Published online 12 September 2014 *Nucleic Acids Research, 2014, Vol. 42, No. 19* ***11849–11864***
doi: 10.1093/nar/gku794

SURVEY AND SUMMARY

RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders

Marzena Wojciechowska†, Marta Olejniczak†, Paulina Galka-Marciniak†, Magdalena Jazurek† and Wlodzimierz J. Krzyzosiak*

Department of Molecular Biomedicine, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704, Poznan, Poland

Received June 16, 2014; Revised August 20, 2014; Accepted August 21, 2014

ABSTRACT

Repeat-associated disorders caused by expansions of short sequences have been classified as coding and noncoding and are thought to be caused by protein *gain-of-function* and RNA *gain-of-function* mechanisms, respectively. The boundary between such classifications has recently been blurred by the discovery of repeat-associated non-AUG (RAN) translation reported in spinocerebellar ataxia type 8, myotonic dystrophy type 1, fragile X tremor/ataxia syndrome and C9ORF72 amyotrophic lateral sclerosis and frontotemporal dementia. This noncanonical translation requires no AUG start codon and can initiate in multiple frames of CAG, CGG and GGGGCC repeats of the sense and antisense strands of disease-relevant transcripts. RNA structures formed by the repeats have been suggested as possible triggers; however, the precise mechanism of the translation initiation remains elusive. Templates containing expansions of microsatellites have also been shown to challenge translation elongation, as frameshifting has been recognized across CAG repeats in spinocerebellar ataxia type 3 and Huntington's disease. Determining the critical requirements for RAN translation and frameshifting is essential to decipher the mechanisms that govern these processes. The contribution of unusual translation products to pathogenesis needs to be better understood. In this review, we present current knowledge regarding RAN translation and frameshifting and discuss the proposed mechanisms of translational challenges imposed by simple repeat expansions.

INTRODUCTION

A particular group of human neurological disorders is associated with expansions of simple repetitive elements within specific genes. This class comprises more than 20 diseases, which have been categorized into coding and noncoding repeat expansion disorders depending on the genetic location of their causative mutations (1–3). The toxicity of the coding repeats located within ATG-initiated open reading frames (ORFs) is typically governed via a protein *gain-of-function* mechanism. This mode of repeat toxicity is found in numerous disorders, including Huntington's disease (HD), spinocerebellar ataxia (SCA) type 1, 2, 3, 6, 7, 8 and 17, dentatorubral-pallidoluysian atrophy (DRPLA) and spinal and bulbar muscular atrophy (SBMA), in which the expression of exonic CAG repeat expansions gives rise to polyglutamine (polyGln)-rich proteins that adversely affect various cellular functions (3). However, the toxicity of coding CAG repeats can also be exerted on the transcript level via an RNA *gain-of-function* mechanism that was initially described for the noncoding repeat expansion disorder myotonic dystrophy type 1 (DM1) (4–8). Also Fragile X-associated tremor ataxia syndrome (FXTAS), myotonic dystrophy type 2 (DM2), SCA31, SCA10, SCA8 and, most recently, amyotrophic lateral sclerosis and frontotemporal dementia (C9ALS/FTD) have been shown to exhibit toxicity via a mutant transcript *gain-of-function* mechanism. In these diseases, the expression of transcripts harboring expansions of particular repeats leads to the formation of nuclear RNA foci that sequester specific RNA-binding proteins, resulting in a loss of their normal function (9–11).

*To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: wlodkrzy@ibch.poznan.pl
†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

Over the past several years, these traditional paradigms of repeat toxicity have been reevaluated, as bidirectional transcription through repeat regions of genes associated with numerous human neurological disorders has been described (12,13) along with non-AUG translation initiation at expanded CAG, CGG and GGGGCC (G_4C_2) repeats (14–22). The discovery of repeat-associated non-AUG (RAN) translation, which indicates that noncoding diseases could involve unexplored yet toxic proteins and that coding diseases are likely to produce unexpected proteins in other frames, has proven how inscrutable these diseases remain. Importantly, novel proteins have been found to accumulate in the relevant tissues of the diseases in both patients and model organisms, suggesting that homopolymeric or dipeptide repeat repeat-associated non-AUG (RAN) translation products might play a role in microsatellite expansion disorders. Nevertheless, their precise contribution to the process of pathogenesis of SCA8, DM1, FXTAS and C9ALS/FTD remains to be determined.

Several mechanisms of translation initiation are possible; however, none seems to explain RAN translation. The cap-dependent scanning mechanism and cap-independent internal ribosome entry site (IRES) mechanisms are the best known types of translation initiation (23,24); however, other initiation mechanisms such as leaky scanning, ribosome shunting or reinitiation are also known (25). The fact that microsatellite expansions do not follow the canonical rules of translation initiation and generate series of homopolymeric or dipeptide repeat proteins (DPRs) in multiple frames indicates the involvement of a novel yet unrecognized process or the altered utilization of known processes. RNA structures formed by the repeats must be considered when deciphering mechanisms of RAN translation because their formation and stability have been shown to affect the abundance of RAN translation products (14). Could such structures be used to trigger an IRES type of mechanism? Could IRES translation-associated factors (ITAFs) that normally help to stabilize RNA structures be more abundant at repeat-formed structures, allowing translation initiation without an AUG start codon? Could alternative start codons occurring within or in close proximity to these repeats permit the noncanonical initiation of protein production? These questions and many others arise when dissecting the novel ambiguity of expanded repeats.

Microsatellite repeat expansions have also been demonstrated to affect the maintenance of a reading frame by challenging the translation elongation machinery. Frameshifting by one nucleotide downstream or one nucleotide upstream during translation across elongated CAG repeats of ATXN3 and HTT transcripts results in the generation of new proteins containing polyAla (SCA3 and HD) and polySer (HD) (26–30). Such ribosome slippage occurs probably at random positions along the expanded CAG repeats in tissues and cells, and what initiates the process remains unknown. Nonetheless, the frequency of frameshifting increases with the number of CAG repeats and seems to depend on the formation of hairpin structures by the repeats. The resulting chimeric mutant proteins accumulate in intranuclear and cytoplasmic aggregates that are toxic on the cellular level, thus contributing to the pathogenesis of SCA3 and HD (26,28,30).

The occurrence of RAN translation and frameshifting challenge the common dogmas that the expansions of CAG, CGG and G_4C_2 repeats located in non-coding sequences do not encode any proteins and that CAG repeat mutations located in polyGln ORFs express proteins only in one frame. These findings raise the possibility that the new proteins resulting from RAN translation and frameshifting can be ubiquitous and contribute to the pathogenesis of these neurological disorders.

In this review, we describe numerous features of the RAN translation of expanded CAG, CGG and G_4C_2 repeats in SCA8, DM1, FXTAS and C9ALS/FTD and present known facts regarding the frameshifting that occurs at expanded CAG repeats in SCA3 and HD. We discuss putative mechanisms of RAN translation initiation and frameshifting during translation elongation and share our thoughts regarding the prevalence and biomedical importance of these repeat-associated phenomena.

BASIC FEATURES OF RAN TRANSLATION

RAN translation of ATXN8 and antisense DMPK CAG repeat transcripts from SCA8 and DM1

In deciphering an ambiguous pathomechanism of SCA8, the Ranum group has found that microsatellite expansions of CAG repeats can express homopolymeric proteins in all three reading frames in the absence of an AUG initiation codon (14). This discovery, called Repeat-Associated Non-AUG translation or RAN translation, instigated a series of further findings from other laboratories that have proven that atypical protein production can be found in the presence of CGG repeats of FXTAS (16) as well as of G_4C_2 and C_4G_2 repeats of ALS/FTD (15,17–22,31,32).

RAN translation, which occurs across long hairpin-forming CAG repeats of ATXN8 and DMPK transcripts, was consistently found in human embryonic kidney (HEK293T) cells and in murine neuroblastoma (N2a) cells (Supplementary Table S1 and Figure S1) and was also demonstrated using CAG transcripts *in vivo* in SCA8 and DM1 human and mouse tissues (Supplementary Table S2). This non-AUG translation leads to the expression of atypical homopolymeric proteins in all three reading frames, producing polyGln in the CAG frame, polyserine (polySer) in the AGC frame and polyalanine (polyAla) in the GCA repeat frame. Furthermore, expanded CUG repeats are likely to be RAN translated in three possible frames (the CUG, UGC and GCU frames), coding, respectively, polyleucine (polyLeu), polycysteine (polyCys) and polyAla (14).

RAN translation was shown to depend on the repeat length, and the abundance of nascent proteins tends to increase with the repeat number. In SCA8, polyAla was the most expressed, followed by polyGln and polySer, and the repeat lengths at which robust levels of these proteins were detected by western blot in HEK293T cells were 73, 42 and 58, respectively (Supplementary Table S1). These three proteins can be co-expressed in a single cell; however, their levels varied significantly among cells, as shown by immunofluorescence analysis. In addition, the intracellular localization of the RAN translation products varies, and polyAla proteins are dispersed in the cytoplasm but may form aggregates in the nucleus. PolyGln forms nuclear aggregates, and

the least abundant polySer forms both cytoplasmic and nuclear aggregates (14).

RAN translation of sense FMR1 CGG repeat transcripts from FXTAS

Two years after the first report on RAN translation, Todd *et al.* showed that this noncanonical protein synthesis also occurs at CGG repeats in FXTAS (16). Non-AUG-initiated proteins were found in transfected mammalian cells (Supplementary Table S1 and Figure S1), in patients' brain tissues and in different FXTAS model organisms, including transgenic mice and flies (Supplementary Table S2). Homoaminoacid RAN translation proteins were expressed in two of three possible reading frames: polyglycine (polyGly) from the GGC repeat frame and polyAla from the GCG frame, while no polyarginine (polyArg) was detected from the reading frame with CGG repeats. Different abundances of these proteins, i.e. predominance of polyGly and an inability to detect polyArg, may reflect a large difference in the initiation frequency in different frames, preferential drop-off of nascent polyArg peptide or alternatively, large differences in cellular stability of polypeptides that were synthesized in different frames.

When tested in transfected cells, RAN translation has been observed either from expanded (55 and 88–160) or normal (30–50) CGG repeat lengths, with a higher abundance of proteins being expressed from short (~30 CGG) rather than from longer (~88 CGG) repeats as determined by western blot (Supplementary Table S1). These results may be a consequence of the decreased efficiency of RAN translation from expanded repeats but may also be caused by the limitations of the methods that were used to detect its products, as concluded by Todd *et al.* (16).

RAN translation of G_4C_2 sense and C_4G_2 antisense C9ORF72 transcripts from ALS/FTD

The latest findings from various groups indicate that RAN translation can also be initiated from expanded hexanucleotide repeats associated with C9ALS/FTD (15,17–19,22). The *C9ORF72* locus of ALS/FTD was shown to be bidirectionally transcribed, and both the sense and the antisense *C9* transcripts with hexanucleotide repeats can express DPRs in all possible reading frames (15,17,19). The DPRs expressed from sense transcripts are polyGlyAla, polyGlyPro and polyGlyArg (18), whereas the DPRs expressed from antisense transcripts are polyProArg, polyGlyPro and polyProAla (15). Furthermore, additional polyProArg and polyGlyPro proteins were expressed from short ORFs of the antisense transcripts due to the presence of AUG codons preceding the C_4G_2 repeats in these reading frames (15).

DPRs of sense and antisense transcripts have been detected in experimental models of C9ALS/FTD, in human postmortem brain tissues (15,18) and in human iPSC-derived neurons (20,21) (Supplementary Tables S1 and S2). These non-AUG-initiated DPRs form intracellular inclusions, and when found in neuronal cells, they co-aggregate (22). The efficiency of the *C9* RAN translation is length-dependent with respect to the hexanucleotide repeats, being more efficient for longer repeats (15,17,22,32). Mori *et al.* has demonstrated the expression of sense strand-derived DPRs in HEK293T cells, and polyGlyAla was expressed from constructs harboring ~38 G_4C_2 repeats. PolyGlyPro was found when the repeat length was extended to 145, whereas no polyGlyArg DPRs were detected from constructs with 11 to 145 G_4C_2 repeats (Supplementary Figure S1) (19). The Ranum group reported that all three DPRs are detectable in HEK293T cells transfected with constructs of only 30 G_4C_2 repeats; however, their expression becomes more abundant with 60 and 120 repeats (15).

