World Journal of Pharmaceutical and Life Sciences

www.wjpls.org

Impact Factor: 8.21 Coden USA: WJPLA7

RNA TRANSLATION IN NEURODEGENERATIVE DISORDERS AN UPDATE ON THE BIOLOGY, AND CELLULAR MECHANISMS

[a]*Divyashree H. B., [b]Priya M. D., [c]Poornima R., [d]P. K. Karibasappa, [e]Kiran Kumar H. B.

[a]Department of Biotechnology, Government Science College, Nrupathunga University, Nrupathunga Road, Bangalore-560001, India.
[b]Department of Zoology and Genetics, Government Science College, Nrupathunga University, Nrupathunga Road, Bangalore-560001, India.
[c]Department of Biotechnology, Government Science College, Nrupathunga University, Nrupathunga Road, Bangalore-560001, India.
[d]Department of Studies in Food Technology, Davanagere University, Davangere-577007 Karnataka, India.
[e]Government Science College, Nrupathunga University, Nrupathunga Road, Bangalore-560001, India.

***Corresponding Author: Divyashree H. B.**

Department of Biotechnology, Government Science College, Nrupathunga University, Nrupathunga Road, Bangalore-560001, India.

DOI: https://doi.org/10.5281/zenodo.23805062

How to cite this Article: [a]*Divyashree H. B., [b]Priya M. D., [c]Poornima R., [d]P. K. Karibasappa, [e]Kiran Kumar H. B. (2026). Rna Translation In Neurodegenerative Disorders An Update on The Biology, And Cellular Mechanisms. World Journal of Pharmaceutical and Life Sciences, 12(11), 85–96.

This work is licensed under Creative Commons Attribution 4.0 International license.

Article Received on 29/09/2026 **Article Revised on 17/10/2026** **Article Published on 01/11/2026**

ABSTRACT

Secondary structures like hairpins, stem loops, and pseudoknots modify the ribosomes' binding, scanning, and recognition of start codons and control the levels of protein expression. These structures can affect translation start depending on their position. Research indicates that somatic and germline mutations, as well as a number of neurological disorders are caused due to these secondary structures. Repeat-associated non-AUG (RAN) translation, which generates dipeptide repeat proteins (DPRs), is the most prevalent genetic cause of several neurological disorders. Model species like yeast, Drosophila, cultured cells, and primary mammalian neurons have provided insights into Biological, cellular and pathological aspects of DRP. Cumulatively, the research indicates reduced cell death and/or survival. Bidirectional transcription in repeat regions of genes associated with several human neurological disorders has been documented. The identification of new proteins accumulating in the relevant tissues of the illnesses in both patients and model organisms suggest that homopolymeric or dipeptide repeat are associated non-AUG (RAN) translation. Spinocerebellar ataxia type 8, myotonic dystrophy type 1, fragile X tremor/ataxia syndrome, C9ORF72 amyotrophic lateral sclerosis, and frontotemporal dementia are few disorders associated are the disorders associated. Recently, there has been a lot of interest in using antisense oligonucleotides (ASOs) to treat the pathology. Thus this area of non-canonical translation and protein toxicity has importance both from basic science and disease perspectives. The current review is an update on the topic covering diverse areas by providing the necessary background.

KEYWORDS: 5' methyl-7-guanosine (m7G); upstream open reading frames (uORFs); TDP-43; AXTN8/ATXN8OS; antisense oligonucleotides (ASOs).

1. Basics of eukaryotic translation and relevance to RAN translation

The sequential construction of elongation-competent 80S ribosomes at mRNA start codons is known as translation initiation. The coordinated action of many eukaryotic initiation factors (eIFs) is required in this complicated process (Alberts et al., 2022). The eIF4F complex recognises the 5' methyl-7-guanosine (m7G) cap on mRNA to initiate the process (De Robertis et al., 2020). The direct cap-binding subunit eIF4E, the scaffolding subunit eIF4G, the DEAD-box RNA helicase eIF4A, and the extra helicase stimulatory factors eIF4B and eIF4H make up the eIF4F complex. The poly-adenosine-binding protein is recognised by eIF4G and binds to the 3' polyA

tail of mRNAs (Alberts et al., 2022). Circularisation of the mRNA and initiation efficiency are the outcomes of the binding (De Robertis et al., 2020). The eIF4F complex, still linked to the m7G cap, is joined by the 43S pre-initiation complex (PIC), comprising of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, and the ternary complex (in turn composed of methionine-conjugated tRNA (tRNAMet) and eIF2 (Kratzat et al., 2020). This joining of the 43S PIC to the eIF4F complex is mediated by an eIF4G-eIF3 interaction. Successful translation of most eukaryotic mRNAs requires the RNA helicase activity of eIF4A in order to resolve RNA-RNA secondary structures adjacent to the m7G cap and prepare a stage for the 43S PIC (O'Sullivan and Fraser 2023). The 43S PIC and elements of the eIF4F complex scan over the 5' UTR in the 5' to 3' direction in the ribosomal scanning model of translation initiation (Kim et al., 2024). Studies suggest eIF4A is needed at this stage to resolve weaker internal secondary structures (Yourik et al., 2017). The 43S PIC scans until it comes across an AUG codon in a favourable Kozak context, (A/G)NNAUGG (Hernández et al., 2024). Together with eIF1A, basepairing between the AUG codon and CAU anti-codon loop on tRNAMet causes eIF1 to be ejected, which raises the stringency of AUG startcodon selection (Fujita et al., 2022). The related GTPase-activating protein (GAP) helps eIF2 hydrolyse its bound GTP (Marintchev et al., 2025). At this stage, the 40S ribosome is dedicated to choosing its start codon and develops a closer bond with the substrate mRNA, which is referred to as the 48S PIC. The 60S ribosomal subunit joins the 40S subunit in the last phases of initiation (Gregory et al., 2019), most of the remaining eIFs are expelled, eIF5B hydrolyses its bound GTP (Nag et al., 2016), and translation elongation starts with the creation of the first peptide bond.

