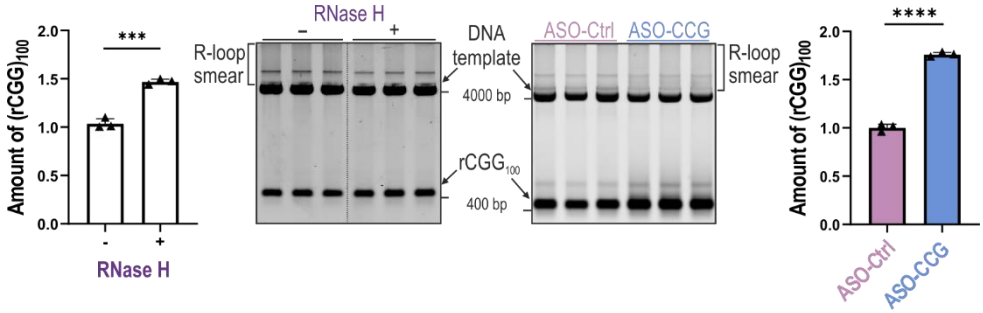


Figure 16. Increase in the efficiency of in vitro transcription in the presence of RNase H or ASO-CCG. *(Left) Quantification of the amount of $rCGG_{100}$ and the corresponding gel in the presence of RNase H; (Right) Quantification of the amount of $rCGG_{100}$ and the corresponding gel in the presence of 2.5 μM ASO-Ctrl or ASO-CCG. $rCGG_{100}$ signal was measured for N = 3 independent samples and normalized to the signal of the DNA template for each lane. The photo of the left gel was cropped. Statistical analysis was based on a two-tailed unpaired Student's t-test; ***, p<0.001; ****, p<0.0001.*

Since it is known that CGG repeats are involved in R-loops formation within *FMR1* 5'UTR I wanted to check whether targeting CGG repeats by antisense oligonucleotides (**ASOs**) would have some, potentially negative, effect on the efficiency of R-loops formation. For this purpose, I used ASO targeting CGG repeats (ASO-CCG) which were designed to act as **steric blockers** and, contrary to previously used Gapmers, did not induce degradation of target RNA. These **l**ocked **n**ucleic **a**cids (**LNA**) had modified **p**ho**s**phorothioate oligonucleotides' linkages (**PS**) that made them resistant to cellular nucleases. ASOs containing exclusively LNA units disable recruitment of RNase H to the RNA/LNA duplex and induction of RNA cleavage (reviewed in [178]). I hypothesized that the transcription efficiency of *FMR1* upon ASO-CCG addition would increase as the result of a decreased amount of formed R-loops or reduced thermodynamic stability of formed R-loops. Indeed, after *in vitro* transcription in the presence of ASO-CCG a significant increase in the $rCGG_{100}$ was observed (***Figure 16, right panel***). However, against the assumption, no reduction of R-loop smear after ASO-CCG treatment could be detected. This would suggest that targeting CGGexp by ASO-CCG did not significantly reduce the amount of R-loop structures but more likely affected their thermodynamic stability leading to increased efficiency of transcription through CGG repeats.

4.1.1.4. ASO-CCG binds directly to both RNA and DNA within R-loops formed over CGG repeats and positively regulates FMR1 transcription

ASO binding influences the detection signal of bound nucleic acids using EtBr staining or other staining methods, as was already shown in another study[179]. Hence, only

transcript yield as the indirect consequence of R-loops regulation by low concentration of ASO-CCG could be analyzed. Due to the aforementioned issues, the *in vitro* detection system has been changed to more quantitatively measure the effect of ASO on the R-loop structures.

To confirm the direct interaction between ASO-CCG and R-loops formed within *FMR1* 5'UTR I performed *in vitro* transcription in the presence of ASO-CCG or ASO-CCG labeled with Cy3 (***Figure 17***). The use of fluorescently labeled ASO-CCG-Cy3 allowed to monitor different RNA and DNA molecules containing CGG repeats if bound with ASO. When *in vitro* transcription was performed in the presence of ASO-CCG-Cy3 the observed signal was coming from the Cy3, thus I knew where ASOs have bound. As it is presented in ***Figure 17*** ASO-CCG-Cy3 binds to three different molecules/complexes in performed reactions: $rCGG_{100}$, ssDNA region containing CGG repeats of the DNA template, and the R-loop structure. A signal from $rCGG_{100}$ and R-loop is sensitive to RNase A (digest all free RNA species), the signal from ssDNA and R-loop is sensitive to DNase treatment, while the signal from R-loop is sensitive to RNase H (digest RNA:DNA hybrids). In this approach the free ASO-CCG-Cy3 is visible on the gel (Cy3-specific filter used in gel scanning), however, free ASO-CCG cannot be detected (EtBr staining). Data obtained in this experiment confirm that ASO-CCG binds to CGG repeats present in both synthesized RNA as well as the coding strand of template DNA. Therefore, the stability of formed R-loops might be affected by ASO-CCG treatment.

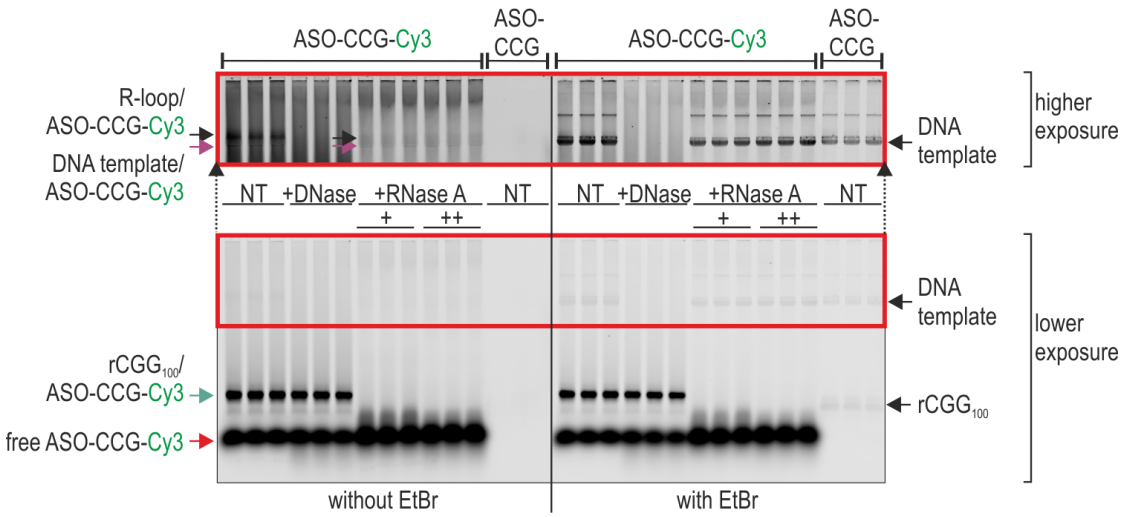


Figure 17. In vitro transcription experiment showing the interaction of ASO-CCG with R-loops and a sense strand of DNA template containing CGG repeats. *In vitro transcription was performed on the ~500 ng of 5'(CGGexp)-GFP(+1) construct digested with AvrII at 37°C for 20*

min in the presence of ASO-CCG or fluorescently labeled ASO-CCG-Cy3 (indicated above the gel). Then samples were digested with either 2 U of DNase TURBO or 1 µl of RNase A („+", 0.3 µg/µl; „++", 5 µg/µl) for 30 min on ice. Control samples were treated with 50% glycerol (NT). To visualize CGG_{100}-containing nucleic acids the agarose gel was firstly scanned using Amersham Typhoon RGB Biomolecular Imager and the fluorescent signal coming from ASO-CCG-Cy3 was detected using Cy3 filter. Then gel was stained with ethidium bromide (0.5 µg/ml) for 20 min and scanned again to visualize DNA templates and all RNA products. The area of gel marked in red is presented above with higher exposition. The experiment was repeated 3 times with similar results.

As the direct interaction between R-loops and ASO-CCG was confirmed I wanted to use ASO-CCG-Cy3 based *in vitro* transcription to establish how R-loops removal would influence the *in vitro* transcription efficiency. In other words how R-loops affect the transcription measured by the $rCGG_{100}$ amount.

In the case of the experiment with ASO-CCG-Cy3 (***Figure 18***), the observed signal came from the Cy3, with or without RNase H pressure. The signal coming from R-loop structures was significantly reduced upon RNase H treatment, which confirms that they are sensitive to RNase H digestion and that the concentration of the used enzyme was sufficient. The addition of RNase H to *in vitro* transcription also significantly increased the amount of $rCGG_{100}$, which confirms that R-loops formed over expanded CGG repeats within *FMR1* 5'UTR have a negative effect on the transcription efficiency.

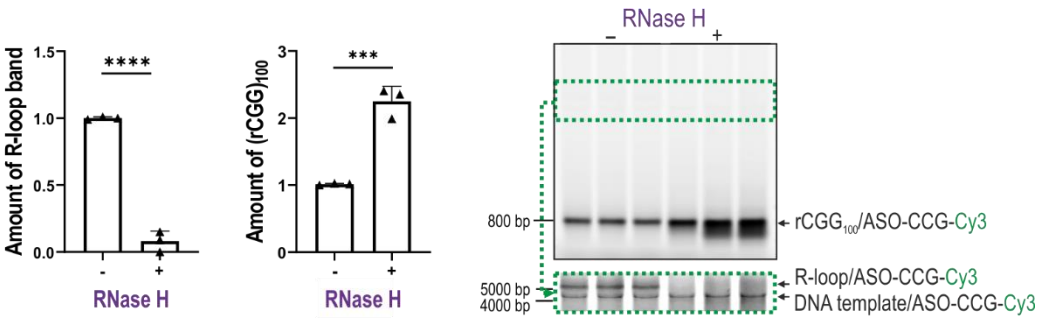


Figure 18. The digestion of R-loops results in the increased transcription of $rCGG_{100}$. *Graphs present quantification of the fluorescent signal coming from fluorescently labeled ASO-CCG-Cy3 bound to either $rCGG_{100}$ or R-loop structure for N = 3 independent samples. Results were normalized to the intensity of the signal coming from a DNA template in the same lane. The area of the gel marked with a green box was exposed to higher laser power and the result is presented below the gel. DNA template was the same as in* ***Figure 16****, but the observed signal came only from partially single-stranded DNA in the region of CGG/CCG repeats or R-loop bound with ASO-CCG-Cy3. Statistical analysis was based on a two-tailed unpaired Student's t-test; ***, $p<0.001$; ****, $p<0.0001$.*

Taken together data presented in this section demonstrate that R-loops are formed within the 5' leader sequence of *FMR1* in *in vitro* conditions and that CGG repeats are involved

in this RNA:DNA hybrid formation. In addition, ASO-CCG binds to RNase H-sensitive R-loops and to the sense strand of DNA template containing CGG repeats, which as a consequence can reduce the stability of formed R-loops leading to an increase in the *FMR1* transcription rate.

Based on these results I proposed a model (***Figure 19***) of how ASO-CCG interact with R-loops and lead to their destabilization followed by increased transcription. The model assumes that R-loops formed over CGG/CCG repeats within the 5'-end of *FMR1* are not canonical R-loop structures since it has been already published by Loomis and co-workers in 2014[126] that displaced sense strand of DNA template is involved in a hairpin or other secondary structure formation. It is also well known that RNA with CGG repeats can form very stable hairpin or G-quadruplex structures[180,181], hence, I proposed a model in which all regions of DNA as well as RNA containing either CGG or CCG repeats fold into stable secondary structures. The hairpin (or other structural conformation e.g. G-quadruplex) formed by CGG repeats within nascent transcript may interact with the structure formed by CGG repeats within displaced DNA strand especially that it has been reported that this region is prone to fold into G-quadruplexes and other secondary structures[182,183,184]. Such interaction can stabilize the R-loop formed in this locus leading to the inhibition of transcription. This effect is probably weakened when transcription occurs in the presence of ASO-CCG due to the destabilization of either one or both structures formed by DNA strand and RNA molecules. This model can explain the increase in transcription efficiency observed in the presence of ASO-CCG.

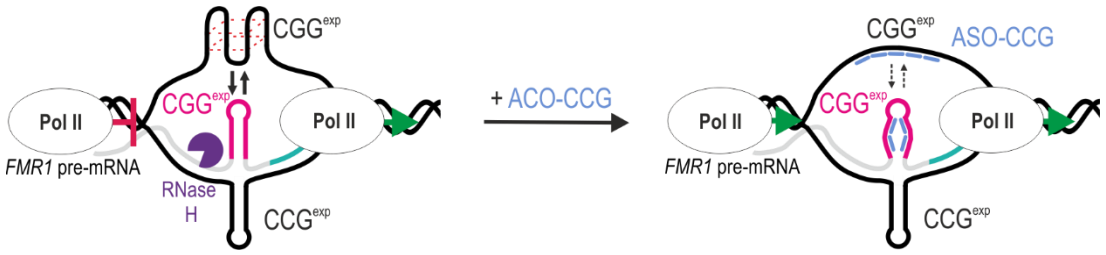


Figure 19. Proposed model of how ASO-CCG invades R-loop structure. *ASO-CCG binds to CGG – containing nucleic acids (RNA and sense strand of DNA template) leading to destabilization of all potentially formed structures by these molecules and weakening the interaction between them. Therefore, it can be supposed that reduced formation of secondary structure/s can endure their negative effect on the transcription and increase its efficiency at both initiation and elongation steps due to ASO-CCG binding.*

4.1.2. *In cellula* study of R-loops formation in 5'-part of *FMR1*

4.1.2.1. Formation of R-loops negatively regulates FMR1 transcription and ASO-CCG abolish this effect in cellular conditions

In vitro system has many advantages as it minimize the effect of other processes operating in the cell on the studied mechanism making analysis simpler. On the other hand, it is also its biggest drawback. Thus, I decided to check whether similar results considering ASO-CCG regulation of R-loops stability and therefore the increase in the *FMR1* yield would be observed *in cellula*.

4.1.2.1.2. Selection of factors regulating R-loops maintenance

In the beginning, I focused on identifying potential factors that would regulate the R-loops formation in cells. The gold standard in experiments studying R-loops removal is RNase H1[131,137,185,172], however, since the RNase H2 has also the capacity of R-loop resolution[186,175] both enzymes have been tested. Another promising factor involved in R-loop regulation was DHX9, which is an RNA helicase A having activity to resolve R-loops[187,188,189,190,176].

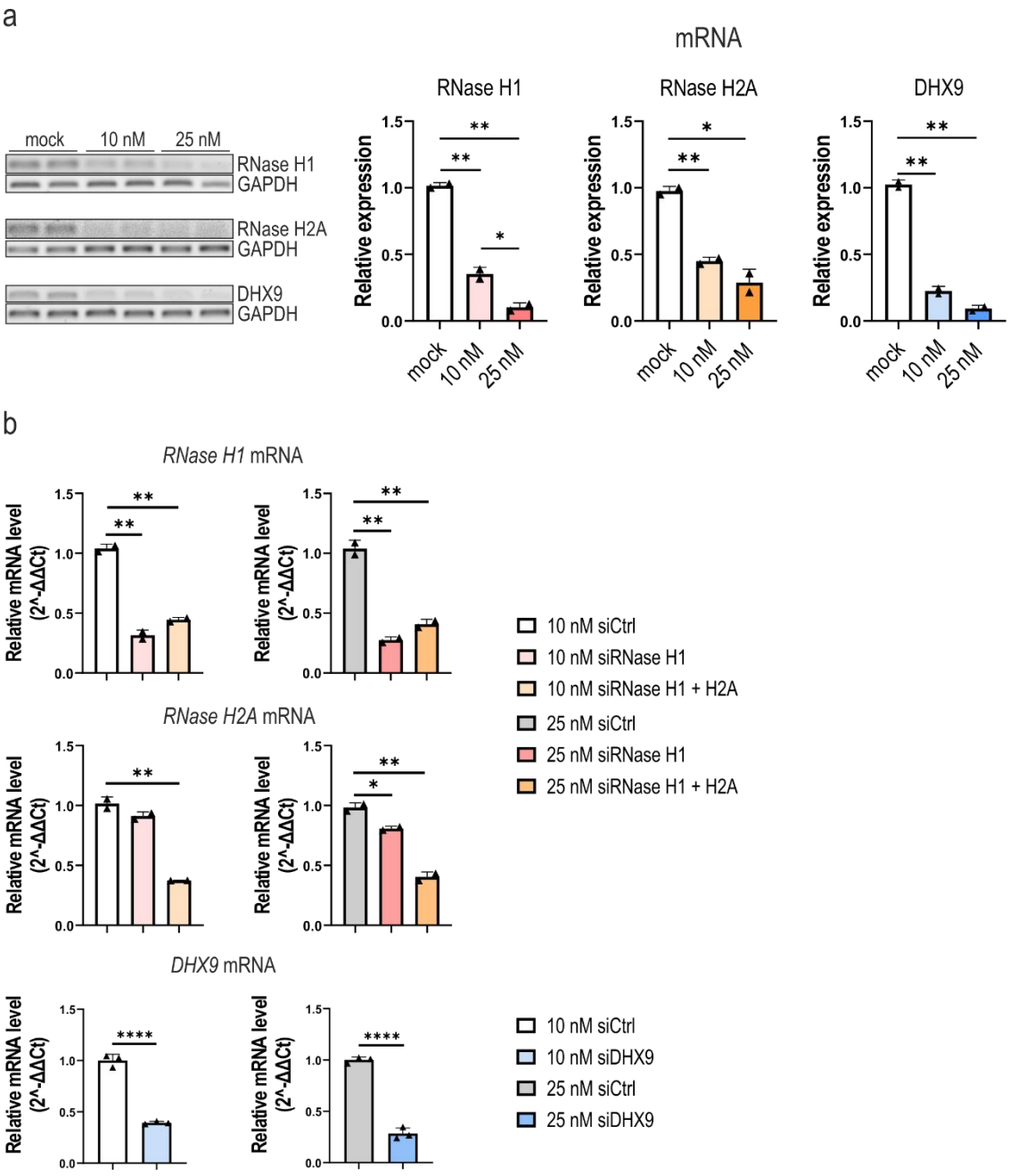


Figure 20. The siRNA knockdown efficiency. a) *CGG^{norm}/- (2) control fibroblasts were transfected with siRNA against RNase H1, RNase H2A, or DHX9 and harvested 48 h post-transfection. RT-PCR analysis was performed with primers specific to mRNA. The values shown in the graphs are the means of N = 2 biologically independent samples, with the SDs;* ***b)*** *The same as in* ***a*** *but for CGG^{exp}/CGG^{exp} fibroblasts. RT-qPCR analysis was performed with primers specific to mRNA. Graphs present results for N = 2 biologically independent samples (n = 2 technical replicates) with SDs for RNase H insufficiency, and N = 3 biologically independent samples (n = 2 technical replicates) with SDs for DHX9 insufficiency. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01, ****, p<0.0001.*

To study the effect of mentioned modulators on the R-loop formation/stability within *FMR1* 5'UTR measured by the yield of *FMR1* mRNA, firstly, the efficiency of RNase H1, RNase H2A, and DHX9 knockdown *via* specific siRNAs has been tested (***Figure***

20). For this purpose, control fibroblasts (CGG^{norm}/- (2); 31 CGG repeats) were transfected with appropriate siRNA and collected after 48 h. The efficiency of knockdown was tested by RT-PCR using primers designed to RNase H1, RNase H2, and DHX9 mRNAs. All three tested siRNAs efficiently reduced the expression of examined R-loops' modulators. Thus, in the following step, I checked how the diminished levels of RNase H1, RNase H2A, and DHX9 would affect R-loops and if R-loops accumulation would occur whether a decrease in the *FMR1* transcription could be observed.

Since CGG repeats are involved in R-loop formation the cell line harboring longer, expanded CGG repeats was chosen for this experiment. Fibroblasts from homozygotic premutation carrier (CGG^{exp}/CGG^{exp}; two alleles containing 60 and 90 CGGs) were transfected with appropriate siRNA (10 nM or 25 nM) and harvested 48 h after transfection. The double knockdown of RNase H1 and RNase H2A was also performed to check if an additive effect on *FMR1* mRNA level would be observed. After total RNA isolation and reverse transcription (RT) the level of *FMR1* mRNA as well as two pre-mRNA regions have been analyzed by RT-qPCR (***Figure 21b***). The siRNA knockdown efficiency (***Figure 20b***) and *FMR1* pre-mRNA and mRNA levels were quantified with the use of primers whose location is presented in ***Figure 21a***.

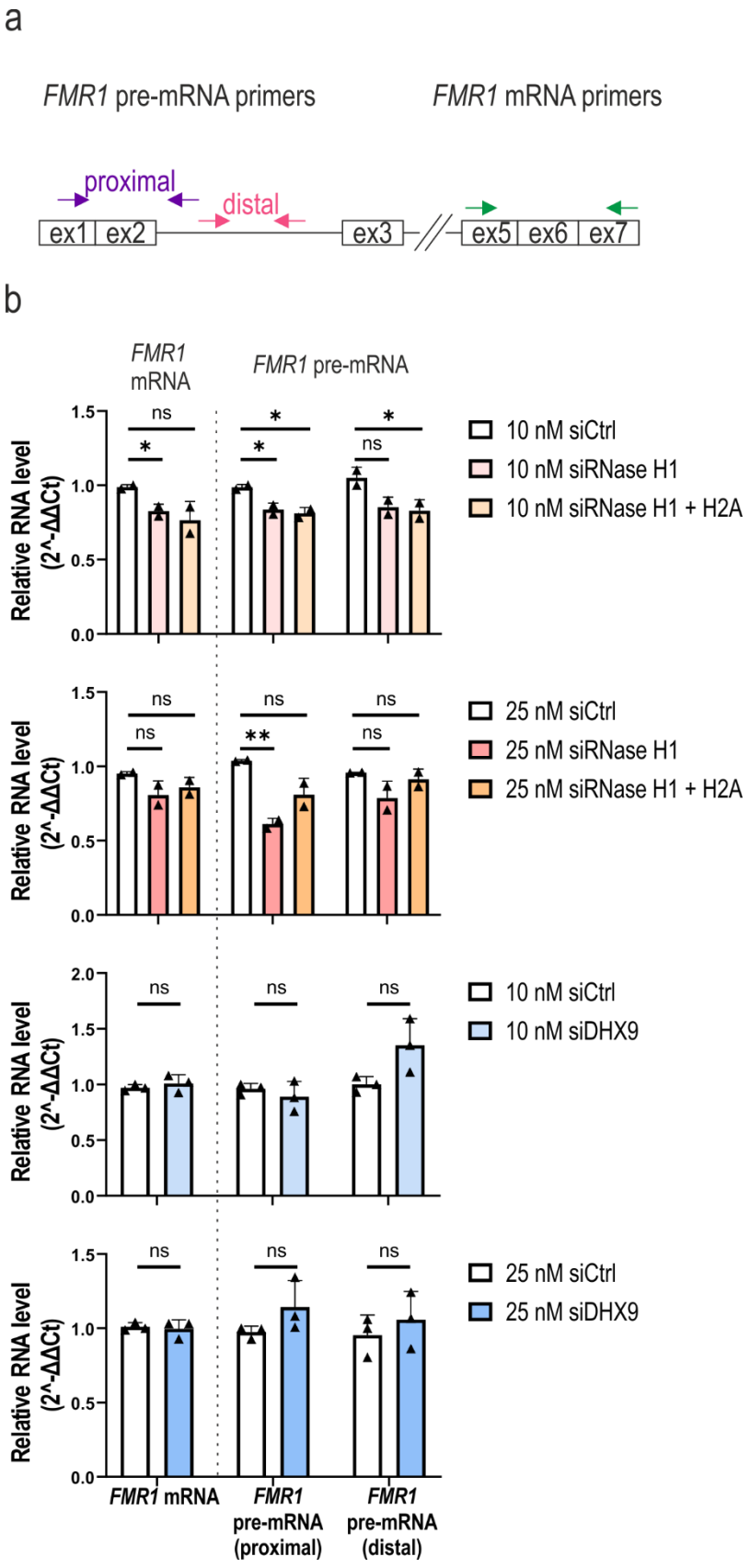


Figure 21. R-loops accumulation in cellula results in a decrease in FMR1 transcription. a) *The localization of primers used for FMR1 pre-mRNA and mRNA analysis. The assay for the proximal part of pre-mRNA enables detection of partially spliced pre-mRNA, after excision of intron 1;* ***b)***

Analysis of the total FMR1 mRNA and pre-mRNA levels in CGGexp/CGGexp fibroblasts, isolated from proband carrying double PM alleles, treated with siRNA against RNase H1, both RNase H1 and RNase H2A or DHX9. Fibroblasts were transfected with ***two concentrations of siRNA*** *and harvested 48 h post-transfection, followed by RNA isolation and RT-qPCR analysis. The efficiency of siRNA knockdown is presented in* ***Figure 20b****; Statistical analysis was based on two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ns, non-significant.*

Obtained results show that only the depletion of RNase H1, which is a factor regulating the R-loop formation, influenced the level of *FMR1* mRNA and pre-mRNA. A decline in the *FMR1* pre-mRNA level suggests that R-loops are indeed formed co-transcriptionally and their accumulation resulting from RNase H1 depletion negatively regulates *FMR1* transcription already at the level of unspliced or partially spliced pre-mRNA. Surprisingly, the double knockdown of RNase H1 and RNase H2A did not show an additive effect on the *FMR1* transcript reduction. The knockdown of DHX9 probably did not affect the R-loop formation within *FMR1* 5'UTR (at least R-loops involved in transcription) since the level of *FMR1* mRNA as well as pre-mRNA remained unchanged. Hence, for further experiments, the RNase H1 was chosen as the factor negatively regulating R-loop maintenance in the *FMR1* locus.

4.1.2.1.3. FMR1 transcription is regulated by ASO-CCG in the context of RNase H1 insufficiency

I hypothesized that R-loops formed over expanded CGG repeats may form a structural block, directly interfering with RNA Polymerase II (Pol II) transcription elongation or initiation, leading to blockage of the further rounds of transcription, and a decrease in the overall transcription efficiency. The effect should correlate positively as the CGG repeats expand, up to 200 CGG repeats (full mutation and *FMR1* silencing).

It has been shown for the Friedreich ataxia syndrome where R-loops form over expanded GAA repeats and trigger transcriptional silencing of *FXN* gene[125] that administration of antisense oligonucleotides targeting GAA repeats restored expression of *FXN* by preventing R-loops formation[191]. Hence, I decided to test whether ASO targeting directly CGG repeats may act similarly and will exert a similar effect on endogenous *FMR1* mRNA and pre-mRNA in human fibroblasts *via* R-loop destabilization and/or prevention of R-loop formation. I used four fibroblast cell lines derived from healthy individuals and FXTAS patients: from two men with a normal range of CGG repeats (20 CGG repeats; (CGG^{norm}/- (1); 31 CGG repeats (CGG^{norm}/- (2)), and from two women carrying alleles with CGGexp (20 and 79 CGG repeats; (CGG^{norm}/CGG^{exp}), and 60 and 90 CGG repeats

(CGG^{exp}/CGG^{exp})). Since R-loops are nucleic acid structures sensitive to RNase H1, which cleaves RNA moiety within RNA:DNA hybrids I decided to measure the level of *FMR1* mRNA and pre-mRNA in the same fibroblasts upon the RNase H1 insufficiency, in other words in conditions mimicking R-loops accumulation. Thus, cells were transfected with siRNA against RNase H1 or siCtrl, and after 24 h second transfection with 200 nM ASO-Ctrl/ASO-CCG was performed. Cells were harvested 48 h post-1st-transfection and total RNA was isolated followed by RT-qPCR analysis.

As presented in ***Figure 22*** ASO-CCG significantly increased the level of *FMR1* mRNA in all cells except those with allele containing short, 20 CGG repeats (CGGnorm/- (1)). In the majority of cells, the level of *FMR1* pre-mRNA was also elevated upon ASO-CCG treatment. These results could represent a sum of a few mechanisms leading to the increase of the *FMR1* level. Nevertheless, based on results from my *in vitro* studies, I assumed that this increase in transcription efficiency was partially caused by ASO–CCG-based prevention and/or destabilization of R-loop structures within the 5' leader sequence of *FMR1*. Hence, unchanged *FMR1* level upon ASO-CCG in cells with 20 CGG repeats suggests that RNAs with such short repeats are not substrates for R-loops formation or that the effect of ASO-CCG is marginal on the formation of this structure.

Therefore I tested whether R-loops accumulation *via* RNase H1 depletion would result in a drop in *FMR1* transcription. In three tested cell lines CGG^{norm}/- (2), CGG^{norm}/CGG^{exp}, and CGG^{exp}/CGG^{exp} treated with siRNA against RNase H1 (grey zone in ***Figure 22***), the increase in the *FMR1* mRNA level induced by ASO-CCG was greatly diminished compared to control siRNA-treated cells. Moreover, the pre-mRNA level was unchanged in these conditions.

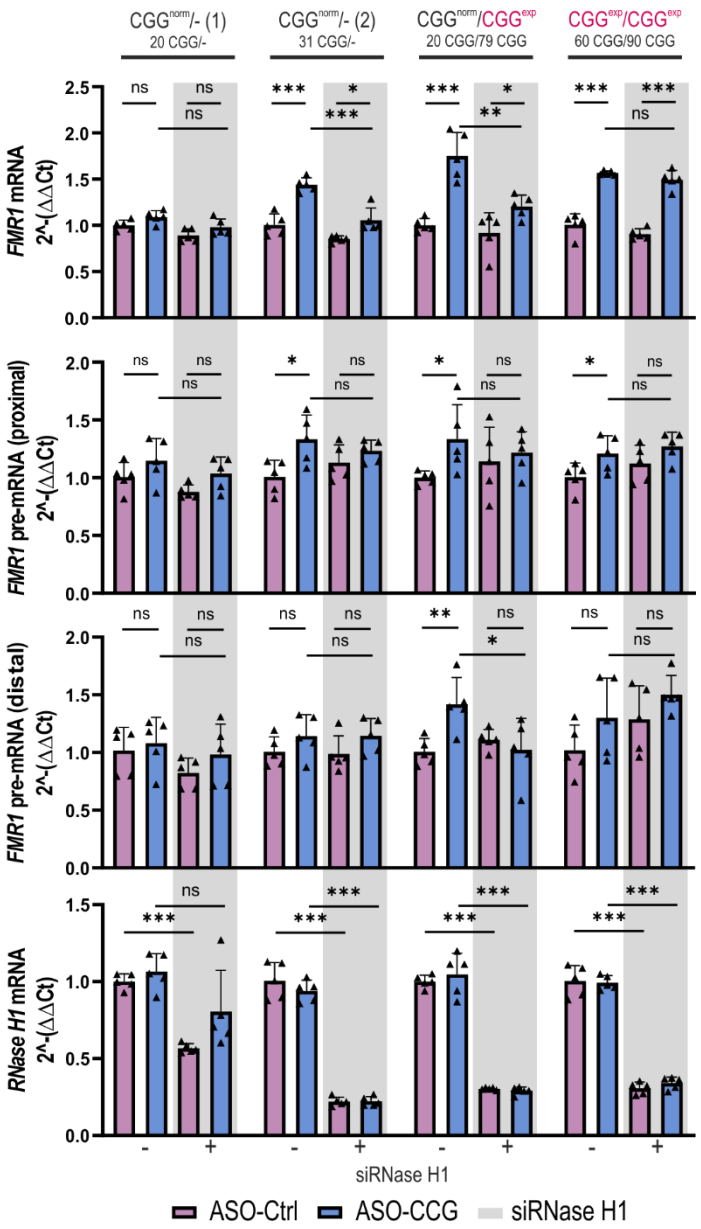


Figure 22. The effect of ASO–CCG on the FMR1 pre-mRNA and mRNA in cellula. *Analysis of the total FMR1 mRNA and pre-mRNA levels in fibroblasts treated with ASOs (RT-qPCR assays were the same as described in* ***Figure 21b****). Fibroblasts were transfected with siRNA against RNase H1 or control siRNA and with 9 nt long, 200 nM ASOs. Graphs present RT-qPCR results for N = 5 biologically independent samples (n = 2 technical replicates) with SDs. RT-qPCR analysis was performed with primers specific for the FMR1 mRNA, two regions of FMR1 pre-mRNA (proximal and distal), and RNase H1 mRNA. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ***, p<0.001; ns, non-significant.*

In conclusion, these results suggest that ASO-CCG may enhance the transcription efficiency of the *FMR1* at least partially by targeting R-loops formed over long CGG repeats, which confirms that these RNA:DNA hybrids negatively regulate *FMR1* transcription.