ARE RAN-TRANSLATED PROTEINS RELEVANT TO THE DISEASES IN WHICH THEY ARE DETECTED?

Despite the great interest in RAN translation, it is still a matter of debate whether the non-AUG-initiated proteins are toxic and can be linked to the pathology of SCA8, DM1, FXTAS and C9ALS/FTD. RAN-translated proteins related to the four disorders are found either in the cytoplasm or in the nucleus as aggregated or diffused deposits, and they may be expressed simultaneously in the cells and co-aggregate into single inclusions. These inclusions can co-exist with nuclear repeat RNA foci but are also found in diseased cells that are void of foci of mutant transcripts (16,17,33). The non-AUG-initiated proteins have been found in cultured cells (Supplementary Table S1), in experimental models of the diseases and in patient tissues (Supplementary Table S2), and they are not detected in healthy controls (14,15,17–19,34,35).

Cellular inclusions containing RAN-translation products are present in tissues that are affected in SCA8, DM1, FXTAS and ALS/FTD. In SCA8, the most abundant polyAla was found to form inclusions in the remaining Purkinje cells in human autopsy samples, whereas in the mouse disease model, the inclusions were found throughout the cerebellum in dendrites and in Purkinje cells, as detected by immunostaining with antibodies to the putative SCA8-polyAla protein (14). In DM1, inclusions of RAN-translation products formed nuclear polyGln aggregates that are detected in patients' blood at higher frequency than in myoblasts and skeletal muscle; in DM1 transgenic mice, such inclusions are more abundant in leukocytes than in cardiac myocytes (14). In FXTAS patients, RAN-translated products were found to form Gly-positive inclusions that accumulate in the hippocampus, cerebellum and cortex, whereas in a mouse model of the disease, polyGly aggregated in the hypothalamus, cortex and brainstem (16). Interestingly, in FXTAS, these inclusions were ubiquitin-positive. Although in SCA8 and DM1, only the most expressed RAN-translation products were found to aggregate, in C9ALS/FTD, all possible DRPs from either sense or antisense transcripts have been detected in inclusions that exhibit dot-like and star-like shapes (Supplementary Table S2). DPRs were present in the patient tissues and in cells of C9ALS/FTD experimental models and were restricted to neurons, being highly abundant in neocortical regions, the cerebellum and the hippocampus, more variable in the striatum and substantia nigra and rare or absent in the medulla and spinal cord (15,17–19,22,31–33,36,37). These aggregates colocalize with inclusions that are positive for

p62 and negative for phospho-TDP-43, a classical molecular feature of *C9ORF72* mutation carriers (19,22,32).

The discovery of RAN translation may challenge the common dogma that the toxicity of noncoding repeat disorders is exclusively caused by mutant transcripts. What are the arguments indicating that RAN-translated products are relevant or irrelevant to neurodegeneration? The majority of data against the toxicity of RAN proteins originate from work on C9ALS/FTD. First, no correlation between the neuroanatomical distribution of DPR inclusions and the degree of degeneration in different regions of the central nervous system in C9ALS/FTD was reported (32), and tissues containing DPRs showed no signs of cell loss or other apparent symptoms of neurodegeneration (31). Second, RAN translation-related pathology was less frequently detected or absent in lower motor neurons in the spinal cord, the area affected in ALS (15,22,32). However, as suggested, this result may be a consequence of increased cell death and atrophy in this region, resulting in a reduced number of cells available for analysis. Additionally, the levels of RAN-translated proteins and their capacity to form aggregates may be lower in the spinal cord (15), or even not observed in iPSC-derived motor neurons (38). Third, despite the outstay of RAN-translated proteins, the toxic phenotype of human C9ORF72 iPSC-derived neurons was rescued upon antisense oligonucleotide (ASO) treatment (20). As reported by Donnelly *et al.*, treatment of these cells with ASO efficiently abrogated known pathogenic features of C9ALS/FTD, such as intranuclear G_4C_2 repeat RNA foci formation, ADARB2 protein sequestration, deregulated gene expression and susceptibility to glutamate toxicity; however, this treatment did not affect the level of the polyGlyPro dipeptide.

The arguments in favor of RAN-translation product toxicity derive primarily from work describing signs of cell death in the presence of non-AUG-originated proteins (14,16). Ranum *et al.* reported that cultured murine neuroblastoma cells transfected with SCA8 expanded CAG repeat constructs expressing polyGln, polyAla and polySer proteins triggered apoptosis, as shown by increased annexin-V staining (14). Furthermore, the authors demonstrated the colocalization of caspase-8, an early indicator of apoptosis, with polyGln aggregates in the skeletal muscles of DM1 patients and in leukocytes from DM1 mice. Enhanced cell death was also observed in mammalian HEK293T cells transfected with G_4C_2 constructs expressing polyProArg and polyGlyPro proteins (14) and in COS7 cells transfected with FXTAS CGG repeat plasmids expressing the polyGly protein (16). Additionally, in a *Drosophila* model of FXTAS, the presence of the RAN-translated polyGly protein was accompanied by rough-eye phenotype, which is a manifestation of toxicity in flies (16).

A very recent study demonstrated that synthetic $(GlyArg)_{20}$ and $(ProArg)_{20}$ dipeptides encoded by sense and antisense C9ORF72 repeats, respectively, are toxic to cells. These dipeptides have ability to enter cells, migrate to nucleus, bind to nucleoli and induce cell death. The ProArg dipeptide toxicity manifests in aberrant pre-mRNA splicing similar to that observed in ALS patients, and in alteration of ribosomal RNA biogenesis (39).

It is also known from HD studies that toxicity of polyQ-expanded Htt depends on its subcellular localization, and the cell nucleus is the primary site of pathogenesis (40,41). The RAN-translated proteins accumulate in the cell nucleus and the question may be asked whether they may trigger similar mechanisms of pathogenesis by recruiting essential nuclear proteins and compromising their functions?

The occurrence of RAN translation across relatively short repeats raises questions about their involvement in the pathogenesis of SCA8, DM1, FXTAS and C9ALS/FTD and may indicate a normal physiological function. According to the hypothesis that protein inclusions may protect cells from toxic misfolded proteins (42,43), it has been postulated that aggregates formed by RAN-translated proteins might not be the primary toxic agents *per se* but that soluble precursors of these proteins might be the causative pathogenic species (15,32). However, determination of the physiological function of RAN-translation products awaits further investigation.

SPECULATED MECHANISMS OF RAN TRANSLATION INITIATION: DO WE KNOW WHERE AND HOW IT STARTS?

The translation of most eukaryotic mRNAs is initiated via the cap-dependent scanning mechanism, which employs the 43S pre-initiation complex loaded with several protein factors. The complex proceeds in a 3′ direction until it encounters the first AUG codon in a Kozak consensus sequence (24). These canonical rules seem not to apply to RAN translation, and it currently remains unclear what forces the synthesis of homoproteins by ribosome and exactly where the process starts. One hypothesized trigger is an RNA structure formed by the repeats that, in fact, features in all RAN-translated sequences reported thus far (Figure 1). As described, the hairpin structure-forming CAG repeats are efficiently RAN translated, in contrast to CAA repeats that do not form hairpins (14). CGG repeats of the FMR1 transcript are also RAN translated; these repeats form hairpins (44) that are able to fold into G-quadruplex structures under specific conditions (45,46). The sense G_4C_2 repeats of *C9ALS/FTD* also fold into G-quadruplexes that are not detected at the antisense C_4G_2 repeats of the *C9*, which, however, may fold into hairpins (47). As these structures show different stability, and because hairpins are less stable than G-quadruplexes, this phenomenon may affect the expression levels of RAN-translation products and their variety from repeat-harboring transcripts (15,17,19).

Because the precise involvement of RNA structures in RAN translation needs further investigation, the question has arisen whether these repeat sequences act as IRES to initiate RAN translation via a cap-independent mechanism. Such a scenario was suggested because hairpin- and G-quadruplex-forming repeats might mimic IRES structures, and the 5′ region preceding the CGG repeats of the FMR1 transcript had previously been identified to function as an IRES and a site of ribosome pausing (48). Based on such factors as dependence on canonical eukaryotic initiation factors (eIFs), the proposed secondary structure, start codon localization and ability of the IRES to function in a cell-free system with or without supplementa-

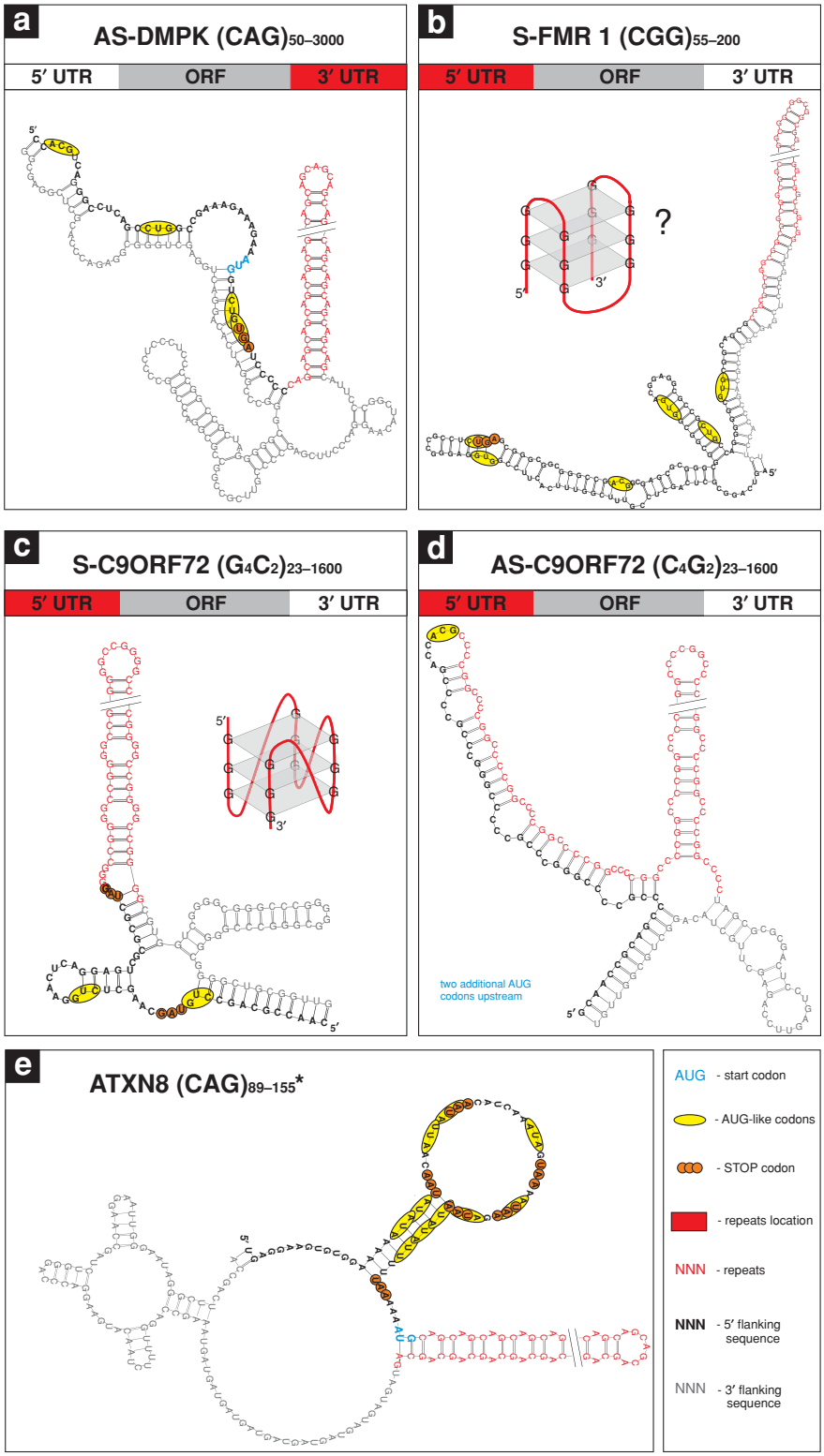

Figure 1. *In silico* structure modeling of transcripts with expanded repeats. Each panel **(a–e)** shows the structure for a transcript with expanded repeats, its orientation (S-sense, AS-antisense) and the number of repeats identified in affected patients: **(a)** DMPK/DM1, **(b)** FMR1/FXTAS, **(c)** and **(d)** C9ORF72/ALS/FTD and **(e)** ATXN8/SCA8. Structures were modeled *in silico* with RNAmetaserver and visualized by Varna software. The modeling was performed for transcripts with 150 repeated nucleotides (50× trinucleotide and 25× hexanucleotide motifs) and, to obtain more relevant results, in two variants of the flanking sequence 200 and 100 nt each side. For figure clarity, we have shortened the hairpins formed by the expanded repeats, as well as the flanking sequence, to ∼50 nt of 5′ flank. In each panel **(a–e)**, the repeat locations within the affected gene are indicated (red box). As indicated in the figure legend, different AUG-like alternative start codons and STOP codons are marked in the sequence preceding the repeats. In **(b)** and **(c)**, despite the hairpin structure, the repeats may form G-quadruplexes. *—Indicates neither ATXN8 nor ATXN8OS contain a conventional ORF.