2. The role of secondary structure in translational initiation

The kinetics and control of translation initiation are greatly influenced by RNA secondary and tertiary structures (Walker et al., 2020). During translational initiation, the secondary structure of mRNA serves as a physical gatekeeper. Protein expression levels are determined by hairpins, stem loops, and pseudoknots, which alter the ribosomes' binding, scanning, and recognition of start codons (Xiang et al., 2023). Depending on where they are located, structures in the 5' UTR can have both positive and negative effects on translation initiation. It has been demonstrated that highly structured regions upstream of an AUG start codon inhibit initiation (Patrasso, et al., 2020), either by preventing the 5'-to-3' translocation of the 43S PIC when placed internally in the 5' UTR (Hinnebusch et al., 2016; Miras et al., 2017) or by blocking the eIF4E-m7G interaction when placed adjacent to the cap. On the other hand, secondary structures downstream of start codons enable initiation at faulty start codons, including non-AUG codons and inadequate Kozak context (Dever et al., 2023). According to research, initiation at non-AUG codons happens less efficiently than at AUG codons in *in vitro* translation systems (Ivanov et al., 2017). However, initiation is significantly increased when secondary structures are present downstream. The process of secondary structure and its functions in translation have been better understood due to recent developments in ribosomal foot printing techniques. Research indicates that a common but previously unknown set of beginning events *in vivo* may be reflected in *in vitro* results. According to Fedorova et al. (2020), there is more evidence of thousands of unexpected translation initiation events, many of which take place at non-AUG codons. This is particularly true for small ORFs upstream of canonical, annotated ORFs in the same mRNA transcript, known as upstream open reading frames (uORFs) (Takahashi et al., 2020). *In vivo* initiation at non-AUG codons appears to play significant regulatory functions in global protein translation and may be partially regulated by transacting eIFs and mRNA-specific cis factors (Andreev et al., 2022). Although the standard technique mentioned above is probably used to translate most eukaryotic mRNAs, research also indicates that there are other atypical pathways. Furthermore, it is observed that mRNA secondary structure influences atypical modes of initiation in a manner comparable to conventional initiation (Tidu and Martin 2023). An internal ribosome entry site (IRES)-mediated pathway is used to translate a variety of viral and cellular RNAs (Marques et al., 2022). IRESes are intricate RNA structures that directly attract eIFs and ribosomal subunits to internal locations in RNA transcripts. Translation initiation avoids the eIF4E-m7G interaction, despite their extreme structural and functional heterogeneity (D'Agostino et al., 2025). Some types of IRESes require the majority of eIFs, whereas others only need the 40S ribosomal subunit and an alanine-conjugated tRNA (Lozano et al., 2018). For instance, the cricket paralysis virus (CrPV) codes for a protein with an N-terminal alanine and uses CCU as an initiation codon (Muhs et al., 2015). IRESes are also present in a number of eukaryotic mRNAs, such as c-Myc, p53, Bcl-2, and others (Sharma et al., 2016). Under a variety of stress circumstances where canonical translation is impeded, translation of these proteins is sustained, suggesting a special mechanism that avoids this general inhibition. The *histone 4* gene mRNA, which seems to be translated via a ribosomal tethering mechanism, provides a second illustration of non-canonical forms of initiation (Hayek et al., 2021). The Histone 4's 5' UTR is shorter than other 5' UTRs (the mouse homolog is nine nucleotides).

The recruitment of the eIF4F complex to a structural element within the ORF is the first step in the translation of histone 4 mRNA. Next, the eIF4F binds to the m7G cap, which is buried within a neighbouring structural element (Martin et al., 2016). After being recruited to eIF4F, the 43S PIC is subsequently moved straight to the AUG start codon.

Research suggests that hairpin formation may have an impact on the effectiveness of translation. Additionally, hairpin formation downstream of an AUG at particular sites results in an increase in translation rate, while hairpins with high thermal stability upstream of the translation start site limit translation by up to 85–95% (Weenink et al., 2018). According to a study by Solodushko and Fouty (2023), there may be a hairpin structure close to the 5' end of the Lc mRNA that suppresses downstream translation in the transient transformation experiment and the rabbit reticulocyte *in vitro* translation system. The existence of RNA secondary structures in the 5′UTR is a significant factor in determining the rate of translation since RNA molecules fold in complicated configurations (Georgakopoulos-Soares et al., 2022). The iron response element (IRE) in the 5′ UTR is a stem-loop structure that best demonstrates the mechanism. An iron-binding protein attaches itself to this hairpin when iron levels are low. As a result, the cell produces no ferritin and the ribosome is prevented from scanning (Wang et al., 2021).

Additionally, the distance between translation-associated motifs, such as the stem-loop and the cap, can be altered by the development of secondary structures (Paek et al., 2015). A relative depletion of G-quadruplexes at the RNA level is suggested by the enrichment of RNA G-quadruplexes at 5′UTRs (Lee et al., 2024), where they exhibit a higher frequency at the template strand. G-quadruplexes in the 5′UTR of mRNAs have been demonstrated to be inhibitory elements, and other studies have demonstrated that they obstruct ribosome recognition (Roy et al., 2024). Additional research also shows that G-quadruplexes at 5′ UTR of eukaryotic genes can enhance translation by favouring IRES recognition. A G-quadruplex motif and two hairpin sequences are present in the IRES of the *FGF-2* gene, which is linked to tissue development and repair, and they facilitate translation in a cap-independent translational program (Zhang et al., 2017). According to a number of earlier research human illnesses frequently include areas that are prone to the creation of secondary structures. Germline and somatic mutations are over-represented in G-quadruplexes (Zhuk et al., 2024). Several neurodegenerative diseases have been linked to the growth of microsatellite repeats at 5′UTRs, which has been linked to abnormal translation. For instance, the hexanucleotide GGGGGC in the first intron of the *C9orf72* gene causes amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (Mizielinska et al., 2025).

We collated literature from online databases Pubmed, Sciencedirect, Google scholar restricting literature from 215 to 2026 with few exceptions. We also used several published reviews from the year 2020-2026. We did not include any specific search criteria or exclude any keyword. The literature was critically evaluated for relevance, quality, and contribution to the topic.

3. Bidirectional transcription and RAN translation

In addition to non-AUG translation beginning at enlarged CAG, CGG, and GGGGCC (G4C2) repeats, bidirectional transcription through repeat regions of genes linked to multiple human neurological illnesses has been reported (Banez-Coronel and Ranum 2019). A homopolymeric or dipeptide repeat repeat-associated non-AUG (RAN) translation is suggested by the discovery of novel proteins accumulating in the pertinent tissues of the disorders in both patients and model organisms (Cleary et al., 2018). Antisense transcripts can be found for many mammalian genes, including the wild-type alleles linked to microsatellite expansion disorders, according to recent assessments of genome-wide transcription (Santos et al., 2022). The matching antisense transcript, which is noncoding and only partially overlaps the sense transcriptional unit, usually in the promoter and 3′ terminal regions, is typically found at a lower copy level than the primary sense strand, which is typically protein coding. It has been proposed that antisense transcription controls gene expression in a number of ways, such as (i) transcriptional interference of the sense strand due to possible collisions between RNA polymerase II holoenzyme complexes transcribing opposite strands; (ii) modulation of gene expression by recruiting different chromatin remodelling complexes that mediate DNA methylation and histone modifications; (iii) masking splice and polyadenylation sites of pre-mRNAs transcribed from sense strands, resulting in alternative splice and poly(A) site selection; (iv) duplex formation between sense and antisense transcripts, which modifies RNA interference, stability, and translation leading to RNA editing (Wight and Werner 2013; Werner, 2014). By changing the relative amounts of sense vs antisense transcripts or by influencing transcriptional start or termination sites, microsatellite repeat expansion may interfere with these regulatory processes (Pinto et al., 2017). It is unknown whether antisense RNAs are transcribed through the enlarged repeat regions of most unstable microsatellite disease genes. One example is SCA8, a dominant SCA with less penetrance, where bidirectional transcription has been shown to produce a highly pure polyQ protein called ataxin 8 from ATXN8 and toxic noncoding RNAs from the antisense transcriptional unit (ATXN8OS) (Kim et al., 2013). Both protein-mediated (polyQ intranuclear inclusions) and RNA-mediated (RNA foci formation and mis-splicing) events have been reported in SCA8. Since the only detectable RNAs were noncoding ATXN8OS transcripts with a CUGexp near the 3′ end, SCA8 was initially believed to be an exclusively RNA-mediated disease, similar to DM1, in contrast to other forms of spinocerebellar ataxia that contain CAGexp in the coding region of the sense strand (Hu, et al., 2017). Richagneux et al. (2026), report that transcription produces CUGexp RNAs that efficiently capture the MBNL1 splicing factors and modify alternative exon usage during RNA processing to favour the synthesis of foetal isoforms. These CUGexp RNA–MBNL1 complexes build up in