4.1.3. ASO-CCG is able to enhance *FMR1* transcription in FXS cells with partially active mutant gene

The expansion of CGG repeats within the *FMR1* 5'UTR to more than 200 leads to fragile X syndrome (FXS) development. FXS patients are characterized by full mutation (FM) which results in complete loss of FMRP production due to *FMR1* hypermethylation. Although, there are few proposed mechanisms that may lead to *FMR1* silencing one is especially reliable and interesting in the light of results described in previous subchapters. Colak and colleagues[146] performed studies in which they presented that: *1)* the methylation of *FMR1* promoter in FXS occurs only in the presence of *FMR1* transcript; *2)* the lack of *FMR1* silencing in normal and premutation carriers is a consequence of the lack of *FMR1* mRNA binding to the promoter; *3)* there is a direct binding between *FMR1* pre-mRNA and *FMR1* gene fragment. Based on these data it has been suggested that the methylation of *FMR1* promoter in FXS patients is correlated with the R-loop formation. The involvement of R-loops in the *FMR1* methylation was also suggested in another study presenting a similar mechanism for FRDA silencing[125].

A growing body of evidence has been presented for the potential restoration of *FMR1* transcription using CRISPR tools[192] and other approaches[156,193]. Results of my study presented above indicate that ASO-CCG binds to the nascent RNA and sense strand of DNA during *in vitro* transcription and invades the R-loop structure leading to increased transcription efficiency. Therefore, I wanted to check whether ASO-CCG may affect the structural organization of expanded repeats from a FM range (above 200 CGGs) involved in R-loops formation and reduce hypermethylation of the *FMR1* locus leading to transcription reactivation in FXS conditions. Hence, I used fibroblasts derived from FXS patients with *FMR1* hypermethylation[168,169]. I assumed, that long ASO-CCG treatment could result in the resolution of formed R-loops, and if they are involved in *FMR1* silencing in FXS the transcription from *FMR1* should be, at least partially, restored.

As a positive control, the cells were firstly treated with a DNA methylation inhibitor, the 5-aza-2'-deoxycytidine (**5-azadC**), which was used in other studies and resulted in the *FMR1* expression restoration on mRNA level in FXS cells up to 25% of control cells[125]. In this study two primary cell cultures were used: the FX08-01 and FX13-01 fibroblasts, which were isolated from FXS male individuals. The *FMR1* CGG length analysis revealed that derived cells had >435 CGG repeats[169] meaning that the length of repeats

could be longer and the threshold of 435 repeats represents the limit of the method utilized for genotyping. Although both fibroblast cells were described to have the reduced *FMR1* mRNA level, and loss of FMRP[169] it turned out that, in this study, contrary to FX13-01 (***Figure 23c***), in FX08-01 fibroblasts no *FMR1* mRNA was detected (***Figure 23b***). Noteworthy, the silencing of *FMR1* transcription in FXS cells correlated with the increase in the number of cell passages has been already observed for FXS hESCs and could explain the discrepancy between published data and my observations. Nevertheless, using two FXS primary cell cultures characterized by different degrees of *FMR1* silencing created an opportunity to study the role of R-loops in *FMR1* silencing in FXS. Importantly, in FX13-01, since *FMR1* transcription is partially active, the co-transcriptional formation of R-loops may occur. Based on presented above *in vitro* studies these structures can be targeted by the ASO-CCG leading to increased transcription of *FMR1*. On the contrary, in the FX08-01, due to loss of *FMR1* transcription, possibly no R-loops are formed.

Therefore, I wanted to study the putative role of R-loops formed over expanded CGG repeats on the *FMR1* silencing in FXS. Firstly, both FXS fibroblasts (FX08-01 and FX13-01) were treated for 7 days with 1μM 5-azadC followed by wash-out steps to establish whether *FMR1* transcription will be activated and if so when remethylation after 5-azadC treatment occurs (***Figure 23a***). As presented in ***Figure 23b-c*** the treatment with 5-azadC significantly reduced methylation of the promoter region of *FMR1* which can be observed as increased transcription of *FMR1* in FX13-01 and reactivation of transcription in FX08-01 cells. The highest level of *FMR1* transcript was observed in the range of 7 to 14 days (1st – 3rd time point) and from 10 to 17 days (2nd to 4th time point) after the beginning of 5-azadC treatment for FX08-01 and FX13-01, respectively.

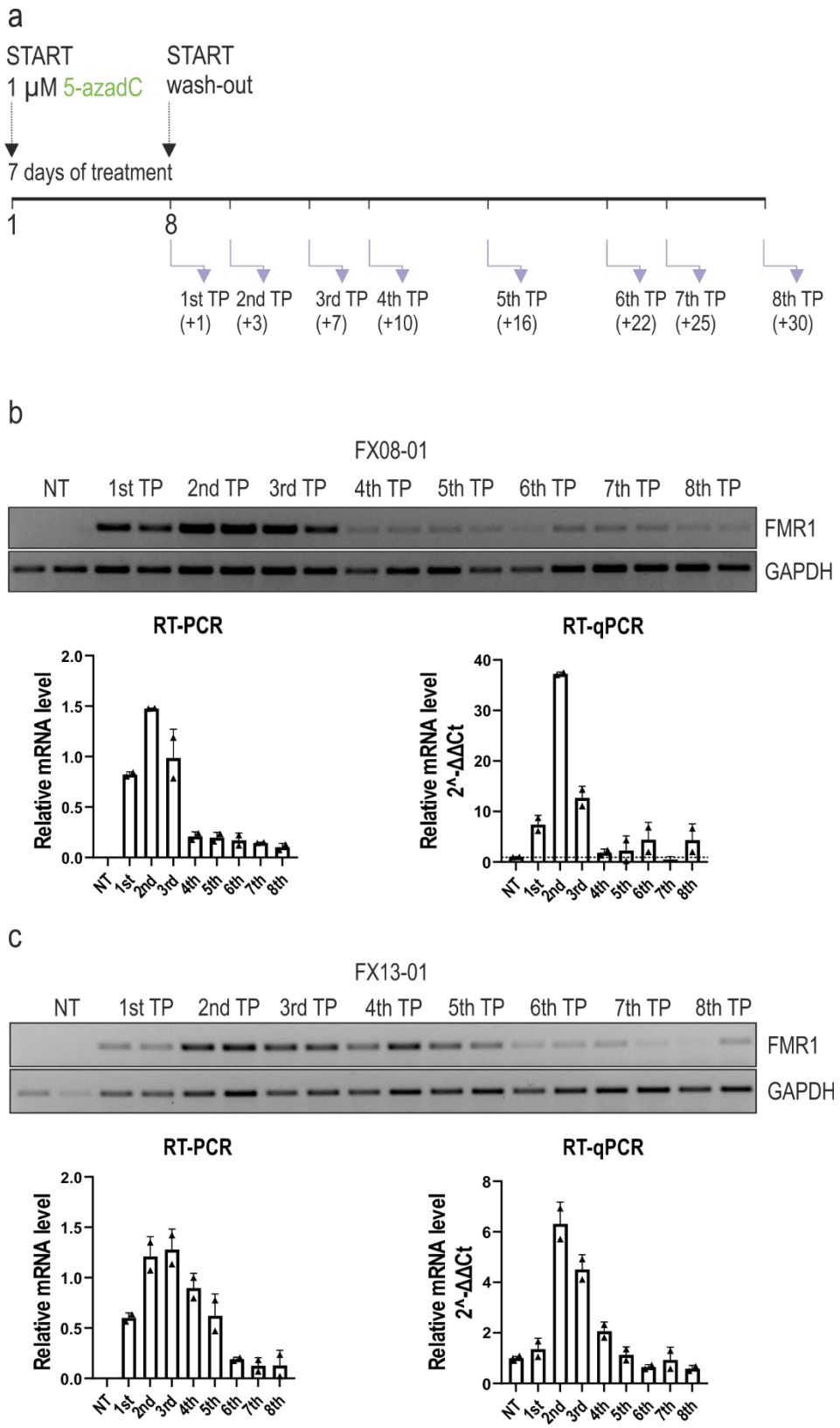


Figure 23. Abolition of the methylation status of FMR1 promoter by 5-azadC treatment in FXS-patients-derived fibroblasts. a) *Scheme of experiment taking into account the 5-azadC treatment and cell collection at appropriate time points (TP). Numbers in brackets correspond to days from the start of wash-out;* ***b)*** *RT-PCR quantification (left) and the corresponding gel (top) for FMR1 mRNA level analysis in FX08-01 cell line; (right) quantification using RT-qPCR analysis. Graphs*

present results for N = 2 biologically independent samples with SDs. Results were normalized to GAPDH; ***c)*** *RT-PCR quantification (left) and the corresponding gel (top) for FMR1 mRNA level analysis in FX13-01 cell line; (right) quantification of RT-qPCR analysis. Graphs present results for N = 2 biologically independent samples with SDs. Results were normalized to GAPDH. For RT-qPCR analyses: n = 2 technical replicates.*

These results confirm that methylation of *FMR1* can be reversed leading to *FMR1* transcription reactivation in both tested cell lines. According to statements that R-loops are involved in the process of *FMR1* methylation in FXS[146,125] I wanted to study whether the destabilization of R-loops structures involved in *FMR1* silencing may also reactivate transcription. To achieve this goal FX08-01 cells were cultured for 2 weeks under the pressure of ASO-Ctrl or ASO-CCG. As the positive control of *FMR1* transcription reactivation part of the cells were treated with 1 μM 5-azadC for 7 days followed by another 7 days of culture in a medium without DNA methylation inhibitor. In the meantime, the appropriate amount of passages (approximately 3) and transfections were performed (***Figure 24a***).

Although my previous results showed that ASO-CCG invades R-loops formed over expanded CGG repeats within *FMR1* locus (premutation range ~100 CGGs), and therefore increases its transcription efficiency *in vitro* (***Figure 16, right panel*** and ***Figure 18***) and *in cellula* (***Figure 22***) no such effect was observed in FXS cells (full mutation; over 435 CGGs), at least no effect on *FMR1* mRNA level. In the case of FX08-01 ASO-CCG treatment was not able to reactivate the *FMR1* transcription (***Figure 24b***).

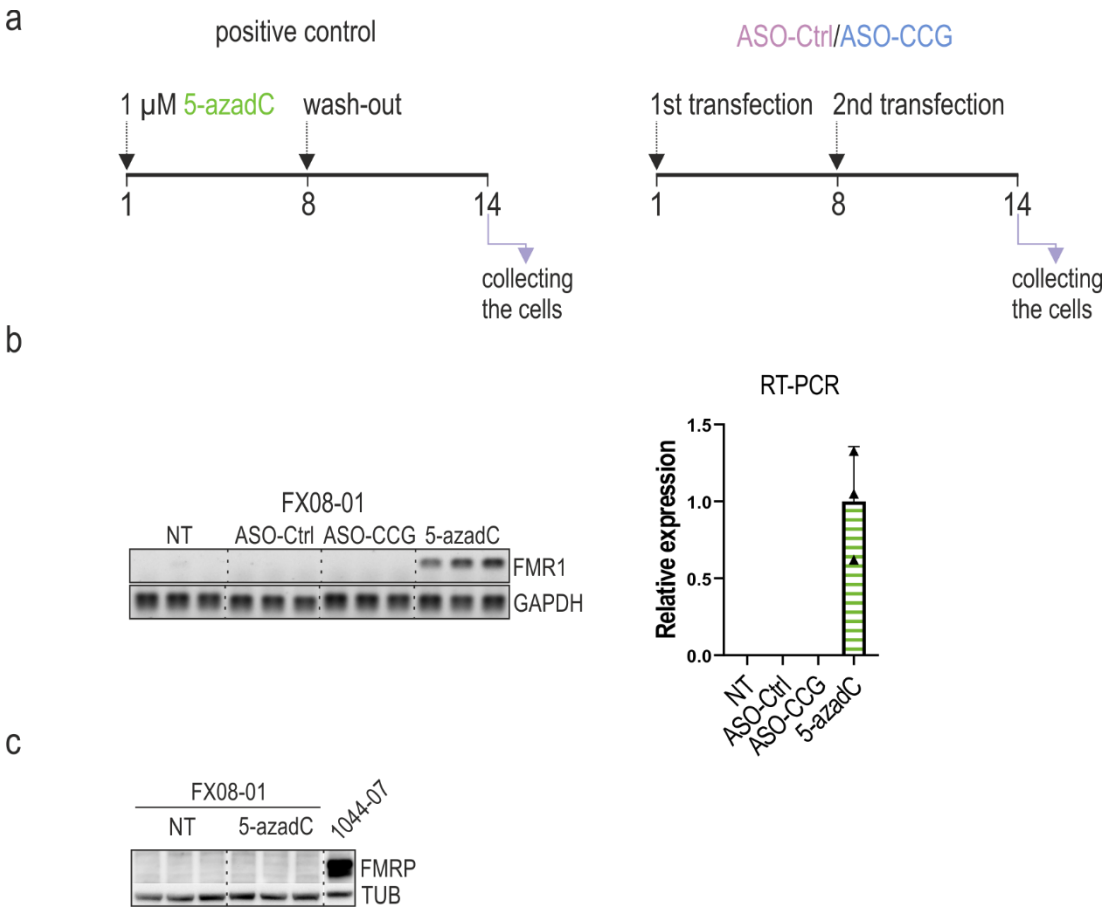


Figure 24. FMR1 mRNA level after long-term ASO-CCG treatment in FXS-patient derived fibroblasts. a) *Scheme of experiment taking into account the 5-azadC treatment (positive control; left) and ASO-Ctrl/ASO-CCG transfections in FX08-01 cells (right). The cells were cultured for 2 weeks with either 5-azadC or with ASO-Ctrl/ASO-CCG. In the case of 5-azadC treatment, cells were cultured for 7 days in the medium supplemented with 1 µM 5-azadC followed by 7 days in the clear medium (wash-out). Fibroblasts treated with ASO were cultured for 2 weeks and in the meantime two transfections with 200 nM ASO-Ctrl or ASO-CCG were performed;* ***b)*** *RT-PCR quantification and the corresponding gel of FMR1 mRNA level in FX08-01 fibroblasts;* ***c)*** *Western blot analysis of FMRP level in untreated and 5-azadC treated FX08-01 fibroblasts. The last lane presents the level of FMRP in untreated fibroblasts derived from FXTAS patient (1044-07), showing a basal level of this protein. Dotted lines mean that the photo of the gel was cropped.*

Observed results might be explained by: *1)* the fact that R-loops formed in FXS cells are not involved in the methylation process, *2)* ASO-CCG are not able to invade these structures, since they are much more stable than R-loops formed over shorter CGG repeats observed in healthy individuals and FXTAS conditions, or *3) FMR1* is already methylated and no transcription occurs, thus no co-transcriptionally formed R-loops are created and the DNA is present in the form of heterochromatin, fully inaccessible for ASO. Since R-loops involvement in *FMR1* methylation has been proven the latter options seemed to be more likely.

Nevertheless, intriguing was the fact that even upon significant transcription reactivation of *FMR1* by 5-azadC no protein was synthesized (***Figure 24c***). To understand the process behind this phenomenon I decided to perform nucleocytoplasmic fractionation on fibroblasts derived from FXTAS patients (1044-07; 97 CGG repeats) and FX13-01 fibroblast derived from FXS patient which was characterized by more than 435 CGG repeats, no FMRP, and, in contrary to FX08-01, with the low level of *FMR1* transcript. Obtained results (***Figure 25a***) suggest that in tested FXS cells the *FMR1* mRNA retains mostly in the nucleus which could partially explain why no protein is produced as a consequence of reduced accessibility of mRNA template for translation. In light of these data, I aimed to check how ASO-CCG administration affects *FMR1* mRNA localization between the nucleus and cytoplasm. Hence, I performed the nucleocytoplasmic fractionation of FX13-01 cells treated with ASO-Ctrl or ASO-CCG and harvested 48 h post-transfection. Both, the nuclear level of *FMR1* mRNA and the cytoplasmic level were considerably increased after ASO-CCG treatment, however, the foldchange between ASO-Ctrl and ASO-CCG was much higher for cytoplasmic fraction than for nuclear which suggests the effect of ASO-CGG on nucleocytoplasmic transport (***Figure 25b***). Nevertheless, it should be noticed that the level of *FMR1* mRNA in the cytoplasmic fraction was much lower than in the nuclear.

These findings suggest that ASO-CCG affects the transcription of *FMR1* alleles and may also impact the nuclear retention, nucleocytoplasmic transport, subcellular localization, and stability of *FMR1* mRNA. These processes jointly may lead to a pronounced increase in *FMR1* mRNA level in the cytoplasm which however did not result in the translation of FMRP protein (total protein lysate from ASO-Ctrl/ASO-CCG treated fibroblasts; ***Figure 25c***). This result can be explained by the fact that this level of *FMR1* mRNA is still very low in cytoplasm and also the structure formed by such long CGG repeats probably completely prevents efficient scanning of PIC to initiate FMRP biosynthesis. Therefore, even if there is a significant increase in the amount of *FMR1* mRNA in the cytoplasm it is not translated, at least to the level detected in western blot analysis.

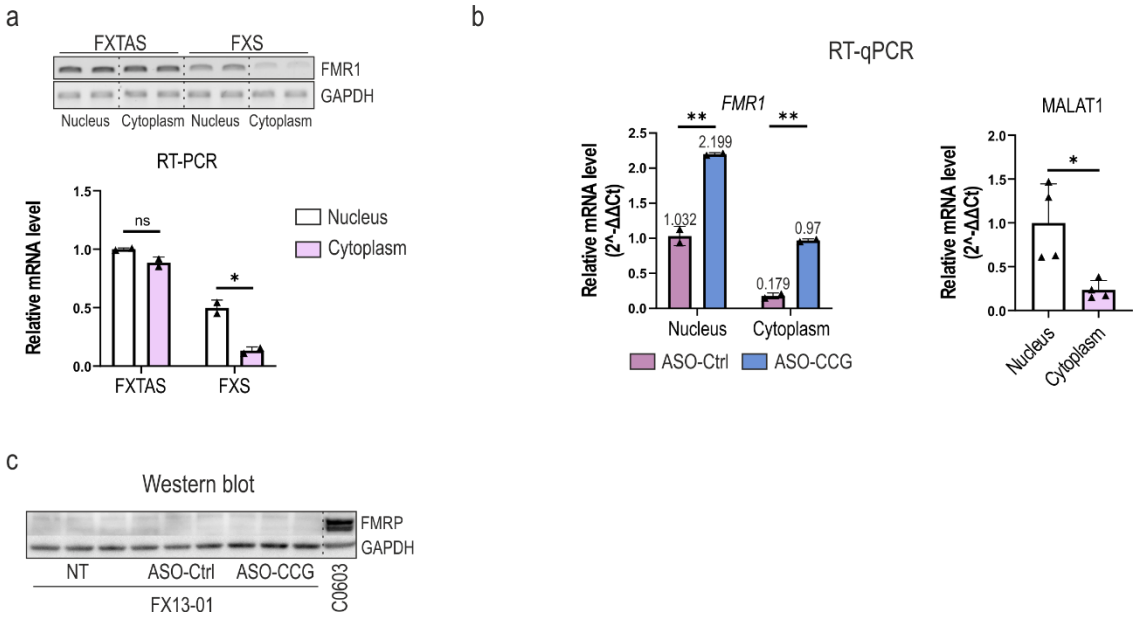


Figure 25. FMR1 mRNA nuclear retention in FXS-patient derived cells. a) *Quantification of RT-PCR analysis and the corresponding gel of FMR1 mRNA level after nucleocytoplasmic fractionation in 1044-07 (FXTAS) and FX13-01 (FXS) cells. Graphs present results for N = 2 biologically independent samples. Results were normalized to GAPDH;* ***b)*** *(Left) RT-qPCR analysis of FMR1 mRNA level after ASO-Ctrl/ASO-CCG treatment in nuclear and cytoplasmic fractions from FXS fibroblasts (FX13-01). Graphs present results for N = 2 biologically independent samples (with n = 2 technical replication each) with SDs. The Mean values are presented above each bar; (Right) The fraction purity was verified by the MALAT1 mRNA level. Results were normalized to GAPDH;* ***c)*** *Western blot analysis of total protein lysates from FX13-01 fibroblasts treated with 200 nM ASO-Ctrl and ASO-CCG for 48 h. The total protein lysate from C0603 control fibroblasts was loaded as the control of FMRP migration. Dotted lines mean that the image of the gel was cropped. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ns, non-significant.*

To sum up, results presented in this section showed that ASO-CCG, contrary to 5-azadC, is unable to reactivate the *FMR1* transcription in the fibroblasts derived from FXS patients with full mutation if the *FMR1* gene is completely silenced. However, ASO-CCG can slightly increase the *FMR1* transcription in FXS cells if the *FMR1* locus is partially active, but only if cells were pre-treated with 5-azadC. In addition, the nucleocytoplasmic fractionation showed that ASO-CCG led to an increase in the *FMR1* mRNA level in the cytoplasm. However, the increase in the *FMR1* mRNA level after either 5-azadC treatment or complex treatment with 5-azadC and ASO-CCG did not result in the detection of FMRP which potentially is a consequence of the nuclear retention of this mRNA and impairment of protein synthesis from *FMR1* mRNA with very long CGG repeats. The effect of the FMRP translation efficiency dependent on the number of CGG repeats will be presented in the further part of this work.

4.2. PRIMARY AND SECONDARY STRUCTURES OF 5'UTR OF *FMR1* mRNA ARE SIGNIFICANT FACTORS IN THE REGULATION OF FMRpolyG SYNTHESIS

FMRpolyG protein which is synthesized through non-canonical RAN translation from mutant *FMR1* mRNA containing expanded CGG repeats in the PM range is one of the pathogenic factors involved in the etiology of fragile X-linked disorders other than FXS. FMRpolyG protein in the context of CGGexp within *FMR1* was first named in 2013[69] as „a protein that contains an N-terminal polyglycine stretch followed by a 42 amino acid (**aa**) carboxyl-terminal domain out of frame with the downstream FMRP start codon''. Although this name is commonly used by other researchers the protein sequence named FMRpolyG may differ between various studies and developed reporter systems. In the case of this dissertation, the name FMRpolyG describes a protein containing the *FMR1*-specific N-terminal sequence of 12 aa followed by polyglycine stretch (composed either by 16 or 85 glycine residues) and the C-terminal sequence (40 aa) containing *FMR1*-specific ex1 region in-frame with Nluc-FLAG protein.

ACG (+1) codon, located 32 nt upstream of CGG repeats, has been established as the main translation initiation site for FMRpolyG. It is located about 89 – 134 nt from the cap structure dependent on the TSS of the *FMR1*. It has been shown that the product of RAN translation through CGG repeats constituted roughly 10% of the total AUG-initiated translation from studied mRNA[69]. It has been also stated that RAN translation of FMRpolyG is 30 – 40 % as efficient as canonical AUG-initiated translation through the repeat (when ACG (+1) near-cognate start codon was changed to AUG codon)[70]. Clearly, the efficiency of RAN translation is lower than the canonical translation.

These studies aimed to establish how different *cis*-regulatory elements regulate the efficiency of FMRpolyG translation initiation. Since this process probably depends on the formation of ribosome initiation complex on the ACG (+1) codon, the following aspects affecting the choice of this near-cognate start codon have been considered: *(i) nucleotide sequence in the vicinity of this codon, and (ii) stable secondary structure of RNA formed by the sequence located downstream of this codon.* Both circumstances can interfere with the speed of ribosome scanning, leading to ribosome pausing or even dissociation of the ribosome from mRNA, therefore, strongly regulating the initiation of FMRpolyG synthesis.

4.2.1. Development of NanoLuciferase reporter assays to study the efficiency of RAN translation of FMRpolyG and canonical translation of FMRP

To determine factors regulating the RAN translation initiation of FMRpolyG in the context of sequential and structural dependencies and to be able to study the RAN translation efficiency I generated a series of constructs with the human *FMR1* 5'UTR. The 5'UTR *FMR1* sequence was fused with a sequence encoding NanoLuciferase (Nluc) with FLAG-tag in two reading frames: *(i)* +1 frame for FMRpolyG, and *(ii)* +0 frame for FMRP equivalent (first 17 aa of FMRP fused with luciferase, as control of AUG-initiated canonical translation). The Nluc reporter was chosen as it has been shown that its activity remains unchanged upon fusion with polyglycine. Additionally, it has been stated that the size of CGG repeats does not affect Nluc mRNA or protein stability[70]. Also, contrary to previously used GFP and mCherry tags[194] Nluc does not affect tagged protein aggregation capacity[195,196,197].

All steps of cloning are described in detail in Methods (***see Methods 3.1.3. "Cloning procedures"***). Briefly, the plasmids had the native 5′ UTR of the *FMR1* gene with 16 CGG or 85 CGG repeats preceded by the near cognate ACG (+1) codon embedded in the natural Kozak sequence context. The *FMR1* 5'UTR was fused with Nluc-FLAG in either FMRpolyG (+1 frame) or FMRP frame (+0 frame). The frameshift, from FMRP to FMRpolyG, was generated by the insertion of a single nucleotide at the end of the *FMR1* sequence (+/-G; ***see Figure 26b and 26d***). Hence FMRP-Nluc-FLAG plasmid encodes FMRP equivalent, translated from AUG (+0) codon, containing first 17 aa of FMRP fused to Nluc-FLAG, and FMRpolyR which is translated from ACG (+0) codon within 5'UTR. The FMRpolyG-Nluc-FLAG encodes for FMRpolyG which can be translated from ACG (+1) and GUG (+1) near-cognate start codons[198,199,200]. The scheme of developed plasmids and the sequence of the whole 5'UTR is presented in ***Figure 26*** below.

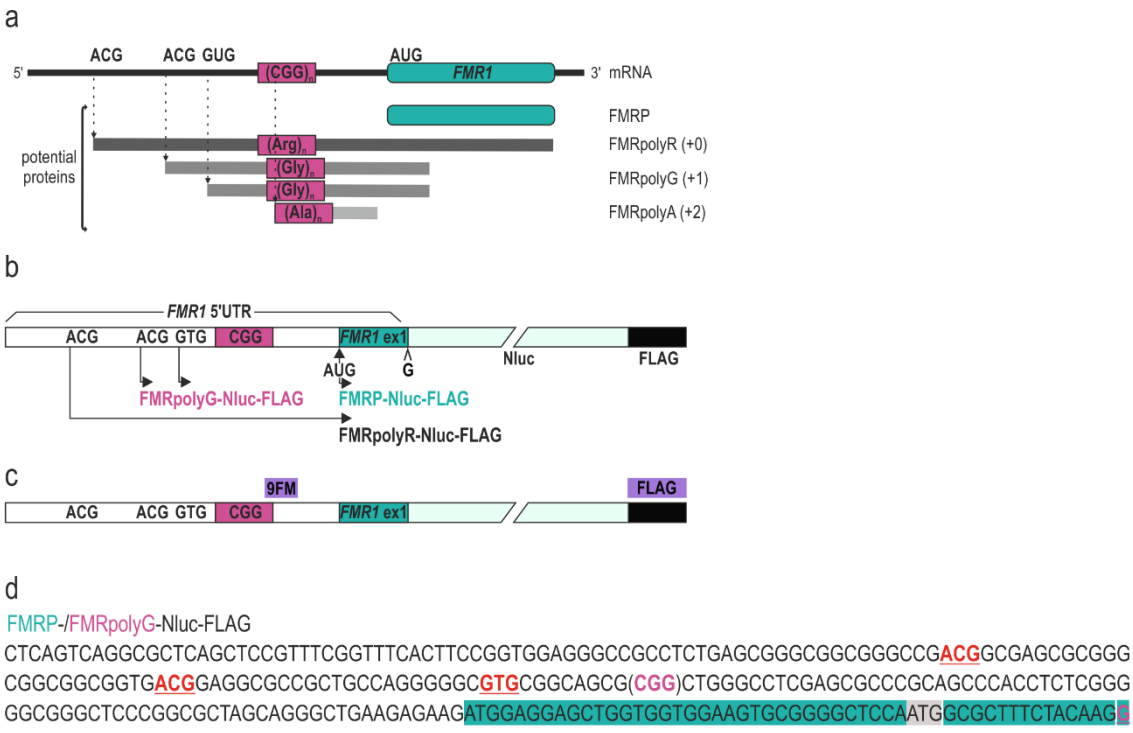


Figure 26. Schematic of FMR1 mRNA with the RAN translation products. a) *FMR1 mRNA is a template for the synthesis of four different proteins: initiation at the AUG codon downstream 5'UTR is used to produce FMRP (in this system measured by the FMRP protein equivalent – 17 aa of FMRP fused to Nluc-FLAG). Three near-cognate start codons in the 5'UTR allow for the initiation of biosynthesis of the following RAN products: FMRpoly(R) (ACG (+0)) and FMRpoly(G) (ACG (+1) and GUG (+1)). Initiation within the CGG repeats generates FMRpoly(A) (GCC (+2));* ***b)*** *Schematic of FMRP/FMRpolyG-Nluc-FLAG constructs;* ***c)*** *Epitopes of 9FM*[76] *(FMRpolyG), and anti-FLAG antibodies (violets) are presented in the context of plasmid scheme;* ***d)*** *Sequence of FMR1 5'UTR cloned into FMRP/FMRpolyG-Nluc-FLAG constructs. Three near-cognate start codons are bolded (red), and CGG repeats are shown in brackets (pink). The insertion of the G nucleotide to generate frame shift to the FMRpolyG frame is bolded (the last nucleotide in the sequence; pink). The FMR1 ex1 sequence encoding for FMRP is marked in green. The ATG codon marked in grey represents the codon that was mutated to AAA only for the FMRpolyG frame as the potential donor for additional proteins (see Methods 3.1.3.1.5. "Additional mutations performed on constructs in the FMRpolyG frame").*

For the majority of designed experiments, I used plasmids with 16 CGG repeats to minimize the effect of the secondary structure formed by long CGG repeats on the FMRpolyG translation initiation. However, in some cases, the usage of a construct with 85 CGG repeats was required to answer the question of what influences more the initiation of RAN translation, the sequence context or the RNA structure.

4.2.1.1. Mutation of GUG (+1) near-cognate start codon shows that ACG (+1) is the major initiation codon for FMRpolyG

As it is presented in ***Figure 27b*** the FMRpolyG may be translated from two near-cognate start codons – ACG (+1) and GUG (+1). I planned to study the effect of nucleotide context in the vicinity of ACG (+1) codon on the RAN translation efficiency, therefore, it was important at the first step to ensure that translation of FMRpolyG will be initiated from a single, particular codon. I performed a mutagenesis of the second near-cognate start codon – GUG (+1), as it was already published that RAN translation of FMRpolyG is initiated in the majority from ACG (+1)[67,69]. The predictions of RNA structure formed by *FMR1* 5'UTR with or without GUG (+1) mutation remained unchanged (***Figure 27a***).

To verify how this mutation affects the RAN translation of the *FMR1* message the HEK-293 cells were seeded on a 48-well and 96-well plate and transfected with appropriate plasmids. Cells for western blot (seeded on a 48-well plate) were collected 24 h post-transfection, sonicated, and run on the polyacrylamide gel in denaturing conditions followed by western blot analysis. Samples for the NanoLuc assay were harvested also after 24 h and prepared as written in the Methods section (***see Methods 3.13. "Nano-Glo Dual-Luciferase Reporter Assay"***). The comparison between protein products of constructs harboring GTG (GUG (+1)) or GTT (GUG (+1) mutated to GUU) seems to confirm results published by other labs that GUG (+1) codon plays a minor role in the initiation of FMRpolyG synthesis. The mutation of GUG→GUU did not affect the level of FMRpolyG detected in the western blot by 9FM antibody (recognizes the C-terminal part of FMRpolyG[76]; ***Figure 26c***). Also, the efficiency of canonical translation measured by the level of FMRP equivalent seemed to be unchanged, visualized both by western blot and NanoLuc assay. The amount of FMRpolyG after GUG→GUU mutation was however slightly higher based on data obtained by NanoLuc assay.