tion, known viral IRES were divided into four groups. Initiation on types I and II typically requires most canonical initiation factors including eIF3, eIF4A and the C-terminal domain of eIF4G in addition to the eIF2-Met-tRNA$_i^{Met}$-GTP ternary complex. In contrast, type III IRES directly attach the 43S complex to the initiation codon independently of eIF4F, eIF4B, eIF1 and eIF1, whereas type IV IRES initiate without eIFs or Met-tRNA$_i^{Met}$ (49,50). Of the four known IRES-initiated translation mechanisms used by viruses, only type IV is both cap-independent and AUG-independent and thus could explain certain features of RAN translation initiation. The three other types of viral IRES mechanisms require an AUG codon and thus fail to explain why RAN translation produces proteins in multiple frames. With all these uncertainties, it seems that an alternative, as yet unknown mechanism may govern translation initiation at expanded RNA repeats.

The exact sites of RAN translation initiation at expansions of CAG, CGG, G_4C_2 and C_4G_2 repeats are unknown and may differ among repeats and reading frames. For CAG repeats in SCA8, RAN translation in the polyGln frame initiates at or near the start of the repeats, independently of an upstream AUG codon, whereas in the GCA frame encoding polyAla, it initiates at various sites throughout the length of the repeats (14). For CGG repeats in FXTAS, this atypical translation was shown to start between the 21st and 12th nucleotide 5′ of the repeats for the polyGly reading frame, was sensitive to stop codons when placed just proximal to the repeats and required an upstream AUG-like codon such as GUG. By contrast, polyAla formation arose independently of AUG-like codons and was insensitive to stop codon insertion 5′ of the repeats (16). For G_4C_2 hexanucleotides in C9ALS/FTD, the exact start of RAN translation is also unknown, but its termination does not necessarily occur on the repeats. As it has been demonstrated for polyGlyAla and polyGlyArg frames, at least a fraction of RAN translation products are extended beyond the repeats into the 3′ region (22). Consequently, the length of some RAN-translated proteins may depend on the position of the stop codon in a particular frame.

The sequences adjacent to repeats may affect the RAN translation ability or efficiency, as demonstrated by Ranum *et al.*, who used HEK293T cells transfected with constructs containing 20 bp of the 5′ endogenous flanking sequence upstream of the CAG repeats at the HD, HDL2, DM1 and SCA3 loci (14). In this system, the products of RAN translation in the polyGln frame, which were abundant in the SCA8 context, were expressed at lower levels in the HD, HDL2, SCA3 and DM1 sequence contexts. In addition, the process of RAN translation of these constructs appeared to be less prevalent when studied in the cell-free rabbit reticulocyte lysate system, indicating specific protein requirements. When tested in a noncellular environment, RAN translation was also shown to be sensitive to the presence of alternative start codons (14). Together, these results suggest the importance of particular RNA structures, AUG-like codons and specific cellular proteins for RAN translation, whose efficiency varies for repeats occurring in different gene contexts and cellular habitats.

PROJECTED MECHANISMS OF RAN TRANSLATION ACROSS EXPANDED SIMPLE REPEATS IN HUMAN NEURODEGENERATIVE DISEASES

Repeat locations in particular transcripts, their structures and flanking sequences

The generation of homopolymeric or dipeptide repeat RAN-translated proteins has thus far been observed at repeats located in noncoding regions of a few genes, including *ATXN8* and 3'UTR of *DMPK* and 5′UTRs of *FMR1* and *C9ORF72*. Is this repeat location preferred for RAN translation? Might repeats genetic location influence RAN initiation, and could this process also occur in repeats located in ORFs?

The RNA structures formed by the repeats are recognized as potential triggers when trying to decode the mechanism of RAN translation initiation. Their formation and stability affect the abundance of RAN-translated products. For longer repeats that have a more stable RNA structure, increased RAN-translated protein levels were detected (14,51). It is known that repeats that undergo RAN translation can form hairpin or G-quadruplex structures, which have different molecular architecture and stability (Figure 1 and Supplementary Figure S2). Our previous *in vitro* analysis of RNA structures formed by different triplet motifs revealed the following order of hairpin stability: CGG > CAG > CUG > CCG (52). The crystallographic structures of CNG repeat duplexes (N = A,U,G,C) showed distortions of their double-stranded helices caused by noncanonical N–N base pairs, implying both similarities and differences in protein binding by repeat hairpins (53–55). In addition, we have demonstrated that in certain repeat containing transcripts, e.g. *HTT*/HD, *AR*/SBMA and *CACNA1*/SCA6, the repeat flanking sequences may contribute to hairpin stability (56). Additionally, the presence of interruptions in repeat tracts strongly influences the stability of the hairpin structure they adopt (57). Could such properties determine RAN translation process intensity in different disease models? The available data show differences in RAN translation efficiency between the *C9ORF72* antisense and sense transcripts, in which C_4G_2 repeats adopt hairpin structures, whereas G_4C_2 repeats may also fold into more stable G-quadruplex structures (47). The intramolecular G-quadruplexes are adopted by RNA motifs having the following consensuses sequence $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ (N represent any nucleotide, 3+ means 3 and more and 1−7 means number ranging from 1 to 7). The stability of these structures depends on several factors including the presence of potassium ions, number of consecutive guanines and separating nucleotides between G-tracts (58,59). Thus RAN translation from the latter transcript adopting G-quadruplex structure seems to be more robust (15,17,22). For this same reason, normal CGG repeats of sense FMR1 transcripts might undergo the noncanonical translation that typically occurs on expanded repeats. Interestingly, the presence of a G-quadruplex in the 5′UTRs was demonstrated to impair cap-dependent translation, promoting cap-independent translation, as G-quadruplexes were revealed to be structural components of some human IRES sequences (60).

Could any other features of transcripts undergoing RAN translation be considered to explain the process of non-canonical protein synthesis? Our *in silico* structure prediction of expanded repeats along with their flanking regions (Figure 1) revealed the presence of modifying factors that may potentially stimulate or inhibit RAN translation. The stimulatory factors may include additional hairpin stabilization by repeat flanking sequences (e.g. additional extension of (CGG)n hairpin by S-FMR1 flanking sequences (Figure 1b)), AUG START codons and various AUG-like codons (CUG, GUG, UUG, AUA, ACG and AUU which differ from AUG by a single base only (61)). Whereas inhibitory factors may comprise STOP codons (UAG, UAA, UGA) in different reading frames. How significant is the presence of alternative START codons for RAN translation initiation, and do STOP codons that may abolish the influence of AUG-like codons in the same frame play a role? These sequences were found to occur with different frequencies in the analyzed transcripts, and their relation to RAN translation efficiency should be carefully examined. At first glance, in the sense *C9* transcript, an alternative START codon upstream of the G_4C_2 repeats that is not followed by STOP in the polyGlyAla frame (Figure 1c) does not affect the abundance of RAN proteins in this frame (15). Similarly, a STOP codon located directly before the G_4C_2 repeats in the polyGlyPro frame does not seem to disturb protein expression in this frame. In the antisense *C9* transcript, we found only one AUG-like codon in close proximity to the hexanucleotide repeats (Figure 1d), but two additional start codons are present >100 nt upstream of the repeats (15). Despite these differences in the frequency of potential stimulatory and inhibitory factors of RAN translation, both sense and antisense *C9* transcripts undergo RAN translation; however, the intensity of this phenomenon is higher for the sense transcript (15,17,22).

Similarly, in a few other transcripts with expanded repeats associated with human neurodegenerative diseases, we observe various representations of modifying factors that may potentially stimulate or inhibit RAN translation (Supplementary Figures S2a–w) regardless of whether the repeats are located in noncoding or coding portions of the diseases-causing genes. Structure prediction and sequence analysis reveal some extra stabilization of repeat hairpins by the flanking sequences but do not show substantial differences in the abundance of AUG-like codons in the analyzed transcripts when compared with transcripts proven to be RAN translated (Figure 1 and Supplementary Figure S2). Interestingly, we have observed a general tendency for a higher abundance of AUG-like codons in the antisense transcripts than in their sense counterparts. These results suggest that other transcripts with expanded and structure-forming repeats may also be subject to RAN translation.

Translation initiation factors in cap-dependent and cap-independent RAN translation and the involvement of repeat binding proteins

Canonical translation initiation is a complex process requiring more than ten initiation factors (Figure 2) (23,24). However, under certain cellular conditions such as viral infection, protein synthesis might be switched to another mode and proceed with minimal involvement of the translation machinery (62). Depending on the mode of RAN translation initiation, different sets of protein factors may be involved (Figure 2). In a cap-dependent scanning mechanism, the utilization of a full set of initiation machinery might be perturbed when the scanning ribosome stalls at a highly structured obstacle of expanded repeats with bound proteins (Figure 2a). Because this stalling could force the scanning ribosome to pause, various proteins attached to structured RNA and proteins of the scanning complex might be rearranged, triggering noncanonical translation initiation at non-AUG codons. It is plausible that such perturbation could involve the eIF1 initiation factors, including eIF1A, which plays a critical role in the fidelity of start codon selection (23,24). Another candidate to consider is eIF2, whose function is to deliver Met-$tRNA_i^{Met}$ to the ribosome. There is evidence of eIF2-independent translation under stress conditions (63–67). As reported, RAN translation does not require an AUG codon, which implies that tRNAs charged with amino acids other than methionine are involved in translation initiation (14). Moreover, one of the translation initiation factors, eIF2D, can direct $tRNA^{Phe}$ to the P-site of the ribosome to initiate translation (65). A distinct set of proteins may be considered in the context of direct recruitment of the ribosome to structures formed by the expanded repeats (Figure 2b). One possible scenario is that the structured expanded repeats bind the small ribosomal subunit in the absence of initiation factors, but RAN translation still requires eIF3 and eIF2-GTP for initiation, as shown for viral IRES type III (49,50). Another scenario involves the steps and factors described for viral IRES type IV (49,50), in which RAN protein synthesis is initiated with minimal involvement of the translation machinery, assuming that the structures formed by expanded repeats are tightly folded and structurally mimic the missing initiation factors. In contrast to viral IRES, their cellular counterparts are less understood, more diverse in their structure and less stable (67). Taken together, the available data remain insufficient to answer the question of whether IRES mechanisms are indeed involved in RAN translation.