nuclear RNA foci, which could be a harmful component or just a defence mechanism (Wang et al., 2018). Numerous studies suggest that this kind of CUGexp toxicity mechanism may also be responsible for SCA8. Loss-of-function mutations in the *muscleblind* gene, the fly homologue of *MBNL1*, exacerbate the progressive retinal neurodegeneration caused by transgenic expression of either the wild type or an enlarged human SCA8 gene in Drosophila (Mutsuddi et al., 2004). Additionally, both SCA8 patients and transgenic mice that express a human AXTN8/ATXN8OS gene with a (CTG)116 expansion mutation (SCA8 BAC-exp) have CUGexp foci in their cerebellar cortex (Purkinje cells, Bergmann glia, molecular layer interneurons) (Daughters et al., 2009). Additionally, MBNL1-regulated alternative splicing of GAT4 pre-mRNA has been associated with increased levels of the GABA transporter GAT4/GABT4 in both human SCA8 and mouse SCA8 BAC-exp brains (Nietz et al., 2025). These findings support a scenario in which MBNL1 proteins are trapped by mutant ATXN8OS CUGexp transcripts, leading to the mis-splicing of MBNL target RNAs that encode proteins crucial for GABAergic inhibition. Remarkably, this RNA is projected to encode a nearly pure polyQ protein, in contrast to other CAGexp expansion disorders. Purkinje cells in human SCA8 and SCA8 BAC-exp mice brains and brain stems exhibit cerebellar polyQ inclusions, identified by monoclonal antibody (mAb) 1C2, which recognises extended polyQ tracts. This observation suggests that antisense transcription takes place via these microsatellite repeats (Owens et al., 2015). Although pure polyQ proteins appear to be extremely toxic, it is unclear whether the levels of ATXN8 transcription and polyQ protein, which cannot be detected by protein blot analysis, are sufficient to cause cellular dysfunction for sense transcription. As a result, susceptible cell populations may be quickly eliminated (Kumar et al., 2023). Expression of CUGexp RNAs for antisense transcription results in loss of cerebellar GABAergic inhibition and RNA gain-of-function effects on RNA splicing (Davenport and Swanson 2025). However, as the SCA8 BAC-exp transgenic animal does not replicate these neurodegenerative alterations, the biological cause of the cerebellar shrinkage seen in SCA8 is still unknown.

The fact that another gene *KLHL1*, is also transcribed antisense to the ATXN8OS gene and that the 5′ ends of these genes overlap in a head-to-head arrangement presents another challenge (Echeverria and Cooper 2012). The limited penetrance linked to SCA8 is one of its perplexing characteristics; some unaffected people have ATXN8/ATXN8OS alleles with up to 1000 CTG repeats (Aydin et al., 2018). It is conceivable that bidirectional transcription via repetitions causes the genes at this locus to be silenced, greatly reducing the amounts of harmful RNAs and proteins. Nevertheless, heterochromatin production at this location has not been reported. The pathophysiology of FXTAS may possibly involve antisense transcription (Derbis et al., 2021).

Through the CCG repeat region, a promoter found in *FMR1* gene intron 2 induces the production of the antisense *ASFMR1* gene. Blood cells from FXTAS patients have higher levels of ASFMR1 transcript, which is spliced, polyadenylated, transported to the cytoplasm, and may be translated into a polyproline protein using an initiation codon upstream of the CUGexp, though there is no experimental proof (Zafarullah et al., 2023). ASFMR1 RNA is coded into an alternatively spliced form and the CGGexp mutation modifies the splicing pattern. *FMR4*, another noncoding gene, starts upstream of the FMR1 start site (Huang G, et al., 2019). Given that both transcriptional units are increased in FXTAS patient leukocytes but muted in fragile X, these two genes seem to share a bidirectional promoter.

4. RAN translation and cellular roles

Recent findings in the field of repeat expanded disease genomics question the prevailing hypothesis that the expansions of CAG, CGG, and G4C2 repeats found in non-coding regions do not encode any proteins. Also the observation that CAG repeat mutations found in polyGln ORFs express proteins only in one frame stands challenged in-light of the latest discoveries. The result of several studies suggest that the novel proteins produced by frameshifting and RAN translation may be widely distributed and play a role in the pathophysiology of neurological diseases (Guo et al., 2021). The cap-dependent scanning mechanism and cap-independent internal ribosome entry site (IRES) procedures are the most common translation initiation mechanisms that explain RAN translation (Haimov et al., 2015; Yang and Wang 2019). However, there are also additional known initiation mechanisms, like leaky scanning, ribosome shunting, or reinitiation (Pooggin and Ryabova 2018). Since the development and stability of the RNA structures created by the repeats have been demonstrated to influence the amount of RAN translation products, these structures must be taken into account when interpreting mechanisms of RAN translation (Wojciechowska, 2014). By putting the translation elongation machinery under strain, repeat expansions have also been shown to have an impact on the maintenance of the reading frame. During translation over extended CAG repeats of *ATXN3* and *HTT* gene transcripts, frameshifting by one nucleotide upstream or downstream produces novel proteins that contain polySer (HD) and polyAla (SCA3 and HD) (Akçimen et al., 2020). The physiological details of this process are still unclear, although such ribosome slippage most likely happens at random locations along the enlarged CAG repeats in tissues and cells.

However, the frequency of frameshifting appears to be dependent on the repetition creation of hairpin structures and rises with the amount of CAG repeats (Wright et al., 2022). The resulting chimeric mutant proteins contribute to the pathogenesis by building up in cytoplasmic and intranuclear aggregates that are harmful to cells.

Thus far, repeats in noncoding regions of a few genes, such as *ATXN8* and the 3'UTR of *DMPK* and the 5' UTRs of *FMR1* and *C9ORF72*, have been shown to produce homopolymeric or dipeptide repeat RAN-translated proteins (Kearse and Todd, 2014). When attempting to decipher the mechanism of RAN translation initiation, the RNA structures created by the repeats are identified as possible triggers (Mariotti et al., 2017). The quantity of RAN-translated products is impacted by their creation and stability. Repeats that undergo RAN translation are known to generate G-quadruplex or hairpin structures, which have distinct molecular stability and architecture (Wang et al., 2019). According to Kiliszek and Rypniewski (2014), the crystallographic structures of CNG repeat duplexes (N = A, U, G, and C) exhibit distortions of their double-stranded helices due to noncanonical N–N base pairs, suggesting both similarities and variations in protein binding by repeat hairpins. Furthermore, research has shown that repeat flanking sequences may contribute to hairpin stability in several repeat-containing transcripts, including as HTT/HD, AR/SBMA, and CACNA1/SCA6 (Heinz A, et al., 2021). Also, the repetitive tract interruptions have a significant impact on the stability of the hairpin form they adopt (Krzyzosiak et al., 2012).