Western blot is a semi-quantitative method while NanoLuc assay is a much more sensitive and accurate approach thus the results generated by both techniques may differ, especially since NanoLuc assay may detect all synthesized proteins which are in frame with Nluc while western blot is limited to the range of proteins with molecular weights of studied proteins. Due to mentioned issues, I was using both techniques as they complement each other.

Nevertheless, the change in the FMRpolyG level in this direction supports the statement that the GUG (+1) codon is negligible in the context of RAN translation initiation. Therefore, all further mutants were developed from the backbone of constructs containing GTG→GTT mutation.

For simplicity, easier understanding, and greater readability the mutated constructs **FMRP/FMRpolyG-Nluc-FLAG-GTT** will be named **WT** in all further analyses.

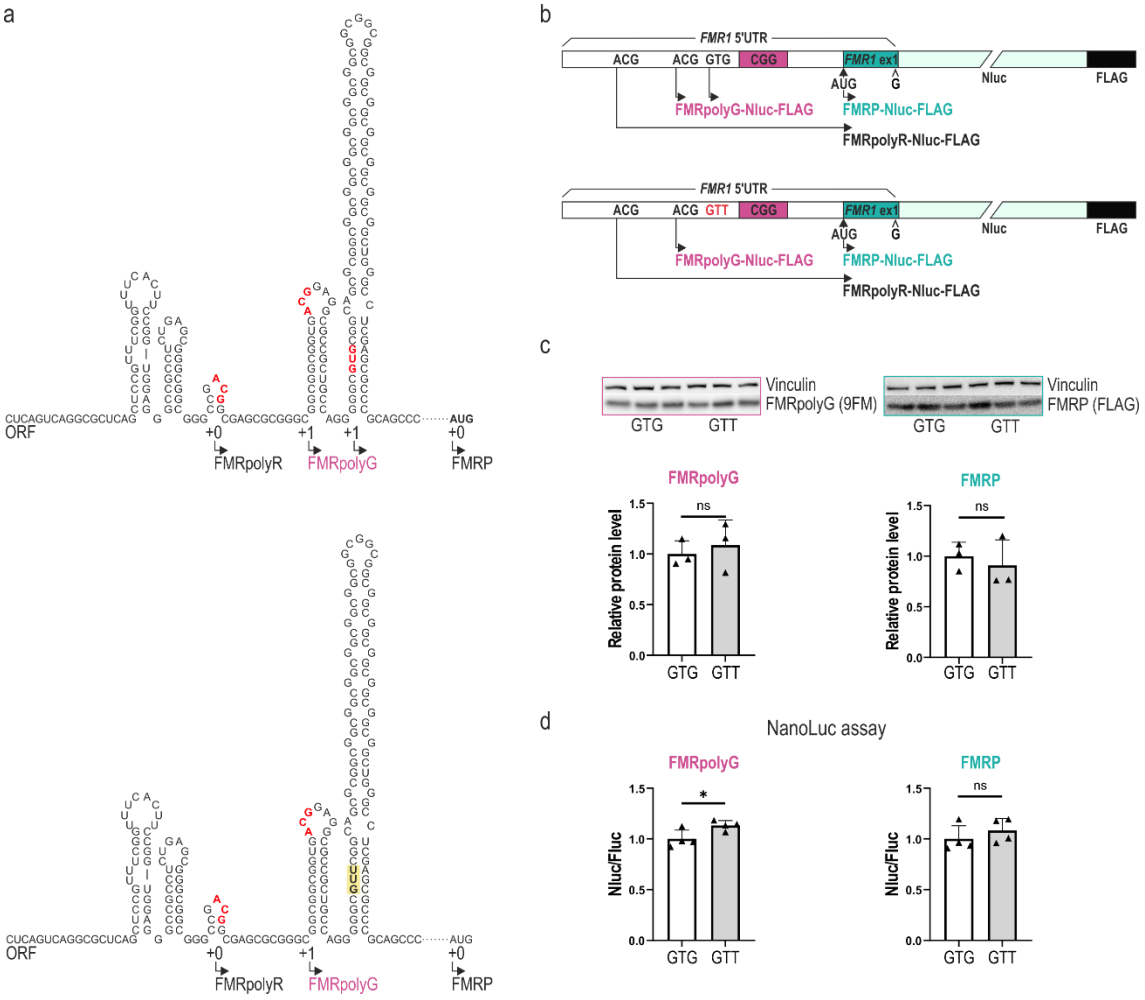


Figure 27. Unification of FMRpolyG RAN translation initiation site. *The GUG (+1) near-cognate start codon was mutated to ensure FMRpolyG RAN translation initiation at a single, particular near-cognate start codon – ACG (+1);* ***a)*** *Predicted structures of FMR1 5'UTR with the native GUG (+1) codon (upper), and GUG→GUU mutation (lower);* ***b)*** *Schematic of cloned constructs: FMRP/FMRpolyG-Nluc-FLAG (upper), and FMRP/FMRpolyG-Nluc-FLAG-GTT (lower;* ***called WT in next analyses****);* ***c)*** *Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody and FMRP equivalent measured by anti-FLAG antibody in context of GUG (+1) codon mutation. Results were normalized to Vinculin. Graphs present results for N = 3 biologically independent samples with SDs;* ***d)*** *Quantification of results from NanoLuc assay. Graphs present results for N = 4 biologically independent samples with*

*SDs. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; ns, non-significant.*

4.2.1.2. Mutation of TIS for FMRpolyR protein influence the level of detected FMRpolyG protein

As has been already mentioned the FMRP-Nluc-FLAG constructs (+0 frame) allow for FMRP equivalent detection as well as a significantly larger protein – the FMRpolyR, which is polyarginine tract fused with part of FMRP and Nluc-FLAG (+0 frame). FMRpolyR is initiated at ACG (+0) codon located 57 nt upstream CGG repeats and 22 nt upstream ACG (+1) codon, and generates an N-terminal polyarginine extension of FMRP. However, this product of RAN translation is not easily detected endogenously as well as in reporter systems[69,70], and constitutes ~2.5% of FMRpolyG which is the most abundant RAN protein translated from CGG repeats within *FMR1*[70]. Interestingly, it has been shown that the FMRpolyR was detected only when the construct had no more than 18 CGGs within the *FMR1* 5'UTR sequence and was attenuated at longer CGG repeats[80] which suggests that the length of CGG repeats has a negative effect on the RAN translation of FMRpolyR protein[70].

Since the TIS of FMRpolyR (ACG (+0)) is located upstream of the major initiation site for FMRpolyG (ACG (+1)) there is a possibility that translation of FMRpolyR affects the FMRpolyG translation. To verify this issue first I performed a mutation of ACG (+0) near-cognate start codon to AAA to confirm that the predicted FMRpolyR protein is translated from this particular codon in the developed construct with 16 CGGs (16FMRP/FMRpolyG-Nluc-FLAG) (***Figure 28a***). The HEK-293 cells were transfected and processed for western blot and NanoLuc analyses as described above in 4.2.1.1. subchapter. The western blot analysis revealed that the observed extra protein migrating above the FMRP protein equivalent was indeed the FMRpolyR protein since the introduced mutation resulted in the loss of protein (***Figure 29a, upper panel***). As expected, the level of FMRP equivalent remained unchanged measured by both western blot (***Figure 29a, upper panel***) and NanoLuc assay (***Figure 29a, lower panel***). Although the introduced mutation did not influence the predicted RNA structure in the region of ACG (+1) codon – TIS for FMRpolyG – (***Figure 28b***) the significant decrease in the FMRpolyG level was observed in the NanoLuc assay analysis (***Figure 29a, lower panel***). Although this result was not confirmed by western blot, some trends in the same direction could be observed. The decrease in the FMRpolyG level suggested that RAN translation

in the polyR-frame contributes to translation in the polyG-frame. Moreover, it has been proven that on *FMR1* CGG repeats the translational frameshift from polyR (+0 frame) to polyG (+1 frame) occurs (R-to-G)[201]. Although this mechanism has been proven for longer CGG repeats (~100 and 25 CGGs) this mechanism may be at play in this case since it was stated that R-to-G frameshift on CGG repeats occurs after incorporating 1-4 arginines[201].

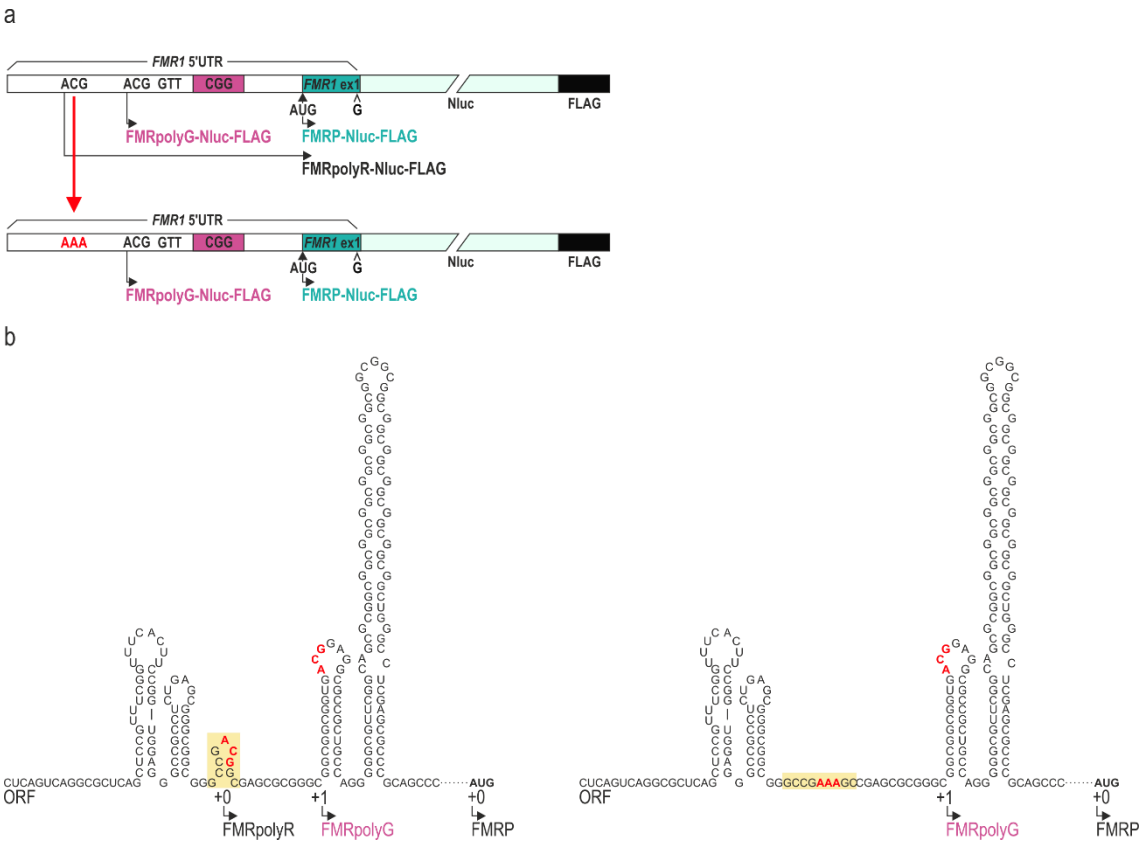


Figure 28. Scheme of constructs designed to confirm the FMRpolyR translation from the FMRP-Nluc-FLAG construct. *The ACG (+0) near-cognate start codon was mutated to AAA to confirm that the observed protein (+0 frame) is the FMRpolyR.* ***a)*** *Schematic of cloned constructs: (upper) FMRP/FMRpolyG-Nluc-FLAG (called WT), and (lower) FMRP/FMRpolyG-ACG(+0)AAA;* ***b)*** *Predicted structures of FMR1 5'UTR with the native ACG (+0) codon (left), and ACG(+0)AAA mutation (right).*

One of the explanations of the RAN translation mechanism assumes that translation initiation at near-cognate start codons is stimulated by the paused translating ribosomes which meet an obstacle of a secondary structure formed by expanded CGG repeats. Thus, a formed queue of stacked ribosomes may favor the initiation of translation of upstream ORFs even if their start codons are embedded in the weak Kozak context sequence or are near-cognate start codons. Therefore, for the aforementioned reasons, in parallel to the constructs with 16 CGG repeats I developed constructs containing longer, 85 CGG

repeats. I assumed that such constructs will allow to compare the effect of the length of CGG repeats on both, the non-AUG-initiated and the AUG-initiated, canonical translation.

I hypothesized that if ribosome queuing occurs then the increase in the initiation rate of RAN proteins, including FMRpolyR, should be observed. According to assumptions, the level of FMRP equivalent remained unchanged measured by both western blot analysis and NanoLuc assay (***Figure 29b***), however, the FMRpolyR protein was no longer detected by the western blot analysis.

This result represented probably the sum of, at least, two mechanisms: *1)* longer CGGs reduce the translational efficiency of whole *FMR1* mRNA therefore the level of translated FMRpolyR is under the detection threshold by western blot, and *2)* there is a negative correlation between the translation of FMRpolyR and the number of CGG repeats as it was already stated[70]. However, the mechanism of ribosome queuing and the increased translation initiation of FMRpolyR can not be excluded at this point.

Nevertheless, these results confirmed that in the Nluc-based reporter system, with 16 CGGs developed in this study, the FMRpolyR protein can be analyzed and that the RAN translation of this protein has a direct effect on the FMRpolyG level, probably due to the polyR-to-polyG frameshift.

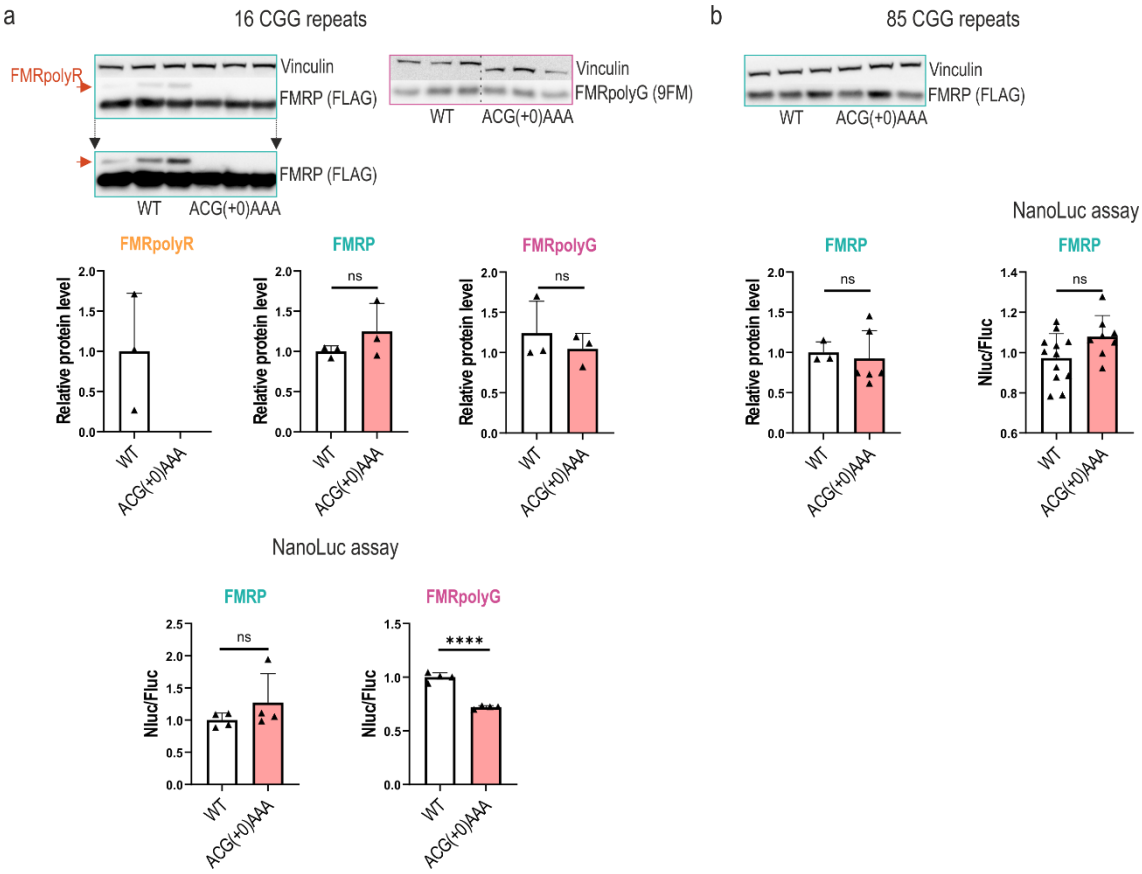


Figure 29. Mutation of ACG (+0) near-cognate start codon for FMRpolyR influences the level of FMRpolyG. a) *<u>16 CGGs</u>: (upper panel), western blot analysis and corresponding quantification of FMRpolyR and FMRP equivalent measured by the anti-FLAG antibody, and FMRpolyG level measured by 9FM antibody. The FMRpolyR produced from FMRP-Nluc-FLAG is marked by an orange arrow. Results were normalized to Vinculin. Graphs present results for N = 3 biologically independent samples with SDs; (lower panel), Quantification of results from NanoLuc assay which detects both, FMRP-equivalent and FMRpolyR proteins, however, due to the dominant level of FMRP-equivalent protein the amount of FMRpolyR is omitted in the analysis. Graphs present results for N = 4 biologically independent samples with SDs;* ***b)*** *<u>85 CGGs</u>: (upper panel), Western blot analysis and corresponding quantification (bottom left) of FMRP equivalent level analyzed as in* ***a****. Graphs present results for N = 3 biologically independent samples with SDs; (bottom right) Quantification of results from NanoLuc assay. Graphs present results for N = 12 and N = 8 biologically independent samples with SDs for WT and ACG(+0)AAA, respectively; Statistical analysis was based on a two-tailed unpaired Student's t-test; ****, p<0.0001; ns, non-significant.*

4.2.1.3. Potential uORFs present in FMR1 5'UTR constructs used in this study

It has been already mentioned that differences in translation efficiency calculated based on the two methods used in this study, western blot and NanoLuc assay analyses, may arise due to the background noise present in NanoLuc assay caused by the presence of additional proteins synthesized in frame with Nluc which are not detected or taken into account during western blot analysis. This noise results from other putative near-cognate

start codons within the *FMR1* 5'UTR as well as within the Nluc sequence. Possible open reading frames for RAN proteins and products of canonical translation that could be encoded by designed constructs are presented in ***Figure 30***, however, these lists contain only computationally predicted ORFs (ORFs starting within Nluc are omitted). It cannot be ruled out that some additional near-cognate start codons would activate as the result of introduced mutations. Such a situation was already observed for Nluc reporter during FMRpolyA studies by Kearse and colleagues[70] where mutating the AUG start codon of the Nluc to GGG produced a truncated protein product translated from initiation at a CUG (in optimal Kozak context) within Nluc sequence encoding Leucine at position 20.

16FMRP-Nluc-FLAG

CTCAGTCAGGCGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGGCCGCCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGG
CGGCGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTGCGGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC
GGCGGCGGCGGCTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGGGGCGGGCTCCCGGCGCTAGCAGGGCTGAAGAGAAGATGGA
GGAGCTGGTGGTGGAAGTGCGGGGCTCCAATGGCGCTTTCTACAAGATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAG
ACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCA
AAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCC
AGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGT
TACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGT
GGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTG
GCGGCTGTGCGAACGCATTCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

ORF1 (MW=28,59 kDa)
ACGGCGAGCGCGGGCGGCGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTGCGGCAGCGCGGCGGCGGCGGCGGCGGCGGCG
GCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGGGGCGGGCTCCCGGCGCTAGCAGGGC
TGAAGAGAAGATGGAGGAGCTGGTGGTGGAAGTGCGGGGCTCCAATGGCGCTTTCTACAAGATGGTCTTCACACTCGAAGATTTCGTTG
GGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTC
CGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGG
CGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTG
GTAATCGACGGGGTTACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTA
ACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGG
AGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

TASAGGGGDGGAAARGRAAARRRRRRRRRRRRRRRLGLERPQPTSRGRAPGASRAEEKMEELVVEVRGSNGAFYKMVFTLEDFVGDWR
QTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVIIPYEGLSGDQMGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMID
YFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLINPDGSLLFRVTINGVTGWRLCERILAGDYKDDDDK-

ORF2 (MW=24,52 kDa)
CTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGGGGCGGGCTCCCGGCGCTAGCAGGGCTGAAGAGAAGATGGAGGAGCTGGTGGT
GGAAGTGCGGGGCTCCAATGGCGCTTTCTACAAGATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTAC
AACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCT
GAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAA
TTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACAT
GATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACA
AAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGA
ACGCATTCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

MGLERPQPTSRGRAPGASRAEEKMEELVVEVRGSNGAFYKMVFTLEDFVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGE
NGLKIDIHVIIPYEGLSGDQMGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLINPDG
SLLFRVTINGVTGWRLCERILAGDYKDDDDK-

ORF3 (MW=22,06 kDa)
ATGGAGGAGCTGGTGGTGGAAGTGCGGGGCTCCAATGGCGCTTTCTACAAGATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGC
GACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCG
ATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATG
GGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACG
GGGTTACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACC
CTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGG
CTGGCGGCTGTGCGAACGCATTCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

16FMRP-Nluc-FLAG

ORF3 cd

MEELVVEVRGSNGAFYKMVFTLEDFVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVIIPYEGLSGDQMGQIEK
IFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLINPDGSLLFRVTINGVTGWRLCERILAGDY
KDDDDK-

ORF4 (MW=21,23 kDa)

GTGCGGGGCTCCAATGGCGCTTTCTACAAGATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCT
GGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCG
GTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTTA
AGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGATCG
ACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTA
TCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCAT
TCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

VRGSNGAFYKMVFTLEDFVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVIIPYEGLSGDQMGQIEKIFKVVYP
VDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLINPDGSLLFRVTINGVTGWRLCERILAGDYKDDDDK-

16FMRpolyG-Nluc-FLAG

CTCAGTCAGGCGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGGCCGCCTCTGAGCGGGCGGCGGGCCG**ACG**GCGAGCGCGGG
CGGCGGCG**GTGACG**GAGGCGCCG**CTG**CCAGGGGGC**GTG**CGGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC
GGCGGCGGCGGCTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGGGGCGGGCTCCCGGCGCTAGCAGGG**CTG**AAGAGAAGATGGA
GGAGCTGGTGGTGGAAGTGCGGGGCTCCA**AAA**GCGCTTTCTACAAG**G**GGGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACA
GACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCC
AAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGC
CAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGG
TTACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGT
GGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTG
GCGGCTGTGCGAACGCATTCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

ORF1 (MW=26,75 kDa)

GTGACGGAGGCGCCGCTGCCAGGGGGC**GTG**CGGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG
GCGGCTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGGGGCGGGCTCCCGGCGCTAGCAGGGCTGAAGAGAAGATGGAGGAGCTG
GTGGTGGAAGTGCGGGGCTCCA**AAA**GCGCTTTCTACAAG**G**GGGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCC
GGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGAT
TGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCG
AAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCC
GAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACG
GCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCT
GTGCGAACGCATTCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

VTEAPLPGGVRQRGGGGGGGGGGGGGGGGWASSARSPPLGGGLPALAGLKRRWRSWWWKCGAPKALSTRGVFTLEDFVGDWRQTAGY
NLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVIIPYEGLSGDQMGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRP
YEGIAVFDGKKITVTGTLWNGNKIIDERLINPDGSLLFRVTINGVTGWRLCERILAGDYKDDDDK-

16FMRpolyG-Nluc-FLAG cd

ORF2 (MW=26,26 kDa)

CTGCCAGGGGGC**GTG**CGGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGGGCCTCGAG
CGCCCGCAGCCCACCTCTCGGGGGCGGGCTCCCGGCGCTAGCAGGGCTGAAGAGAAGATGGAGGAGCTGGTGGTGGAAGTGCGGG
GCTCCAAAGCGCTTTCTACAAGGGGGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACC
AAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAA
AATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTG
GTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATT
TCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCGAC
GAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGG
CGGGAGATTACAAGGATGACGACGATAAGTAA

LPGGVRQRGGGGGGGGGGGGGGGGWASSARSPPLGGGLPALAGLKRRWRSWWWKCGAPKALSTRGVFTLEDFVGDWRQTAGYNLDQ
VLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVIIPYEGLSGDQMGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGI
AVFDGKKITVTGTLWNGNKIIDERLINPDGSLLFRVTINGVTGWRLCERILAGDYKDDDDK-

ORF3 (MW=22,73 kDa)

CTGAAGAGAAGATGGAGGAGCTGGTGGTGGAAGTGCGGGGCTCCAAAGCGCTTTCTACAAGGGGGGTCTTCACACTCGAAGATTTCG
TTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGT
GTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAG
CGGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCACA
CTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCAC
TGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCA
ACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGGAGATTACAAGGATGACGACGATAAGTAA

LKRRWRSWWWKCGAPKALSTRGVFTLEDFVGDWRQTAGYNLDQVLEQGGVSSLFQNLGVSVTPIQRIVLSGENGLKIDIHVIIPYEGLSGDQ
MGQIEKIFKVVYPVDDHHFKVILHYGTLVIDGVTPNMIDYFGRPYEGIAVFDGKKITVTGTLWNGNKIIDERLINPDGSLLFRVTINGVTGWRLCER
ILAGDYKDDDDK-

Figure 30. Potential open reading frames predicted within FMRP/FMRpolyG-Nluc-FLAG constructs. *The sequence of FMR1 5'UTR fused to Nluc-FLAG for both reading frames is presented. ACG (+0), ACG (+1), and GUG (+1) (showed as GTG on DNA sequence) near-cognate start codons are bolded in red, and other putative near-cognate start codons are bolded in black. The GUG (+1) codon (grey) is mutated in constructs to GUU. The sequence of CGG repeats (for simplicity only 16 CGGs are shown) is marked in pink, the FMR1 ex1 sequence is marked in green while Nluc fused to FLAG are marked in blue and black, respectively. Mutations specific to the FMRpolyG frame within FMR1 ex1 are marked in red.*

4.2.2. Translation of different reading frames of *FMR1* mRNA is CGG repeat length-dependent

FMR1 mRNA level has been shown to increase, in the premutation range, as the CGG repeats expands[35,33,30,202]. However, the increase in the number of CGG repeats was also correlated with the gradual impairment of *FMR1* translation efficiency of major ORF and reduced level of FMRP[34,30,35,36]. On the other hand, the RAN reporter mRNAs with expanded repeats have been shown to be translated at higher levels than those with normal lengths of CGG repeats[70].

Although, the level of FMRpolyG protein may depend on several factors the number of CGG repeats seems to be one of the primary and most important factor that regulates the

FMR1 mRNA translation. Interestingly, structural studies revealed that *FMR1* flanking regions of CGG repeats (14 nt upstream and 17 nt downstream CGGs) are involved in the stabilization of hairpin structure formed by CGG repeats, regardless of the number of repeats[181]. Nevertheless, as presented in ***Figure 31a*** the difference in the thermodynamic stability of RNA hairpin structures predicted to be formed by 16 and 85 CGG repeats is fourfold. Of course, it can not be without impact on the RAN translation which utilizes a scanning mechanism of initiation. Taking into consideration the cap-dependent mechanism of RAN translation initiation the stable secondary structures in the *FMR1* 5'UTR can potentially inhibit the ribosome scanning directly affecting the initiation step of translation. Therefore, the question that I wanted to address was, how the different size of CGG repeats regulates the translation of the *FMR1* message in designed model systems.

HEK-293 cells were seeded on appropriate vessels and transfected with constructs carrying either 16 or 85 CGG repeats. Cells were collected 24 h post-transfection for NanoLuc assay analysis and RNA isolation and 48 h post-transfection for western blot. Obtained results showed that levels of both FMRpolyG and FMRP proteins were higher if they were translated from *FMR1* mRNA with 16 CGG repeats (***Figure 31b***) suggesting that *FMR1* translation decreases while CGG repeats increase regardless open reading frame. However, this impairment of translation is, at least partially, a consequence of a diminished level of *FMR1* mRNA carrying 85 CGG repeats (***Figure 31d***). As expected, the level of translated FMRP was greater from *FMR1* mRNA with shorter CGG repeats (***Figure 31c***). In conclusion, these results present that differences in FMRpolyG level in the designed model system result from changes in transcription but mostly translation.

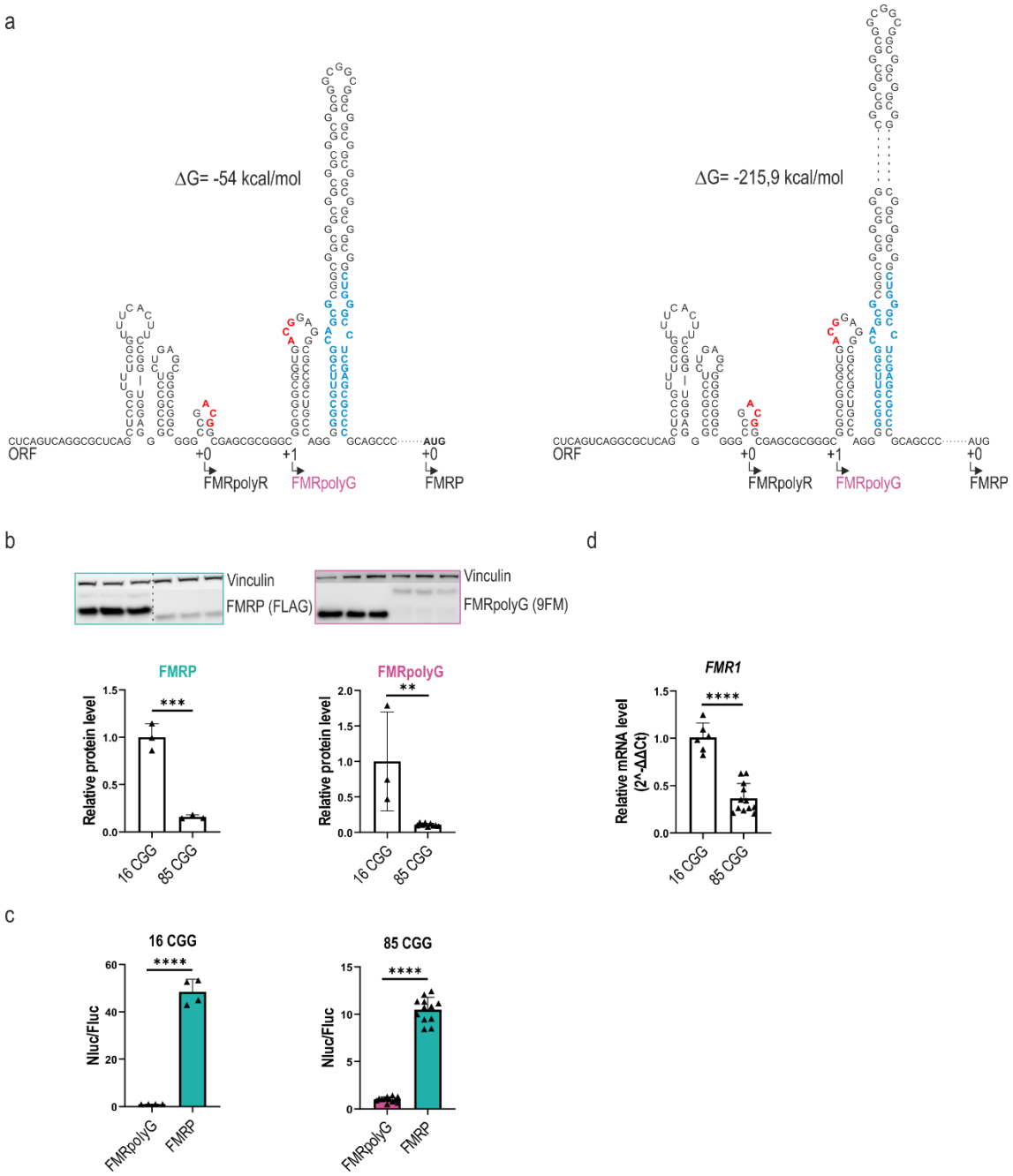


Figure 31. The level of proteins translated from FMR1 is influenced by the size of CGG repeats. ***a)*** *Predicted structures of FMR1 5'UTR with 16 CGG repeats (left) and 85 CGG repeats (right). The predicted ΔG values are presented. The flanking regions responsible for hairpin stabilization are marked in blue;* ***b)*** *Changes in the FMRP equivalent and FMRpolyG levels depending on the CGG repeats size. The graph presents western blot results for N = 3 biologically independent samples (except FMRpolyG 85 CGG repeats where N = 9) with SDs. The left gel was cropped. Results were normalized to Vinculin;* ***c)*** *Results from NanoLuc assay presenting differences in FMRpolyG and FMRP expression depending on the different sizes of CGG repeats;* ***d)*** *Relative FMR1 mRNA level. Results were normalized to GAPDH. The graph presents RT-qPCR results for N = 6 and N = 12 biologically independent samples (with n = 2 technical replication each) with SDs for constructs with 16 and 85 CGG repeats, respectively. Statistical analysis was based on a two-tailed unpaired Student's t-test; **, p<0.01; ***, p<0.001; ****, p<0.0001.*

4.2.3. Kozak context sequence influences the initiation of RAN translation of FMRpolyG

The Kozak consensus sequence is one of the crucial factors playing a role in the AUG-initiated translation. The Kozak sequence context in vertebrates is GCCRCC**ATG**G where R is a purine (A or G). Positions -3 and +4 are considered the most important in regulating the efficiency of the translation initiation process. However, even a weak Kozak context sequence might not necessarily result in a weak translation. It happens because other mechanisms like leaky scanning, re-initiation or internal initiation of translation, or stable secondary RNA structures present near the start codon may also play a role in the regulation of translation initiation.