Similarly, it should be established whether repeat binding proteins, such as MBNL, which are known to localize in both the nucleus and cytoplasm (68–70), play a role in RAN translation. The question of whether these proteins participate directly in the initiation event or facilitate recruitment of key initiation factors to the expanded repeats still needs to be answered.

Other factors that might be involved in RAN translation initiation

Could global translation attenuation and cellular stress conditions permit RAN translation? (71–74). It is known that IRES-dependent translation occurs preferentially under cellular stress conditions such as growth, mitosis, apoptosis or viral infection, in which global cap-dependent translation is compromised (62) and is accompanied by the proteolysis of eIF4G and changes in the phosphorylation levels of eIF4E-binding proteins and eIF2A (24,75). Furthermore, stress might trigger post-translational modifications

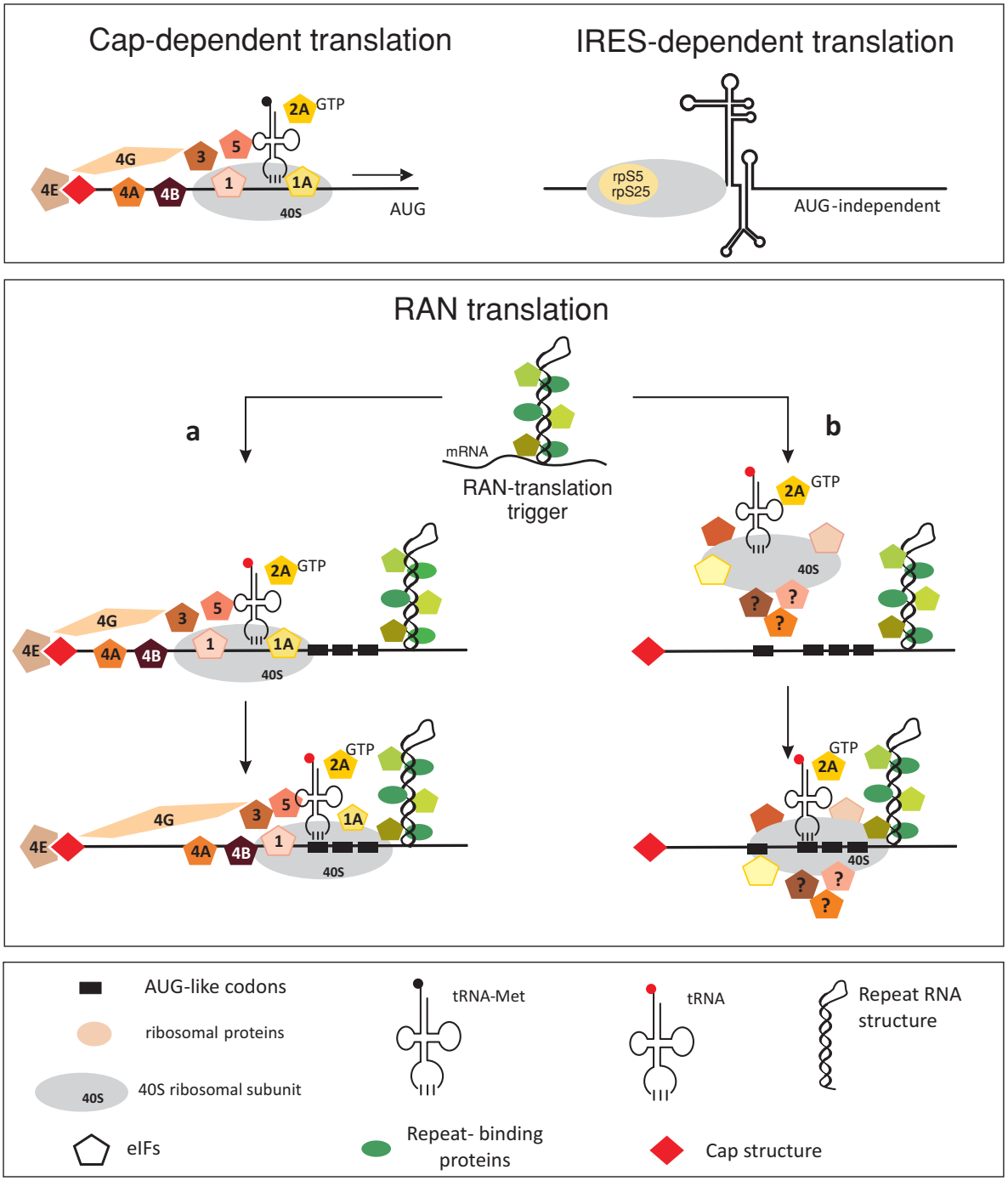

Figure 2. Speculated Mechanisms of RAN Translation Initiation. There are at least two translation initiation mechanisms: a canonical cap-dependent mechanism described in Eukaryotes and a cap-independent type IV IRES-dependent mechanism described in Dicistroviridae viruses. In the cap-dependent mechanism of translation initiation, the initiator tRNA is delivered to the ribosome as a ternary complex with eIF2 and GTP, and then, eIF1, eIF1A, eIF5 and eIF3 initiation factors promote binding to form a 43S pre-initiation complex. This complex binds to the mRNA through its interaction with the cap-binding complex, which consists of eIF4E, eIF4G, eIF4A and eIF4B factors, and begins scanning for the start codon. In cap-independent/IRES-dependent translation initiation, highly structured type IV IRES, which does not require initiation factors, operates essentially as an all RNA-based ribosome recruitment apparatus; the IRES position of pseudoknot I mimics the initiator tRNA codon/anti-codon interaction that results in an initiation at the non-AUG codon. This type of translation initiation requires the interaction between IRES and ribosomal proteins RPS25 and RPS5. RAN translation is initiated by repeat-forming RNA structures. These structures could bind various proteins, including specific repeat binding proteins, translation initiation factors and other regulatory proteins. The riboprotein complex may resemble a translation initiation unit and trigger translation initiation in the absence of an AUG start codon. Thus, RAN translation might occur via a cap-dependent scanning mechanism **(a)** and/or as a result of IRES-like ribosome recruitment **(b)**. In **(a)**, the scanning ribosome approaches a repeat hairpin that results in the stalling of the pre-initiation complex. This forces an unspecific interaction between hairpin proteins and the pre-initiation complex and/or dissociation of some translation initiation factors that may result in noncanonical translation initiation at non-AUG codons. In **(b)**, a direct recruitment of the ribosome to hairpin-associated protein complex results in unspecific translation initiation, similar to what is proposed in **(a)**, with the use of codons other than AUG.

of ITAFs, affecting both their subcellular localization and IRES-binding affinity, thus modulating their activity (67).

It has been reported that the expression of transcripts harboring expanded CUG, CCUG, CAG and G_4C_2 repeats can cause cellular stress conditions mediated through the inhibition of protein translation, as demonstrated in DM1 and DM2 (71,72), and by triggering nucleolar stress, as shown in SCA3, HD and C9ALS/FTD (76–78). Such stress conditions could affect the post-translational modification of proteins involved in translation initiation, leading to faulty functioning, which is then reflected in altered RNA-binding affinity that can trigger RAN protein synthesis.

CAG REPEAT EXPANSIONS OF SCA3 AND HD CAN CHALLENGE TRANSLATION ELONGATION VIA RIBOSOMAL FRAMESHIFTING

A decade before RAN translation was discovered, Rouleau *et al.* proposed that frameshifting within expanded CAG repeats of the *ATXN3* gene resulted in the production of a polyAla-containing protein, which forms intranuclear inclusions contributing to toxicity (27). Six years later Davies and Rubinsztein demonstrated that this phenomenon is not restricted to SCA3 and occurs also in translation of the *HTT* gene, which is responsible for HD (29). Frameshifting one nucleotide downstream ((−1) frame) or one nucleotide upstream ((+1) frame) during translation results in the generation of a new protein, and in addition to the polyGln encoded by the (0) CAG frame, the polyAla and polySer proteins can be produced from the (−1) GCA frame and the (+1) AGC frame, respectively. Do they add novel toxic elements to the established toxicity of the expanded polyGln proteins, thus participating in polyGln disorder pathogenesis? What is the role of CAG repeats in the mechanism of ribosomal frameshifting? These and other emerging questions are addressed in this chapter.

Ribosomal frameshifting occurs both *in vitro* and *in vivo*

Translational frameshifting within the ATXN3 and HTT transcripts occurs both *in vitro* and *in vivo.* Frameshifted species containing polyAla tracts have been detected in the brain and lymphoblastoid cells of SCA3 patients as well as in transgenic *Drosophila* models and in various mammalian cells that have been transfected either with full-length or truncated *ATXN3* sequence with expanded CAG repeats (Supplementary Table S3 and Figure S3) (26–28). In the case of HD, frameshift products that contain polyAla or polySer tracts were observed in postmortem brains of HD patients and in transgenic HD mice (HD-N171-N82Q) (29). Only polyAla-containing proteins were observed in other studies where two different cell lines, neuroblastoma N2a and epithelial HeLa cells, were transfected with the CAG repeat of *Htt* exon 1 (Supplementary Table S3) (30). Immunodetection with specific C-terminal antibodies accompanied by Western blotting was mostly used to detect abnormal SCA3 proteins. In the case of HD, valuable information concerning the mechanism of frameshifting was also obtained with the use of MALDI-TOF mass spectrometry (MS) (30). Frameshifting occurred with an equal stochastic probability at any CAG codon within the expanded CAG repeat, giving rise to proteins with various hybrid polyGln/polyAla stretches (30). As shown by MALDI-TOF-MS analysis, frameshifted proteins formed in N2a cells expressing Htt51Q contained Gln:Ala ratios ranging from Gln1:Ala49 to Gln48:Ala2 in their repeat tract. This finding indicates that a (−1) frameshift can occur at any CAG codon within the CAG stretch, and its frequency is modulated by the number of CAG codons to be translated (30).

Factors that may be responsible for frameshifting

The frequency of frameshifting along the expanded CAG tract of ATXN3 and HTT is dependent on repeat length, and the formation of intranuclear inclusions containing frameshifted proteins increases with the number of repeats (26,27,30). In SCA3 patients, protein aggregates were observed with ≥75 CAGs, whereas in transfected COS7 cells, they were abundant after protein expression from a construct containing 82 CAGs and rare from a construct containing at 37 CAGs. In the case of the shortest constructs, with 14 CAGs, whether frameshifting products were observed depended on the detection method (Supplementary Table S3 and Figure S3) (27). Interestingly, not all polyGln-positive cells contained frameshifted proteins; however, the frequency of such cells increased dramatically as the CAG repeat length increased (26). Flanking sequences do not influence frameshifting within the mutant *ATXN3* transcript (27).

Frameshifting frequency also depends on the experimental model, with higher frequencies in transfected cells compared with *in vivo* models. Frameshifted products were observed in intranuclear aggregates within the caudate/putamen of brains from HD patients and in the cortex of HD transgenic mice (29). However, frameshift-positive aggregates represented a mere 4% of the total huntingtin-positive inclusions, as demonstrated in the mouse cortex via double immunofluorescence. As a result of occurrence at low level, no frameshift products were detected on western blots of murine brains. Frameshifted huntingtin was also not detected in stable, inducible PC12 cells even after expression of an *Htt* exon 1 Q74 fragment for 1 month, when the vast majority of cells contained Htt-Gln aggregates as detected by immunocytochemistry (Supplementary Table S3). This result suggests that huntingtin frameshifting is time-dependent and occurs at low levels. When N2a cells stably expressing Htt exon 1 with Q65 fragments were additionally transfected with constructs containing 51 CAG repeats, the frequency of frameshifting was >60% (30).