The C9ORF72 antisense and sense transcripts exhibit different RAN translation efficiency, with C4G2 repetitions adopting hairpin structures and G4C2 repeats possibly folding into more stable G-quadruplex structures (Liu, et al., 2022). The presence of modifying elements that could either stimulate or hinder RAN translation was discovered using *in silico* structure prediction of enlarged repeats and their flanking regions. Extra hairpin stabilisation by repeat flanking sequences, such as extra (CGG)n hairpin extension by S-FMR1 flanking sequences (AUG START codons and different AUG-like codons (CUG, GUG, UUG, AUA, ACG, and AUU), are among the stimulatory factors (Lee et al., 2024). In contrast, STOP codons (UAG, UAA, and UGA) in various reading frames may be inhibitory factors (Loughran et al., 2023). Both sense and antisense C9 transcripts undergo RAN translation despite these variations in the frequency of possible stimulatory and inhibitory elements of RAN translation; the sense transcript experiences the phenomena more intensely (Zu et al., 2013). Regardless of whether the repeats are found in noncoding or coding regions of the disease-causing genes, different representations of modifying factors that may potentially stimulate or inhibit RAN translation have been found in a few other transcripts with expanded repeats linked to human neurodegenerative diseases.

5. Dipeptide repeat proteins (DPRs) and insights from model studies

The most common hereditary aetiology of a number of neurological illnesses is repetition-associated non-AUG (RAN) translation, which produces dipeptide repeat proteins (DPRs) (Schmitz et al., 2021). The hippocampus, cerebellum, neocortex, and thalamus have the highest burden of DRPs, despite the fact that their distribution across the brain varies greatly (Schludi, 2015). PolyGA appears to be the most abundant DRP in a number of brain regions evaluated using many different antibodies followed by polyGP and polyGR. The DRPs formed only from antisense transcripts (polyPA and polyPR) tend to be the least abundant. As shown in a number of other neurodegenerative proteinopathies, DRP inclusion formation is neuroprotective while the soluble DRPs oligomers cause toxicity (Hao et al., 2025; Tolar et al., 2021). As an alternative, Edbauer and Haas (2016) suggests a "amyloid-like" mechanism of toxicity for the DRPs, whereby the buildup of DRPs sets off a series of processes that cause TDP-43 to mislocalize and aggregate in neurons that are specifically sensitive. Five different DPR species are produced when translation takes place throughout all sense and antisense transcript reading frames (Wilson et al., 2019). The most prevalent DPR species is Poly-GA (glycine-alanine), which causes moderate toxicity that is dependent on dosage and duration and forms cytoplasmic aggregates (Steffke et al., 2024). According to studies, neither Poly-GP (glycine-proline) nor Poly-PA (proline-alanine) exhibits any appreciable toxicity (Jafarinia et al., 2020). Specifically, glycine-alanine (GA) and glycine-arginine (GR) DRPs are generated from the sense strand (Nelson, et al., 2024), while the proline-alanine (PA) and proline-arginine (PR) arise from the antisense strand (Hall et al., 2026). RAN translation of both strands produces the two proteins that contain glycine-proline (GP) (Al-Turki et al., 2026). Phase separation properties are altered by arginine-rich DPRs (poly-PR and poly-GR), which interact improperly with cellular RNA and disturb biomolecular condensates (Park et al., 2023). DPRs cause endoplasmic reticulum (ER) stress, stress granule formation, and global translation suppression by overwhelming protein degradation mechanisms, including as the ubiquitin-proteasome system, autophagy, and lysosomes (Verde et al., 2026; Smeele et al., 2024). The pathogenic effect of the DPR is determined by the compartment in which it builds up. Nucleocytoplasmic protein trafficking is hampered by cytoplasmic poly-GA clumps. Research indicates that RanGAP, a regulator of nucleocytoplasmic transport, is directly impacted by GGGGCC repeats (Freibaum et al., 2015). This implies that both the repeat-containing RNA and the arginine-containing DRPs are involved in this manner of toxicity. In contrast to solely cytoplasmic aggregates, nuclear or nucleolar localisation typically causes severe problems in transcription and translation (Yang et al., 2015). Protein production and nucleolar protein quality control are hampered by nuclear poly-GA, which is more hazardous (Frottin et al., 2021). Studies *in vitro* and *in vivo* have demonstrated that DRP expression alone can cause neurodegeneration, despite the fact that their distribution throughout the brain thus begging the question of their precise role in disease. The insights into DRP investigations have been made possible by model species such as yeast, Drosophila, and through cultured cells and primary mammalian neurons.

Figure –1 illustrates the repeats and proteins/histology identified from patients affected with Neuro degenerative diseases. Studies collectively point to decreased survival and/or cell death (White et al., 2019). Research indicates that RNA foci are formed by transgenic flies that express different lengths of GGGGCC repeats with stop codons in all three reading frames (He et al., 2020). However, flies with pure repetitions experience severe cell death and create polyGR and polyGP proteins (Yuva-Aydemir et al., 2018). When overexpressed in mouse models, both polyGR and polyPR create intranuclear aggregates that damage the nucleoli (Cook et al., 2020). However, pathological investigation indicate patient brain tissue does not contain nucleolar DRPs (Manal et al., 2017). PolyGR proteins are drawn into cytoplasmic polyGA inclusions when co-expressed with polyGA (Radwan et al., 2020), indicating that nucleolar stress may not be a major cause of toxicity in patients. Alternatively, Cicardi et al. 2023 identified nucleolar polyPR-positive inclusions in spinal cord tissue from C9 ALS patient, and suggest that the high toxicity of nucleolar polyPR results in increased vulnerability of neurons. Studies in flies and yeast suggest that PolyGR proteins can also mediate toxicity through impairment of the Notch pathway (Papsdorf et al., 2015). Although comparatively less toxic than the DPR the non-arginine-containing DRP proteins when expressed in isolation also appear important to neurodegeneration as evidenced from culture studies (). Adult flies expressing exclusively GA-100 DRPs within neurons demonstrate significantly reduced survival (Catterson et al., 2023), and expression of polyGA in primary mammalian neurons causes increased toxicity through impairment of the ubiquitin-proteasome system (Angelopoulou et al., 2023), induction of ER stress (Bobak et al., 2015), and sequestration of Unc119, a trafficking protein (Fehlhaber et al., 2023). In *C9orf72* ALS/FTD models, toxic gain-of-function effects stem from both the translated DPRs (Beckers et al., 2024) and the accumulation of repetitive G_4C_2 RNA foci (Frottin, et al., 2021), which often act additively or synergistically to compromise cellular health.

6. RNA translation in neurodegenerative diseases: CGG repeats in FXTAS; translation at GGGGCC/GGCCCC repeats and Huntington disease (HD).