Nevertheless, the recognition of the AUG start codon strongly depends on the nucleotide context present in the vicinity of that codon. In the case of non-AUG translation that context may have an even higher influence on the codon recognition and therefore on the efficiency of translation initiation. Thus, it implies that the initiation of RAN translation could be sensitive to nucleotides at positions that not affect the initiation at the AUG codon.

Therefore, I asked how the optimal context of the Kozak sequence is crucial for the efficient initiation of FMRpolyG translation. To analyze this issue six mutants of Kozak sequence context for ACG (+1) near-cognate start codon in both reading frames, FMRpolyG and FMRP, were generated. Three of them were designed to weaken the sequence context (Kozak1, Kozak2, and Kozak3; ***Figure 32a***), and another three to make the context stronger (Kozak4, Kozak4b, Kozak5) (***Figure 33a***). Kozak1 had a mutation at the -3 position (G→T), Kozak2 at the +4 position (G→A), and Kozak3 was a double mutant and had mutations at both -3 and +4 positions (G→T and G→A, respectively). The mutation design was based on the data describing sequence requirements for translation initiation at non-AUG start codons[203].

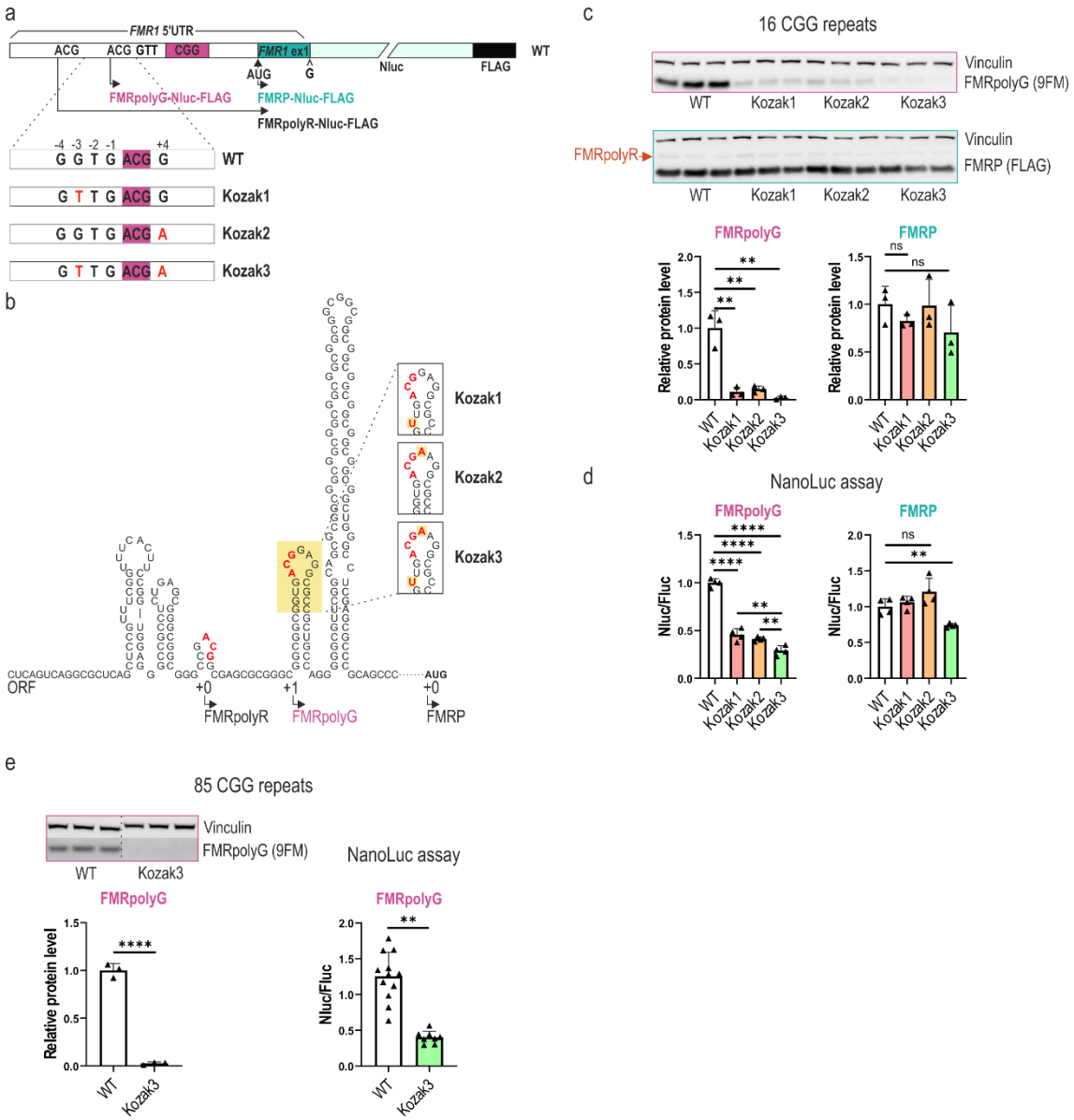


Figure 32. The efficiency of FMRpolyG translation from ACG (+1) near-cognate start codon embedded within a weak context of Kozak sequence. *Three mutants of ACG (+1) Kozak sequence were designed to make the context weaker – Kozak1, Kozak2, and Kozak3;* ***a)*** *Scheme of cloned constructs. Kozak1-3 mutants were developed for 16 CGG repeats, while Kozak3 was additionally cloned for 85 CGG repeats;* ***b)*** *Predicted structure of FMR1 5'UTR with introduced mutations. The structure prediction is presented for FMR1 5'UTR with 16 CGGs;* ***c)*** *16 CGGs: Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody and FMRP equivalent measured by the anti-FLAG antibody. The FMRpolyR produced from 16FMRP-Nluc-FLAG is marked by an orange arrow. Results were normalized to Vinculin. Graphs present results for N = 3 biologically independent samples with SDs;* ***d)*** *16 CGGs: Quantification of results from NanoLuc assay. Graphs present results for N = 4 biologically independent samples with SDs;* ***e)*** *85 CGGs: (Left) Western blot analysis and corresponding quantification of FMRpolyG level analyzed as in* ***c****. Graphs present results for N = 3 biologically independent samples with SDs; (right) Quantification of results from NanoLuc assay. Graphs present results for N = 12 and N = 8 biologically independent samples with SDs for WT and Kozak3, respectively;*

*Statistical analysis was based on a two-tailed unpaired Student's t-test; **, p<0.01; ****, p<0.0001; ns, non-significant.*

As it is presented in ***Figure 32b*** predicted structures of *FMR1* 5'UTR after introducing Kozak1, Kozak2, and Kozak3 mutations do not change and only single mismatches are present in the Kozak1 and Kozak3 mutants. Thus, the main factor responsible for observed changes in the FMRpolyG level was indeed the changed ACG (+1) Kozak sequence context. Results from the western blot (***Figure 32c***) demonstrated that weakening the Kozak context for ACG (+1) strongly decreased the initiation of RAN translation. The additive effect of a double mutant was also visible since the level of FMRpolyG was almost undetectable in 16FMRpolyG-Kozak3. As expected, these changes did not influence the canonical translation initiation measured by synthesis of FMRP equivalent as well as FMRpolyR (***Figure 32c***, orange arrow). Although results generated from the NanoLuc assay had the same trends, the depth of fold change for FMRpolyG was smaller compared to western blot results which probably resulted from the translation of some additional proteins in frame with Nluc-FLAG (***see Figure 9***, western blot results for FMRpolyG with anti-FLAG antibody).

Nevertheless, it can be concluded that the weakening of ACG (+1) Kozak sequence context had an enormous negative effect on the RAN translation initiation at this near-cognate start codon. Simultaneously, the canonical translation from the AUG codon for the FMRP protein remained unchanged.

Encouraged by these results I wanted to test whether, a strong secondary structure formed by 85 CGG repeats will be able to force the initiation at ACG (+1) near-cognate start codon lying in the weakest Kozak context – Kozak3, and therefore increase the level of FMRpolyG. To achieve this goal I cloned the 85FMRpolyG-Kozak3 construct and used it to transfect HEK-293 cells. As presented in ***Figure 32e*** the hairpin structure formed by longer, 85 CGG, repeats did not change the level of translated FMRpolyG protein which suggests that the sequence composition is the major factor regulating the efficiency of RAN translation initiation which in this particular case is weakened to a similar extent by different number of CGG repeats.

In the next step, the same analysis was performed for mutants designed to make the ACG (+1) Kozak sequence context stronger than the wild-type (Kozak4, Kozak4b, and Kozak5; ***Figure 33a***). Kozak4 and Kozak4b had mutations at the -4 position (G→C, and G→A, respectively), however, the Kozak4 had a distinctly better context than the

Kozak4b. Kozak5 had mutations at the -4 (G→C) and at the -2 (T→C) positions. Similarly to above mentioned Kozak mutants, these three new mutations did not influence the structural predictions of mutated *FMR1* 5'UTR, except single mismatches in the Kozak4 and Kozak5 mutants (***Figure 33b***). These mutations were also designed based on the results obtained by high-throughput analysis performed by Diaz de Arce and colleagues[203].

The native ACG (+1) Kozak sequence has quite optimal context (G**G**TGACG**G**, G at -3 and +4 positions), therefore, I focused on the mutation of less important positions from the codon recognition point of view. Therefore for these mutants positions -4 and -2 were selected. Kozak5 contained almost the best ACG (+1) Kozak sequence context – C**G**CGACG**G** (mutated nucleotides are underlined). The one possible stronger context would introduce an additional ACG codon (C**A**CGACG**G**).

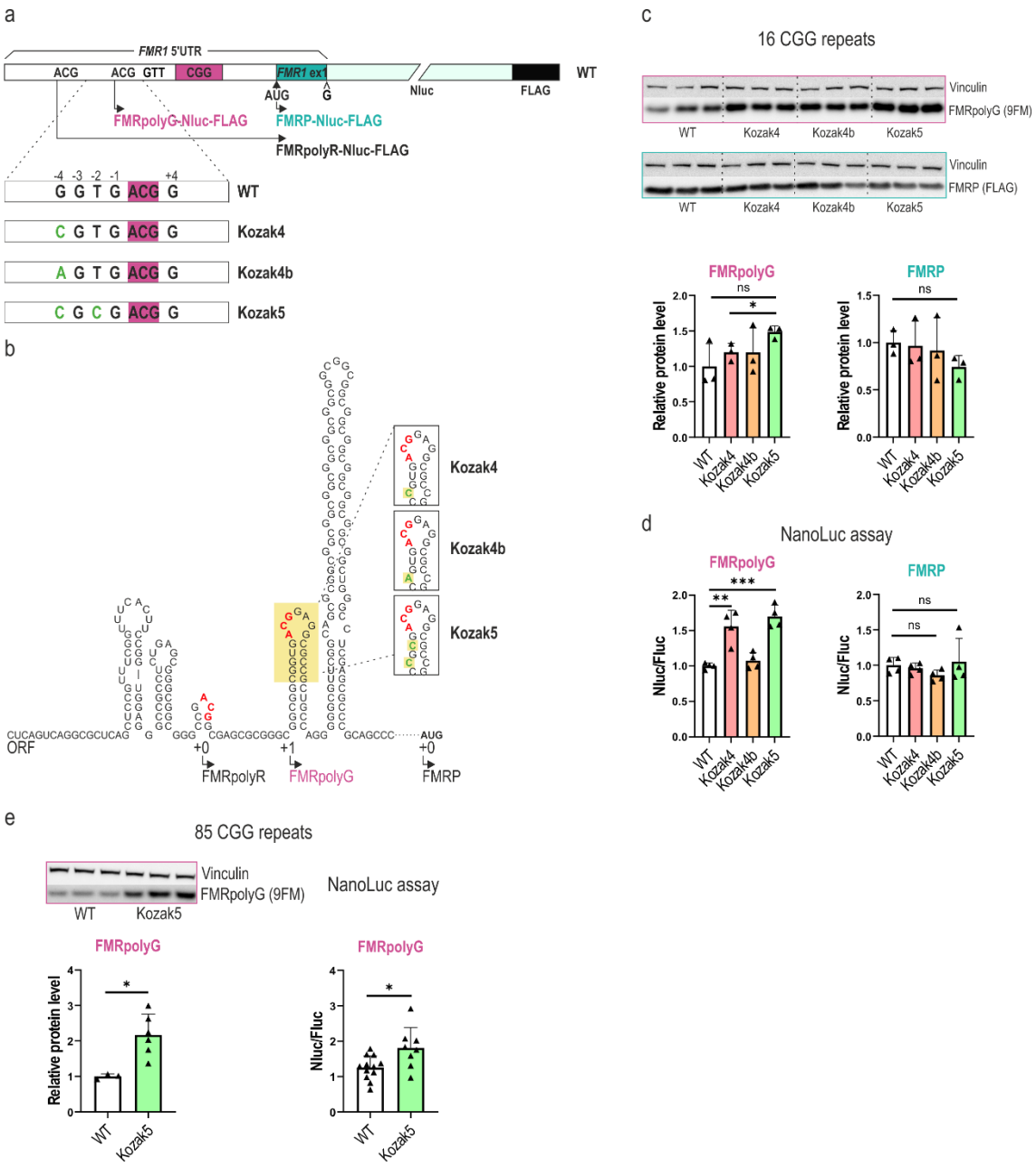


Figure 33. The efficiency of FMRpolyG translation initiation from ACG (+1) codon embedded within a strong context of Kozak sequence. *Three mutants of ACG (+1) Kozak sequence were designed to make the context stronger – Kozak4, Kozak4b, and Kozak5;* ***a)*** *Scheme of cloned constructs. Kozak4-5 mutants were developed for 16 CGG repeats, while Kozak5 was additionally cloned for 85 CGG repeats;* ***b)*** *Predicted structure of FMR1 5'UTR with introduced mutations. The structure prediction is presented for FMR1 5'UTR with 16 CGGs;* ***c)*** *16 CGGs: Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody and FMRP equivalent measured by the anti-FLAG antibody. The FMRpolyR produced from 16FMRP-Nluc-FLAG is omitted since it was almost undetectable. Results were normalized to Vinculin. Graphs present results for N = 3 biologically independent samples with SDs;* ***d)*** *16 CGGs: Quantification of results from NanoLuc assay. Graphs present results for N = 4 biologically independent samples with SDs;* ***e)*** *85 CGGs: (Left) Western blot analysis and corresponding quantification of FMRpolyG level analyzed as in* ***c****. Graphs present results for N = 3 and N = 6 biologically independent samples with SDs for WT and Kozak5, respectively. Results were*

*normalized to Vinculin, (right) Quantification of results from NanoLuc assay. Graphs present results for N = 12 and N = 8 biologically independent samples with SDs for WT and Kozak5, respectively; Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ***, p<0.001; ns, non-significant.*

Data obtained by NanoLuc assay (***Figure 33d***) suggest that improvement of ACG (+1) Kozak sequence context by Kozak4 and Kozak5 mutant had a positive effect on the FMRpolyG RAN translation initiation. A similar trend could be observed in the results generated by western blot analysis where the Kozak5 mutant had also the strongest positive effect on the FMRpolyG level (***Figure 33c***). However, due to the high standard deviation in the WT samples, the observed change was not statistically significant.

The fold change of the observed effect was much smaller in comparison to Kozak1-3 mutants which results directly from the fact that native ACG (+1) is in the optimal Kozak context so introduced mutations could make the context better only to a small extent. According to assumptions, performed mutations did not affect the efficiency of translation initiation at AUG codon for the FMRP equivalent. Also, no stronger increase in the FMRpolyG level was observed for 85 CGG repeats (***Figure 33e***). This suggests that similarly to the Kozak3 mutant, the sequence context is crucial for RAN translation initiation.

4.2.4. Other near-cognate start codons within the 5'UTR of *FMR1* are effective in RAN translation initiation

Many near-cognate start codons located upstream to the main ORF constitute the translation initiation site for upstream open reading frames and may strongly impair the translation of the main ORF. Hence, to avoid translation inhibition, near-cognate start codons are usually embedded in a weak Kozak sequence context that enables leaky scanning of the ribosome and the translation initiation at the downstream, main ORF. However, this is not what is observed for ACG (+1) near-cognate start codon which is embedded in the optimal Kozak context, albeit the FMRP level remains almost unchanged.

Near-cognate start codons have been shown to initiate translation at frequencies of ~ 0 – 10% of AUG codon in the optimal Kozak sequence context[204,205]. To test how efficiently RAN translation initiation will occur at different near-cognate start codons embedded in the same Kozak sequence context I designed four mutants containing the following mutations of ACG (+1) codon: ACG→CTG, ACG→GTG, ACG→AAA, and

ACG→ATG (***Figure 34a***). Since introduced mutations were concerning the sequence region which was predicted to fold into a loop, no changes in the structure of these mutated *FMR1* 5'UTR were observed (***Figure 34b***).

According to available data[205,206] the CUG codon should present the highest efficiency of codon utilization. Indeed, data from the western blot confirmed that mutation of ACG (+1) codon to CUG increased the translation initiation of FMRpolyG (***Figure 34c***). Interestingly, this mutation affected also the initiation at the AUG codon for the synthesis of the FMRP equivalent. As expected, the mutation of the ACG (+1) codon to GUG also positively regulated the RAN translation, albeit to a smaller extent in comparison to the ACG→CTG mutation. However, no change in the case of FMRP equivalent level was observed. According to assumptions, mutation of ACG→AAA resulted in almost complete loss of FMRpolyG protein with the simultaneous slight, but not statistically significant increase in the translation of the FMRP equivalent.

On the other hand, the mutation of ACG→ATG led to a high increase in RAN translation initiation measured by the level of FMRpolyG protein and complete loss of the equivalent of FMRP protein. Since data obtained by NanoLuc assay (***Figure 34d***) roughly presented the same results the difference in fold change of ACG→ATG mutant in the FMRpolyG frame arose directly from the underestimation of western blot data resulting from the membrane burnout in ACG→ATG samples.

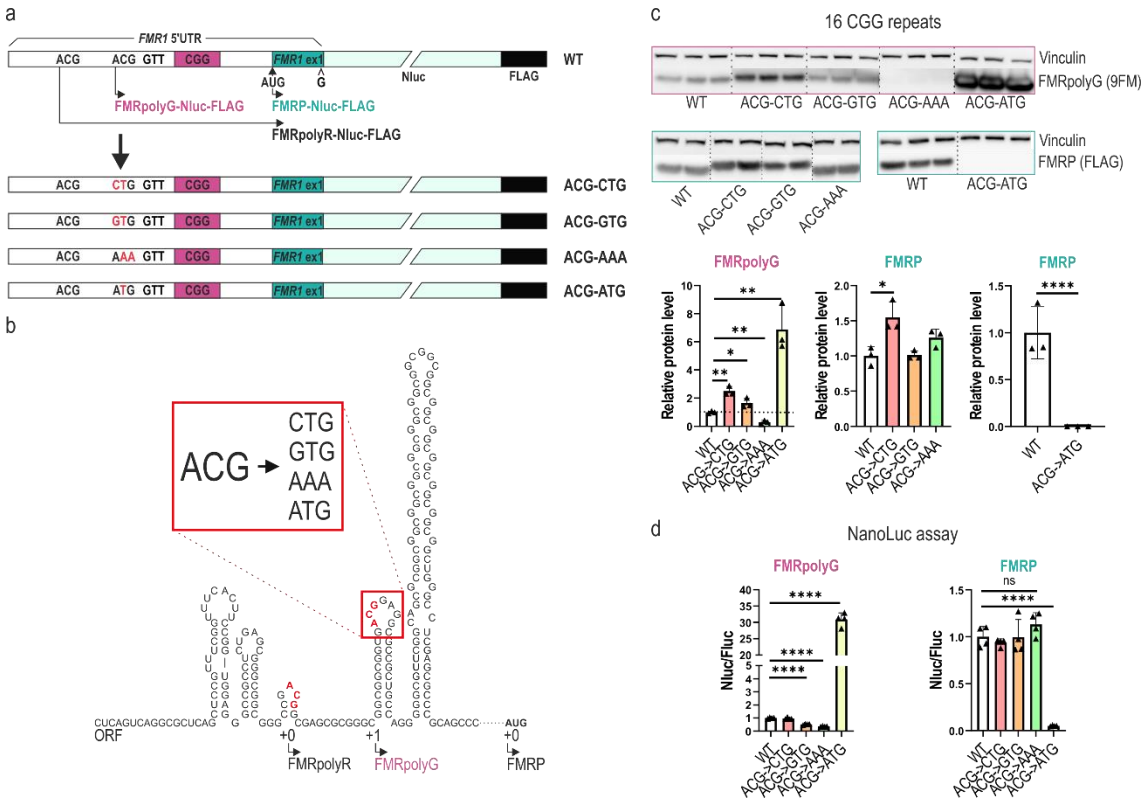


Figure 34. The efficiency of FMRpolyG translation initiation at different near-cognate start codons. *Four mutants of ACG (+1) near-cognate start codon has been designed: ACG→CUG, ACG→GUG, ACG→AAA, and ACG→AUG.* ***a)*** *Scheme of cloned constructs;* ***b)*** *Predicted structure of FMR1 5'UTR with different near-cognate start codons;* ***c)*** *Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody and FMRP equivalent measured by anti-FLAG antibody. The FMRpolyR produced from 16FMRP-Nluc-FLAG is omitted since it was almost undetectable. Results were normalized to Vinculin. Graphs present results for N = 3 biologically independent samples with SDs. Gels were cropped;* ***d)*** *Quantification of results from NanoLuc assay. Graphs present results for N = 4 biologically independent samples with SDs. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ****, p<0.0001; ns, non-significant.*

Although studied here artificial systems are different from what is happening in the cell they were designed to study the relationship between the RAN and AUG-initiated translation from the *FMR1* mRNA and provide information about the regulation of these processes from the mechanistic point of view.

Based on these results it can be concluded that other near-cognate start codons, like CUG and GUG, in optimal Kozak sequence context, are effective in RAN translation initiation. The extensive growth in the efficiency of RAN translation initiation had a direct effect on the signal loss of the equivalent of FMRP protein. Observed dependence emerged from the fact that both proteins are translated from the same mRNA. To better visualize this issue I proposed a model which describes the mechanism behind these results (***Figure***

35). The model assumes that in wild-type *FMR1* mRNA (***Figure 35a***) a 43S preinitiation ribosome complex scans the mRNA searching for an appropriate – AUG – start codon. When the AUG codon is found the 80S ribosome starts the synthesis of FMRP. In this scenario almost always translation starts at the AUG codon for FMRP ORF, and only very few ribosomes will initiate translation at the ACG (+1) near-cognate start codon for RAN translation. However, if the ACG (+1) near-cognate start codon is mutated to AUG (***Figure 35a'***) due to extremely high AUG codon utilization translation will initiate almost always at the mutated codon. Therefore, very few 43S preinitiation complexes will be able to initiate translation at the native AUG start codon for FMRP synthesis as the result of a leaky scanning process.

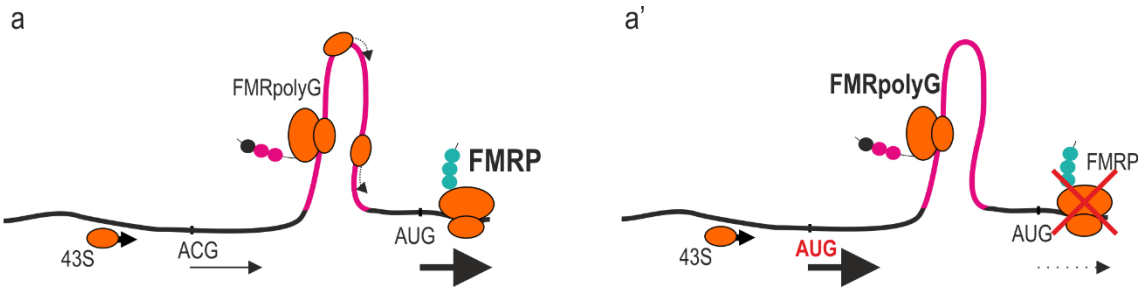


Figure 35. Mechanistic insight of FMRpolyG and FMRP translation initiation model in ACG→ATG mutant. a) *43S preinitiation complex (43S) scans the mRNA in the 5' to 3' direction to initiate the translation at the AUG codon. Due to the RAN translation initiation, only a small percentage of scanning 43S ribosomes will initiate FMRpolyG protein synthesis at the ACG (+1) near-cognate start codon;* ***a')*** *When ACG (+1) near-cognate start codon is mutated to the AUG codon, the scanning 43S will preferentially start the translation at this codon since it will be the first AUG codon found in the mRNA. The FMRP synthesis from the downstream AUG codon will be strongly impaired.*

4.2.5. Localization of the ACG (+1) near-cognate start codon within the 5'UTR of *FMR1* influences the level of FMRpolyG

Since many near-cognate start codons within *FMR1* 5'UTR may act as FMRpolyG translation start site I wanted to establish whether randomly localized ACG codon in +1 frame would be efficient in RAN translation initiation. Therefore, I designed two mutants: rACG1 with the ACG (+1) near-cognate start codon located 33 nt upstream of the native ACG (+1) codon, and rACG2 with the ACG (+1) near-cognate start codon located 9 nt upstream of the native ACG (+1) (***Figure 36a***). In both cases, the native ACG (+1) near-cognate start codon remained unchanged. To introduce these rACG1 and rACG2 mutations I did one point mutations thus no extra nucleotides were added to the sequence.

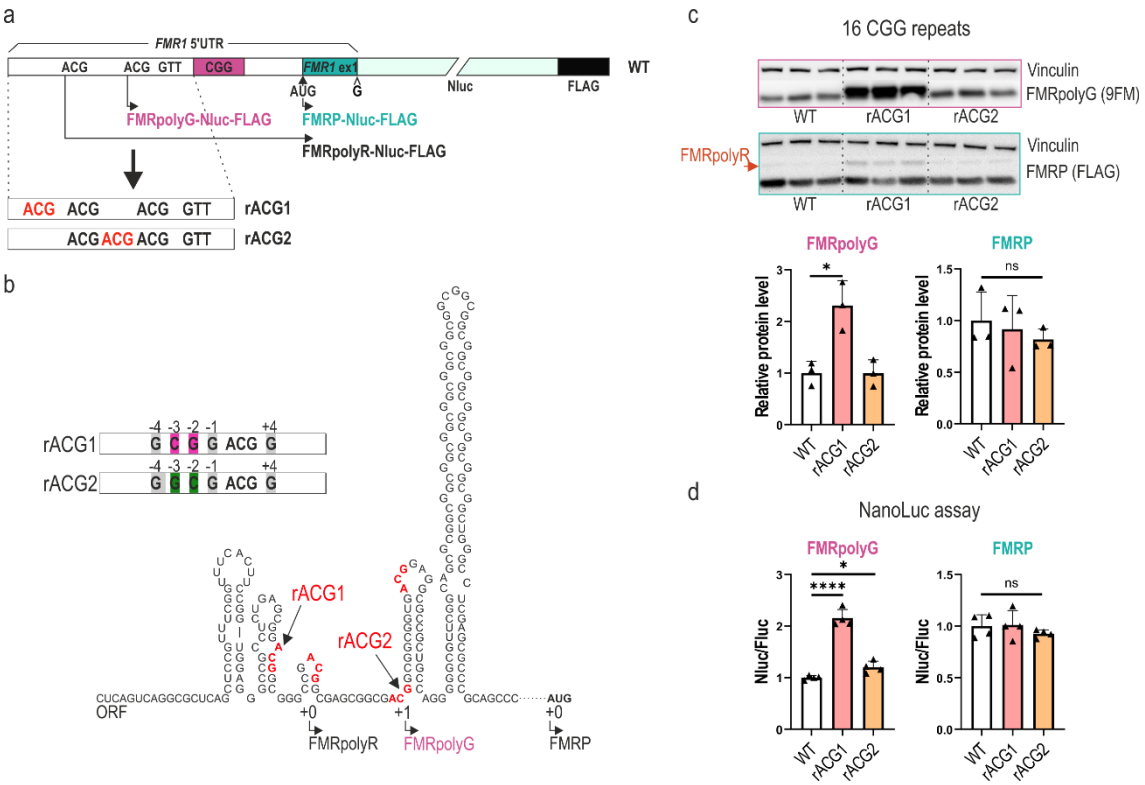


Figure 36. The sequence-dependent and secondary RNA structure-dependent efficiency of FMRpolyG translation initiation from different ACG (+1) near-cognate start codons introduced into FMR1 5'UTR. a) *Scheme of rACG1 and rACG2 constructs;* ***b)*** *Predicted structure of FMR1 5'UTR with introduced mutations. Both mutations are presented simultaneously. The Kozak sequence context for each ACG (+1) near-cognate start codon is presented: the weaker context of the rACG1 (upper case), and the stronger context of the rACG2 (lower case);* ***c)*** *Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody and FMRP equivalent measured by the anti-FLAG antibody. The FMRpolyR produced from 16FMRP-Nluc-FLAG is marked by an orange arrow. Results were normalized to Vinculin. Graphs present results for N = 3 biologically independent samples with SDs;* ***d)*** *Quantification of results from NanoLuc assay. Graphs present results for N = 4 biologically independent samples with SDs. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; ****, p<0.0001; ns, non-significant.*

As presented in ***Figure 36b*** designed mutations did not affect the predicted general structure of *FMR1* 5'UTR. The amount of FMRP equivalent measured by western blot and NanoLuc assay remained unchanged (***Figure 36c-d***). The level of FMRpolyG however was significantly elevated in the rACG1 mutant. The western blot analysis revealed that, as expected, two proteins were produced: one translated from the new rACG1 (+1) codon and the other from the natively present ACG (+1). The quantification of signal intensities suggested that localization of the ACG (+1) codon within the rACG1 mutant was very effective since the level of FMRpolyG was twice as high as in the control. The translation of FMRpolyR seemed to be correlated with the increased

FMRpolyG translation in the rACG1 mutant which is in line with the results obtained in 4.2.1.2. subchapter. Interestingly, the ACG (+1) codon within the rACG2 mutant seemed to be inactive (***Figure 36c-d***) as only FMRpolyG translated from the native ACG (+1) codon was observed. Surprisingly, the observed results were inconsistent with the strength of the Kozak sequence context of the rACG1 and rACG2 mutants. As presented in ***Figure 36b***, the rACG1 mutant was characterized by a weaker Kozak consensus sequence than rACG2, which seemed to be inactive. Due to these discrepancies, I took a closer look at the structural determinants that would affect the choice of ACG (+1) codon and translation initiation efficiency.