It was demonstrated that (−1) frameshifting does not occur within long glutamine coding CAA repeats unlike the case for long CAG repeats. Therefore, no alterations in nuclear morphology were observed upon expression of long CAA repeats, while COS 1 cells transfected with the (CAG)78 repeat displayed an altered morphology and exhibited vacuolization that was preferentially associated with cytoplasmic protein accumulation (26). Consistent with these data, Stochmanski *et al.* determined that in a *Drosophila* model of SCA3, replacing the CAG92 repeat

tract with CAA96 eliminated overt phenotypic anomalies that were characterized by disruption of morphology and pigmentation of the eyes (28). Both CAG and CAA repeats encode glutamine and differ only in their ability to form hairpin structures, which may suggest that this factor is required for frameshifting.

Frameshifted proteins are cytotoxic and affect polyQ aggregation kinetics

In contrast to RAN translation, there are more premises that polyAla-containing proteins produced during frameshifting are cytotoxic. In both SCA3 and HD, frameshifted proteins are present in cells harboring polyQ aggregates and contribute to cell toxicity by changing cell morphology and inducing cell death. They accumulate in intranuclear inclusions and can also be found as perinuclear and cytoplasmic aggregates that colocalize with ubiquitin, a well-known component of protein inclusions in polyGln disorders (73).

Frameshifted polyAla proteins apparently add to polyGln toxicity in cells with expanded CAG repeats. In contrast to the slow, progressive accumulation that has been detected for CAG/Gln constructs, transfection with GCA/Ala constructs results in an early and rapid accumulation of Ala-containing products and a more severe phenotype. The presence of a transgene expressing the polyAla tract without the context of the ataxin-3 protein is sufficient for aggregate formation, which produces a toxic phenotype. It has been shown that (−1) frameshifting events and the concomitant production of polyAla-containing ataxin-3 are key factors that contribute to toxic eye phenotype development, as observed in the *Drosophila* model of SCA3. The *in vivo* expression of the polyGln-containing ataxin-3 alone was not sufficient to cause a degenerative phenotype in the fly (28).

Frameshifted (−1) Htt products containing polyGln/polyAla tracts interacted with nonframeshifted HttGln proteins and caused marked changes in their aggregation with expanded polyGln both *in vivo* and *in vitro*, and this effect depended strongly on the Gln:Ala ratio (30). *In vitro*-formed inclusions (in N2a and HeLa cells) were either ring-shaped and surrounded the core huntingtin (Htt103Gln) aggregates or were dense and punctate, which appeared in the vicinity of the ring inclusions but did not fuse with them. Both types of inclusions contained frameshifted species and were frequently located in the vicinity of cytoplasmic or nuclear membranes. Certain perinuclear aggregates formed local hollows in the nuclear membrane, altering its structure (30).

What is the role of the frameshifted species generated in SCA3 and HD, and are they relevant to the pathogenesis of these diseases? Previously, both polySer- and polyAla-containing proteins were shown to be modifiers of mutant huntingtin toxicity (79). It was also proposed that translational frameshifting within expanded CAG stretches may contribute to the observed variations in HD course and onset, leading to cell-selective neurotoxicity on both the cellular and organismal levels. This phenomenon could at least partially account for the degeneration of selective neuronal populations in patients, although the disease-related polyQ proteins are widely expressed. Furthermore, tissue-specific variations in polyQ length may contribute to cell-selective pathogenesis because the CAG stretch is several repeats longer in the brain than in other tissues even within a single individual (80,81), and the frequency of frameshifting increases with repeat length. Taken together, frameshifted proteins are important factors that contribute to polyGln aggregate toxicity.

Mechanisms that may govern translational frameshifting

Since the discovery of CAG repeat frameshifting, important questions regarding its frequency, mechanism and role in pathogenesis have arisen. These questions include the following: What triggers the frameshifting? Does it depend on the length of the CAG repeat encoding the polyQ stretch? Is it cell-type specific? Is frameshifting limited only to SCA3 and HD? How do the frameshifted species modulate the aggregation of the remaining pool of proteins with expanded polyQ tracts? What is the precise effect of the frameshifted proteins on the well-acknowledged toxicity of polyGln-bearing proteins?

Programmed ribosomal frameshifting (PRF) is a common mechanism that increases the protein-coding capacity of small genomes such as viruses and mitochondria. Frameshifting events were also observed in the case of some mammalian genes (82,83). Three elements within mRNA have been shown to be required for the efficient induction of (−1) PRF: a 7-nt slippery sequence (84), a short 5–12 nt spacer and a stimulatory structure that can be a pseudoknot, hairpin structure, G-quadruplex (85–89) or mRNA duplex formed with the antisense oligonucleotide (90). The nature of the slippery sequence facilitates slippage of the tRNAs within the P and A sites on the mRNA, whereas the stimulatory structure provides an energetic barrier that causes an elongating ribosome to pause on the slippery sequence. Several models have been proposed to explain the mechanism of (−1) PRF (91–94), and it seems that structural elements of mRNA play an important role in this process, e.g. the stability of the first 3 or 4 bp at the base of the hairpin stem was shown to be critical for shifting the ribosome by these structures (87,95). Jacks *et al.* proposed 'the two tRNA simultaneous slippage model,' in which the shift occurs after delivery of the aa-tRNA to the A site but before peptidyl transfer (91). A downstream stimulatory motif inhibits the EF2-driven movement of the ribosome. The resulting tension is relieved by unpairing of the aa-tRNA and peptidyl-tRNAs from the mRNA, slipping of the mRNA by one base and re-pairing of the tRNAs in the (−1) frame (91,92,96). A second model proposed by Weiss *et al.* assumes that (−1) PRF occurs during the translocation step (85,97). Briefly, after peptide bond formation, the deacetylated-tRNA and the peptidyl-tRNAs occupy, respectively, the P/E and A/P sites. They can unpair from the mRNA due to the physical barrier formed by the stimulatory structure, move, and re-pair in the new reading frame. In the third model, the (−1) PRF is triggered by incomplete translocation and depends on the interactions of tRNAs with the A, P and E sites on the ribosome (93).

It remains to be established whether frameshifting on CAG repeats results from an incomplete 2-base translo-

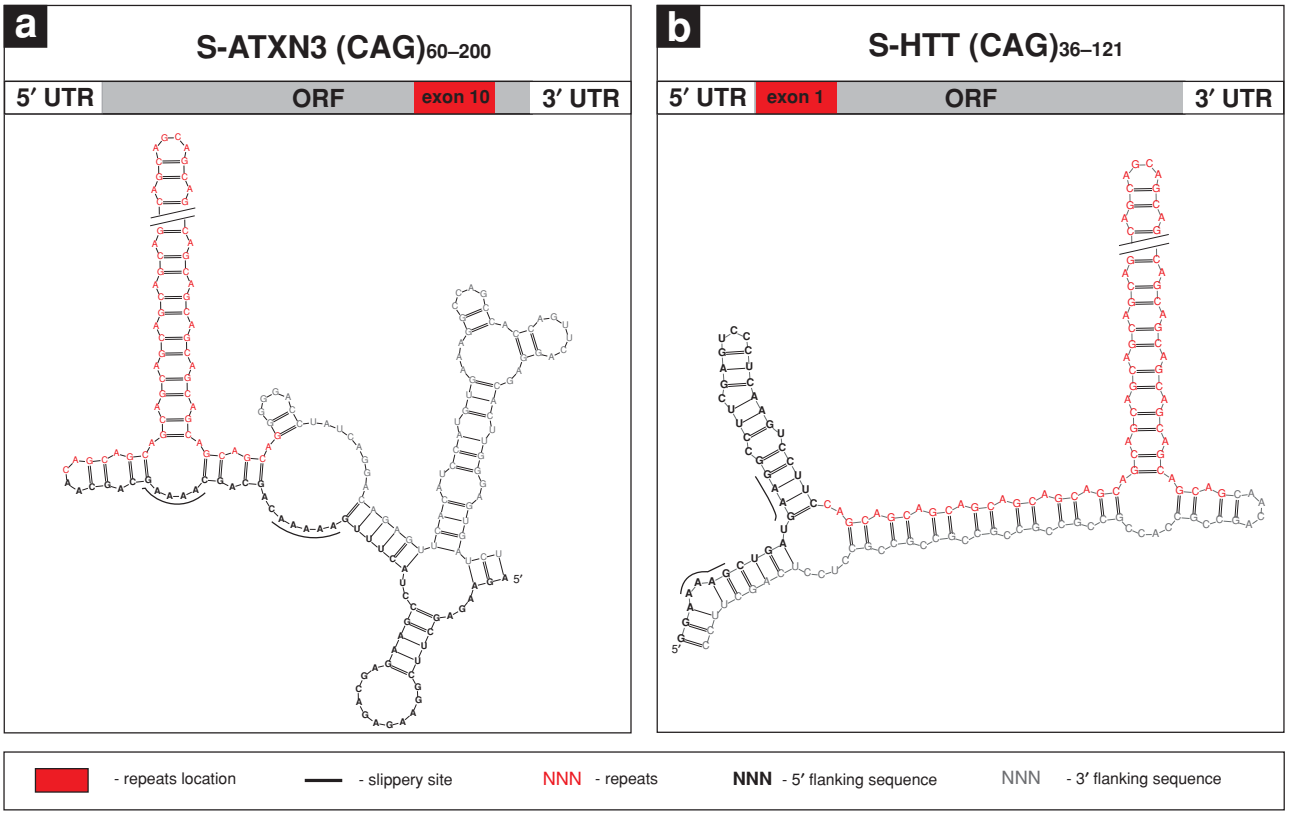

Figure 3. *In silico* structure modeling of transcripts with expanded CAG repeats in ATXN3 and HTT. Structures adopted by **(a)** ATXN3/SCA3 and **(b)** HTT/HD sense transcripts with expanded CAG repeats were modeled *in silico* with Mfold and visualized by Varna software. The modeling was performed for the transcript with 50 CAG repeats and, to obtain more relevant results, in two variants of the flanking sequence 200 and 100 nt each side. For figure clarity, we have shortened the hairpins formed by expanded repeats, as well as the flanking sequence, to ∼50 nt of 5′ flank. Potential frameshifting sites are indicated by a black line.

cation event ((+2) frameshifting) or ribosome slippage by one base in the 5′ direction ((−1) frameshifting). Both these processes generate the same GCA frame encoding alanine; however, different nomenclature used by the authors may be confusing. Because frameshifting has been shown to occur in both SCA3 and HD, the most straightforward explanation is that the CAG repeat sequences of, respectively, ATXN3 and HTT are particularly susceptible to such repositions of reading frames, as experimental evidence demonstrated a higher frequency of frameshifting on longer repeats of both transcripts (26,28–30). However, as the structural requirements of RNA that induce frameshifting are not sufficiently known, it seems plausible that the hairpin structures formed by elongated CAG repeats of ATXN3 and HTT (Figure 3) (98–100) may serve as frameshifting stimulatory sequences, leading to ribosome pausing (Figure 4). Pseudoknots are a common type of stimulatory motif, but hairpin structures are also used by some viruses, e.g. HIV-1 (101,102). During translation elongation on expanded CAG repeat tracts, multiple hairpin structures may fold and unfold by moving the ribosomal complex. In investigating mechanisms of ribosomal frameshifting within repeated sequences, it would be interesting to determine whether the influence of CAA interruptions which are known to alter structure of CAG repeat hairpins (57) affects the frameshifting frequency.

Another explanation of frameshifting might rely on ribosomal slippage and the occurrence of slippery sequences within ATXN3 and HTT transcripts (Figure 3) (103). However, as frameshifting occurs stochastically at any CAG codon within the repeat stretch, as demonstrated in HD (30), it seems unlikely that slippery sequences proceeding a repeat tract play a mechanistic role as a trigger of changes in reading frames. Is it possible that CAG repeats themselves act as unusual slippery sequences? Viral slippery sequences composed of the heptameric motif N NNW WWH (IUPAC) enable repairing of the nonwobble bases of both the aa-tRNAs and peptidyl-tRNAs with the (−1) codons (104). In light of this fact, it remains a mystery what factors would help to maintain interactions between the frameshifted 5′GCA3′ codon and the 5′CUG3′ anticodon, where no Watson–Crick pairs would be possible (Figure 4) (103).