FXTAS is a late-onset neurodegenerative illness caused by the amplification of CGG repeats in FMR1's 5' untranslated region (Ishiura, et al., 2025). The number of repeats in unaffected people is fewer than 45. Whereas in FXTAS it is 55 to 200 repeats, this is known as "premutation" range. Premutation repeat expansions promote transcription of FMR1 mRNA containing these repeats (Dischler et al., 2025). In contrast, expansions with more than 200 repeats cause transcriptional silence of the FMR1 gene, resulting in the loss of FMR1 mRNA and the Fragile X protein, FMRP (Brykczynska et al., 2016). Clinically, FXTAS is distinguished by action tremors, ataxia, parkinsonism, and cognitive impairment, while pathologically by both neuronal and non-neuronal ubiquitinated inclusions in the cerebral cortex, brainstem, and cerebellum is observed (Shickman et al., 2018). CGG-expanded FMR1 5' UTR facilitates RAN translation initiation (Malik, et al., 2021). In the absence of an AUG start codon, initiation within the 5' UTR occurs in at least two reading frames: the GGC (+1) frame that generates a polyglycine product (FMRpolyG) (Niewiadomska, et al., 2026), and the GCG (+2) frame that yields a polyalanine product (FMRpolyA) (Linsalata, et al., 2019). FMRpolyG accumulates in ubiquitinated inclusions in human tissue and cell lines (Friedman-Gohas et al., 2020), in toxicity, cellular models (Buijsen et al., 2014) and animal models (Sellier, et al., 2017). FMRpolyG-positive inclusions have been discovered in ovarian stromal cells in FXTAS mice models (Buijsen et al., 2015) and in FXPOI patient (Friedman-Gohas et al., 2020) implying that FMRpolyG expression is associated with other Fragile X-related clinical characteristics. The creation of a FMRpolyG-green fluorescent protein (GFP) fusion reporter construct with 3', 5', or 88 CGG repetitions suggests that RAN translation can occur in the absence of pathogenic expansions (Rodriguez and Todd 2021). Inserting a stop codon exactly upstream of the CGG repeats prevents production of FMRpolyG reporters, demonstrating that RAN translation of FMRpolyG begins upstream of the repeats (Todd et al., 2013). Further mutational study demonstrates that initiation in this frame can occur at several upstream near-AUG codons in the human 5'UTR (Chaldebas, et al., 2026). This shows that RAN translation can begin in the GCG frame within the repetitions. These findings suggest the fascinating prospect that RAN translation of the same sequence may differ mechanistically across reading frames. Rresearch evidences implicate FMR1 5' UTR has IRES activity (Dobson et al., 2008). Inserting the FMR1 5' UTR between two ORFs in a plasmid-based bicistronic reporter does not prevent translation of the second ORF (Ludwig et al., 2011). Reporter experiments show that monocistronic reporters with the 5' UTR of FMR1 were less cap-dependent than reporters with the 5' UTR of β-globin, as expected if IRES was used (Cheng W, et al., 2018). According to studies, a considerable fraction of FMRP is translated using a cap-dependent process, emphasising the importance of the eIF4E-m7G interaction in both cells and *in vitro* translation systems (Borden and Volpon 2020). RAN translation of CGG-expanded FMR1 reporters is cap-dependent across many repeat reading frames. Furthermore, presence of sufficient free m7G cap, which binds to and sequesters eIF4E, prevents CGG RAN translation of both the FMRpolyG and FMRpolyA reporters is reported (Reyes et al., 2023). *In silico* modelling and *in vitro* investigation reveals that consecutive CGG repeats create a stable hairpin shape (Kiliszek and Rypniewski, 2014) which has a considerable impedance to scanning 43S PICs (Tutak, et al., 2024). As with ribosomal scanning increasing the length of CGG repeats lowers the expression of a downstream AUG triggered reporter

(Kearse et al., 2016). Further evidence for a scanning mechanism comes from the observation of initiation at a near-AUG codon in an artificial hairpin put in the 5' UTR in place of the CGG repeats (Rovozzo et al., 2016). Secondary structures increase initiation at optimally positioned non-AUG codons by causing scanning 43S PICs to stall. It is plausible that a similar mechanism may be used by RAN translation on CGG-expanded FMR1 mRNA to produce FMRpolyG (Zhang et al., 2020). Commencement of CGG RAN translation requires mRNA secondary structures. Accordingly, the relative effectiveness of RAN translation in both the polyglycine and polyalanine reading frames in RNA transfected cells increases with the quantity of CGG repetitions in reporter constructs (Rodriguez et al., 2020).In the glycine reading frame, this is accompanied by a reduction in the requirement for any particular near-AUG codon upstream of the repeat, indicating compromised start codon selection fidelity (Boivin et al., 2026). The m7G cap on FMR1 mRNA is bound by the eIF4F complex and 43S PIC (Kapur M,et al., 2017) this complex then moves downstream through the 5' UTR until it comes across secondary structure made up of either the surrounding intrinsic sequence of the 5' UTR or CGG repeats. FMRpolyG and FMRpolyA are produced when ribosomal stalling causes aberrant translation initiation at nonAUG codons either upstream of or within the repeat in the +1 and +2 frames (Zhang et al., 2022).

The common known cause of ALS and FTD is the C9orf72 GGGGCC/GGCCCC hexanucleotide repeat expansions (Mizielinska et al., 2025). ALS is the most common type of motor neuron disease, affecting about 2–4/100,000 people (Liščić 2015), and is characterised by progressive paralysis. About 20 out of every 100,000 people between the ages of 45 and 65 suffer from FTD, and it is the second most prevalent kind of presenile dementia. The three clinical syndromes of FTD are behavioural variants, semantic dementia, and progressive nonfluent aphasia (Liscic 2020). Furthermore, the majority of ALS patients and the most prevalent form of FTD have TDP-43-positive inclusions in their neurons and glia (Babazadeh, et al., 2023). The GGGGCC repeat, which is found in the promoter region of isoform 2 and the first intron of transcript isoforms 1 and 3, is enlarged in C9 ALS/FTD patients from 2–25 repetitions in healthy persons to over a thousand repeats (Balendra and Isaacs 2018). In mutation carriers, both the sense and antisense strands of C9orf72 are transcribed, producing GGGGCC and GGCCCC-repeat carrying RNAs (Błaszczyket al., 2024). With the sense RNA repeat producing a G-quadruplex and hairpin *in vitro* (Wang et al., 2018*)* and the antisense RNA repeat recently demonstrated to take an A-form-like double helix *in vitro* (Castro et al., 2020) both of these enlarged repeat sequences are expected to generate extremely stable RNA secondary structures. RAN translation at GGGGCC/GGCCCC repeats in a 5' leader context depends on repeat length; with longer repeats resulting in robust DRP generation (He et al., 2020). Additionally, research suggests that the repeat-length required for initiation varies depending on the reading frame and sequence context. Immunocytochemistry (ICC) has identify all DRPs with as little as 30 or 40 repeats when GGGGCC and GGCCCC repeats are positioned downstream of a synthetic sequence (Quaegebeur, et al., 2020). PolyGA is produced similarly with as little as 38 repeats when GGGGCC repeats are positioned downstream of 113 nucleotides from intron1 (Castelli, et al., 2021). While polyGP and polyPR are found in cells expressing 66 GGCCCC repetitions downstream of 99 native nucleotides (Thompson et al., 2024), antisense DRPs likewise demonstrated distinct length requirements for detection when positioned downstream of original sequence. There are still a number of unresolved questions about the mechanism of RAN translation at GGGGCC and GGCCCC repetitions. The precise RNA species that undergo RAN translation in patients with C9 repeat expansion are still unclear. RAN-translated GGGGCC repeats in patients may originate from aberrant disease-specific transcripts produced by transcriptional pausing (Jazurek-Ciesiolka, et al., 2020), a retained intron (Niblock et al., 2016), or a spliced intron in a lariat (Miller et al., 2025). The delineation of the role of repeats in fly models implicate lower RAN translation and toxicity when compared to repeats inserted into a 5' leader sequence, indicating that spliced lariats containing GGGGCC repeats are not be effectively used to generate DRPs (Jiang, et al., 2022). When Drosophila reared at high temperatures, DRP synthesis from the intronic repeat becomes sufficient to cause toxicity, suggesting that an intronic environment can allow pathological RAN translation under specific circumstances (Nitta and Sugie 2022).