It revealed that, according to predictions, rACG1 was located at a much more optimal distance from the short but quite stable RNA hairpin structure (containing the native ACG (+1) codon) than rACG2 which was engaged in the formation of the stem of this hairpin structure (***Figure 36b***). Although this short hairpin was predicted to have only ΔG=−14 kcal/mol this structure could be stabilized by adjacent downstream hairpin formed by 16 CGG repeats which was predicted to have ΔG=−54 kcal/mol. Thus, the scanning 43S preinitiation complex could be affected and slowed down leading to ACG (+1) near-cognate start codon selection, even if it was embedded within a weak Kozak context sequence. This mechanism I believe applies to the rACG1 mutant where the ACG (+1) codon was located 23 nt upstream of the beginning of this short hairpin. As ribosome profiling data present this is the distance that can be considered as the length of mRNA covered by the ribosome during translation which is roughly estimated as 30 nt[207]. However, ribosome footprints vary by organism, cell type, and experimental protocol[208]. Thus, the scanning ribosome could be more eager to recognize near-cognate start codons if they are positioned in the middle of the large ribosome subunit. Nevertheless, the distance between the ACG (+1) of the rACG2 mutant seemed to be too short to position the ribosome at optimal orientation, which is the P-site at the ACG codon. However, these conclusions are based on the presumptions resulting from the secondary RNA structure predictions, which have to be verified by the experimental procedures.

Hence, I assumed that the difference in the distance between studied ACG (+1) codons in rACG1 and rACG2 mutants and the structural blockage that could slow down the scanning 43S ribosome could be the main factor influencing the efficiency of translation initiation at the particular ACG (+1) codon.

4.2.6. Distance between native ACG (+1) and downstream stable RNA secondary structure significantly affects the RAN translation initiation

Based on the results obtained for rACG1 and rACG2 mutants I decided to test a hypothesis that the distance between the translation start site and the stable secondary RNA structure influences the efficiency of FMRpolyG biosynthesis, especially that similar mechanisms for regulation of AUG-initiated translation have been stated before for different mRNAs[121,209]. This dependence was expected to be particularly important for non-AUG initiation which is exceptionally sensitive to conditions that slow down or pause the progression of scanning of 43S ribosome.

To achieve this goal I designed four mutants, on the backbone of constructs with 16 CGG repeats, which have inserted an artificial hairpin-forming sequence at different regions downstream of native ACG (+1) near-cognate start codon (***Figure 37a***). The sequence of the structure was selected from a previously published study[167] and the necessary modifications were performed (***see Methods 3.1.3.2.4. "Constructs with additional hairpin forming sequence"***). The Gibbs free energy of the predicted hairpin was ΔG=–46 kcal/mol while the downstream hairpin formed by 16 CGG repeats with stabilizing flanking regions had ΔG=–54 kcal/mol. Therefore, the stability of the artificial hairpin was similar to the structure formed by the repeats. The sequence-forming-structure was introduced 2-, 6-, 14- and 20 nt downstream of native ACG (+1) near-cognate start codon, and developed constructs were named Hairpin2nt, Hairpin6nt, Hairpin14nt, and Hairpin20nt, respectively. The predictions of mutated *FMR1* 5'UTRs secondary structures are presented in ***Figure 37c-f***.

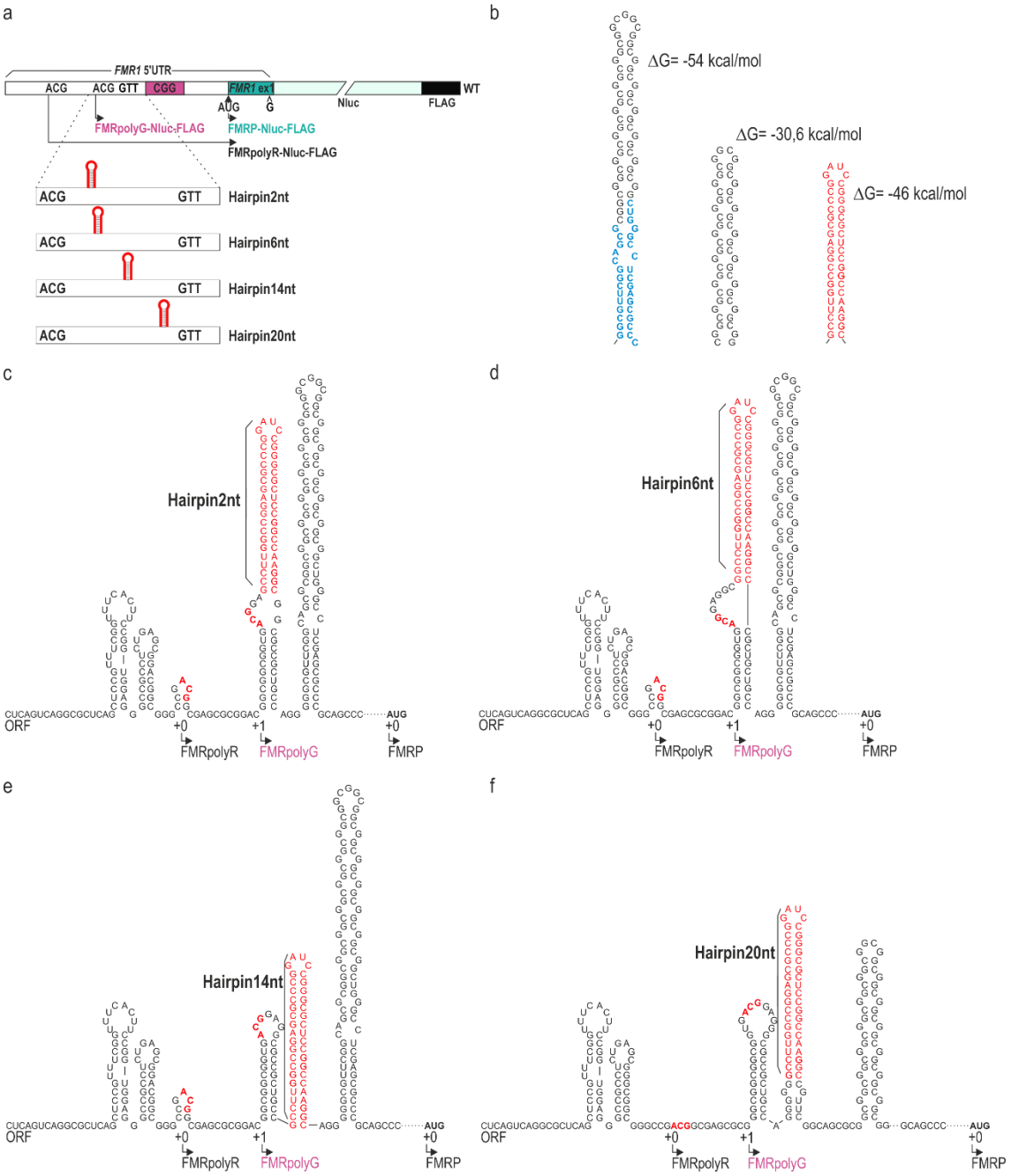


Figure 37. Predictions of RNA secondary structure formed by FMR1 5'UTRs with inserted artificial hairpin structures. a) *Scheme of developed constructs;* ***b)*** *Predicted structure and Gibbs free energy of (left) hairpin formed by 16 CGG repeats with stabilizing flanking regions, (middle) the hairpin formed by 16 CGG repeats, and (right) the hairpin formed by the artificial sequence-forming-structure;* ***c-f)*** *Predicted structures of FMR1 5'UTR containing 16 CGGs with introduced artificial hairpins.*

As presented in ***Figure 38a-b*** the distance of 2 nt and 6 nt between the ACG (+1) near-cognate start codon and the hairpin structure had a highly negative effect on the FMRpolyG translation initiation at that codon. However, extending this distance to 14 nt and 20 nt resulted in an extremely high increase in the FMRpolyG level. According to

structure predictions, the hairpin formed by the Hairpin20nt mutant was located 16 nt downstream of the ACG (+1). Therefore, it could result in better positioning of ribosome site P at the ACG codon than the structure formed by Hairpin14nt which is suggested by the obtained results (***Figure 38a-b***).

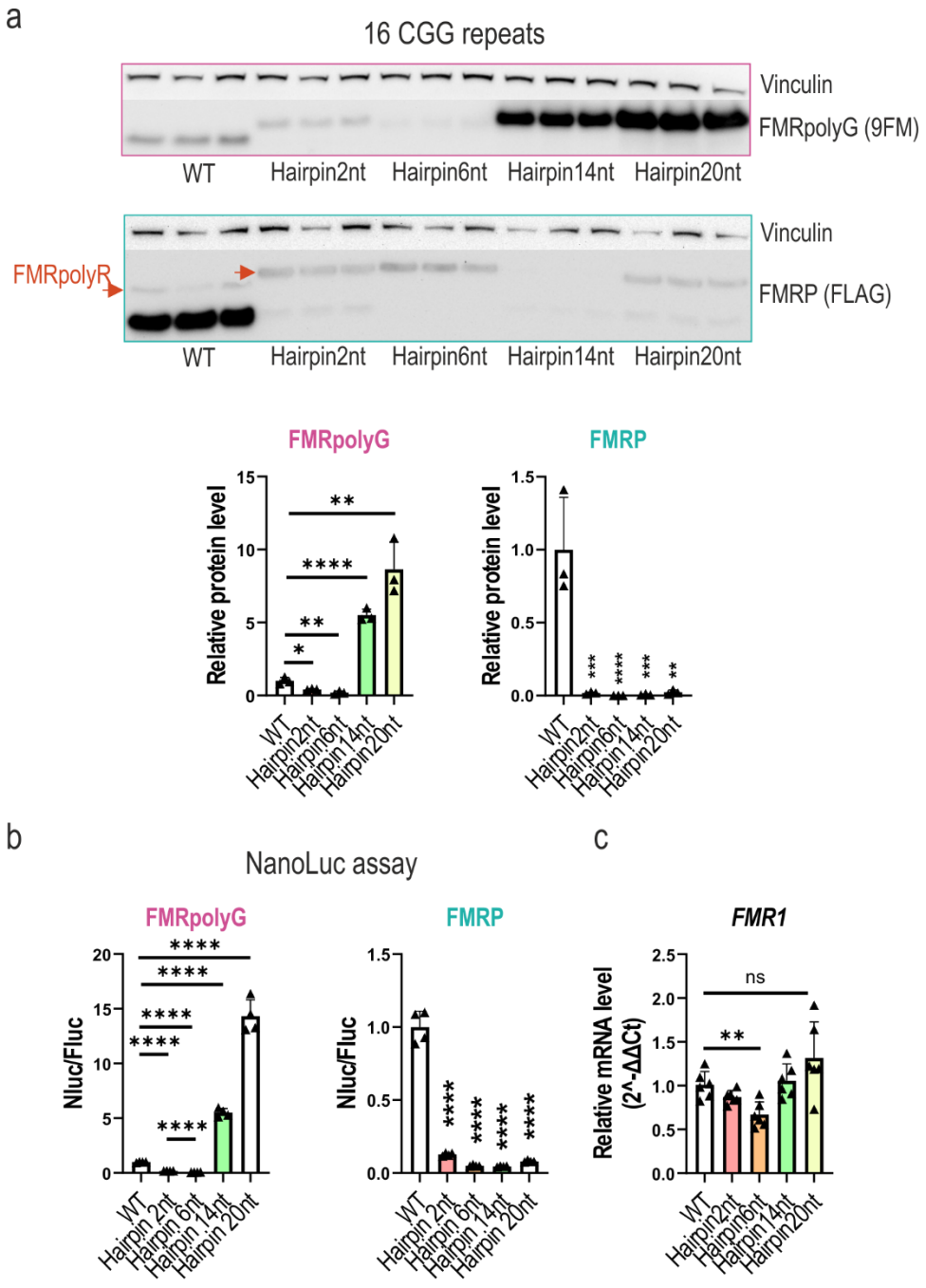


Figure 38. Stable RNA secondary structure downstream ACG (+1) may either decrease or increase the FMRpolyG synthesis. a) *Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody and FMRP equivalent measured by the anti-FLAG antibody. The FMRpolyR produced from 16FMRP-Nluc-FLAG is marked by an orange arrow. Results were normalized to Vinculin. Graphs present average results from N = 3 biologically*

independent samples with SDs; ***b)*** *Quantification of results from NanoLuc assay. Graphs present results for N = 4 biologically independent samples with SDs;* ***c)*** *Relative FMR1 mRNA level. Results were normalized to GAPDH. The graph presents results for N = 3 biologically independent samples (each with n = 2 technical replicates) with SDs; Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001; ns, non-significant.*

The analysis of the *FMR1* mRNA level revealed that the amount of transcript generated from different plasmids is even confirming that observed changes resulted from the translation efficiency, not from the altered transcription/RNA stability. The only statistically significant change was observed for Hairpin6nt which showed a small decrease in the *FMR1* level. Thus, the observed decrease in the level of FMRpolyG and FMRP equivalent translated from this mRNA might result, at least partially, directly from the reduced mRNA level. Nevertheless, the efficiency of FMRP equivalent translation was impaired in all tested mutants. It could result partially from the general decrease in mutant *FMR1* mRNA translation efficiency or more likely due to the presence of an additional stable hairpin adjacent to the hairpin formed by CGG repeats which led to the 43S ribosome dissociation during scanning. Hence, similarly to ACG→ATG mutant, only a limited number of ribosomes, due to leaky scanning or re-initiation, could start translation at the AUG codon for FMRP synthesis. The model of this proposed mechanism is illustrated in ***Figure 39***. The inhibiting role of complex secondary structures within *FMR1* 5'UTR on the downstream ORF could also explain the almost total loss of FMRP translated from Hairpin14nt, which based on structure prediction had the most structured 5'UTR and could most efficiently block the incoming ribosomes.

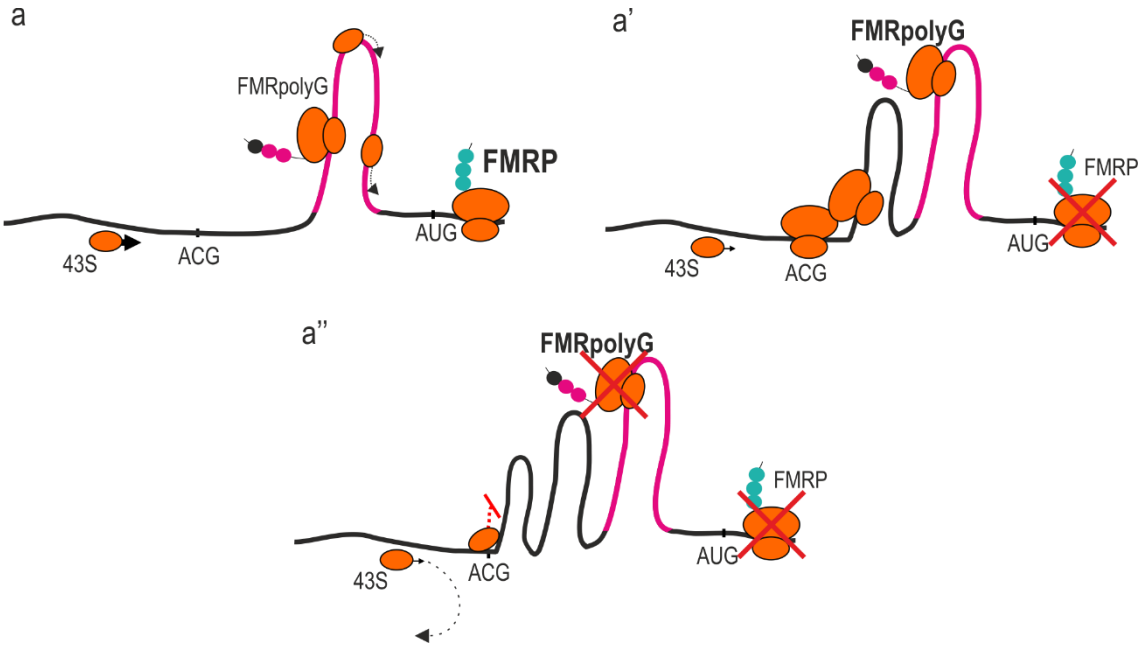


Figure 39. Model of translation initiation at highly structured FMR1 5'UTR. a) *43S preinitiation complex (43S) scans the mRNA in the 5' to 3' direction to initiate the translation at the AUG codon. Due to the RAN translation initiation, only a small percentage of scanning 43S ribosomes will initiate FMRpolyG protein synthesis at the ACG (+1) near-cognate start codon;* ***a')*** *When the structural obstacle is too stable for preinitiation complex then 43S scanning ribosome can dissociate from the mRNA or only partially resolve the secondary RNA structure. Thus, very few or even none scanning 43S ribosomes will reach the AUG start codon for FMRP ORF. However, due to forced, by the RNA structure, translation initiation at near-cognate start codon the complete 80S ribosome will be formed and resolve the structure to process with RAN translation of FMRpolyG;* ***a")*** *When the 5'UTR is highly structured and formed complex secondary structures can not be resolved by the 43S ribosome or even by the 80S ribosome then the complete translation inhibition of particular mRNA may occur.*

In the next step, I designed another mutant with an extended distance between the ACG (+1) near-cognate start codon and the hairpin structure formed by CGG repeats (***Figure 40a-b***). I wanted to establish whether a distantly located structure would have any effect on the translation initiation at the ACG (+1). The sequence 5' CACACACACACACACACA 3' was predicted to not form any secondary RNA structure. I added a CAlinker sequence (18 nt long) 14 nucleotides downstream ACG (+1) in FMRpolyG-Nluc-FLAG constructs with 16 and 85 CGGs. Hence, the distance between the ACG (+1) near-cognate start codon and the hairpin structure was extended from 17 nt to 35 nt.

I expected that the initiation of FMRpolyG translation would be reduced, however, the level of FMRpolyG depletion was surprising (***Figure 40c, left and middle panels***). The constant level of *FMR1* mRNA (***Figure 40c, right panel***) suggested that the observed loss of FMRpolyG protein was directly resulting from ineffective translation initiation. Interestingly, the FMRpolyG depletion was much less significant for mutant with 85 CGG repeats which could suggest that in the case of long repeats, and more stable RNA structure (ΔG=–215,9 kcal/mol) the ribosome queuing mechanism of translation initiation may be at play.

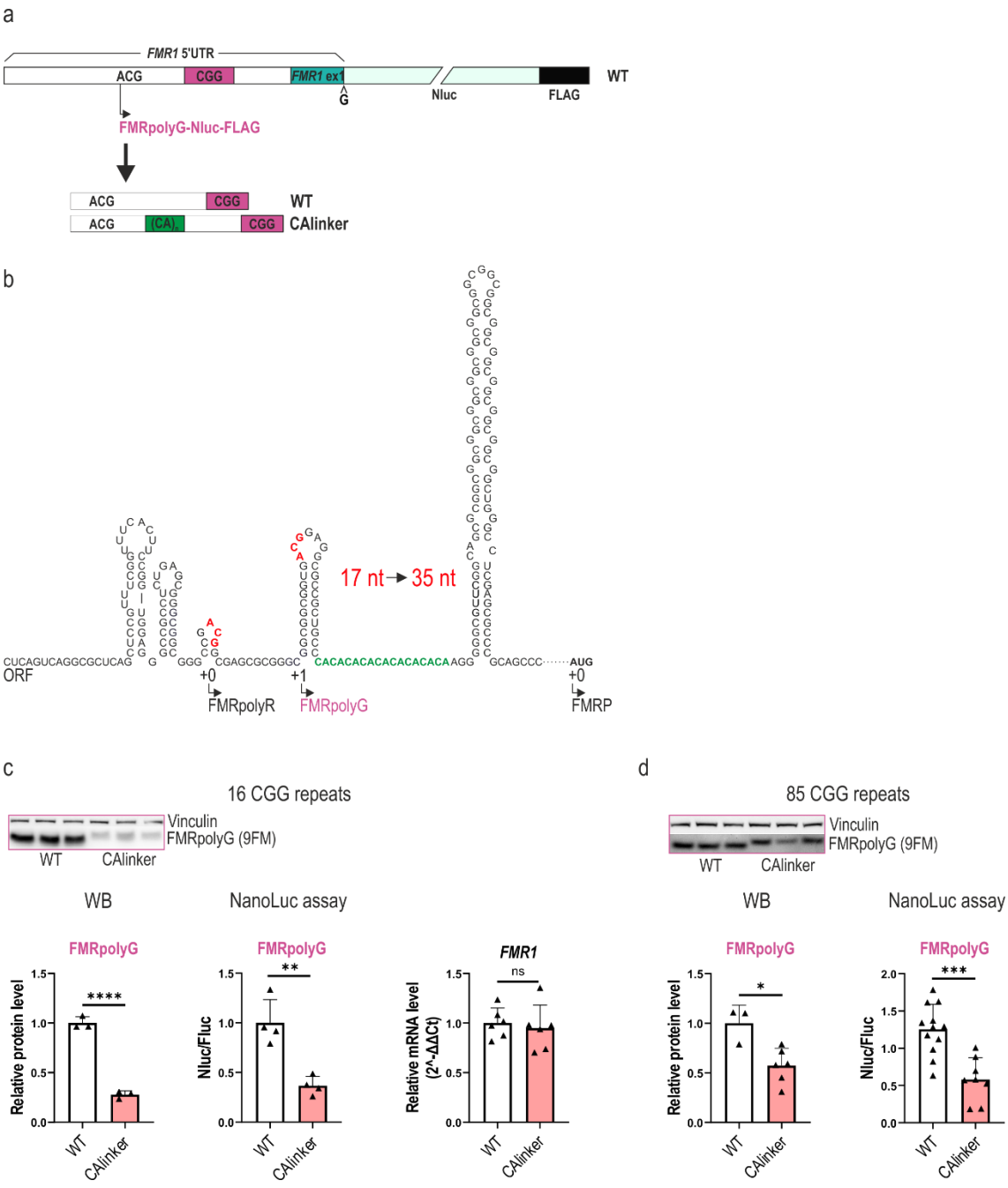


Figure 40. Extended distance between the ACG (+1) near-cognate start codon and CGG hairpin structure has a negative effect on the FMRpolyG translation initiation. a) *Scheme of CAlinker construct;* ***b)*** *Predicted structure of FMR1 5'UTR with introduced mutation. The sequence of the linker is marked in green. The structure prediction is presented for FMR1 5'UTR with 16 CGGs;* ***c)*** *16 CGGs: (Left) Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody. Results were normalized to Vinculin. The graph presents results for N = 3 biologically independent samples with SD, (middle) Quantification of results from NanoLuc assay. The graph presents results for N = 4 biologically independent samples with SDs, (right) Relative FMR1 mRNA level. Results were normalized to GAPDH. The graph presents results for N = 3 biologically independent samples (each with n = 2 technical replicates) with SDs;* ***d)*** *85 CGGs: The same as in* ***c****, albeit without relative FMR1 mRNA level analysis. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001; ns, non-significant.*

Finally, I wanted to verify how the initiation of RAN translation would be regulated by the binding of antisense oligonucleotides (ASOs) within different regions of CGG repeats involved in the formation of CGG repeat hairpin structure either at its 5' or 3' site. I used the ASOs targeting flanking regions of CGG repeats (ASO1 and ASO3) which, as it was confirmed, are responsible for hairpin stabilization[181]. Chosen ASOs were 20-nucleotide-long **steric blockers** and were exclusively composed of 2'MOE units. The target sequences are marked in ***Figure 41a***.

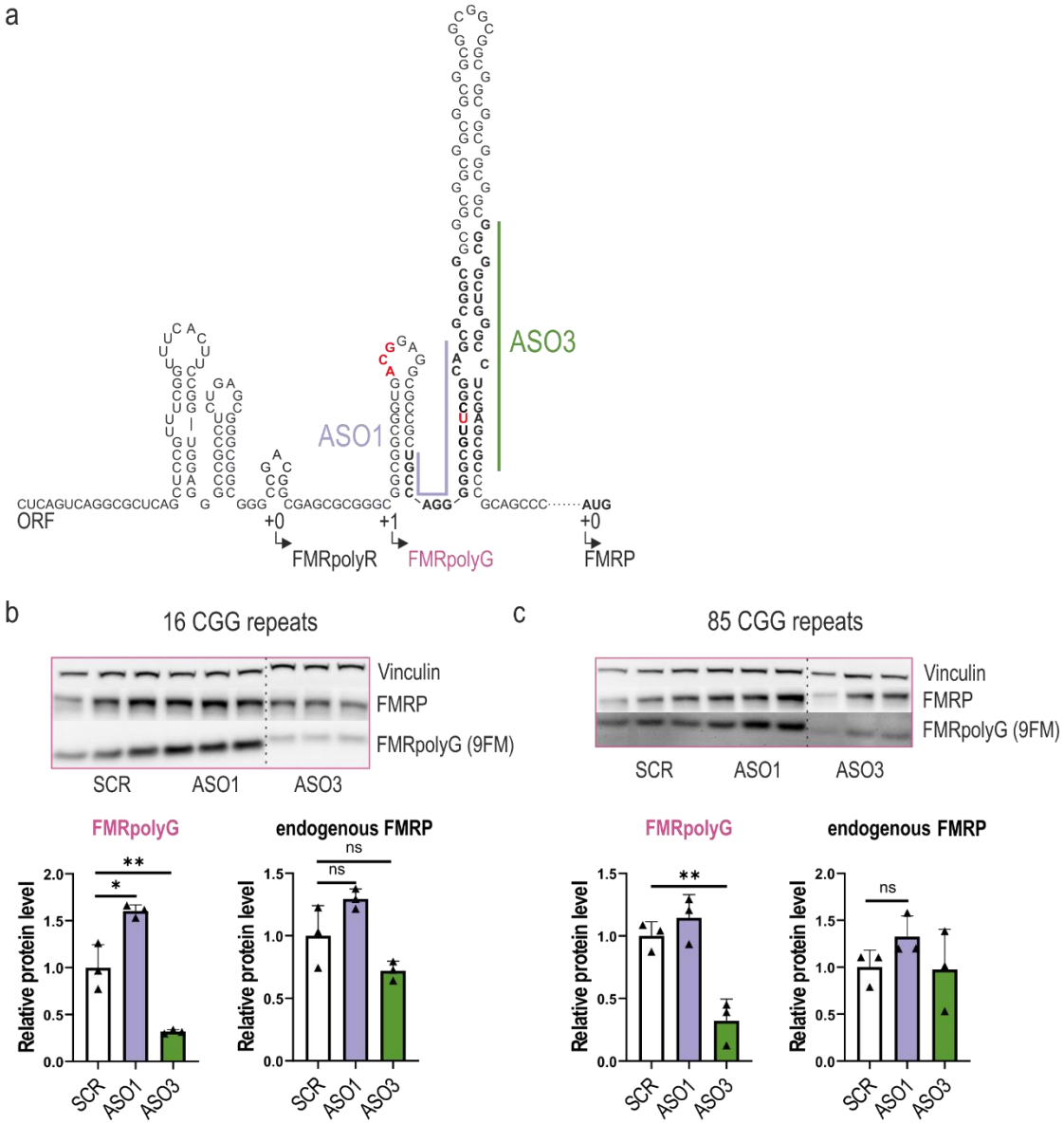


Figure 41. ASOs targeting the flanking regions of the CGG hairpin structure regulate the initiation of FMRpolyG translation. a) *Predicted structure of FMR1 5'UTR with marked target sequences for ASO1 and ASO3 binding. The mismatch of one nucleotide in the region of GUG→GUU mutation is marked in red;* ***b)*** *16 CGGs: Western blot analysis and corresponding quantification of FMRpolyG level measured by 9FM antibody and endogenous FMRP level measured by anti-FMRP antibody. Results were normalized to Vinculin. The graph presents results for N = 3 biologically independent samples with SD;* ***c)*** *85 CGGs: The same as in* ***b****. Gels*

*were cropped. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; **, p<0.01; ns, non-significant.*

HEK-293 cells were transfected with appropriate constructs and after 3 h the transfection of 200 nM ASOs was performed. Cells were harvested 48 h after the second transfection and proceeded with western blot analysis. For this experiment, the following plasmids have been used: 16FMRpolyG-Nluc-FLAG and 85FMRpolyG-Nluc-FLAG (called WT in this dissertation).

The ASO3, targeting the 3' site of the flanking region of CGG repeats, resulted in a significant depletion in the FMRpolyG synthesis detected in both models with short (***Figure 41b***) and long (***Figure 41c***) CGG repeats. On the other hand, treatment with ASO1 led to an increase in the FMRpolyG translation, however, only for short CGG repeats the increase was statistically significant. The level of endogenous FMRP level remained unchanged. The inconsistent effect of ASO1 between two models differing in the number of CGG repeats suggests that the observed change in the FMRpolyG level resulted from the mechanism connected with the stability of RNA hairpin structure. On the other hand, the change after ASO3 treatment suggests that the observed result was independent of the number of CGG repeats. Therefore, the possible change in the distance between ACG (+1) near-cognate start codon and the hairpin formed by CGG repeats could be the reason for the observed results.

In ***Figure 42*** I presented how the binding of ASO1 and ASO3 may affect the structure of *FMR1* 5'UTR and therefore change the efficiency of RAN translation initiation at the ACG (+1) near-cognate start codon. Both ASOs could partially resolve hairpin structures in the region of their binding. In the case of ASO1, it could be expected that the binding of steric blocker could directly slow down the PIC scanning or even lead to the PIC stalling and therefore increase the initiation at near-cognate start codon. The distance between the ACG (+1) near-cognate start codon and the RNA region with associated ASO1 was 22 nt, and my previous experiments (***Figure 38***) suggested that such distance from steric blockage was sufficient to force initiation at near-cognate start codon. However, this increase was not observed when a very stable hairpin structure was formed (85 CGGs, ***Figure 41c***) possibly due to the strong decrease in the general *FMR1* translation efficiency. Targeting the 3' site of the CGG flanking region by ASO3 would instead extend the distance between the ACG (+1) near-cognate start codon and the stem of the CGG hairpin, from 17 nt to 30 nt. Hence, the strong depletion in the FMRpolyG

level was observed independently from the number of CGG repeats. Even though the ribosome queuing on the long repeats is possible, the positive effect on the translation initiation at ACG (+1) near-cognate start codon forced by paused ribosomes may be invisible. This may be due to the generally negative effect of long CGG repeats on the translational efficiency of *FMR1* mRNA.

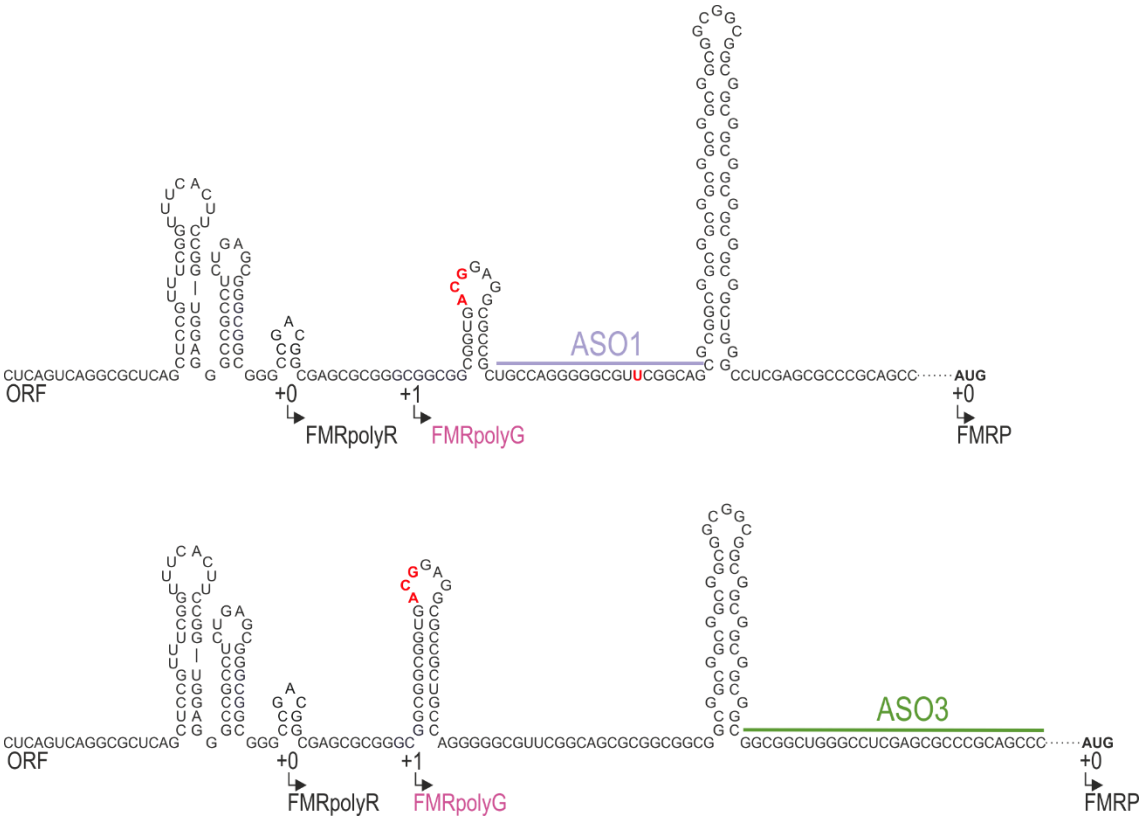


Figure 42. Proposed model of structural change of FMR1 5'UTR due to ASO1 and ASO3 binding. *The scheme presents the 5'UTR sequence with 16 CGG repeats. The mismatch of one nucleotide in the region of GUG→GUU mutation for ASO1 binding is marked in red.*

5. DISCUSSION

Expansion of unstable CGG repeats within 5'UTR of *FMR1* may lead to the development of many different fragile X-linked syndromes. Importantly, depending on the size of CGGexp, different disorders, characterized by distinct molecular pathomechanisms and completely different clinical manifestations can be distinguished. Nevertheless, the toxic *FMR1* mRNA seems to be the core of each pathomechanism involved in both fragile X-associated tremor/ataxia syndrome and fragile X syndrome.