Recently, new experimental evidence on the mechanism of frameshifting at expanded CAG repeats has been published (30). The formation of various transframe-encoded species generated *via* frameshifting at long CAG repeats in HD is proposed to be a consequence of a depletion of the charged glutaminyl-transfer RNA tRNAGln-CUG that pairs exclusively with the CAG codons. It has been shown that the amount of tRNAGln decreases with increasing length of the encoded CAG sequences (hungry codons), and

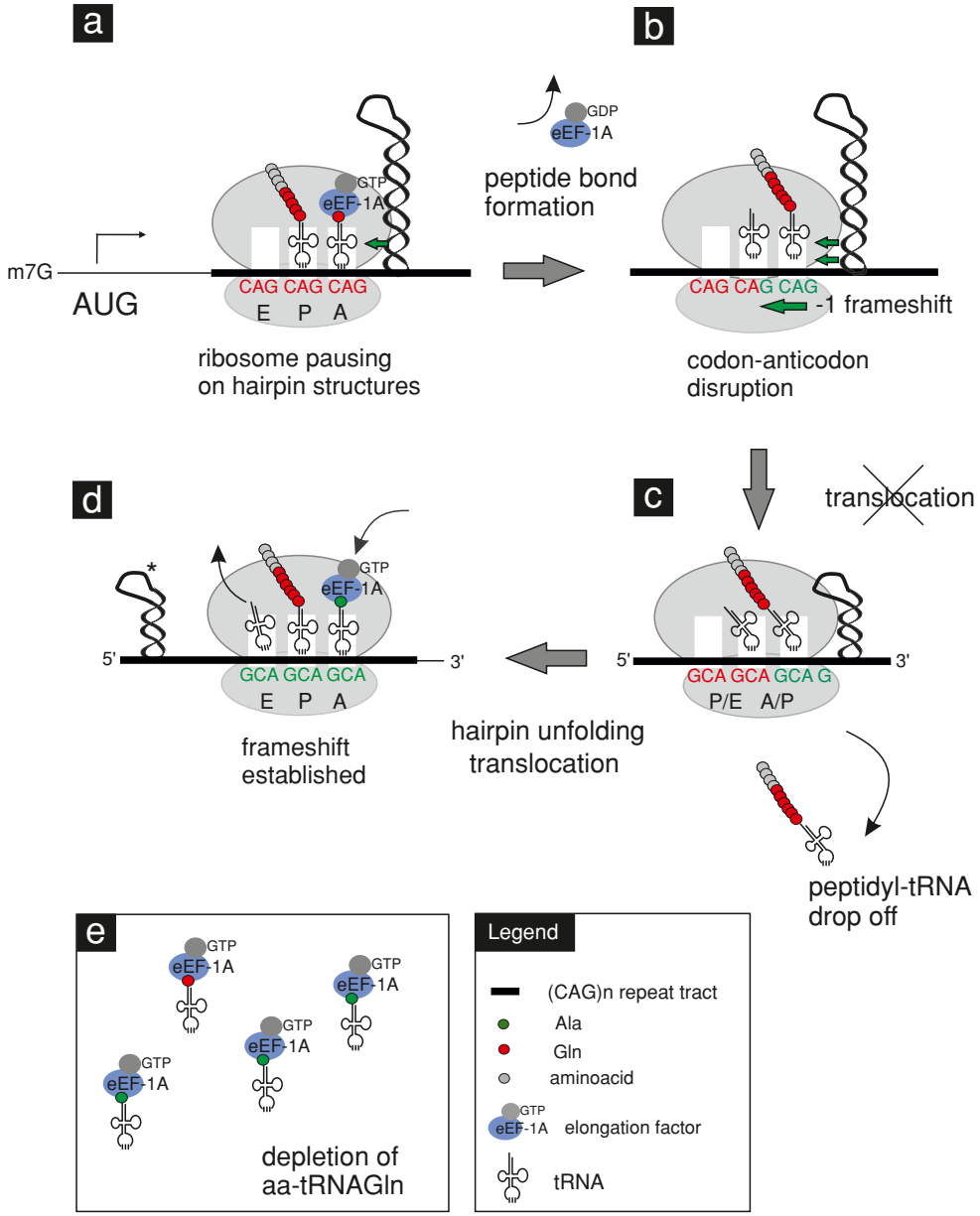

Figure 4. Speculated mechanisms of ribosomal frameshifting at CAG repeats in SCA3 and HD. Translation elongation is a cyclic process of mRNA-dependent peptide bond formation with the use of ribosomes, aminoacyl-tRNAs, energy and specific protein factors. At the start of each cycle, the complex of correct aminoacyl-tRNA and the elongation factor eEF-1AGTP are positioned in the ribosome decoding center (A site). **(a)** During this step, the ribosome P site binds the initiator Met-tRNA or peptidyl-tRNA. After codon recognition, followed by GTP hydrolysis, the amino acid from aminoacyl-tRNAs is transferred to the C-terminus of the growing polypeptide chain. **(b)** The reaction leaves an uncharged tRNA in the ribosomal P site and the new peptidyl-tRNA in the A site. Normally, during the next step, the ribosome translocates one codon forward on the mRNA, and deacylated tRNA is released from the exit (E) site of the ribosome. During translation elongation of expanded CAG-containing transcripts, the ribosomal complex may pause on the hairpin structure formed by the elongated CAG repeat tract. Accommodation of aa-tRNA may pull the downstream mRNA into the ribosome, thus creating tension between the hairpin and slippery site. This leads to the dissociation of peptidyl-tRNA and repairs mRNA in the (−1) frame (GCA). **(c)** The weak interaction between the codon-anticodon GCA:CUG may result in peptidyl-tRNA drop off. **(d)** After unwinding of the hairpin, elongation proceeds in a new frame. **(e)** An increased demand for glutaminyl-tRNAGln-CUG may be the additional factor that stimulates changes in the reading frame. * - refolding hairpin.

in N2A cells expressing 65 or 103 CAG repeats, a significant increase in the frequency of frameshifting was observed from, respectively, 60 to 80%. Moreover, the frameshifting frequency varies among different cell lines. In N2a cells and HeLa cells, which differ in their intrinsic concentration of tRNAGln-CUG, the higher frameshifting frequency occurs in the N2a cells, which have a lower amount of tRNAGln-CUG than the HeLa cell line (Supplementary Table S3). Interestingly, the concentration of tRNAGln-CUG also differs among different brain areas, and as shown in HD mice, the tRNAGln-CUG concentration is lower in the striatal and hippocampal tissues than in the cortical and cerebellar regions (30). This fact raises the interesting question of whether tRNAGln-CUG levels also vary among different brain tissues in HD patients, thus contributing to the cell-selective pathology of the disease. Because the most prominent early effects in HD have been observed in the striatum, even though striatal neurons do not selectively express higher levels of HTT mRNA (105), this possibility could provide a mechanistic explanation of the selective pathology and importantly disclose a new therapeutic target. However, this issue remains to be examined.

The mechanism of frameshifting in SCA3 and HD seems to be more similar to viral −1 PRF since mostly polyAla-containing proteins are generated. However, we cannot exclude +1 PRF events, as observed in *Escherichia coli* and yeasts (106–108), because polySer-containing proteins were detected in HD (29). The +1 PRF is also enhanced by the presence of 'hungry codons' in the A site (109,110). This information would support the idea suggesting that depletion of cognate charged tRNAGln-CUG is a cause of frameshifting within Htt transcript, however polySer proteins resulting from +1 frameshifting were much rarely detected in this study (30).

Whether elongated CAG repeats act as slippery or stimulatory sequences or both remains to be determined. Nonetheless, the information gathered thus far seems to confirm the role of CAG repeats in generating frameshifted proteins because the frequency of frameshifted events increases with the length of the CAG tract; long CAA repeats that do not form hairpins do not give rise to frameshifted proteins; frameshifting events can occur at every position within the CAG tract; frameshifting is cell- and tissue-type dependent and (v) frameshifted proteins are toxic to cells and affect polyQ aggregation kinetics. It is possible that frameshifting is a more widespread phenomenon and also occurs on the expanded CAG repeats involved in SCA1, 2, 3, 6, 7, 17, DRPLA and SBMA. Experiments with the use of dedicated antibodies (allowing the detection of Ala ((−1) frame) and Ser ((+1) frame)-containing proteins) could clarify this matter.

CONCLUDING REMARKS

As discussed in this review, mounting evidence demonstrates that simple repeat expansions causing several human neurological diseases trigger noncanonical translation initiation and elongation, and two unusual processes, RAN translation and frameshifting, are recognized across the repeats. Protein products of RAN translation and frameshifting were detected in patient tissues and in mouse and *Drosophila* models of the disorders in addition to numerous human cell lines after transfection with specific constructs. However, the methods used to study these processes and detect their products have limitations that give raise to concerns and further questions. For example, it cannot be excluded that some antibodies used to detect specific RAN translation or frameshifting products may recognize both entities. Thus, the results obtained with commonly used immunological methods need to be confirmed by other methods such as MS. Strong promoters used in genetic constructs may explain the higher frequency of frameshifting observed in cellular models compared with patient tissues as well as detection of RAN translation and frameshifting products in disease models expressing nonpathological numbers of repeats.

Although RAN translation and frameshifting are linked to distinct translation steps, they demonstrate several common features. The most evident is the requirement for structured repeat tracts, dependence on the repeated sequence length, production of proteins in various frames and formation of intracellular inclusions. Among their distinctive feature is the fact that RAN translation has been demonstrated only at sequences that were considered previously to be noncoding, whereas evidence for frameshifting on expanded repeats has been described for coding sequences. In spite of the fact that some features regarding RAN translation and frameshifting have been already revealed, their exact mechanism remains elusive.

Another unresolved issue is whether RAN translation and frameshifting products are toxic to cells. In the absence of more direct evidence, the inclusion bodies harboring RAN and frameshifted proteins detected in patient tissues, in addition to the decreased viability of cells transfected with constructs expressing RAN-translated and frameshifted proteins, is only suggestive of their toxicity.

Finally, it is yet to be determined whether RAN translation is a unique phenomenon or noncanonical translation that gives rise to novel polypeptides and short proteins is more frequent in cells. Recent studies demonstrate that the human proteome contains a multitude of short ORF-encoded peptides (SEPs), which share certain similarities with RAN-translated proteins (111,112). Certain SEPs have a regulatory function and affect protein expression by controlling translation re-initiation at downstream initiation codons and by activating the nonsense-mediated mRNA decay pathway (113). On the other hand, various cryptic peptides, products of alternative translation initiation, act as alternative sources of antigens for major histocompatibility complex class I components (114,115). Whether such functions could be performed by RAN translation and frameshifting products is another issue to be addressed.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

FUNDING

National Science Centre [2012/06/A/NZ1/00094 to W.J.K.], European Regional Development Fund within the Innovative Economy Programme [POIG.01.03.01-30-098/08 to W.J.K.], Polish Ministry of Science and Higher

Education [N N401 572140 to M.W., N N302 633240 to M.O.].
Conflict of interest statement. None declared.