A CAG-repeat expansion in the first coding exon of the Huntingtin gene, *HTT* causes Huntington disorder (Ayyildiz et al., 2023). Normal-sized repeats are usually in the 20s, with 100% penetrance at repeat sizes of 40 or above and a minimal threshold of disease as more than 35 (Arning and Nguyen 2022). This annotated ORF's CAG repeat codes for a polyglutamine stretch that can be broken down into a smaller peptide by protease cleavage or alternative splicing (Stoyas and La Spada 2018). Numerous studies indicate that extensive (often more than 60) CAG repeats alone or in conjunction with HTT exon 1 are adequate to cause toxicity in illness models in mice and flies (Ast et al., 2025; Martí 2016). Evidence, however, also points to the native functions of the Huntington protein as a contributing factor to the molecular and clinical characteristics seen in some HD mice models and in individuals with Huntington's disease. Krans, et al. 2025 study offers a strong proof of the RAN translation in Huntington's illness. The study found notable staining in the striatum and cerebellar tissues in the brains of Huntington disease patients and in a mouse model of the disease using antibodies produced against the anticipated C-terminal regions. RAN-translated proteins polyleucine and polycysteine products are produced from an antisense strands created by the

repeat in the CUG orientation (Wojciechowska et al., 2014). Further, it is observed that RAN translation in nearby frames from CMV-driven, plasmid-derived transcripts is unaffected by the presence of an AUG in the polyglutamine (Goodman and Bonini N2019). The repeat sequence itself and/or its secondary structure may be crucial for RAN initiation, as no RAN translation was seen from constructs with CAA repeats or when the codon used to create a homopolymeric protein was changed throughout the sequence (Glineburg, et al., 2018). The results at the HTT locus are internally consistent with earlier research on the translation of CAG RAN in Spinocerebellar Ataxia type 8 (SCA8). In the typical sequence context of SCA8, an AUG initiation codon in ATXN8 is located just one codon above the repetition (Sakamoto et al., 2026). Also, the translation of the polyglutamine-containing protein is not stopped by mutating that start codon to AAG (Perez, et al., 2021). Studies reveal that RAN translation of products containing alanine and serine are not prevented by inserting an AUG and V5 tag above the repeat or placing a stop codon just upstream of the repeat in the glutamine reading frame (Wright et al., 2022). Similar to the HD sequence context, RAN translation was prevented in all reading frames when the CAG repeat in the SCA8 leader sequence was changed to CAA. However, RAN polyglutamine translation is not supported when the upstream sequence from another CAG repeat expansion disorder, Huntington Disease Like 2 (HDL2) is included. Whereas, Polyalanine expression, on the other hand, are consistently preserved in both sequence contexts. Regardless of the upstream sequence, attempts to replicate CAG RAN translation in an *in vitro* rabbit reticulocyte lysate system showed a total absence of polyalanine synthesis (Krans et al., 2015). Furthermore, additional AUG start codons close to the cognate start codon was necessary for the *in vitro* translation of polyglutamine or polyserine products from these constructs, similar to what was observed for RAN translation at CGG repeats (Scoles et al., 2015). However, in cells transfected with CAG repeats, this requirement for a close AUG codon was eliminated (Malik et al., 2021). Thus, it seems that the upstream sequence, repeat length, and cellular context affect RAN translation efficiency for the CAG repeat in a variety of disease-specific sequence settings. Future research employing quantitative techniques will be required to answer this question empirically, even if there are indications that similar sequence-dependent effects might exist in Huntington's disease. An N-terminal methionine is mostly used for initiation at non-AUG codons *in vitro* (Tharp et al., 2021). Methionine would continue to be the first amino acid of RAN products if the commencement of RAN translation proceeds in a similar manner. However, the available data on this is conflicting. Since N-terminal methionines are frequently proteolytically removed and replaced by an acetyl group, it is uncertain if the N-terminal amino acid used at start is methionine or alanine, which is used in IRES-based translation (Lee et al., 2016). Since several years frameshifting from the polyglutamine reading frame into the polyalanine or polyserine reading frames (Hu and Ng 2012) has been proposed as a possible mechanism underlying some features of neurotoxicity in animal models of Huntington disease and SCA3. Mass spectroscopy and 2D electrophoresis were employed to show that at least some of the HTT polyalanine product is the consequence of -1 frameshifting (Girstmair, et al.., 2013). On the other hand, sequence of AUG-driven constructs that were amino and carboxy-terminally tagged to imitate exon 1 of Huntington disease in transfected cells, no frameshifting products was seen (Wieland et al., 2026).

Table 1: List of few Human diseases with bidirectional transcription, repeats and pathologic mechanism involved.

Disease	Gene/Repeat	Pathologic mechanism	Sense/anti-sense reads	Reference
Dentorubral pallidoluysian atrophy (DRPLA)	DRPLA (CAG)3 – 36/49 – 88	Protein GOF (polyQ)	741/27	Batra, et al., 2010
Huntington disease-like 2 (HDL2)	JPH3 (CAG)6 – 28/40 – 59	RNA GOF (CUGexp)	31/188	
Myotonic dystrophy type 2 (DM2)	ZNF9 (CCTG)10 – 26/75 to 11 000	RNA GOF (CCUGexp)	2091/0	
Spinal and bulbar muscular atrophy (SBMA)	AR (CAG)9 – 36/40 – 55	Protein GOF (polyQ)	103/3	
Spinocerebellar ataxia type 2 (SCA2)	ATXN2 (CAG)13 – 33/34	Protein GOF (polyQ)	759/13	
Spinocerebellar ataxia type 6 (SCA6)	CACNA1A (CAG)4 – 18/20 – 29	Protein GOF/LOF (?)	242/48	
Spinocerebellar ataxia type 10 (SCA10)	ATXN10 (ATTCT)10 – 29/800 – 4500	Unknown	1560/9	
Spinocerebellar ataxia type 31 (SCA31)	TK2 (TGGAA)0/≥110 RNA	GOF (UGGAAexp)	412/4	
Oculopharyngeal muscular dystrophy (OPMD)	PABPN1 (GCG)10/12 – 17	Protein GOF (polyA)	292/1	

Table 2: A list of RAN proteins identified in neurological diseases.