5.1. R-LOOP FORMED OVER EXPANDED CGG REPEATS WITHIN *FMR1* 5'UTR IS DRUGGABLE TARGET FOR ANTISENSE OLIGONUCLEOTIDES IN FXTAS BUT ONLY PARTIALLY IN FXS

The formation of R-loops, among others, is a common feature of mentioned trinucleotide expansion disorders. *In vitro* studies performed in this work concerning the effect of the R-loop formation on the transcription of *FMR1* containing ~100 CGG repeats demonstrated that **R-loops formed over rCGG repeats** have a negative effect on the transcription of *FMR1* which is rescued *via* RNase H treatment. Additionally, the use of ASO-CCG, binding to CGG repeats within nascent RNA and non-template DNA strand, resulting in reduced stability of RNA:DNA hybrids, positively regulated the *FMR1* transcription. These results were also confirmed *in cellula* in fibroblasts derived from FXTAS patients. Of note, the effect was CGG repeat length-dependent.

Loomis and colleagues[126] analyzed the *FMR1* 5'UTR sequence in the context of features for R-loop formation. They found that despite CGG repeats there are other regions that fulfill the requirements for R-loop formation. Among others, the presence of G-clusters described as R-loop "seeding points" upstream CGG repeats were identified.

Results from the genome-wide mapping studies presented that R-loops are abundant at promoters of RNA Pol II-transcribed genes[129,131,144,132]. At the CpG-containing promoters, including the *FMR1* promoter, R-loop may facilitate transcription *via* the protection of underlying DNA from methylation[129,133]. This mechanism is in agreement with the fact that DNA methyltransferases (DNMT) poorly bind to RNA:DNA hybrids[133] and could explain, at least partially, the increased *FMR1* mRNA level in the PM alleles. In line with that, it has been shown that *FMR1* promoter in PM carriers showed almost two times higher acetylation of histones H3 and H4 compared to normal alleles[210]. However, R-loops may also constitute the promoters for Pol II transcribed genes by themselves[145].

Also, it has been proven that expansion of CGG repeats shifts the usage of TSS in the upstream direction[38,39] thus more G-clusters can be utilized. As was already mentioned these G-clusters constitute the initiation points for R-loops formation. In consequence, the more CGG repeats the more R-loops can be formed. Indeed, Loomis and co-workers presented that upon increased transcription rate more R-loops within *FMR1* 5'UTR were formed[126]. The authors also suggested that R-loops forming at the CGGexp may result in a more open chromatin structure and thus contribute to the increase of *FMR1* mRNA in FXTAS patients. Although it may be possible, no evidence has been presented so far.

Importantly, R-loops formed over expanded triplet repeats may be different from R-loops at the 3'-ends of genes and R-loops formed over CpG islands-containing promoters. R-loops over expanded repeats form a structural block, directly interfering with Pol II transcription elongation and influencing the transcription efficiency[125,136,140]. In accordance, it was shown that R-loops formed over expanded GAA repeats within the *FXN* gene led to decreased transcription which was rescued *via* over-expression of exogenous RNase H1[125]. Authors suggested that R-loops negatively regulate transcription by interfering with Pol II. In line with that, it was also presented in *in vitro* studies and in yeast that co-transcriptionally formed R-loops may lead to impairment of transcription elongation[172,211].

Based on the presented data I hypothesize that 5'UTR of the *FMR1* gene is a template for **two types of R-loops**, and their function may be further affected by the number of CGG repeats. R-loops formed at CpG island promoter of *FMR1* (herein ***Promoter-R-loops***) may lead to increased transcription of *FMR1*, while those formed over CGG repeats (herein ***CGG-R-loops***) have adverse effect and result in a reduction of transcription efficiency due to extremely stable RNA:DNA duplex stabilized by DNA structure formed on sense strand. Importantly, I assume that both structures play regulatory functions in normal *FMR1* alleles. However, when the CGG repeats expand the capability to resolve *CGG-R-loops* may be hardly disturbed leading to deleterious downstream effects.

According to obtained results, I assume that the observed increase in transcription efficiency upon RNase H treatment *in vitro* or transcription decrease upon RNase H1 depletion *in cellula* is either a sum of effects performed by both, however, acting contradictory, types of R-loops or that *CGG-R-loops* formed over expanded CGG repeats have a greater impact on the *FMR1* transcription than *Promoter-R-loops*. Nevertheless,

the experiments performed with ASO-CCG allowed to exclusively verify the effect of *CGG-R-loops* on the *FMR1* transcription.

In vitro transcription experiments showed that CGG trinucleotide repeats alone are able to form R-loops[212]. Experiments with ASO-CCG-Cy3 (***Figures 17 and 18***) confirmed that in a model developed in this study the *CGG-R-loops* are indeed formed. It was also presented that ssDNA non-template strand displaced during R-loop formation is involved in hairpin formation[126]. Such structure could reduce DNA:DNA duplex reannealing behind the transcription complex, and thus increase *CGG-R-loop* formation and/or stability. Additionally, since the preference to form a quadruplex by the CGG sense strand and the transcript has been confirmed[182] it is possible thus the structures will interact with each other and stabilize themselves aiding R-loop formation as I presented in the proposed model describing this phenomenon in ***Figure 19***. Importantly, this phenomenon will apply to the number of CGG repeats above a particular threshold as mentioned structures formed by short CGG repeats would not be long enough to reach each other and interact with.

Interestingly, studies published by Groh and co-workers[125] confirmed that R-loops are formed within *FMR1* 5'UTR containing both short and long CGG repeats suggesting their regulating role in the *FMR1* transcription. However, as CGG repeats expand the defects in mRNA processing can result in a *CGG-R-loop*-dependent activation of the DNA damage response as a consequence of stabilization of R-loops on expanded CGG repeats. In this regard, the activation of the double-stranded DNA-break repair pathway, but only in highly transcribed expanded CGG repeats, was reported[143].

In conclusion, R-loops formed within *FMR1* 5'UTR, containing the premutation range of CGG repeats, studied in this work are R-loops formed by CGG repeats since the ASO-CCG confirmed the specificity of the RNaseH-sensitive sequence (***Figure 18***). However, the involvement of *Promoter-R-loops* in the observed results can not be excluded. Nevertheless, the observed decrease in the *FMR1* transcription efficiency can be explained by the impairment of Pol II driven by the structural obstacle formed by stable and durable *CGG-R-loops*. In addition, the RNA:DNA hybrids are potentially stabilized *via* the secondary structures formed on the non-template DNA strand as the CGG repeats expand. Therefore, the ASO-CCG would affect the strength of this interaction and positively regulate transcription.

Of note, the positive effect on the transcription initiation performed by the *Promoter-R-loops* is still possible, however, due to the stacked *CGG-R-loops* increased transcription initiation of *FMR1* mRNA may lead to increased harmful *CGG-R-loop* formation and thus no elevation in *FMR1* level will be observed.

The structural complexities within the promoter region of *FMR1* may be involved in the transition of *FMR1* from an active state in PM to a silenced state in FM. As it was mentioned stable R-loops, which are not transient, may lead to DNA damage which has been associated with aberrant DNA methylation[213]. Thus, R-loops formed within *FMR1* FM alleles, that are longer and more stable, may account for the different effects on transcriptional rate and protein expression than R-loops formed at PM alleles. Although DNA methyltransferase 1 (DNMT1) was mentioned to poorly bind to RNA:DNA hybrids it has been proven that this DNMT recognizes structured DNA as a substrate for methylation. Therefore, the hairpin structures formed by FM expanded CGG and CCG repeats within nascent RNA and DNA strands may recruit DNMT1 and leads to gene silencing. Of note, the DNA hairpin formation within a non-template strand of *CGG-R-loop* in *FMR1* has been shown[126]. On the other hand, it was suggested that *CGG-R-loops* which are not efficiently resolved can disrupt the protective function of *Promoter-R-loops* against methylation and drive the silencing of *FMR1*[126]. A similar mechanism has been already confirmed[129]. As a result, *CGG-R-loop* may further promote the loss of active chromatin marks within the *FMR1* promoter leading to the *FMR1* transcription silencing observed in FXS cells.

The involvement of R-loops formed over expanded CGG repeats in *FMR1* methylation in FXS has been also confirmed by others[125,146]. Recently, it has been proven that R-loops trigger the silencing of *FMR1* and are formed before the heterochromatin marks appear[125]. Following the suggestion that R-loops may recruit chromatin repressive marks to the *FMR1* promoter the enrichment of H3K9me2 at R-loop containing pause region of β-actin gene has been reported[173]. Groh and co-workers[125], to test the role of R-loops in the *FMR1* methylation performed transcription reactivation by 5-azadC which resulted in the expression of *FMR1* mRNA in FXS cells at the level of 25% of control cells. The authors suggested that if inhibition of DNA methylation results in only partial *FMR1* reactivation thus formed R-loops may prevent the full reactivation. Although it is feasible, the possibility that transcription impairment through the long CGG repeats occurs should also be addressed. Nevertheless, the authors showed that an increase in R-loop signal from

expanded GAA repeats is correlated with the increase in repressive chromatin marks and subsequent repression of the *FXN* gene in Friedreich Ataxia patients.

In this work, two FXS primary cells, FX08-01 and FX13-01, derived from males, characterized by partial and full silencing of *FMR1* transcription have been used. The obtained results presented that transcription of *FMR1* activation by 5-azadC treatment was possible in both FXS cells. However, return to the silenced state has been started 7 – 16 days after 5-azadC removal from the media suggesting that co-transcriptionally formed *CGG-R-loops* recruit again heterochromatin marks to the *FMR1* promoter. Although a significant increase in the *FMR1* mRNA level was observed, no FMRP protein was produced.

Till now different approaches targeting distinct events in the *FMR1* silencing pathway have been tested to reactivate transcription of silenced *FMR1* in cellular FXS models. Kumari and colleagues tested different histone methyl-transferase (HMT) inhibitors of H3K9 methylation in patient-derived cells[164]. They found that, among others, chaetocin was able to reactivate the *FMR1* transcription however the effect was quite small. This fungal toxin specifically inhibits the enzymatic activities of various histone methyltransferases[214]. The effectiveness of chaetocin was enlarged when cells were pretreated with 5-azadC suggesting that prior demethylation of DNA is essential for optimal transcription reactivation. Interestingly, this small molecule was also effective at delaying the re-methylation of *FMR1* after 5-azadC treatment of FXS cells. However, the mechanism behind these delays has not been identified. In another work, Kumari[193] tested the inhibitors of EZH2 (Enhancer Of Zeste 2 Polycomb Repressive Complex 2 Subunit) which is a H3K27 trimethyltransferase. The 1a small compound was previously tested by Colak[146] and was suggested to inhibit R-loops formation within *FMR1* 5'UTR. Additionally, the 1a molecule was presented to successfully bind with the RNA secondary structure formed by the CGG repeats[215] through the binding to the repeating G-G internal loops in the rCGG hairpin[146]. Results obtained by Kumari, although confirmed the effectiveness of *FMR1* transcription reactivation by 1a compound in cells pretreated with 5-azadC and the delayed silencing after 5-azadC withdrawal, however, the authors undermined the hypothesis about preventing R-loop formation by 1a molecule. Of note, in both studies, the authors did not present the FMRP levels.

Since the application of demethylating agents was sufficient to reactivate *FMR1* transcription only transiently and more importantly usually these agents were toxic and presented significant off-target activity the more accurate method using the Clustered Regularly Interspaced Palindromic Repeats-CRISPR associated protein 9 (CRISPR-Cas9) system have been utilized[216]. The deletion of expanded CGG repeats in FXS patient-derived cells has been utilized by two separate groups[161,217]. In both studies, although the experimental design was different, efficient *FMR1* transcription followed by FMRP synthesis was achieved. Interestingly, a modified CRISPR-Cas9 system has been also used for targeted modulation of gene expression. Therefore, the activation of the *FMR1* gene using a dCas9 (endonuclease-deficient; deadCas9) fused to multiple domains of the VP16 transcription activator has been reported[192]. The authors presented that upon the usage of guide RNAs targeting DNA region containing ~ **800 CGG** repeats the significant activation of *FMR1* transcription in FXS hESCs and patient-derived Neural Progenitor Cells (NPCs) was achieved. However, contrary to human cell lines with normal repeat sizes, a significant increase in FMRP production in FXS hESC cells was not observed. A similar system using the DNA methylation modification enzyme Tet was utilized to demethylate the CGG repeats in the pathological *FMR1* locus[162]. The authors presented that demethylation of allele containing ~ **450 CGG** repeats led to the hypomethylation of the CGG locus, increased acetylation of H3K27 and trimethylation of H3K4, and simultaneously reduced trimethylation of H3K9 at the *FMR1* promoter. The synthesis of FMRP in FXS iPS cells was restored to ~ 70% of the level in control cells. Interestingly, in the study performed by Lee and co-workers[160] the FMRP restoration was induced by CpG demethylation and R-loop formation executed by dCas9 with single guide RNA targeting CGG repeats. The experimental design was based on the fact that aberrant R-loops trigger DNA damage signals, and are then resolved by the MSH2/MMR repair pathway. Resolution of the R-loops, as expected, led to the CGG repeat contraction, followed by *FMR1* reactivation, and restoration of FMRP translation. Finally, the positive feedback loop illustrating obtained data was proposed. The authors suggested that *de novo* formation of R-loops will open the chromatin and promote *FMR1* transcription which in turn will further enhance the R-loop formation. The number of CGG repeats in the tested lines varied between **300** and **600** and the rescue of FMRP translation constituted ~ 50% of FMRP level in the control cells.

Since R-loops are the triggers for *FMR1* methylation in FXS I wanted to test whether targeting R-loops by ASO-CCG, which as I know from performed in this work *in vitro* studies invade *CGG-R-loop* structures, would inhibit methylation and lead to the *FMR1* transcription reactivation. The lack of *FMR1* transcription after 14 days of ASO-CCG treatment (***Figure 24***) may result from the fact that ASO-CCG can not invade such long and stable *CGG-R-loops* which are formed within FM alleles. Alternatively, it is feasible that due to the total loss of transcription within the *FMR1* locus and the heterochromatization of the DNA, no R-loops can be formed. On the other hand, studies performed by Colak and colleagues[146] showed that small compound 1a that selectively binds to the repeating G-G internal loops in the rCGG hairpin can only prevent the silencing of *FMR1* in FXS hESCs since the application of 1a to already silenced *FMR1* promoter did not reverse the silencing. The same mechanism might stand behind the utilization of ASO-CCG on already methylated *FMR1* locus, thus the inhibition of *FMR1* methylation *via* 5-azadC followed by ASO-CCG treatment could answer whether ASO can delay/inhibit the re-methylation of *FMR1* promoter and constitute the potential therapeutic approach for FXS. The preliminary experiment regarding this aspect has been already performed. As presented in ***Figure 43*** there is a trend concerning the increased level of *FMR1* after ASO-CCG treatment in comparison to ASO-Ctrl treated FXS cells, however, due to the low number of biological samples the difference is statistically significant only in one time point, No. 4th. Additionally, the effect of ASO dilution due to cell divisions together with a wider tested time window should be addressed in future experiments.

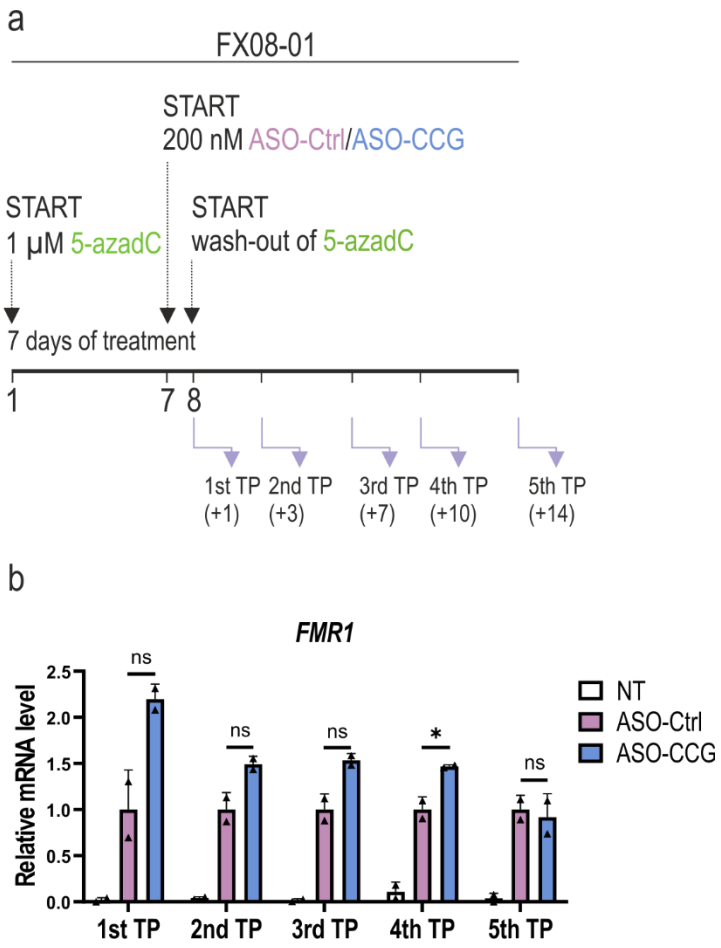


Figure 43. FMR1 mRNA level after ASO-CCG treatment of FXS-patient derived fibroblasts with FMR1 transcription reactivated by 5-azadC. a) *Scheme of experiment taking into account the 5-azadC treatment and ASO-Ctrl/ASO-CCG transfections in FX08-01 cells. The cells were cultured for 7 days in the medium supplemented with 1 µM 5-azadC followed by 7 days in the clear medium (wash-out), however, fibroblasts were treated with ASO on the 7th day of culture and were cultured for another 7 days;* ***b)*** *RT-PCR quantification of FMR1 mRNA level in FX08-01 fibroblasts. Graphs present results for N = 2 biologically independent samples with SDs. Statistical analysis was based on a two-tailed unpaired Student's t-test; *, p<0.05; ns, non-significant. TP – time point.*

Based on the results from *FMR1* reactivation in FXS cells performed in this project and by other groups[164,193,161,217, 192,162,160] I hypothesize that the FMRP translation after reactivated *FMR1* transcription depends, at least partially, on the number of CGG repeats within *FMR1* 5'UTR. Both fibroblast cells, FX08-01 and FX13-01, used in this study were characterized to possess more than 435 CGGs[169], however, the exact number of repeats is unknown, therefore even if the transcription is reactivated the transcript may retain in the nucleus due to impaired nucleocytoplasmic transport. It would be in line with

the observation of nuclear retention of *DMPK* transcripts responsible for myotonic dystrophy type 1 pathomechanism[218]. Indeed, the nuclear retention of FM transcripts and their involvement in the R-loop formation has been reported[193] and this mechanism can contribute to the limited FMRP expression observed in cells with reactivated transcription with larger FM alleles. Alternatively, if *FMR1* mRNA is transported to the cytoplasm, even if only in part, the extremely stable secondary structure formed by CGGs may inhibit the translation from the downstream AUG start codon for FMRP production due to disturbed 43S ribosome scanning. As a consequence of both proposed processes, the FMRP translation can be inhibited, thus the tool leading to increased FMRP synthesis should be taken into account in further studies. However, the toxicity driven by the rCGGexp can not be omitted, hence approaches resulting in CGG repeat contraction seem to be better therapeutic solutions.

The results presented in ***Figure 25a*** support my suggestion that the lack of FMRP translation in FXS cells, even when the *FMR1* mRNA is present, may result from the nuclear retention of the transcript. The treatment with ASO-CCG significantly increased the level of available *FMR1* mRNA in the cytoplasm (***Figure 25b***), however, still no FMRP was synthesized (***Figure 25c***). Next to the impairment of translation, the possibility that the level of *FMR1* mRNA is too low to produce the FMRP above the threshold of detection *via* western blot should also be addressed in further studies.

Since our group has already shown that an increase in the level of *FMR1* pre-mRNA and nuclear mRNA in FXTAS is relatively low in ASO-CCG-treated cells[177] I assumed that the significant increase of *FMR1* mRNA in the cytoplasm (***Figure 25b***) is caused mainly by the elevated stability of mRNA, perhaps due to reduced efficiency of translation of both FMRpolyG and FMRP. A similar effect was previously described for many other genes[219,220].

Taken together, R-loops formed within *FMR1* 5'UTR are important regulatory structures that control the transcription of *FMR1* in normal alleles. However, as the CGG repeats expand, the more stable *CGG-R-loops* are formed which may have different deleterious effects on the protective *Promoter-R-loops*, *FMR1* transcription, and chromatin state within *FMR1* locus. The ability of *CGG-R-loops* to trigger the *FMR1* silencing makes them an attractive target for putative therapeutic approaches. However, understanding how cells distinguish the regulatory/useful R-loops from harmful/toxic R-loops is an

extremely important question that needs to be answered before effective treatment utilization.

Summing up, my research showed for the first time that:

1. Antisense oligonucleotides composed of short CCG repeats (ASO-CCG) can successfully invade *CGG-R-loop* structures formed *in vitro* in the 5'-leader sequence of the *FMR1* and lead to the increase of the *FMR1* transcription in the PM range of CGG repeats;
2. Treatment of FXS cells containing partially active *FMR1* with ASO-CCG induced an increase of *FMR1* mRNA level and elevated the pool of this mRNA in the cytoplasm, which could be considered as potential therapeutic strategy.

5.2 PRIMARY AND SECONDARY STRUCTURES OF 5'UTR OF *FMR1* mRNA ARE SIGNIFICANT FACTORS IN THE REGULATION OF FMRpolyG SYNTHESIS

Expanded CGG repeats in the PM range present in *FMR1* 5'UTR trigger initiation of RAN translation through an AUG-independent mechanism and production of toxic FMRpolyG protein and other mono-amino acids tract-containing proteins. Non-AUG initiation enables multiple translation initiations from the same mRNA. According to *FMR1* RAN translation of transcript in sense direction, three possible RAN proteins, in three reading frames, can be synthesized. However, the differences in the efficiency of translation, half-life, and accumulation rate result in distinct levels of detected RAN proteins. Thus, the most common product of RAN translation from mutant *FMR1* is FMRpolyG followed by FMRpolyA and FMRpolyR detected at ~37% and ~2.5% relative to FMRpolyG, respectively[69,67,70]. Interestingly, it has been suggested that the major differences in translation efficiency of these RAN proteins result not from initiation but from elongation because the mutations of near-cognate start codons for all free reading frames to AUG did not change significantly the ratio between those RAN products[201]. However, the effect of surrounding sequences on the efficiency of particular codon utilization seems to be omitted by the authors of previously published studies. The observed results may directly arise from the fact that AUG start codons are less likely to be sensitive to the sequence context than the near-cognate start codons. Therefore, no change in the distribution of RAN proteins after near-cognate start codons mutation may result not from the equal efficiency of translation initiation but from the fact that AUG

codons with strong Kozak context are not as vulnerable for sequence context regulation as non-AUG codons are. According to that possibility, differences in the RAN translational initiation at different reading frames modulated by the surrounding sequence were presented for expanded CAG repeats in Spinocerebellar Ataxia Type 8 (SCA 8)[221]. Nonetheless, the impairment of translation elongation due to, for instance, a lower abundance of the tRNAs decoding the particular codons is more than likely to occur.

Emerging studies revealed that non-AUG translation of FMRpolyG may initiate at different near-cognate start codons, mainly at ACG and GUG located 32 nt and 8 nt upstream of the CGG repeats, respectively, however, the ACG near-cognate start codon has been stated to be the most widely utilized[70,69,67]. Indeed, results presented in this work confirmed that mutation of the second putative FMRpolyG near-cognate start codon – GUG (8 nt upstream CGGs) – did not significantly affect the level of this RAN product (***Figure 27c-d***). Interestingly, Todd and co-workers showed that a single mutation of any of the potential alternative start codons of FMRpolyG within *FMR1* 5'UTR did not eliminate the synthesis of FMRpolyG[69]. This would suggest that various near-cognate codons upstream to the CGG repeats can be utilized, therefore the mutation of the single one is not sufficient to inhibit/weaken translation, probably due to the compensation effect. In line with that another study analyzing the CGGexp-mediated impairment in FMRP translation demonstrated that blocking near-cognate start codons for RAN proteins with ASOs increased FMRP levels however this effect was not visible when single RAN initiation sites were blocked[80].

Furthermore, the results presented in ***Figure 9*** seem to confirm this phenomenon. Although the results published by Kearse and colleagues of experiments with the use of similar constructs based on the nLuc reporter suggested to not producing additional products in +1 frame, the results presented by Wright and co-workers[201] referring to and using the same plasmids stated that the FMRpolyG produced from constructs "exhibits a multi-band pattern". Therefore, I hypothesize that RAN translation of mutant *FMR1* can initiate at various near-cognate start codons within 5'UTR and the utilization of particular near-cognate start codon depends on factors such as genetic background or cellular conditions. Hence, the results from the transient transfection with the same plasmids but at different cells may differ significantly as distinct near-cognate start codons can be used. Given the foregoing, changing the sequence within *FMR1* 5'UTR by introducing particular mutations can also affect the start codon utilization.

Despite the mentioned ACG and GUG near-cognate start codons the CUG triplet located 20 nt upstream CGGs showed the enhanced read density at data from the study utilizing the Herringtonine (translational inhibitor) to stall ribosomes at the initiation sites[222]. Interestingly, mutation of this CUG codon was presented in another study to increase the initiation of FMRpolyG translation[70] (*Figure S4B*). Of note, this CUG near-cognate start codon was predicted also in the analysis performed in this work as a potential TIS for FMRpolyG (***Figure 9***). Furthermore, it has been also suggested that FMRpolyG can be translated directly from the CGG repeats[201,223,70] since the UAG stop codon introduced just 3 codons upstream CGGs did not inhibit entirely the FMRpolyG synthesis[70]. On the contrary, no translation of FMRpolyG was observed by another group when the entire region upstream CGGs was deleted[67]. Hence, the seemingly initiated FMRpolyG translation at CGG repeats could result either from the frameshift from other reading frames not inhibited by the stop codon in frame with FMRpolyG or from the fact that the native *FMR1* 5'UTR sequence is an important regulator of FMRpolyG translation initiation even when it occurs directly at CGG repeats. Taken together, due to the highly GC-rich character of the *FMR1* 5'UTR sequence many near-cognate start codons may be used for initiation of translation as a consequence of the slowdown of 43S PIC and the ratio of codon utilization between different cells and conditions can be highly variable. Taking into account what was mentioned above I would like to highlight that those are the reasons why I did not want to mutate all putative near-cognate start codons present within *FMR1* 5'UTR. To study the mechanism and dependence of RAN translation initiation I believe that the sequence of *FMR1* 5'UTR should be as close to the native one as possible. In line with that, developing in the future the constructs with the native *FMR1* 5'UTR under the control of the *FMR1* promoter could be very informative.

FMRP-Nluc-FLAG plasmid encoding FMRP equivalent is also a donor of FMRpolyR protein which is translated from ACG (+0) codon located 57 nt upstream CGG repeats (***Figure 26b***). Although the initiation of FMRpolyR translation was confirmed to initiate at this ACG near-cognate start codon the studies performed by Todd and colleagues using the GFP fusion reporter system showed that no protein in +0 frame was generated[69]. Furthermore, such FMRP protein with N-terminal extension was not detected in the samples from FXTAS patients suggesting that this RAN product is translated under the threshold of detection or it is quickly degraded. Additionally, no FMRpolyR protein was detected by western blot in HeLa cells or at *in vitro* translation at normal or expanded

repeat lengths[70]. On the contrary, the protein was synthesized and detected by western blot when the *FMR1* 5'UTR did not contain CGG repeats. Hence, the authors suggested that the length of CGG repeats may have a negative effect on the FMRpolyR translation. In another study conducted by Rodriguez and colleagues, it has been shown that a reporter system based on the nLuc allows for the FMRpolyR detection by western blot if the protein contains up to 18 CGGs[80]. The results presented in this dissertation demonstrate that FMRpolyR can be readily detected by western blot when *FMR1* 5'UTR contains 16 CGG repeats (***Figure 29***), however, the threshold of FMRpolyR detection can be even lower and require to be established. The discrepancy between these results and data presented by other groups probably arises from the diverse sensitivity of the used reporters and differences in the *FMR1* 5'UTR sequence as most of the groups introduced many restriction sites during cloning. Nevertheless, the reporter system developed for this dissertation constitutes an elegant tool for studying FMRpolyR translation.

The mutation of FMRpolyR near-cognate start codon (ACG (+0)) was introduced (***Figure 28***) to confirm that the observed protein is indeed the RAN product in this reading frame. The significant decrease in the FMRpolyG level, detected by NanoLuc assay (***Figure 29a, bottom panel***), suggested that RAN translation in the polyR-frame contributes to translation in the polyG-frame. Indeed, it has been proven that on *FMR1* CGG repeats the translational frameshift from polyR-frame to polyG-frame (R-to-G) occurs[201]. Interestingly, the frameshift was detected in all three reading frames of RAN translation *in vitro* - 0, +1, and +2, and it has been proven that the mechanism of frameshift is dependent on the secondary structure formed downstream to TIS by the CGG repeats that might slowdown ribosomal elongation[201]. In line with that, it has been presented that frameshift in +1 reading frame is usually increased by the elevated ribosome pause time, amino acid charge, and the availability of the amino-acyl-tRNAs[224,225]. Importantly, the ribosome pausing can be further affected by the lower abundance of the tRNAs decoding the arginine codons (relative to the tRNAs decoding glycine codons) thus the frameshift from R-to-G can be favored *in vivo*[224,226]. Of note, *in vivo,* only R-to-G frameshift was found despite the detection of all frameshift events in *in vitro* conditions suggesting that the +1 frameshift is favored within the studied HEK-293 cells[201].

To sum up, obtained results confirmed that in the nLuc-based reporter system with 16 CGGs, the FMRpolyR protein can be analyzed and that the RAN translation of this protein has a direct effect on the FMRpolyG level, probably due to the polyR-to-polyG

frameshift. Simultaneously, the FMRpolyR protein is not detected when *FMR1* 5'UTR contains 85 CGG repeats (***Figure 29b***) which may result from the: *(i)* suggested negative effect of the CGG repeats on the FMRpolyR translation, *(ii)* arise directly from the decreased efficiency of translation of the entire *FMR1* due to expanded CGG repeats, or *(iii)* be the consequence of R-to-G frameshift that could be much more abundant on longer CGG repeats thus all/almost all FMRpolyR protein would be involved in frameshift process. This phenomenon may also explain why the FMRpolyR protein is not detected in samples from FXTAS patients.