REFERENCES

1. Lopez Castel,A., Cleary,J.D. and Pearson,C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.*, **11**, 165–170.
2. Ranum,L.P. and Day,J.W. (2002) Dominantly inherited, non-coding microsatellite expansion disorders. *Curr. Opin. Genet. Dev.*, **12**, 266–271.
3. Orr,H.T. and Zoghbi,H.Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci.*, **30**, 575–621.
4. Hsu,R.J., Hsiao,K.M., Lin,M.J., Li,C.Y., Wang,L.C., Chen,L.K. and Pan,H. (2011) Long tract of untranslated CAG repeats is deleterious in transgenic mice. *PLoS One*, **6**, e16417.
5. Ho,T.H., Savkur,R.S., Poulos,M.G., Mancini,M.A., Swanson,M.S. and Cooper,T.A. (2005) Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell Sci.*, **118**, 2923–2933.
6. Li,L.B., Yu,Z., Teng,X. and Bonini,N.M. (2008) RNA toxicity is a component of ataxin-3 degeneration in Drosophila. *Nature*, **453**, 1107–1111.
7. Wang,L.C., Chen,K.Y., Pan,H., Wu,C.C., Chen,P.H., Liao,Y.T., Li,C., Huang,M.L. and Hsiao,K.M. (2011) Muscleblind participates in RNA toxicity of expanded CAG and CUG repeats in Caenorhabditis elegans. *Cell. Mol. Life Sci.*, **68**, 1255–1267.
8. Mykowska,A., Sobczak,K., Wojciechowska,M., Kozlowski,P. and Krzyzosiak,W.J. (2011) CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Res.*, **39**, 8938–8951.
9. Wojciechowska,M. and Krzyzosiak,W.J. (2011) Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum. Mol.Genet.*, **20**, 3811–3821.
10. Galka-Marciniak,P., Urbanek,M.O. and Krzyzosiak,W.J. (2012) Triplet repeats in transcripts: structural insights into RNA toxicity. *Biol. Chem.*, **393**, 1299–1315.
11. Gendron,T.F., Belzil,V.V., Zhang,Y.J. and Petrucelli,L. (2014) Mechanisms of toxicity in C9FTLD/ALS. *Acta Neuropathol.*, **127**, 359–376.
12. Moseley,M.L., Zu,T., Ikeda,Y., Gao,W., Mosemiller,A.K., Daughters,R.S., Chen,G., Weatherspoon,M.R., Clark,H.B., Ebner,T.J. *et al.* (2006) Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.*, **38**, 758–769.
13. Batra,R., Charizanis,K. and Swanson,M.S. (2010) Partners in crime: bidirectional transcription in unstable microsatellite disease. *Hum. Mol. Genet.*, **19**, R77–R82.
14. Zu,T., Gibbens,B., Doty,N.S., Gomes-Pereira,M., Huguet,A., Stone,M.D., Margolis,J., Peterson,M., Markowski,T.W., Ingram,M.A. *et al.* (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 260–265.
15. Zu,T., Liu,Y., Banez-Coronel,M., Reid,T., Pletnikova,O., Lewis,J., Miller,T.M., Harms,M.B., Falchook,A.E., Subramony,S.H. *et al.* (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E4968–E4977.
16. Todd,P.K., Oh,S.Y., Krans,A., He,F., Sellier,C., Frazer,M., Renoux,A.J., Chen,K.C., Scaglione,K.M., Basrur,V. *et al.* (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*, **78**, 440–455.
17. Gendron,T.F., Bieniek,K.F., Zhang,Y.J., Jansen-West,K., Ash,P.E., Caulfield,T., Daughrity,L., Dunmore,J.H., Castanedes-Casey,M., Chew,J. *et al.* (2013) Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.*, **126**, 829–844.
18. Ash,P.E., Bieniek,K.F., Gendron,T.F., Caulfield,T., Lin,W.L., Dejesus-Hernandez,M., van Blitterswijk,M.M., Jansen-West,K., Paul,J.W. 3rd, Rademakers,R. *et al.* (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, **77**, 639–646.
19. Mori,K., Weng,S.M., Arzberger,T., May,S., Rentzsch,K., Kremmer,E., Schmid,B., Kretzschmar,H.A., Cruts,M., Van Broeckhoven,C. *et al.* (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science*, **339**, 1335–1338.
20. Donnelly,C.J., Zhang,P.W., Pham,J.T., Heusler,A.R., Mistry,N.A., Vidensky,S., Daley,E.L., Poth,E.M., Hoover,B., Fines,D.M. *et al.* (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron*, **80**, 415–428.
21. Almeida,S., Gascon,E., Tran,H., Chou,H.J., Gendron,T.F., Degroot,S., Tapper,A.R., Sellier,C., Charlet-Berguerand,N., Karydas,A. *et al.* (2013) Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol.*, **126**, 385–399.
22. Mori,K., Arzberger,T., Grasser,F.A., Gijselinck,I., May,S., Rentzsch,K., Weng,S.M., Schludi,M.H., van der Zee,J., Cruts,M. *et al.* (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol.*, **126**, 881–893.
23. Jackson,R.J., Hellen,C.U. and Pestova,T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol*, **11**, 113–127.
24. Sonenberg,N. and Hinnebusch,A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, **136**, 731–745.
25. Firth,A.E. and Brierley,I. (2012) Non-canonical translation in RNA viruses. *J. Gen. Virol.*, **93**, 1385–1409.
26. Toulouse,A., Au-Yeung,F., Gaspar,C., Roussel,J., Dion,P. and Rouleau,G.A. (2005) Ribosomal frameshifting on MJD-1 transcripts with long CAG tracts. *Hum. Mol. Genet.*, **14**, 2649–2660.
27. Gaspar,C., Jannatipour,M., Dion,P., Laganiere,J., Sequeiros,J., Brais,B. and Rouleau,G.A. (2000) CAG tract of MJD-1 may be prone to frameshifts causing polyalanine accumulation. *Hum. Mol. Genet.*, **9**, 1957–1966.
28. Stochmanski,S.J., Therrien,M., Laganiere,J., Rochefort,D., Laurent,S., Karemera,L., Gaudet,R., Vyboh,K., Van Meyel,D.J., Di Cristo,G. *et al.* (2012) Expanded ATXN3 frameshifting events are toxic in Drosophila and mammalian neuron models. *Hum. Mol. Genet.*, **21**, 2211–2218.
29. Davies,J.E. and Rubinsztein,D.C. (2006) Polyalanine and polyserine frameshift products in Huntington's disease. *J. Med. Genet.*, **43**, 893–896.
30. Girstmair,H., Saffert,P., Rode,S., Czech,A., Holland,G., Bannert,N. and Ignatova,Z. (2013) Depletion of cognate charged transfer RNA causes translational frameshifting within the expanded CAG stretch in huntingtin. *Cell Rep.*, **3**, 148–159.
31. Mann,D.M., Rollinson,S., Robinson,A., Bennion Callister,J., Thompson,J.C., Snowden,J.S., Gendron,T., Petrucelli,L., Masuda-Suzukake,M., Hasegawa,M. *et al.* (2013) Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathol. Commun.*, **1**, 68.
32. Mackenzie,I.R., Arzberger,T., Kremmer,E., Troost,D., Lorenzl,S., Mori,K., Weng,S.M., Haass,C., Kretzschmar,H.A., Edbauer,D. *et al.* (2013) Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol.*, **126**, 859–879.
33. Cooper-Knock,J., Walsh,M.J., Higginbottom,A., Robin Highley,J., Dickman,M.J., Edbauer,D., Ince,P.G., Wharton,S.B., Wilson,S.A., Kirby,J. *et al.* (2014) Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*, **137**, 2040–2051.
34. Murray,M.E., Bieniek,K.F., Banks Greenberg,M., DeJesus-Hernandez,M., Rutherford,N.J., van Blitterswijk,M., Niemantsverdriet,E., Ash,P.E., Gendron,T.F., Kouri,N. *et al.* (2013) Progressive amnestic dementia, hippocampal sclerosis, and mutation in C9ORF72. *Acta Neuropathol.*, **126**, 545–554.
35. Pletnikova,O., Sloane,K.L., Renton,A.E., Traynor,B.J., Crain,B.J., Reid,T., Zu,T., Ranum,L.P., Troncoso,J.C., Rabins,P.V. *et al.* (2014) Hippocampal sclerosis dementia with the C9ORF72

hexanucleotide repeat expansion. *Neurobiol. Aging*, **35**, 2419.e17–2419.e21.
36. Proudfoot,M., Gutowski,N.J., Edbauer,D., Hilton,D.A., Stephens,M., Rankin,J. and Mackenzie,I.R. (2014) Early dipeptide repeat pathology in a frontotemporal dementia kindred with C9ORF72 mutation and intellectual disability. *Acta Neuropathol.*, **127**, 451–458.
37. Liu,E.Y., Russ,J., Wu,K., Neal,D., Suh,E., McNally,A.G., Irwin,D.J., Van Deerlin,V.M. and Lee,E.B. (2014) C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol.*, doi:10.1007/s00401-014-1286-y.
38. Sareen,D., O'Rourke,J.G., Meera,P., Muhammad,A.K., Grant,S., Simpkinson,M., Bell,S., Carmona,S., Ornelas,L., Sahabian,A. *et al.* (2013) Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci. Transl. Med.*, **5**, 208ra149.
39. Kwon,I., Xiang,S., Kato,M., Wu,L., Theodoropoulos,P., Wang,T., Kim,J., Yun,J., Xie,Y. and McKnight,S.L. (2014) Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science*, pii:1254917.
40. Landles,C., Sathasivam,K., Weiss,A., Woodman,B., Moffitt,H., Finkbeiner,S., Sun,B., Gafni,J., Ellerby,L.M., Trottier,Y. *et al.* (2010) Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease. *J Biol Chem*, **285**, 8808–8823.
41. Ross,C.A. and Tabrizi,S.J. (2011) Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol.*, **10**, 83–98.
42. Arrasate,M., Mitra,S., Schweitzer,E.S., Segal,M.R. and Finkbeiner,S. (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature*, **431**, 805–810.
43. Wolfe,K.J. and Cyr,D.M. (2011) Amyloid in neurodegenerative diseases: friend or foe? *Semin. Cell Dev. Biol.*, **22**, 476–481.
44. Napierala,M., Michalowski,D., de Mezer,M. and Krzyzosiak,W.J. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.*, **33**, 451–463.
45. Schaeffer,C., Bardoni,B., Mandel,J.L., Ehresmann,B., Ehresmann,C. and Moine,H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *Embo J.*, **20**, 4803–4813.
46. Khateb,S., Weisman-Shomer,P., Hershco-Shani,I., Ludwig,A.L. and Fry,M. (2007) 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.
47. Reddy,K., Zamiri,B., Stanley,S.Y., Macgregor,R.B. Jr. and Pearson,C.E. (2013) The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.*, **288**, 9860–9866.
48. Chiang,P.W., Carpenter,L.E. and Hagerman,P.J. (2001) The 5′-untranslated region of the FMR1 message facilitates translation by internal ribosome entry. *J. Biol. Chem.*, **276**, 37916–37921.
49. Kieft,J.S. (2008) Viral IRES RNA structures and ribosome interactions. *Trends Biochem. Sci.*, **33**, 274–283.
50. Filbin,M.E. and Kieft,J.S. (2009) Toward a structural understanding of IRES RNA function. *Curr. Opin. Struct. Biol.*, **19**, 267–276.
51. Krzyzosiak,W.J., Sobczak,K., Wojciechowska,M., Fiszer,A., Mykowska,A. and Kozlowski,P. (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.*, **40**, 11–26.
52. Sobczak,K., Michlewski,G., de Mezer,M., Kierzek,E., Krol,J., Olejniczak,M., Kierzek,R. and Krzyzosiak,W.J. (2010) Structural diversity of triplet repeat RNAs. *J. Biol. Chem.*, **285**, 12755–12764.
53. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2009) Structural insights into CUG repeats containing the 'stretched U-U wobble': implications for myotonic dystrophy. *Nucleic Acids Res.*, **37**, 4149–4156.
54. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2010) Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res.*, **38**, 8370–8376.
55. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res.*, **39**, 7308–7315.
56. Jasinska,A., Michlewski,G., de Mezer,M., Sobczak,K., Kozlowski,P., Napierala,M. and Krzyzosiak,W.J. (2003) Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Res.*, **31**, 5463–5468.
57. Sobczak,K. and Krzyzosiak,W.J. (2005) CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. *J. Biol. Chem.*, **280**, 3898–3910.
58. Payet,L. and Huppert,J.L. (2012) Stability and structure of long intramolecular G-quadruplexes. *Biochemistry*, **51**, 3154–3161.
59. Wong,H.M. and Huppert,J.L. (2009) Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.*, **5**, 1713–1719.
60. Bugaut,A. and Balasubramanian,S. (2012) 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*, **40**, 4727–4741.
61. Peabody,D.S. (1989) Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.*, **264**, 5031–5035.
62. Martinez-Salas,E., Lozano,G., Fernandez-Chamorro,J., Francisco-Velilla,R., Galan,A. and Diaz,R. (2013) RNA-binding proteins impacting on internal initiation of translation. *Int. J. Mol. Sci.*, **14**, 21705–21726.
63. Pestova,T.V., de Breyne,S., Pisarev,A.V., Abaeva,I.S. and Hellen,C.U. (2008) eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *Embo J.*, **27**, 1060–1072.
64. Terenin,I.M., Dmitriev,S.E., Andreev,D.E. and Shatsky,I.N. (2008) Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nat. Struct. Mol. Biol.*, **15**, 836–841.
65. Dmitriev,S.E., Terenin,I.M., Andreev,D.E., Ivanov,P.A., Dunaevsky,J.E., Merrick,W.C. and Shatsky,I.N. (2010) GTP-independent tRNA delivery to the ribosomal P-site by a novel eukaryotic translation factor. *J. Biol. Chem.*, **285**, 26779–26787.
66. Skabkin,M.A., Skabkina,O.V., Dhote,V., Komar,A.A., Hellen,C.U. and Pestova,T.V. (2010) Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev.*, **24**, 1787–1801.
67. Komar,A.A. and Hatzoglou,M. (2011) Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle*, **10**, 229–240.
68. Miller,J.W., Urbinati,C.R., Teng-Umnuay,P., Stenberg,M.G., Byrne,B.J., Thornton,C.A. and Swanson,M.S. (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *Embo J.*, **19**, 4439–4448.
69. Houseley,J.M., Wang,Z., Brock,G.J., Soloway,J., Artero,R., Perez-Alonso,M., O'Dell,K.M. and Monckton,D.G. (2005) Myotonic dystrophy associated expanded CUG repeat muscleblind positive ribonuclear foci are not toxic to Drosophila. *Hum. Mol. Genet.*, **14**, 873–883.
70. Fardaei,M., Larkin,K., Brook,J.D. and Hamshere,M.G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.*, **29**, 2766–2771.
71. Huichalaf,C., Sakai,K., Jin,B., Jones,K., Wang,G.L., Schoser,B., Schneider-Gold,C., Sarkar,P., Pereira-Smith,O.M., Timchenko,N. *et al.* (2010) Expansion of CUG RNA repeats causes stress and inhibition of translation in myotonic dystrophy 1 (DM1) cells. *Faseb J.*, **24**, 3706–3719.
72. Huichalaf,C., Schoser,B., Schneider-Gold,C., Jin,B., Sarkar,P. and Timchenko,L. (2009) Reduction of the rate of protein translation in patients with myotonic dystrophy 2. *J. Neurosci.*, **29**, 9042–9049.
73. Botta,A., Malena,A., Loro,E., Del Moro,G., Suman,M., Pantic,B., Szabadkai,G. and Vergani,L. (2013) Altered Ca2+ homeostasis and endoplasmic reticulum stress in myotonic dystrophy type 1 muscle cells. *Genes (Basel)*, **4**, 275–292.
74. Ikezoe,K., Nakamori,M., Furuya,H., Arahata,H., Kanemoto,S., Kimura,T., Imaizumi,K., Takahashi,M.P., Sakoda,S., Fujii,N. *et al.* (2007) Endoplasmic reticulum stress in myotonic dystrophy type 1 muscle. *Acta Neuropathol.*, **114**, 527–535.