Transcript	RAN protein	Protein accumulation in vivo	Reference
FXPOI Sense	PolyGly	FXPOI RAN PolyGly accumulates in the ovarian stromal cells of a FXPOI patient and FXTAS mice. Mice show and increase number of primordial follicules at 40 weeks and polyGly aggregates in the pituitary gland. PolyGly accumulates as intranuclear inclusions that are ubiquitin-positive	Banez-Coronel and Ranum 2019
SCA8 Sense	PolySer	Detected in the cerebellum of SCA8 patients and in the cortex, hippocampus, brain stem, and cerebellum of SCA8 BAC transfenic mice. In mice, polySer accumulates as dot-like agreggates in an age-dependent manner that correlates with the mouse motor dysfunction phenotype. PolySer preferentially accumulates in white matter regions that show demyelination and axonal degeneration	
SCA8 Antisense	PolyAla	Accumulates in the cerebellum of SCA8 mice and in the cerebellum of SCA8 patients. RAN SCA8 PolyAla is detected in the soma and dendrites of Purkinje cells	
DM1 Antisense	PolyGln	Accumulates as nuclear aggregates in DM1 cardiac myocytes and leukocytes. PolyGln inclusions are also found in myoblasts and skeletal muscle at low frequency. DM1 mice carrying 55 CTG repeats show polyGln staining in leukocytes in cardiac tissue	
DM2 Sense	LPAC	Primarily accumulates as small cytoplasmic or perinuclear aggregates in neurons, but also occasionally in glial cells, in brain regions showing abundant macrophages. Cortex, hippocampus, and striatum are positive for LPAC accumulation	
SCA31 Sense	WNGME	Detected in the eye imaginal discs of a SCA31 Drosophila model and in the cerebellum of SCA31 patients, where it accumulates as dot-like granules in the cell bodies and dendrites of Purkinje cells. In flies, WNGME levels correlate with increased eye degeneration	
FECD Sense	PolyCys	Detected as nuclear immunostaining in the corneal endothelium of FECD patients carrying ≥55 repeats	
FECD Antisense	PolyGln	Detected by western blot in fibroblast derived from FECD patients that are homozygous for the repeat expansion	

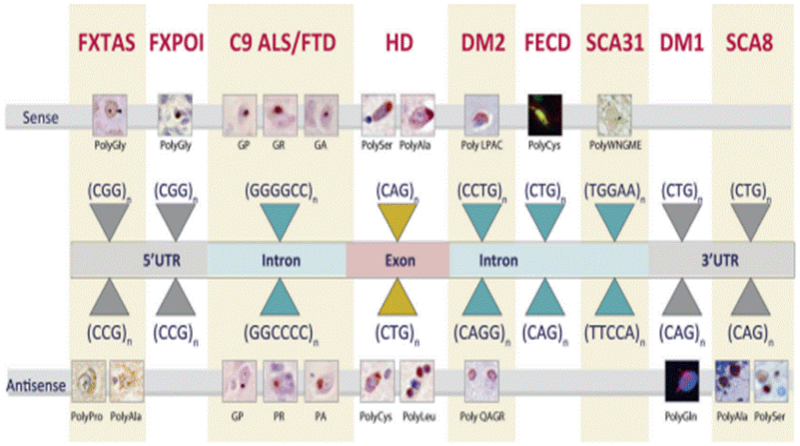

Figure –1.Various proteins identified from patients affected with Neuro degenerative diseases. Banez-Coronel and Ranum 2019

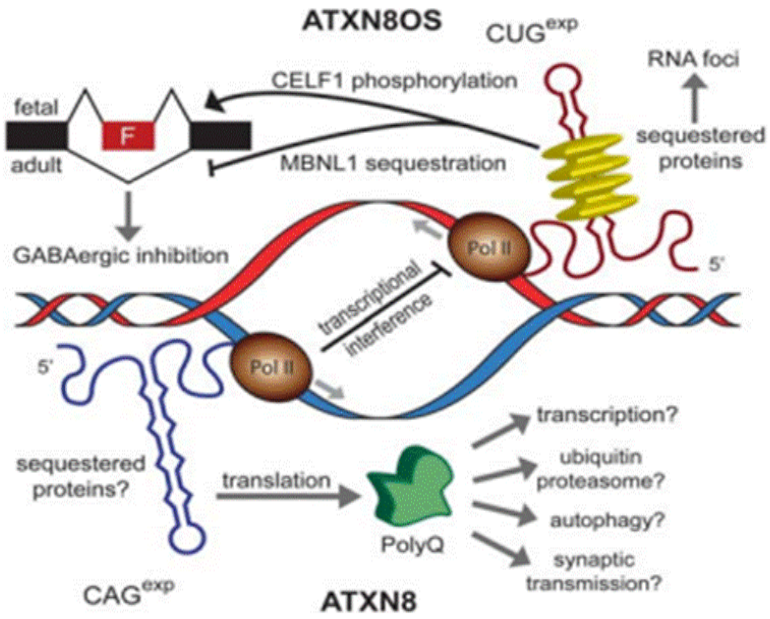

Figure-2. Mechanisms of bidirectional transcription in SCA8. Batra, et al., 2010

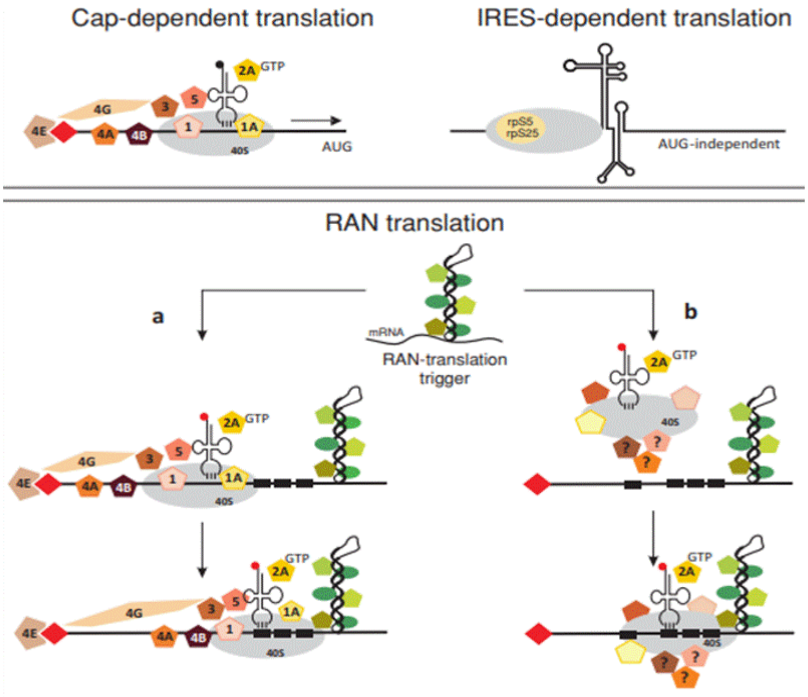

Figure-3.Mechanisms of RAN mediated translation initiation. Wojciechowska et al., 2014