The length of CGG repeats within *FMR1* 5'UTR seems to be a crucial factor regulating RAN translation. Intriguingly, the GC-rich character and highly structured *FMR1* 5'UTR with expanded CGG repeats should theoretically inhibit the translation since the formation of a strong secondary structure within the 5'UTR was reported to inhibit ribosomal scanning[107,227,228]. The hairpin formed by 16 CGG repeats in developed constructs has estimated Gibbs free energy of ΔG=−54 kcal/mol which according to the literature data is sufficient to greatly inhibit translation initiation by posing a substantial energetic barrier for scanning 40S ribosome[227,229]. The most frequent size of the normal allele (30 CGGs) has estimated Gibbs free energy of ΔG=−50 to −60 kcal/mol (depending on the number of interruptions)[35]. Paradoxically, FXTAS patients present only slightly reduced levels of FMRP, although the increase in the number of CGG repeats was correlated with the gradual impairment of *FMR1* translation efficiency[34,30,35,36]. On the other hand, the level of mutant *FMR1* mRNA in the premutation range has been reported to correlate positively with the expanded CGG repeats[35,33,30,202] suggesting that an increased transcription rate of *FMR1* may constitute a compensatory effect to lowered FMRP level. Surprisingly, in reporter systems, the equimolar RAN reporter mRNAs with expanded repeats have been shown to be translated at higher levels than those with normal lengths of CGG repeats[70]. These observations suggest that other mechanisms are involved to circumvent the stability of the CGG hairpin and to allow for *FMR1* translation through the repeats. Interestingly, the translation of the CGG hairpin sequence seems to be regulated also by the sequence in the vicinity of the structure[230]. Therefore, other structures within the *FMR1* 5'UTR can modulate the stability of secondary structure and/or accessibility of the CGG hairpin sequence, or alternatively, other *trans*-factors which can modulate the hairpin structure are recruited to allow for translation through CGG repeats.

The comparison of constructs with 16 CGGs and 85 CGGs in this work demonstrated that the longer CGG repeats strongly affected the translation of both FMRpolyG as well as FMRP equivalent (***Figure 31***). Surprisingly, although all constructs possessed the same CMV promoter uneven levels of *FMR1* mRNA from constructs with short and long repeats were detected. This suggests that the observed reduction of FMRpolyG and FMRP equivalent levels results from alterations in both transcription and translation impairment. Of note, it has been reported that CGG repeats within *FMR1* 5'UTR, however also in the context of heterologous promoters, affect the transcription of reporter gene[231].

Transient transfections of cells with plasmids containing *FMR1* 5'UTR with 99 CGG repeats led to the ~50% reduction in translation efficiency of reporter mRNA when compared to the constructs with 30 CGGs[101,35]. The plasmids containing 16 CGGs within the *FMR1* 5'UTR did not significantly change the translation efficiency[104], however, conflict results were presented in another study where 16 CGG repeats reduced the translation efficiency by ~20%[230]. Chen and colleagues showed that CGG repeats reduce the translation of reporter firefly mRNA only when *FMR1* 5'UTR contained more than 30 CGG repeats[35]. The scope of inhibition was directly correlated with the increasing number of CGG repeats. Surprisingly, they noticed a nearly 2-fold increase in the reporter translation for CGG repeats in the range between 0 and 30 (including construct with 16 CGGs). Of note, one of the used cell lines in the reported study was HEK-293 which was also used in my experiments. Similar to the experiments performed in this work, the level of reporter mRNA was not aligned before translation thus observed differences in the level of firefly reporter appear to arise from both transcription and translation changes. However, contrary to my studies, the RT-PCR analysis revealed that the message levels of reporter luciferase were shown to increase as CGG repeats expanded[35].

Khateb and co-workers presented the suppressed translation of the reporter firefly luciferase by the premutation CGG tracts[101]. The authors showed that *in vitro* transcription of *FMR1* 5'UTR driven by the T7 promoter was only slightly affected by the length of CGG repeats. On the contrary, the translation was highly affected by the expansion of CGG repeats. The transcription of *FMR1* 5'UTR with 99 CGGs, relative to sequence without CGG repeats, was decreased by 33%. Translation of equimolar *FMR1* mRNA with 99 CGG repeats was lowered by 72% in comparison to the mRNA with no CGGs.

In another study, it was presented that the FMRpolyG-GFP level was decreasing as the CGG repeats were increasing[69]. Primerano and co-workers provided direct evidence that premutation *FMR1* alleles possess an inhibitory role in the *FMR1* translation by analysis of the polysome profiles in the lymphoblastoid cell lines derived from PM carriers (97–195 CGG repeats)[30]. They observed that an increase in the *FMR1* mRNA and a decrease in the FMRP level were correlated with the increase in the number of CGG repeats. In line with that, studies performed by Ludwig and colleagues showed that CGG repeats inhibit *FMR1* translation initiation in a length-dependent manner[230].

The level of FMRP is unchanged or only slightly reduced in FXTAS patients' cells. Till now, few mechanisms have been proposed to explain this phenomenon. One hypothesis assumed that scanning through the CGG repeats within *FMR1* mRNA requires RNA unwinding by RNA helicases and when they are recruited the scanning of the ribosome followed by initiation at the canonical AUG start codon for FMRP is possible[69]. The same authors suggested that ribosomes translating FMRpolyG through the CGG repeats can terminate translation behind the hairpin structure and re-initiate the synthesis at the AUG start codon. Also, some groups agreed that the observed increase in the *FMR1* mRNA in PM carriers is a part of the feedback loop that is activated to maintain the FMRP level at a physiological level.

Although in the nLuc-based reporter system developed in this project, similarly to other studies, a reduction in translation efficiency of sequence with longer CGG repeats is observed at that point it is not possible to define whether there is a linear correlation between the length of CGG repeats and the level of studied proteins. Thus, the constructs with a various range of CGG repeats will be developed, as well as a construct with no CGG repeats, to characterize the system more thoroughly. Nevertheless, a reduction in the level of FMRP equivalent observed in the developed system may result from the impairment of the translation of the entire *FMR1*. In accordance, it has been reported that even 30 CGG repeats affect the efficiency of FMRP translation in reporter systems[35,230]. Importantly, the techniques utilized in this study (western blot and NanoLuc assay) provide information only on the level of translated protein however the ratio between nuclear and cytoplasmic fraction of *FMR1* mRNA is omitted in the analysis. It is feasible that next to the transcription impairment of long CGG repeats the reduced level of FMRP equivalent protein arise also due to partial retention of *FMR1* transcript with 85 CGG repeats within the nucleus (***Figure 25a***). Therefore, a smaller amount of mRNA would be

accessible for translation in the cytoplasm. Taking this possibility into account, the nucleocytoplasmic fractionation of HEK-293 cells after transient transfection with appropriate constructs will be performed to verify this issue.

Surprisingly, the comparison of FMRpolyG level produced from the construct with 16 and 85 CGG repeats, contrary to expectations, showed decreased level of RAN products from longer CGG repeats. Interestingly, since the changes in the FMRpolyG detection have been reported this observation may constitute a false-negative result. Sellier and co-workers demonstrated that depending on the fusion partner the threshold of FMRpolyG detection varies[67]. Thus, when FMRpolyG was fused to a small FLAG tag the RAN product was detected only with expanded CGG repeats (over 60 to 70 CGGs). On the contrary, fusion to bigger tags such as GFP (~25 kDa) allowed for FMRpolyG detection with short stretches of CGG repeats (30 CGGs) or even without any CGG repeats. This observation is in line with the fact that usually small peptides, translated from uORFs, are hardly detectable and fusion with large tags leads to increased cellular stability and easier detection using western blot[232]. In consequence, the difference in the FMRpolyG detection may result from the stabilizing effect of the large tag (herein GFP) on the FMRpolyG thus the differences in the level of RAN translation are blurred. It may also provide an explanation why the increase in the FMRpolyG is observed as CGG repeats expands in FXTAS patients, where protein is not tagged, and why this correlation is lost in reporter systems. Furthermore, these results highlight two issues: *(i)* the translation initiation of FMRpolyG can occur at near-cognate start codons independently of the size of CGG repeats, and *(ii)* depending on the size of the fused tag the threshold of FMRpolyG detection differs significantly. Importantly, although the authors did not put attention to this issue in the discussion there was also a difference in the correlation between FMRpolyG level and the number of CGG repeats depending on the fusion partner. In other words, when FMRpolyG-FLAG was utilized the positive correlation between the level of FMRpolyG and CGG size was observed, however, when the FMRpolyG was fused to GFP no correlation was observed. The level of RAN product was unchanged or even reduced. Noteworthy, the fusion of FMRpolyG with Nluc and FLAG tag in my constructs may result in a loss of observed positive correlation between the number of CGG repeats and the level of FMRpolyG. On the other hand, Kearse and colleagues demonstrated conflict results that FMRpolyG-Nluc-FLAG reporter mRNA with expanded CGG repeats, in equimolar mRNA transfected HeLa cells, were translated

at higher levels than reporters with a normal length of CGGs. However, the authors also showed that **the synthesis of FMRpolyG appeared to be repeat-independent in *in vitro* conditions**[70]. These observations suggest that depending on the molecular background and design of the experiment the relation between the number of CGG repeats and the translation efficiency of FMRpolyG can be misinterpreted. Thus, to verify this issue and confirm the tag-dependent effect on the FMRpolyG level measured by western blot the set of plasmids containing different numbers of CGG repeats embedded in the native *FMR1* 5'UTR sequence fused to FLAG tag (already available in our laboratory) will be used for transient transfection of HEK-293 and HeLa cells. Simultaneously, the same experiments as presented in this dissertation (***Figure 31***) will be conducted in HeLa cells to verify the effect of genetic background and therefore accessible modulators on the FMRpolyG synthesis.

As I presented above the regulation of *FMR1* RAN translation is an extremely complex, multilayer, and still very elusive mechanism. Hence, understanding how RAN translation works from the mechanistic point of view is essential if the therapeutic approaches targeting RAN translation are to develop.

Experiments concerning the strength of ACG (+1) Kozak sequence context on the initiation of FMRpolyG translation demonstrated that, as expected, the near-cognate start codon is highly sensitive to changes in the surrounding sequence. Studies based on the reporter systems revealed that nucleotides at positions +5 and +6 are also important for efficient non-AUG initiation[233,234]. Recently, FACS-seq-based studies revealed that optimal sequence context enables the initiation at non-AUG start codon with an efficiency comparable to the AUG codon[203]. Results presented in this dissertation demonstrated that the nucleotide context in the vicinity of ACG (+1) near-cognate start codon of FMRpolyG is crucial for the RAN translation initiation. Mutants designed to weaken the Kozak context sequence, namely Kozak1, Kozak2, and Kozak3, confirmed that positions –3 and +4 are essential for efficient translation initiation at ACG (+1) near-cognate start codon within *FMR1* 5'UTR with short CGG repeats (***Figure 32a-d***). These results are coherent with high-throughput analysis of TIS motifs utilizing non-AUG start codons which presented that ACG codon is critically dependent on the guanine nucleotide at +4 position[203]. Although, +4G is also known to be optimal and well conserved for AUG initiation, in contrary to AUG the adenine at +4 position is not tolerated and leads to a strong decrease in ACG near-cognate start codon utilization which is presented by

Kozak2 mutant. Furthermore, the position −3 is crucial for both AUG and non-AUG-dependent initiation, however, as already mentioned, the non-AUG codons are more sensitive to changes in the surrounding sequence. Hence, the Kozak1 mutant resulted in a strong decrease in FMRpolyG level. As expected, the additive effect of both mutations, demonstrated by the Kozak3 mutant, was observed. Since performed mutations strongly affected the efficiency of FMRpolyG translation one could expect that the level of AUG-initiated translation measured by FMRP equivalent would increase. However, as presented in ***Figure 31c***, the ratio between FMRpolyG and FMRP equivalent produced from the same mRNA suggests that even total loss of FMRpolyG synthesis would rather not increase reasonably the level of FMRP equivalent. Interestingly, the Kozak3 mutant with 85 CGG repeats (***Figure 32e***) did not result in the increased level of RAN product, in comparison to construct with 16 CGGs (***Figure 32c-d***), suggesting that more stable secondary structure downstream near-cognate start codon, although, embedded in very poor Kozak sequence context, can not increase the utilization of studied codon.

According to expectations, mutants of the Kozak sequence designed to make the sequence context stronger (Kozak4, Kozak4b, and Kozak5) resulted in a moderate increase in the FMRpolyG level. Mutation of guanine at –4 position to cytosine (–4G→C) in Kozak4 mutant led to an increase in the efficiency of initiation of FMRpolyG translation (***Figure 33c-d***) since cytosine at this position constitutes the most optimal nucleotide for ACG codon[203]. Kozak4b (–4G→A), however, did not result in an unequivocal increase in the FMRpolyG level as adenine at the –4 position is less optimal than cytosine. Although the effect of cytosine at position –2 is rather negligible, together with the mutation at position –4 resulted in the additive increase in the efficiency of FMRpolyG synthesis. Interestingly, the foldchange of increase in FMRpolyG level from Kozak5 construct containing 85 CGG repeats (***Figure 33e***) was higher than this observed for construct with 16 CGG repeats (***Figure 33c-d***) which may suggest that stable secondary structure formed by expanded CGG repeats support the utilization of ACG near-cognate start codon when it is embedded in the optimal sequence context. Such an outcome can be achieved by a few mechanisms including ribosome queuing and/or elongated dwell time of the scanning ribosome.

As different near-cognate codons within *FMR1* 5'UTR can be utilized as initiators of FMRpolyG synthesis the correlation between the particular codon and the efficiency of

RAN translation initiation was established. Mutation of ACG (+1) near-cognate start codon to CUG (ACG→CUG) presented that the CUG triplet was the strongest codon and resulted in the highest level of FMRpolyG (***Figure 34c-d***). These observations are coherent with data from other studies demonstrating that indeed the CUG codon is the strongest near-cognate codon[203]. Mutation of ACG→AAA led to almost complete loss of FMRpolyG synthesis, however, similarly to results for Kozak3, no change in the level of FMRP equivalent was observed since RAN translation of FMRpolyG with 16 CGG repeats in my system constitute roughly ~2% of the AUG-initiated canonical translation of FMRP equivalent (***Figure 31c***). The basal level of signal detected by NanoLuc assay for ACG→AAA mutant may suggest that either other proteins, in-frame with Nluc, were synthesized, as presented in ***Figure 9*** and ***Figure 30***, or that another, however, weak, near-cognate start codon has been activated thus FMRpolyG was still translated. Interestingly, it has been presented that the GUG codon directly adjacent to ACG (+1) codon (***see Figure 26d***) is an active site of FMRpolyG translation[223], however, in the case of ACG→AAA mutation, the sequence context of the GUG codon would be weakened due to the adenine at +4 position and could result in such low level of FMRpolyG protein. On the other hand, another option would assume that the observed signal came from the frameshift of polyR to polyG reading frame. On the contrary, the mutation of ACG→AUG codon, according to the first-AUG rule[98,235,236], resulted in an extremely strong increase in the level of FMRpolyG protein and abolished translation of FMRP equivalent which was also observed in similar studies conducted by other groups[69,67]. As presented in ***Figure 35***, the loss of translation of FMRP equivalent results from the fact that uORF of FMRpolyG is translated at efficiently utilized AUG codon. Hence, no ribosome will initiate translation at the downstream AUG start codon of FMRP if the start codon for RAN translation is efficiently utilized.

Data obtained from experiments with mutations of the Kozak context sequence demonstrated that the sequence surrounding the near-cognate start codon is crucial for the efficient initiation of RAN translation. Also, depending on the strength of the sequence context it was presented that secondary structure downstream TIS may have diverse influence on the efficiency of codon utilization. In light of these observations, the results from rACG1 and rACG2 mutants turned out to be extremely interesting and surprising. Although, both studied near-cognate start codons were embedded in similar contexts of Kozak sequences (rACG1 GCGGACGG; rACG2 GGCGACGG) the strength of the

context of the rACG1 mutant was weaker due to the cytosine at −3 position (***Figure 36b***). In comparison, the rACG2 mutant had the most optimal, guanine at this position. Nevertheless, despite optimal Kozak sequence context, the efficiency of translation initiation at ACG near-cognate start codon from rACG2 was close to zero while FMRpolyG translated from studied codon from rACG1 resulted in ~2-fold increase in the level of RAN product (***Figure 36c-d***). As other factors, than sequential dependencies, had to modulate the initiation at those near-cognate start codons I took a closer look at the predicted secondary structures formed in the vicinity of introduced mutations (***Figure 36b***). It revealed that rACG1 was located at a more optimal distance (23 nt) from the short hairpin structure (containing the native ACG (+1) codon) than rACG2 which was engaged in the formation of the stem of this hairpin. Thus, the distance between the ACG (+1) of the rACG2 mutant seemed to be too short to position the ribosome at optimal orientation, which is the P-site at the ACG codon. Although ribosome profiling data presented that the distance which can be considered as the length of mRNA covered by the ribosome during translation is roughly estimated as 30 nt[207] it has been proven that ribosome footprints can vary significantly depending on the organism, cell type, and experimental protocol[208]. Thus, it is possible that the increase in the efficiency of FMRpolyG translation initiation at ACG near-cognate start codon from rACG1 mutant results from the more optimal positioning of the scanning ribosome. Namely, the presence of secondary structure downstream studied TIS, stabilized by adjacent CGG hairpin, could *(i)* position ACG near-cognate start codon in the middle of the large ribosome subunit, and *(ii)* increase the dwell time of the ribosome due to hairpin unwinding and directly lead to increase in the codon utilization in rACG1 mutant. Therefore, the structural dependencies would be the main factor in the regulation of translation initiation in that case.

In 1990 Marilyn Kozak performed an elegant study providing the very first clues about the role of secondary structures downstream TIS on the efficiency of translation initiation[121]. She presented that depending on the distance between the start codon and the stable hairpin structure the initiation can be either enhanced or reduced. The optimal location of the hairpin 14-16 nt downstream of the start codon has been established. As already mentioned, based on the RNase protection assays it was presented that this length corresponds to the sequence length positioning the ribosome P-site close to the start

codon. Hence the scanning ribosome could be more eager to initiate translation at a particular codon.

To verify whether ACG (+1) near-cognate start codon utilization within the *FMR1* 5'UTR could be modulated similarly the constructs harboring artificial hairpin structure with the Gibbs free energy close to the one predicted for hairpin formed by 16 CGG repeats were tested (***Figure 37a-b***). The Gibbs free energy of the predicted hairpin was ΔG=–46 kcal/mol that according to the available data is sufficient to greatly inhibit translation initiation by posing a substantial energetic barrier for scanning 43S ribosome[227,229]. The results revealed that both Hairpin14nt and Hairpin20nt resulted in a great increase in the FMRpolyG level however, contrary to the assumptions, the latter one had a more positive impact (***Figure 38a-b***). The structure predictions revealed that the hairpin formed by the Hairpin20nt mutant was in fact located 16 nt, instead of the designed 20 nt, downstream of the ACG (+1) (***Figure 37e***). Therefore, it could result in better positioning of ribosome site P at the ACG codon than the structure formed by Hairpin14nt which was reflected by the obtained results. As expected the mutants with hairpins located close to ACG (+1) near-cognate start codon, namely Hairpin2nt and Hairpin6nt, abolished the FMRpolyG translation initiation at that codon probably by preventing the ribosome from accessing the TIS. The RT-qPCR analysis proved that observed differences in the FMRpolyG levels result directly from the translation efficiency as the *FMR1* mRNA level, expecting Hairpin6nt, was unchanged (***Figure 38c***). Intriguingly, the efficiency of FMRP equivalent translation was impaired in all tested mutants. I proposed that the presence of an energetically stable obstacle upstream of the FMRP AUG codon led to the ribosome dissociation, hence, similarly to ACG→ATG mutant, only a limited number of ribosomes, due to leaky scanning or re-initiation, could start translation at the TIS of FMRP. The model of this proposed mechanism is illustrated in ***Figure 39***, and it could partially explain the observed inhibition of the translation of downstream FMRP ORF in the Hairpin14nt mutant which based on structure prediction had the most structured 5'UTR and could most efficiently block incoming ribosomes.

Interestingly, the translation of FMRpolyR seems to be also regulated by the introduced hairpin structures. Hairpin2nt and Hairpin6nt, according to predictions, may lengthen and stabilize the hairpin structure with native ACG (+1) near-cognate start codon whose stem is located 12 nt downstream the ACG (+0) near-cognate start codon for FMRpolyR. Therefore due to ribosome pausing the ACG (+0) near-cognate start codon can be more

efficiently utilized by the scanning 43S PIC. On the other hand, the distance between the ACG (+0) and the hairpin structure within Hairpin14nt and Hairpin20nt mutants is the same, however, probably due to more structured 5'UTR the level of FMRpolyR differs significantly. Even though dependencies driving the initiation of FMRpolyR translation at ACG (+0) near-cognate start codon are beyond the scope of this work the developed reporter-based system could be successfully used to explore this thread.

The discussed above results concerning the distance between the near-cognate start codon and the stable secondary RNA structure confirmed that structural obstacles may regulate the efficiency of codon utilization. In light of these data, the increased distance between ACG (+1) near-cognate start codon and hairpin formed by CGG repeats should result in decreased efficiency of FMRpolyG translation. Indeed, the CAlinker mutant (***Figure 40***) resulted in a very strong loss of FMRpolyG level (***Figure 40c***) which implies that the ACG (+1) near-cognate start codon embedded in the native *FMR1* 5'UTR sequence constitutes a weak TIS which is mostly overlooked by the scanning ribosome. Thus, the strength of codon utilization is mainly modulated by the stable downstream secondary structure formed by CGG repeats. Importantly, the foldchange of FMRpolyG level reduction in construct with 85 CGG repeats (0.6) was substantially smaller than this observed for CAlinker mutant with short CGG repeats (0.3) supporting the statement about the involvement of secondary structure in the FMRpolyG translation initiation. The increased ACG (+1) utilization forced by longer CGG repeats would result from the phenomenon of ribosome queuing, therefore despite the enlarged distance the stacked/queued ribosomes would mimic the structural obstacle and increase the initiation at ACG (+1). Noteworthy, the non-AUG initiation is exceptionally sensitive to conditions that slow down or pause the progression of scanning of 43S ribosome. In line with that, it has been shown that when translation elongation was inhibited by cycloheximide the level of non-AUG translation was elevated[237] which could directly result from the queuing of preinitiation complexes/ribosomes. To verify this thesis the experiments utilizing the ribosome profiling on HEK-293 cells transfected with mentioned constructs would be beneficial to define the ribosome position on mRNA. Simultaneously, the polyribosome fractionation would provide additional information about translation dynamics through *FMR1* 5'UTR.

Treatment strategies for neurodegenerative diseases using antisense steric blockers have been successfully tested in animal models and human clinical trials[238]. I used two ASO

steric blockers composed exclusively of 2'MOE units to target flanking regions of CGG repeats (ASO1 and ASO3) which, as it was confirmed, are responsible for CGG hairpin stabilization[181]. Since previously tested ASO-CCG, targeting directly CGG repeats within a hairpin structure, reduced the translation of FMRpolyG[177] I wanted to verify how ASO targeting different regions of CGG hairpin would affect the FMRpolyG translation. Although both ASO1 and ASO3 may result in partial unwinding of the hairpin formed by CGG repeats (within the region of stabilizing flanking sequences) the effect on the RAN translation initiation was largely opposite. Obtained results present that binding of ASO to the 5' site of the flanking region of CGG repeats increased the level of FMRpolyG, however only from *FMR1* with short CGG repeats (***Figure 41b-c***) suggesting that the positive effect of ASO1 on the RAN translation is eclipsed by the stable secondary RNA structure. On the contrary, the strong depletion in the FMRpolyG synthesis after ASO3 treatment was observed independently from the number of CGG repeats.

In ***Figure 42*** I presented how the binding of ASO1 and ASO3 may affect the structure of *FMR1* 5'UTR and therefore change the efficiency of RAN translation initiation at the ACG (+1) near-cognate start codon. Both ASOs could partially resolve hairpin structures in the region of their binding. In the case of ASO1, it could be expected that the binding of steric blocker could slow down the PIC scanning or even lead to the PIC stalling and therefore increase the initiation at near-cognate start codon. The distance between the ACG (+1) near-cognate start codon and the beginning of ASO1 was 22 nt, and previous experiments (***Figure 38***) suggested that such distance was efficient to force initiation at the near-cognate start codon. However, this increase was not observed when a very stable hairpin structure was formed (85 CGGs) possibly due to the strong decrease in the general *FMR1* translation efficiency. Targeting the 3' site of the CGG flanking region by ASO3 would instead extend the distance between the ACG (+1) near-cognate start codon and the stem of the CGG hairpin, from 17 nt to 30 nt. Hence, the strong depletion in the FMRpolyG level was observed independently from the number of CGG repeats. Even though the ribosome queuing on the long repeats is possible, the positive effect on the translation initiation at ACG (+1) near-cognate start codon forced by paused ribosomes may be invisible. This may be due to the generally negative effect of long CGG repeats on the translational efficiency of *FMR1* mRNA.

The results described in this part of the study presented for the first time that:

1. The nucleotide context in the vicinity of ACG (+1) near-cognate start codon of FMRpolyG open reading frame is crucial for the RAN translation initiation;

2. The initiation of FMRpolyG biosynthesis is strongly regulated by the distance between the ACG (+1) near-cognate start codon and the stable secondary RNA structure;

3. There is an interplay between the sequence and structure formed within 5'UTR of *FMR1* mRNA which jointly modulate the efficiency of FMRpolyG translation initiation.

6. LIST OF FIGURES

7. LIST OF TABLES

8. BIBLIOGRAPHY

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **412**, 565–566 (2001).

2. Malik, I., Kelley, C. P., Wang, E. T. & Todd, P. K. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607 (2021).

3. Depienne, C. & Mandel, J. L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* **108**, 764–785 (2021).

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

5. Kunst, C. B. *et al.* FMR1 in global populations. *Am. J. Hum. Genet.* **58**, 513–522 (1996).

6. Fernandez-Carvajal, I. *et al.* Screening for expanded alleles of the FMR1 gene in blood spots from newborn males in a Spanish population. *J. Mol. Diagnostics* **11**, 324–329 (2009).

7. Hagerman, P. J. & Hagerman, R. J. The Fragile-X Premutation: A Maturing Perspective. *Am. J. Hum. Genet.* **74**, 805–816 (2004).

8. Oostra, B. A. & Willemsen, R. A fragile balance: FMR1 expression levels. *Hum. Mol. Genet.* **12**, 249–257 (2003).

9. Hall, D. A. In the Gray Zone in the Fragile X Gene: What are the Key Unanswered Clinical and Biological Questions? *Tremor and Other Hyperkinetic Movements* **4**, 208 (2014).

10. Loesch, D. Z. *et al.* Transcript levels of the intermediate size or grey zone fragile X mental retardation 1 alleles are raised, and correlate with the number of CGG repeats. *J. Med. Genet.* **44**, 200–204 (2007).

11. Terracciano, A. *et al.* Expansion to full mutation of a FMR1 intermediate allele over two generations. *Eur. J. Hum. Genet.* **12**, 333–336 (2004).

12. Monaghan, K. G., Lyon, E. & Spector, E. B. ACMG standards and guidelines for

fragile X testing: A revision to the disease-specific supplements to the standards and guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics and Genomics. *Genet. Med.* **15**, 575–586 (2013).

13. Darnell, J. C. & Richter, J. D. Cytoplasmic RNA-binding proteins and the control of complex brain function. *Cold Spring Harb. Perspect. Biol.* **4**, 1–17 (2012).

14. Costa-Mattioli, M. & Klann, E. Translational Control Mechanisms in Synaptic Plasticity and Memory. *Curated Ref. Collect. Neurosci. Biobehav. Psychol.* **61**, 311–328 (2016).

15. Parus, J. L., Kuc, G. & Kierzek, J. Determination of lead and silver in copper blister by isotope excited X-ray fluorescence. *J. Radioanal. Chem.* **44**, 189–197 (1978).

16. Munoz, D. G. *et al.* Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **58**, 987–988 (2012).

17. James, A. *et al.* Fragile X Premutation Tremor/Ataxia Syndrome: Molecular, Clinical, and Neuroimaging Correlates. *Am. J. Hum. Genet.* **72**, 869–878 (2003).

18. Sullivan, A. K. *et al.* Association of FMR1 repeat size with ovarian dysfunction. *Hum. Reprod.* **20**, 402–412 (2005).

19. Aishworiya, R. *et al.* Fragile X-Associated Neuropsychiatric Disorders (FXAND) in Young Fragile X Premutation Carriers. *Genes (Basel).* **13**, (2022).

20. Hagerman, R. J. *et al.* Fragile X-Associated Neuropsychiatric Disorders (FXAND). *Front. Psychiatry* **9**, 1–9 (2018).

21. Tabolacci, E., Palumbo, F., Nobile, V. & Neri, G. Transcriptional reactivation of the FMR1 Gene. A possible approach to the treatment of the fragile X syndrome. *Genes (Basel).* **7**, 1–16 (2016).

22. Greco, C. M. *et al.* Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain* **129**, 243–255 (2006).

23. Marsha Mailick Seltzer1,*, Mei Wang Baker1, Jinkuk Hong1, Matthew Maenner1, Jan Greenberg1, and D. M. Prevalence of CGG Expansions of the FMR1 Gene in a US Population-Based Sample. *Am J Med Genet B Neuropsychiatr Genet.* **159**, 589–597 (2012).

24. Hantash, F. M. *et al.* FMR1 premutation carrier frequency in patients undergoing routine population-based carrier screening: Insights into the prevalence of fragile X syndrome, fragile X-associated tremor/ataxia syndrome, and fragile X-associated primary ovarian insufficiency i. *Genet. Med.* **13**, 39–45 (2011).

25. Owens, K. M. *et al.* FMR1 premutation frequency in a large, ethnically diverse population referred for carrier testing. *Am. J. Med. Genet. Part A* **176**, 1304–1308 (2018).

26. F, T. *et al.* FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Med.* **4**, 100 (2012).

27. Rodriguez-Revenga, L. *et al.* Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *Eur. J. Hum. Genet.* **17**, 1359–1362 (2009).

28. Randi Hagerman, MD1 and Paul Hagerman, MD, P. Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. *Lancet Neurol.* **12**, 786–798 (2013).

29. Kenneson, A., Zhang, F., Hagedorn, C. H. & Warren, S. T. 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 (2001).

30. Primerano, B. *et al.* Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. *Rna* **8**, 1482–1488 (2002).

31. Allen, E. G., He, W., Yadav-Shah, M. & Sherman, S. L. A study of the distributional characteristics of FMR1 transcript levels in 238 individuals. *Hum. Genet.* **114**, 439–447 (2004).

32. Tassone, F., Hagerman, R. J., Chamberlain, W. D. & Hagerman, P. J. Transcription of the FMR1 gene in individuals with fragile X syndrome. *Am. J. Med. Genet. - Semin. Med. Genet.* **97**, 195–203 (2000).

33. Tassone, F. *et al.* Elevated levels of FMR1 mRNA carrier males: A new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* **66**, 6–15 (2000).

34. Feng, Y. *et al.* Translational suppression by trinucleotide repeat expansion at FMR1. *Science (80-.).* **268**, 731–734 (1995).

35. Chen, L. S., Tassone, F., Sahota, P. & Hagerman, P. J. The (CGG)n repeat element within the 5′ untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of a downstream reporter. *Hum. Mol. Genet.* **12**, 3067–3074 (2003).

36. Ludwig, A. L. *et al.* CNS expression of murine fragile X protein (FMRP) as a function of CGG-repeat size. *Hum. Mol. Genet.* **23**, 3228–3238 (2014).

37. Tassone, F. *et al.* Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *Rna* **13**, 555–562 (2007).

38. Tassone, F. *et al.* Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Res.* **39**, 6172–6185 (2011).

39. Beilina, A., Tassone, F., Schwartz, P. H., Sahota, P. & Hagerman, P. J. Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum. Mol. Genet.* **13**, 543–549 (2004).