75. Reineke,L.C., Cao,Y., Baus,D., Hossain,N.M. and Merrick,W.C. (2011) Insights into the role of yeast eIF2A in IRES-mediated translation. *PLoS One*, **6**, e24492.
76. Tsoi,H. and Chan,H.Y. (2013) Expression of expanded CAG transcripts triggers nucleolar stress in Huntington's disease. *Cerebellum*, **12**, 310–312.
77. Tsoi,H., Lau,T.C., Tsang,S.Y., Lau,K.F. and Chan,H.Y. (2012) CAG expansion induces nucleolar stress in polyglutamine diseases. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 13428–13433.
78. Haeusler,A.R., Donnelly,C.J., Periz,G., Simko,E.A., Shaw,P.G., Kim,M.S., Maragakis,N.J., Troncoso,J.C., Pandey,A., Sattler,R. *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, **507**, 195–200.
79. Berger,Z., Davies,J.E., Luo,S., Pasco,M.Y., Majoul,I., O'Kane,C.J. and Rubinsztein,D.C. (2006) Deleterious and protective properties of an aggregate-prone protein with a polyalanine expansion. *Hum. Mol. Genet.*, **15**, 453–465.
80. Gonitel,R., Moffitt,H., Sathasivam,K., Woodman,B., Detloff,P.J., Faull,R.L. and Bates,G.P. (2008) DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 3467–3472.
81. Kennedy,L., Evans,E., Chen,C.M., Craven,L., Detloff,P.J., Ennis,M. and Shelbourne,P.F. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.*, **12**, 3359–3367.
82. Clark,M.B., Janicke,M., Gottesbuhren,U., Kleffmann,T., Legge,M., Poole,E.S. and Tate,W.P. (2007) Mammalian gene PEG10 expresses two reading frames by high efficiency -1 frameshifting in embryonic-associated tissues. *J. Biol. Chem.*, **282**, 37359–37369.
83. Manktelow,E., Shigemoto,K. and Brierley,I. (2005) Characterization of the frameshift signal of Edr, a mammalian example of programmed -1 ribosomal frameshifting. *Nucleic Acids Res.*, **33**, 1553–1563.
84. Brierley,I., Jenner,A.J. and Inglis,S.C. (1992) Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.*, **227**, 463–479.
85. Namy,O., Moran,S.J., Stuart,D.I., Gilbert,R.J. and Brierley,I. (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, **441**, 244–247.
86. Baril,M., Dulude,D., Steinberg,S.V. and Brakier-Gingras,L. (2003) The frameshift stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot. *J. Mol. Biol.*, **331**, 571–583.
87. Yu,C.H., Noteborn,M.H., Pleij,C.W. and Olsthoorn,R.C. (2011) Stem-loop structures can effectively substitute for an RNA pseudoknot in -1 ribosomal frameshifting. *Nucleic Acids Res.*, **39**, 8952–8959.
88. Yu,C.H., Teulade-Fichou,M.P. and Olsthoorn,R.C. (2014) Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res.*, **42**, 1887–1892.
89. Endoh,T., Kawasaki,Y. and Sugimoto,N. (2013) Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew. Chem. Int. Ed. Engl.*, **52**, 5522–5526.
90. Howard,M.T., Gesteland,R.F. and Atkins,J.F. (2004) Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *RNA*, **10**, 1653–1661.
91. Jacks,T. and Varmus,H.E. (1985) Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science*, **230**, 1237–1242.
92. Harger,J.W., Meskauskas,A. and Dinman,J.D. (2002) An "integrated model" of programmed ribosomal frameshifting. *Trends Biochem. Sci.*, **27**, 448–454.
93. Leger,M., Dulude,D., Steinberg,S.V. and Brakier-Gingras,L. (2007) The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. *Nucleic Acids Res.*, **35**, 5581–5592.
94. Liao,P.Y., Choi,Y.S., Dinman,J.D. and Lee,K.H. (2011) The many paths to frameshifting: kinetic modelling and analysis of the effects of different elongation steps on programmed -1 ribosomal frameshifting. *Nucleic Acids Res.*, **39**, 300–312.
95. Mouzakis,K.D., Lang,A.L., Vander Meulen,K.A., Easterday,P.D. and Butcher,S.E. (2013) HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res.*, **41**, 1901–1913.
96. Plant,E.P., Jacobs,K.L., Harger,J.W., Meskauskas,A., Jacobs,J.L., Baxter,J.L., Petrov,A.N. and Dinman,J.D. (2003) The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA*, **9**, 168–174.
97. Weiss,R.B., Dunn,D.M., Shuh,M., Atkins,J.F. and Gesteland,R.F. (1989) E. coli ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.*, **1**, 159–169.
98. Busan,S. and Weeks,K.M. (2013) Role of context in RNA structure: flanking sequences reconfigure CAG motif folding in huntingtin exon 1 transcripts. *Biochemistry*, **52**, 8219–8225.
99. de Mezer,M., Wojciechowska,M., Napierala,M., Sobczak,K. and Krzyzosiaka,W.J. (2011) Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res.*, **39**, 3852–3863.
100. Michlewski,G. and Krzyzosiak,W.J. (2004) Molecular architecture of CAG repeats in human disease related transcripts. *J. Mol. Biol.*, **340**, 665–679.
101. Dulude,D., Baril,M. and Brakier-Gingras,L. (2002) Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res.*, **30**, 5094–5102.
102. Gaudin,C., Mazauric,M.H., Traikia,M., Guittet,E., Yoshizawa,S. and Fourmy,D. (2005) Structure of the RNA signal essential for translational frameshifting in HIV-1. *J. Mol. Biol.*, **349**, 1024–1035.
103. Wills,N.M. and Atkins,J.F. (2006) The potential role of ribosomal frameshifting in generating aberrant proteins implicated in neurodegenerative diseases. *RNA*, **12**, 1149–1153.
104. Kollmus,H., Hentze,M.W. and Hauser,H. (1996) Regulated ribosomal frameshifting by an RNA-protein interaction. *RNA*, **2**, 316–323.
105. Landwehrmeyer,G.B., McNeil,S.M., Dure,L.S.t., Ge,P., Aizawa,H., Huang,Q., Ambrose,C.M., Duyao,M.P., Bird,E.D., Bonilla,E. *et al.* (1995) Huntington's disease gene: regional and cellular expression in brain of normal and affected individuals. *Ann. Neurol.*, **37**, 218–230.
106. Craigen,W.J. and Caskey,C.T. (1986) Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature*, **322**, 273–275.
107. Belcourt,M.F. and Farabaugh,P.J. (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. *Cell*, **62**, 339–352.
108. Morris,D.K. and Lundblad,V. (1997) Programmed translational frameshifting in a gene required for yeast telomere replication. *Curr. Biol.*, **7**, 969–976.
109. Kawakami,K., Pande,S., Faiola,B., Moore,D.P., Boeke,J.D., Farabaugh,P.J., Strathern,J.N., Nakamura,Y. and Garfinkel,D.J. (1993) A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. *Genetics*, **135**, 309–320.
110. Lindsley,D. and Gallant,J. (1993) On the directional specificity of ribosome frameshifting at a "hungry" codon. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 5469–5473.
111. Slavoff,S.A., Mitchell,A.J., Schwaid,A.G., Cabili,M.N., Ma,J., Levin,J.Z., Karger,A.D., Budnik,B.A., Rinn,J.L. and Saghatelian,A. (2013) Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.*, **9**, 59–64.
112. Ma,J., Ward,C.C., Jungreis,I., Slavoff,S.A., Schwaid,A.G., Neveu,J., Budnik,B.A., Kellis,M. and Saghatelian,A. (2014) Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *J. Proteome Res.*, **13**, 1757–1765.
113. Somers,J., Poyry,T. and Willis,A.E. (2013) A perspective on mammalian upstream open reading frame function. *Int. J. Biochem. Cell Biol.*, **45**, 1690–1700.
114. Apcher,S., Daskalogianni,C., Lejeune,F., Manoury,B., Imhoos,G., Heslop,L. and Fahraeus,R. (2011) Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11572–11577.
115. Starck,S.R., Ow,Y., Jiang,V., Tokuyama,M., Rivera,M., Qi,X., Roberts,R.W. and Shastri,N. (2008) A distinct translation initiation mechanism generates cryptic peptides for immune surveillance. *PLoS One*, **3**, e3460.