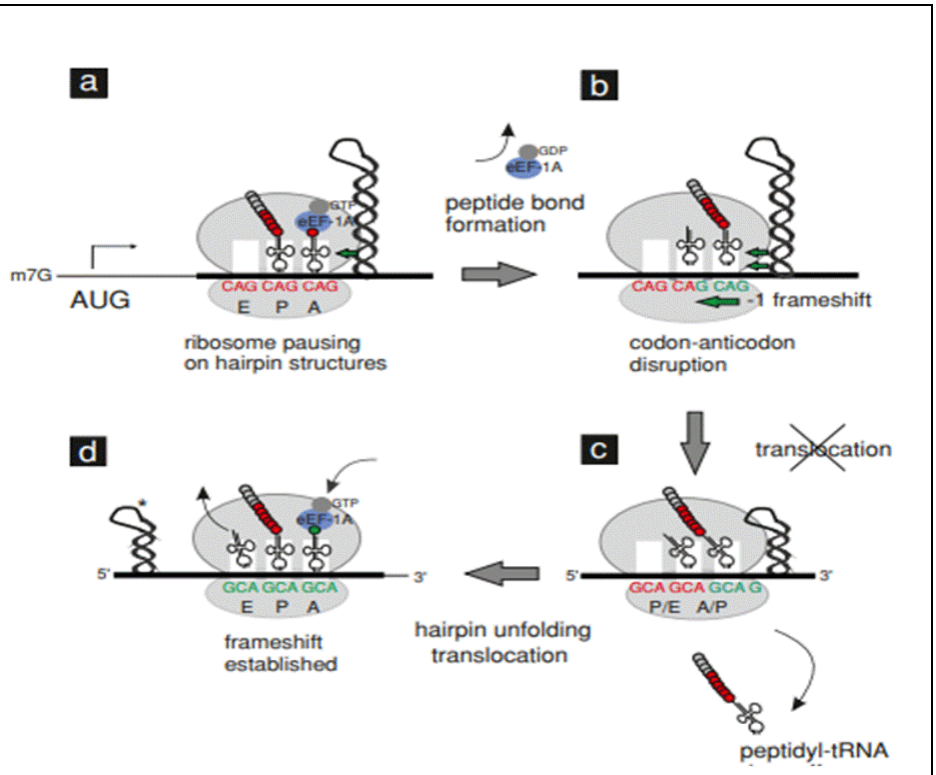

Figure-4. Mechanisms of ribosomal frameshifting at CAG repeats

7. DISCUSSION

Hazardous proteins can be produced by protein translation in the context of nucleotide repeat expansions using a cellular mechanism known as RAN translation. Anticipation in neurodegenerative illnesses is explained by variable growth of simple sequence repeats. The present understanding of repeat-associated disorders is the result of pathogenesis models and investigations intended to identify the molecular events impacted by these expansions. Extended homopolymeric amino acid tracts, frequently polyglutamine or polyQ, and detrimental protein gain-of-function effects are the outcomes of the expansions in protein-coding areas. RNA-mediated toxicity, on the other hand, results from expansions in noncoding areas. For example in diseases like HD and spinocerebellar ataxias (SCA1, 2, 3, 6, 7, and 17) where the expansion mutation occurs in the coding region, (CAG)n expansions (CAGexp) cause the synthesis of polyglutamine (polyQ) tracts in their corresponding proteins, which build up in intracellular, frequently ubiquinated inclusions (). A mechanism of protein gain-of-function leads to toxicity. Numerous cellular processes, such as transcriptional regulation, the ubiquitin proteasome system, autophagy, and synaptic transmission, have been shown to be disrupted by polyQ-containing proteins. For disorders where microsatellite expansions are located in noncoding regions, RNA gain-of-function models have been suggested as an alternative. For example in myotonic dystrophy type 1 (DM1), transcription of CTGexp mutations in the DMPK gene's 3′-untranslated region (3′-UTR) results in stable RNA hairpin structures that act as triggers and scaffolds to change the activities of specific alternative splicing factors, such as MBNL1 and CUGBP1, respectively.

Protein gain-of-function and RNA gain-of-function pathways are assumed to be responsible for repeat-associated diseases resulting from small sequence expansions, which have been categorised as coding and noncoding, respectively. Recent reports of repeat-associated non-AUG (RAN) translation in spinocerebellar ataxia type 8, myotonic dystrophy type 1, fragile X tremor/ataxia syndrome, C9ORF72 amyotrophic lateral sclerosis, and frontotemporal dementia have blurred the lines between these classifications. This noncanonical translation can begin in numerous frames of CAG, CGG, and GGGGCC repetitions of the sense and antisense strands of disease-relevant transcripts without the need for an AUG start codon. Although the exact method of translation initiation is yet unknown, RNA structures created by the repetitions have been proposed as potential triggers. Since frameshifting has been identified across CAG repeats in spinocerebellar ataxia type 3 and Huntington's disease, templates containing microsatellite expansions have also been demonstrated to impede translation elongation. By changing the relative amounts of sense vs antisense transcripts or by influencing transcriptional start or termination locations, microsatellite repeat expansion may interfere with these regulatory processes. Nine diseases caused by repeat expansion mutations have now been linked to RAN translation. RAN translation is

permissive for a variety of microsatellite patterns, such as tri-, tetra-, penta-, and hexanucleotide repeat expansions, indicating that RAN translation may be a common mechanism in the majority of microsatellite expansion disorders.

Research avenues in this area include finding new RAN-positive diseases and novel RAN protein motifs by developing improved techniques to detect RAN products. It is necessary to create novel *in vivo* models that permit both sense and antisense expression in order to evaluate and comprehend the potential and constraints of different therapeutic approaches. Significant non- and near-AUG started translation across the transcriptome is suggested by data from ribosome profiling datasets (). Therefore, aberrant normal non-canonical initiation pathways that result in harmful proteins but otherwise serve useful purposes in other contexts may be reflected in RAN translation. If RAN translation is to be used as a therapeutic target, it will be essential to define these normal processes and their involvement in neuronal biology. There has been a lot of interest in using antisense oligonucleotides (ASOs) to break down expansion RNAs. ASOs offer a platform for the degradation of many disease-causing expansion transcripts, which would simultaneously lower the quantities of RAN protein expressed from those transcripts. Creation of medications that stop RAN protein connections and/or function is another potential avenue. Additional research is required to develop strategies that directly target individual or sets of RAN proteins for clearance through ubiquitination or the use of antibodies.

The spectrum of knowledge in the area of protein translation in repetitive illnesses is still in its early stages, and there are still a number of unsolved questions. One of them is whether initiation also takes place in the CGG (+0, polyarginine) frame or on antisense FMR1 transcripts, and what particular protein or neuronal components are crucial for RAN translation at CGG repeats. Next, it is yet unknown if the mechanisms behind RAN translation are the same for all repetition types, reading frames, and sequence contexts. Finally, another question is does repeat length influence bidirectional expression of the affected locus? and is it possible that toxic RNAs and proteins are coordinately produced from the same expansion mutation or do some cell types preferentially synthesize one over the other?.