40. Pieretti, M. *et al.* Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817–822 (1991).

41. O'Donnell, W. T. & Warren, S. T. A decade of molecular studies of fragile X syndrome. *Annu. Rev. Neurosci.* **25**, 315–338 (2002).

42. Myrick, L. K. *et al.* Fragile X syndrome due to a missense mutation. *Eur. J. Hum. Genet.* **22**, 1185–1189 (2014).

43. Wöhrle, D., Hirst, M. C., Kennerknecht, I., Davies, K. E. & Steinbach, P. Genotype mosaicism in fragile X fetal tissues. *Hum. Genet.* **89**, 114–116 (1992).

44. Nolin, S. L., Glicksman, A., Houck, G. E., Brown, W. T. & Dobkin, C. S. Mosaicism in fragile X affected males. *Am. J. Med. Genet.* **51**, 509–512 (1994).

45. Pretto, D. *et al.* Clinical and molecular implications of mosaicism in FMR1 full mutations. *Front. Genet.* **5**, 1–11 (2014).

46. Hansen, R. S., Gartler, S. M., Scott, C. R., Chen, S. H. & Xlaird, C. M. Methylation analysis of CGG sites in the CpG Island of the human FMR1 gene. *Hum. Mol. Genet.* **1**, 571–578 (1992).

47. Stöger, R., Kajimura, T. M., Brown, W. T. & Laird, C. D. Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. *Hum. Mol. Genet.* **6**, 1791–1801 (1997).

48. McConkie-Rosell, A. *et al.* Evidence that methylation of the FMR-1 locus is responsible for variable phenotypic expression of the fragile X syndrome. *Am. J. Hum. Genet.* **53**, 800–809 (1993).

49. Mikaeili, H., Sandi, M., Bayot, A., Al-Mahdawi, S. & Pook, M. A. FAST-1 antisense RNA epigenetically alters FXN expression. *Sci. Rep.* **8**, 1–11 (2018).

50. Tao Zu1, 2, John D. Cleary1, 2, Yuanjing Liu1, 2, Monica Bañez-Coronel1, 2, Jodi L. Bubenik1, 2, Fatma Ayhan1, 2, Tetsuo Ashizawa1, 3, 6, 7, Guangbin Xia1, 3, 6, H. Brent Clark8, Anthony T. Yachnis4, Maurice S. Swanson1, 2, 5, and L. P. W. R. RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. *Neuron* **13**, 1292–1305 (2017).

51. Chung, D. W., Rudnicki, D. D., Yu, L. & Margolis, R. L. A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Hum. Mol. Genet.* **20**, 3467–3477 (2011).

52. Mori, K. *et al.* Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol.* **126**, 881–893 (2013).

53. Ladd, P. D. *et al.* An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum. Mol. Genet.* **16**, 3174–3187 (2007).

54. Khalil, A. M., Faghihi, M. A., Modarresi, F., Brothers, S. P. & Wahlestedt, C. A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PLoS One* **3**, (2008).

55. Alvarez-Mora, M. I. *et al.* Evaluation of FMR4, FMR5 and FMR6 Expression Levels as Non-Invasive Biomarkers for the Diagnosis of Fragile X-Associated

Primary Ovarian Insufficiency (FXPOI). *J. Clin. Med.* **11**, 1–10 (2022).

56. Zafarullah, M., Li, J., Tseng, E. & Tassone, F. Structure and Alternative Splicing of the Antisense FMR1 (ASFMR1) Gene. *Mol. Neurobiol.* **60**, 2051–2061 (2023).

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

58. Fernández, J. J. *et al.* Gene expression profiles in the cerebellum of transgenic mice over expressing the human FMR1 gene with CGG repeats in the normal range. *Genet. Mol. Res.* **11**, 467–483 (2012).

59. Sofola, O. A. *et al.* 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 (2007).

60. Peng Jin1,*, Ranhui Duan1,*, Abrar Qurashi1, Yunlong Qin1, Donghua Tian3, Tracie C. Rosser1, Huijie Liu1, Yue Feng3, and Stephen T. Warren1, 2. Pur α binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of Fragile X Tremor/ Ataxia Syndrome. *Neuron* **16**, 556–564 (2007).

61. Sellier, C. *et al.* Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J.* **29**, 1248–1261 (2010).

62. Sellier, C. *et al.* 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 (2013).

63. Paronetto, M. P., Achsel, T., Massiello, A., Chalfant, C. E. & Sette, C. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* **176**, 929–939 (2007).

64. Chawla, G. *et al.* Sam68 Regulates a Set of Alternatively Spliced Exons during Neurogenesis. *Mol. Cell. Biol.* **29**, 201–213 (2009).

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

66. Tassone, F. *et al.* Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome. *J. Med. Genet.* **41**, 1–3 (2004).

67. Sellier, C. *et al.* Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron* **93**, 331–347 (2017).

68. Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256–271 (2006).

69. Todd, P. K. *et al.* CGG Repeat Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome Peter. *Neuron* **78**, (2014).

70. Kearse, M. G. *et al.* CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol. Cell* **62**, 314–322 (2017).

71. Green, K. M. *et al.* RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat. Commun.* **8**, (2017).

72. Tabet, R. *et al.* CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat. Commun.* **9**, 1–14 (2018).

73. Ivanov, I. P., Firth, A. E., Michel, A. M., Atkins, J. F. & Baranov, P. V. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* **39**, 4220–4234 (2011).

74. Jongens, R. W. B. and T. A. A non-canonical start codon in the Drosophila fragile X gene yields two functional isoforms. *Neuroscience* **181**, 48–66 (2011).

75. Krans, A., Skariah, G., Zhang, Y., Bayly, B. & Todd, P. K. Neuropathology of RAN translation proteins in fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol. Commun.* **7**, 1–17 (2019).

76. Buijsen, R. A. *et al.* 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**, 1–5 (2014).

77. Ariza, J. *et al.* A Majority of FXTAS Cases Present with Intranuclear Inclusions Within Purkinje Cells. *Cerebellum* **15**, 546–551 (2016).

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

79. Kong, H. E., Zhao, J., Xu, S., Jin, P. & Jin, Y. Fragile X-associated tremor/ataxia syndrome: From molecular pathogenesis to development of therapeutics. *Front. Cell. Neurosci.* **11**, 1–11 (2017).

80. Rodriguez, C. M. *et al.* A native function for RAN translation and CGG repeats in regulating Fragile X protein synthesis. *Nat. Neurosci.* **23**, 386–397 (2020).

81. Dubińska-Magiera, M. *et al.* Xenopus LAP2β protein knockdown affects location of lamin B and nucleoporins and has effect on assembly of cell nucleus and cell viability. *Protoplasma* **253**, 943–956 (2016).

82. Garcia-Arocena, D. *et al.* Fibroblast phenotype in male carriers of FMR1 premutation alleles. *Hum. Mol. Genet.* **19**, 299–312 (2009).

83. Greco, C. M. *et al.* Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain* **125**, 1760–1771 (2002).

84. Michael R. Hunsaker, Claudia M. Greco, Marian A. Spath, Arie P. T. Smits, Celestine S. Navarro, Flora Tassone, Johan M. Kros, Lies-Anne Severijnen, R. F. B. J. H. Widespread non-central nervous system organ pathology in fragile X premutation carriers with fragile X-associated tremor/ ataxia syndrome and CGG knock-in mice Michael. *Acta Neuropathol. Commun.* **122**, 467–479 (2011).

85. Ma, L. *et al.* Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome. *Acta Neuropathol. Commun.* **7**, 1–26 (2019).

86. Ali Entezam1, Rea Biacsi1, Bonnie Orrison2, Tapas Saha1, 3, Gloria E. Hoffman4, Ed Grabczyk1, 5, Robert L. Nussbaum2, 6, and K. U. Regional FMRP deficits and large repeat expansions into the full mutation range in a new Fragile X premutation mouse model. *Gene* **15**, 125–134 (2007).

87. Berman, R. F. & Willemsen, R. Mouse models of fragile X-associated tremor ataxia. *J. Investig. Med.* **57**, 837–841 (2009).

88. Brouwer, J. R. *et al.* CGG-repeat length and neuropathological and molecular correlates in a mouse model for fragile X-associated tremor/ataxia syndrome syndrome. **107**, 1671–1682 (2008).

89. Mei Qina, Ali Entezamb, Karen Usdinb, Tianjian Huanga, Zhong-Hua Liua, Gloria E. Hoffmanc, and C. B. S. A mouse model of the fragile X premutation: effects on behavior, dendrite morphology, and regional rates of cerebral protein synthesis. *Neurobiol Dis* **42**, 85–98 (2011).

90. Hunsaker, M. R., Wenzel, H. J., Willemsen, R. & Berman, R. F. Progressive Spatial Processing Deficits in a Mouse Model of the Fragile X Premutation. *Behav. Neurosci.* **123**, 1315–1324 (2009).

91. Van Dam, D. *et al.* Cognitive decline, neuromotor and behavioural disturbances in a mouse model for fragile-X-associated tremor/ataxia syndrome (FXTAS). *Behav. Brain Res.* **162**, 233–239 (2005).

92. Richards, R. I. & McLeod, C. J. RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Chemtracts* **18**, 153–158 (2005).

93. Castro, H. *et al.* Selective rescue of heightened anxiety but not gait ataxia in a premutation 90CGG mouse model of Fragile X-associated tremor/ataxia syndrome. *Hum. Mol. Genet.* **26**, 2133–2145 (2017).

94. Oh, S. Y. *et al.* 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 (2015).

95. Asamitsu, S. *et al.* CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Sci. Adv.* **7**, 1–14 (2021).

96. Hoem, G. & Johansen, T. The FMRpolyGlycine Protein Mediates Aggregate Formation and Toxicity Independent of the CGG mRNA Hairpin in a Cellular Model for. **10**, 1–18 (2019).

97. Salcedo-Arellano, M. J., Dufour, B., McLennan, Y., Martinez-Cerdeno, V. & Hagerman, R. Fragile X syndrome and associated disorders: Clinical aspects and

pathology. *Neurobiol. Dis.* **136**, 104740 (2020).

98. Kozak, M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123 (1978).

99. Kozak, M. Evaluation of the “scanning model” for initiation of protein synthesis in eucaryotes. *Cell* **22**, 7–8 (1980).

100. Nadel, Y., Weisman-Shomer, P. & Fry, M. The fragile X syndrome single strand D(CGG)(n) nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.* **270**, 28970–28977 (1995).

101. Khateb, S., Weisman-Shomer, P., Hershco-Shani, I., Ludwig, A. L. & Fry, M. The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res.* **35**, 5775–5788 (2007).

102. Wang, J. *et al.* Rapid 40S scanning and its regulation by mRNA structure during eukaryotic translation initiation. *Cell* **185**, 4474-4487.e17 (2022).

103. Dobson, T., Kube, E., Timmerman, S. & Krushel, L. A. Identifying intrinsic and extrinsic determinants that regulate internal initiation of translation mediated by the FMR1 5′ leader. *BMC Mol. Biol.* **9**, 1–13 (2008).

104. Chiang, P. W., Carpenter, L. E. & Hagerman, P. J. The 5′-Untranslated Region of the FMR1 Message Facilitates Translation by Internal Ribosome Entry. *J. Biol. Chem.* **276**, 37916–37921 (2001).

105. Choi, J.-H. *et al.* hnRNP Q Regulates Internal Ribosome Entry Site-Mediated fmr1 Translation in Neurons . *Mol. Cell. Biol.* **39**, (2019).

106. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).

107. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2850–2854 (1986).

108. Kozak, M. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature* **308**, 241–246 (1984).

109. Kozak, M. Context effects and inefficient initiation at non-AUG codons in

eucaryotic cell-free translation systems. *Mol. Cell. Biol.* **9**, 5073–5080 (1989).

110. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger rNAS. *Nucleic Acids Res.* **15**, 8125–8148 (1987).

111. Hinnebusch, A. G. Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem. Sci.* **42**, 589–611 (2017).

112. Pisarev, A. V. *et al.* Specific functional interactions of nucleotides at key-3 and+4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* **20**, 624–636 (2006).

113. Chew, G. L., Pauli, A. & Schier, A. F. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat. Commun.* **7**, 1–10 (2016).

114. Sachs, M. S. & Geballe, A. P. Downstream control of upstream open reading frames. *Genes Dev.* **20**, 915–921 (2006).

115. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORF s are prevalent translational repressors in vertebrates . *EMBO J.* **35**, 706–723 (2016).

116. Gunišová, S., Hronová, V., Mohammad, M. P., Hinnebusch, A. G. & Valášek, L. S. Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. *FEMS Microbiol. Rev.* **42**, 165–192 (2018).

117. Fritsch, C. *et al.* Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* **22**, 2208–2218 (2012).

118. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).

119. Cloutier, P. *et al.* Upstream ORF-Encoded ASDURF Is a Novel Prefoldin-like Subunit of the PAQosome. *J. Proteome Res.* **19**, 18–27 (2020).

120. Akimoto, C. *et al.* Translational repression of the McKusick-Kaufman syndrome transcript by unique upstream open reading frames encoding mitochondrial proteins with alternative polyadenylation sites. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, 2728–2738 (2013).

121. Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8301–8305 (1990).

122. Kochetov, A. V. *et al.* AUG hairpin: Prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics* **8**, 3–9 (2007).

123. Lindsey D. Goodman1, Nancy M. Bonini1, 2. Repeat-associated non-AUG (RAN) translation mechanisms running into focus for GGGGCC-repeat associated ALS/FTD. *Prog Neurobiol.* **183**, 139–148 (2019).

124. Brcic, J. & Plavec, J. NMR structure of a G-quadruplex formed by four d(G4C2) repeats: insights into structural polymorphism. *Nucleic Acids Res.* **46**, 11605–11617 (2018).

125. Groh, M., Lufino, M. M. P., Wade-Martins, R. & Gromak, N. R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLoS Genet.* **10**, (2014).

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

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

128. Roy, D., Zhang, Z., Lu, Z., Hsieh, C.-L. & Lieber, M. R. Competition between the RNA Transcript and the Nontemplate DNA Strand during R-Loop Formation In Vitro: a Nick Can Serve as a Strong R-Loop Initiation Site. *Mol. Cell. Biol.* **30**, 146–159 (2010).

129. Price, S. J. and T. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **23**, 1–7 (2008).

130. Lakhani, C. M. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **176**, 139–148 (2019).

131. Chen, L. *et al.* R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol. Cell* **68**, 745-757.e5

(2017).

132. Nadel, J. *et al.* RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics and Chromatin* **8**, 1–19 (2015).

133. Grunseich, C. *et al.* Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Mol. Cell* **69**, 426-437.e7 (2018).

134. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).

135. Castellano-Pozo, M. *et al.* R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* **52**, 583–590 (2013).

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

137. Reddy, K. *et al.* Processing of double-R-loops in (CAG)•(CTG) and C9orf72 (GGGGCC)•(GGCCCC) repeats causes instability. *Nucleic Acids Res.* **42**, 10473–10487 (2014).

138. Reddy, K. *et al.* Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.* **39**, 1749–1762 (2011).

139. Diab, M. A. *et al.* The g-rich repeats in FMR1 and C9orf72 loci are hotspots for local unpairing of DNA. *Genetics* **210**, 1239–1252 (2018).

140. Belotserkovskii, B. P., Shin, J. H. S. & Hanawalt, P. C. 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 (2017).

141. Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R. & Chédin, F. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi–Goutières syndrome. *Elife* **4**, 1–21 (2015).

142. Sanz, L. A. *et al.* Prevalent, Dynamic, and Conserved R-Loop Structures Associate

with Specific Epigenomic Signatures in Mammals. *Mol. Cell* **63**, 167–178 (2016).

143. Hoem, G. *et al.* CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS. *Hum. Mol. Genet.* **20**, 2161–2170 (2011).

144. Belotserkovskii, B. P., Tornaletti, S., D'Souza, A. D. & Hanawalt, P. C. R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair (Amst).* **71**, 69–81 (2018).

145. Tan-Wong, S. M., Dhir, S. & Proudfoot, N. J. R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol. Cell* **76**, 600-616.e6 (2019).

146. Dilek Colak1, Nikica Zaninovic2, Michael S. Cohen1,*, Zev Rosenwaks2, Wang-Yong Yang3, Jeannine Gerhardt4, Matthew D. Disney3, and Samie R. Jaffrey1, †. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile x syndrome. *Science (80-.).* **176**, 139–148 (2014).

147. Avitzour, M. *et al.* FMR1 epigenetic silencing commonly occurs in undifferentiated fragile X-affected embryonic stem cells. *Stem Cell Reports* **3**, 699–706 (2014).

148. De Esch, C. E. F. *et al.* Epigenetic characterization of the FMR1 promoter in induced pluripotent stem cells from human fibroblasts carrying an unmethylated full mutation. *Stem Cell Reports* **3**, 548–555 (2014).

149. Pietrobono, R. *et al.* Molecular dissection of the events leading to inactivation of the FMR1 gene. *Hum. Mol. Genet.* **14**, 267–277 (2005).

150. Tabolacci, E. *et al.* Epigenetic analysis reveals a euchromatic configuration in the FMR1 unmethylated full mutations. *Eur. J. Hum. Genet.* **16**, 1487–1498 (2008).

151. Naumann, A., Hochstein, N., Weber, S., Fanning, E. & Doerfler, W. A Distinct DNA-Methylation Boundary in the 5′- Upstream Sequence of the FMR1 Promoter Binds Nuclear Proteins and Is Lost in Fragile X Syndrome. *Am. J. Hum. Genet.* **85**, 606–616 (2009).

152. Coffee, B., Zhang, F., Ceman, S., Warren, S. T. & Reines, D. Histone Modifications Depict an Aberrantly Heterochromatinized FMR1. *Am. J. Hum. Genet.* **71**, 923–932 (2002).

153. El-Osta, A. FMR1 silencing and the signals to chromatin: A unified model of transcriptional regulation. *Biochem. Biophys. Res. Commun.* **295**, 575–581 (2002).

154. Usdin, K. *et al.* Repeat-mediated genetic and epigenetic changes at the FMR1 locus in the Fragile X-related disorders. *Front. Genet.* **5**, 1–16 (2014).

155. Kumari, D. & Usdin, K. The distribution of repressive histone modifications on silenced FMR1 alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum. Mol. Genet.* **19**, 4634–4642 (2010).

156. Kumari, D., Gazy, I. & Usdin, K. Pharmacological reactivation of the silenced FMR1 gene as a targeted therapeutic approach for fragile X syndrome. *Brain Sci.* **9**, 1–18 (2019).

157. Abu Diab M, Mor-Shaked, Cohen E, Cohen-Hadad Y, Ram O, Epsztejn-Litman S, E. R. The G-rich repeats in FMR1 and C9orf72 loci are hotspots for local unpairing of DNA. *Genetics* 1239–1252 (2018) doi:10.1534/genetics.118.301672.

158. Chen, X. *et al.* Hairpins are formed by the single DNA strands of the fragile X triplet repeats: Structure and biological implications. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5199–5203 (1995).

159. Rodden, L. N., Rummey, C., Dong, Y. N. & Lynch, D. R. Clinical Evidence for Variegated Silencing in Patients with Friedreich Ataxia. *Neurol. Genet.* **8**, 1–7 (2022).

160. Lee, H. G. *et al.* Site-specific R-loops induce CGG repeat contraction and fragile X gene reactivation. *Cell* **186**, 2593-2609.e18 (2023).

161. Xie, N. *et al.* Reactivation of FMR1 by CRISPR/Cas9-mediated deletion of the expanded CGG-repeat of the fragile X chromosome. *PLoS One* **11**, 1–12 (2016).

162. Liu, X. S. *et al.* Rescue of Fragile X syndrome neurons by DNA methylation editing of the FMR1 gene. *Cell* **172**, 979–992 (2018).

163. Park, C. Y. *et al.* Reversion of FMR1 Methylation and Silencing by Editing the Triplet Repeats in Fragile X iPSC-Derived Neurons. *Cell Rep.* **13**, 234–241 (2015).

164. Kumari, D., Sciascia, N. & Usdin, K. Small molecules targeting H3K9 methylation prevent silencing of reactivated FMR1 alleles in fragile X syndrome patient

derived cells. *Genes (Basel).* **11**, (2020).

165. Hunt, J. F. V. *et al.* High throughput small molecule screen for reactivation of fmr1 in fragile x syndrome human neural cells. *Cells* **11**, 1–12 (2022).

166. Ochman, H., Gerber, A. S. & Hartl, D. L. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623 (1988).

167. Paek, K. Y. *et al.* Translation initiation mediated by RNA looping. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1041–1046 (2015).

168. Rovozzo, R. *et al.* CGG repeats in the 5'UTR of FMR1 RNA regulate translation of other RNAs localized in the same RNA granules. *PLoS One* **11**, 1–23 (2016).

169. Doers, M. E. *et al.* IPSC-derived forebrain neurons from FXS individuals show defects in initial neurite outgrowth. *Stem Cells Dev.* **23**, 1777–1787 (2014).

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

171. Gagnon, K. T., Li, L., Janowski, B. A. & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc.* **9**, 2045–2060 (2014).

172. Aguilera, A. & Huertas, P. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721 (2003).

173. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol. Cell* **42**, 794–805 (2011).

174. Aguilera, A. & García-Muse, T. R Loops: From Transcription Byproducts to Threats to Genome Stability. *Mol. Cell* **46**, 115–124 (2012).

175. Lockhart, A. *et al.* RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep.* **29**, 2890-2900.e5 (2019).

176. Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep.* **23**, 1891–1905 (2018).

177. Derbis, M. *et al.* Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring expanded CGG repeats. *Nat. Commun.* **12**, 1–17 (2021).

178. Gagliardi, M. & Ashizawa, A. T. The challenges and strategies of antisense oligonucleotide drug delivery. *Biomedicines* **9**, (2021).

179. Piao, X., Wang, H., Binzel, D. W. & Guo, P. and LNA in the context of Phi29 pRNA 3WJ. 67–76 (2018) doi:10.1261/rna.063057.117.The.

180. Zumwalt, M., Ludwig, A., Hagerman, P. J. & Dieckmann, T. Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the Fragile X Mental Retardation 1 (FMR1) gene. *RNA Biol.* **4**, 93–100 (2007).

181. Napierala, M., Michalowski, D., de Mezer, M. & Krzyzosiak, W. J. Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.* **33**, 451–463 (2005).

182. Ajjugal, Y., Kolimi, N. & Rathinavelan, T. Secondary structural choice of DNA and RNA associated with CGG/CCG trinucleotide repeat expansion rationalizes the RNA misprocessing in FXTAS. *Sci. Rep.* **11**, 1–17 (2021).

183. Usdin, K. & Woodford, K. J. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res.* **23**, 4202–4209 (1995).

184. Fry, M. & Loeb, L. A. The fragile X syndrome d(CGG)(n) nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4950–4954 (1994).

185. Yan, Q. & Sarma, K. MapR: A Method for Identifying Native R-Loops Genome Wide. *Curr. Protoc. Mol. Biol.* **130**, 1–12 (2020).

186. Amon, J. D. & Koshland, D. RNase H enables efficient repair of R-loop induced DNA damage. *Elife* **5**, 1–20 (2016).

187. Matsui, M. *et al.* USP42 enhances homologous recombination repair by promoting R-loop resolution with a DNA–RNA helicase DHX9. *Oncogenesis* **9**, (2020).

188. Yuan, W. *et al.* TDRD3 promotes DHX9 chromatin recruitment and R-loop resolution. *Nucleic Acids Res.* **49**, 8573–8591 (2021).

189. Chakraborty, P., Huang, J. T. J. & Hiom, K. DHX9 helicase promotes R-loop

formation in cells with impaired RNA splicing. *Nat. Commun.* **9**, (2018).

190. Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst).* **10**, 654–665 (2011).

191. Li, L., Matsui, M. & Corey, D. R. Activating frataxin expression by repeat-targeted nucleic acids. *Nat. Commun.* **7**, 1–8 (2016).

192. Haenfler, J. M. *et al.* Targeted Reactivation of FMR1 Transcription in Fragile X Syndrome Embryonic Stem Cells. *Front. Mol. Neurosci.* **11**, 1–17 (2018).

193. Kumari, D. & Usdin, K. Sustained expression of FMR1 mRNA from reactivated fragile X syndrome alleles after treatment with small molecules that prevent trimethylation of H3K27. *Hum. Mol. Genet.* **25**, 3689–3698 (2016).

194. Derbis, M., Konieczny, P., Walczak, A., Sekrecki, M. & Sobczak, K. Quantitative evaluation of toxic polyglycine biosynthesis and aggregation in cell models expressing expanded CGG repeats. *Front. Genet.* **9**, 1–13 (2018).

195. Travis J. Nelsona, Jia Zhaoa, and C. I. S. Utilizing split-NanoLuc luciferase fragments as luminescent probes for protein solubility in living cells. *Methods Enzym.* **622**, 55–66 (2019).

196. Jia Zhao, Travis J. Nelson, Quyen Vu, Tiffany Truong, and C. I. S. Self-Assembling NanoLuc Luciferase Fragments as Probes for Protein Aggregation in Living Cells. *ACS Chem Biol.* **11(1)**, 132–138 (2016).

197. Nelson, T. J., Liang, S. & Stains, C. I. A Luminescence-Based System for Identification of Genetically Encodable Inhibitors of Protein Aggregation. *ACS Omega* **5**, 12974–12978 (2020).

198. Das, A. T., Harwig, A. & Berkhout, B. The HIV-1 Tat Protein Has a Versatile Role in Activating Viral Transcription. *J. Virol.* **85**, 9506–9516 (2011).

199. Moritz, B., Becker, P. B. & Göpfert, U. CMV promoter mutants with a reduced propensity to productivity loss in CHO cells. *Sci. Rep.* **5**, 1–8 (2015).

200. van den Elzen, A. M. G., Watson, M. J. & Thoreen, C. C. mRNA 50 terminal sequences drive 200-fold differences in expression through effects on synthesis,

translation and decay. *PLoS Genet.* **18**, 1–24 (2022).

201. Wright, S. E. *et al.* CGG repeats trigger translational frameshifts that generate aggregation-prone chimeric proteins. *Nucleic Acids Res.* **50**, 8674–8689 (2022).

202. Tassone, F. *et al.* Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *Am. J. Med. Genet.* **94**, 232–236 (2000).

203. De Arce, A. J. D., Noderer, W. L. & Wang, C. L. Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Res.* **46**, 985–994 (2018).

204. Fedorova, A. D., Kiniry, S. J., Andreev, D. E., Mudge, J. M. & Baranov, P. V. Thousands of human non-AUG extended proteoforms lack evidence of evolutionary selection among mammals. *Nat. Commun.* **13**, (2022).

205. Cao, X. *et al.* Non-AUG start codons: expanding and regulating the small and alternative ORFeome. *Exp Cell Res* **391**, 1–12 (2021).

206. Kearse, M. G. & Wilusz, J. E. Non-AUG translation: A new start for protein synthesis in eukaryotes. *Genes Dev.* **31**, 1717–1731 (2017).

207. Gloria A. Brar and Jonathan S. Weissman. Ribosome profiling reveals the what, when, where, and how of protein synthesis. *Nat Rev Mol Cell Biol.* **16**, 651–664 (2015).

208. Douka, K., Agapiou, M., Birds, I. & Aspden, J. L. Optimization of Ribosome Footprinting Conditions for Ribo-Seq in Human and Drosophila melanogaster Tissue Culture Cells. *Front. Mol. Biosci.* **8**, 1–12 (2022).

209. Kochetov, A. V. *et al.* AUG↔irpin: Prediction of a downstream secondary structure influencing the recognition of a translation start site. *BMC Bioinformatics* **8**, 3–9 (2007).

210. Todd, P. K. *et al.* Histone deacetylases suppress cgg repeat-induced neurodegeneration via transcriptional silencing in models of Fragile X Tremor Ataxia Syndrome. *PLoS Genet.* **6**, 1–17 (2010).

211. Tous, C. & Aguilera, A. Impairment of transcription elongation by R-loops in

vitro. *Biochem. Biophys. Res. Commun.* **360**, 428–432 (2007).

212. Reddy, K. *et al.* Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.* **39**, 1749–1762 (2011).

213. Cuozzo, C. *et al.* DNA damage, homology-directed repair, and DNA methylation. *PLoS Genet.* **3**, 1144–1162 (2007).

214. Greiner, D., Bonaldi, T., Eskeland, R., Roemer, E. & Imhof, A. Identification of a specific inhibitor of the histone methyltransferase su(Var)3-9. *Nat. Chem. Biol.* **1**, 143–145 (2005).

215. Matthew D. Disney1,*, Biao Liu1, Wang-Yong Yang1, Chantal Sellier3, Tuan Tran1, 2, Nicolas Charlet-Berguerand3, and J. L. C.-D. A Small Molecule that Targets r(CGG)exp and Improves Defects in Fragile X-Associated Tremor Ataxia Syndrome. *ACS Chem Biol.* **7**, 1711–1718 (2012).

216. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346**, (2014).

217. Park, C. Y. *et al.* Modeling and correction of structural variations in patient-derived iPSCs using CRISPR/Cas9. *Nat. Protoc.* **11**, 2154–2169 (2016).

218. Davis, B. M., Mccurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7388–7393 (1997).

219. Mauger, D. M. *et al.* mRNA structure regulates protein expression through changes in functional half-life. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24075–24083 (2019).

220. Wu, Q. *et al.* Translation affects mRNA stability in a codon-dependent manner in human cells. *Elife* **8**, 1–22 (2019).

221. Zu, T. *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 260–265 (2011).

222. Nicholas T. Ingolia1, 3, 4, Liana F. Lareau2, and J. S. W. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity of Mammalian Proteomes. *Cell* **147**, 789–802 (2011).

223. Zhang, Y. *et al.* Mechanistic convergence across initiation sites for RAN translation in fragile X associated tremor ataxia syndrome. *Hum. Mol. Genet.* **31**, 2317–2332 (2022).

224. Hong, S. *et al.* Mechanism of tRNA-mediated +1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 11226–11231 (2018).

225. Simms, C. L., Yan, L. L., Qiu, J. K. & Zaher, H. S. Ribosome Collisions Result in +1 Frameshifting in the Absence of No-Go Decay. *Cell Rep.* **28**, 1679–1689 (2019).

226. Atkins, J. F., Gesteland, R. F., Reid, B. R. & Anderson, C. W. Normal tRNAs promote ribosomal frameshifting. *Cell* **18**, 1119–1131 (1979).

227. Kozak, M. Circumstances and Mechanisms of Inhibition of Translation by Secondary Structure in Eucaryotic mRNAs. *Mol. Cell. Biol.* **9**, 5134–5142 (1989).

228. Pelletier, J. & Sonenberg, N. Insertion mutagenesis to increase secondary structure within the 5′ noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* **40**, 515–526 (1985).

229. Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**, 19867–19870 (1991).

230. Anna L Ludwig1, John WB Hershey1, and Paul J Hagerman1, 2, A. Initiation of translation of the FMR1 mRNA occurs predominantly through 5′end-dependent ribosomal scanning. *J Mol Biol.* **407**, 21–34 (2011).

231. Müller-Hartmann, H. *et al.* The human 20-kDa 5'-(CGG)(n)-3'-binding protein is targeted to the nucleus and affects the activity of the FMR1 promoter. *J. Biol. Chem.* **275**, 6447–6452 (2000).

232. Aspden, J. L. *et al.* Extensive translation of small open reading frames revealed by poly-ribo-seq. *Elife* **3**, 1–19 (2014).

233. Grünert, S. & Jackson, R. J. The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. *EMBO J.* **13**, 3618–3630 (1994).

234. Boeck, R. & Kolakofsky, D. Positions +5 and +6 can be major determinants of the

efficiency of non-AUG initiation codons for protein synthesis. *EMBO J.* **13**, 3608–3617 (1994).

235. Kozak, M. The scanning model for translation: An update. *J. Cell Biol.* **108**, 229–241 (1989).

236. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208 (1999).

237. Kearse, M. G. *et al.* Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev.* **33**, 871–885 (2019).

238. Kathleen M. Schoch and Timothy M. Miller. Antisense oligonucleotides: Translation from mouse models to human neurodegenerative diseases. *Neuron* **94**, 1056–1070 (2017).