Non-ATG–initiated translation directed by microsatellite expansions

Tao Zu,a,b,c, Brian Gibbensa,b,c,1, Noelle S. Dotya,b,c,1, Mário Gomes-Pereira2, Aline Huguetb,1, Matthew D. Stoned,g,1, Jamie Margolisb,c,1, Mark Peterson9, Todd W. Markowski10, Melissa A. C. Ingrama,b,c,1, Zhenhong Na, Colleen Forsterj, Walter C. Lowk, Benedikt Schoserd, Nikunj V. Somiaa,b,1, H. Brent Clarkc,i,k, Stephen Schmechel, Peter B. Bitterman9, Geneviève Gourdon, Maurice S. Swanson1, Melinda Moseleya,b,c,1, and Laura P. W. Runam,b,c,2,3

Departments of aGenetics Cell Biology and Development, bBiochemistry, Molecular Biology and Biophysics, cMedicine, dNeurosurgery, Laboratory Medicine and Pathology, and eNeurology, Institute of Human Genetics, fInstitute of Translational Neuroscience, and gCenter for Mass Spectrometry and Proteomics, University of Minnesota Medical School, Minneapolis, MN 55455; jDepartment National de la Santé et de la Recherche Médicale U781, Université Paris Descartes, Hôpital Necker Enfants Malades, 75015 Paris, France; kDepartment of Neurology, Friedrich-Baur Institute, Ludwig Maximilians University Munich, 80336 Munich, Germany; and lDepartment of Molecular Genetics and Microbiology, University of Florida College of Medicine, Gainesville, FL 32610

Edited by C. Thomas Caskey, University of Texas-Houston Health Science Center, Houston, TX, and approved November 12, 2010 (received for review September 9, 2010)

Trinucleotide expansions cause disease by both protein- and RNA-mediated mechanisms. Unexpectedly, we discovered that CAG expansion constructs express homopolymeric polyglutamine, polyalanine, and polyserine proteins in the absence of an ATG start codon. This repeat-associated non-ATG translation (RAN translation) occurs across long, hairpin-forming repeats in transfected cells or when expansion constructs are integrated into the genome in lentiviral-transduced cells and brains. Additionally, we show that RAN translation across human spinocerebellar ataxia type 8 (SCA8) and myotonic dystrophy type 1 (DM1) CAG expansion transcripts results in the accumulation of SCA8 polyalanine and DM1 polyglutamine expansion proteins in previously established SCA8 and DM1 mouse models and human tissue. These results have implications for understanding fundamental mechanisms of gene expression. Moreover, these toxic, unexpected, homopolymeric proteins now should be considered in pathogenic models of microsatellite disorders.

RNA gain-of-function effects have been reported for CUG and CAG expansion RNAs (5, 6).

Both RNA and protein mechanisms appear to operate in spinocerebellar ataxia type 8 (SCA8) because the CTCG-CAG expansion mutation is expressed in both the CUG (ataxin 8 opposite strand, AXN8OS) and CAG (ataxin 8, ATXN8) directions. ATXN8 expansion transcripts express polyGln protein from an ATG-initiated ORF, and both polyGln protein (7) and AXN8OS CUGEXP transcripts (8) accumulate in affected cells.

Results

To understand the role of the ATXN8 polyGln protein in SCA8, we mutated the only ATG initiation codon located 5′ of the CAG expansion on an ATXN8 (A8) minigene. Unexpectedly, we found this mutation did not prevent expression of the polyGln expansion protein in transfected cells (Fig. 1A). Sequence analysis showed that neither full-length nor spliced transcripts, which are expressed at approximately equal ratios from the (−)ATG minigene, are predicted to contain an AUG initiation codon (Fig. S1A). To test if non-ATG translation also could occur in other frames, a triply-tagged A8 minigene, A8(3*KOAT-op)-3TI1, was generated by inserting a 6X STOP codon cassette (two stops in each frame) upstream of the CAGEXP and three different C-terminal tags to monitor protein expression in all frames [i.e., CAG glutamine (Gln); AGC serine (Ser); and GCA alanine (Ala)] (Fig. 1B). Surprisingly, although transcripts generated from this tagged construct were confirmed to lack initiator AUG codons (Fig. S1B) by RT-PCR, tagged polyGln, polyAla, and polySer proteins were expressed (Fig. 1B) in transfected cells.

The polyGln expansion proteins migrated at one or more discrete molecular weights, polyAla as a high molecular weight smear with a faint laddering pattern seen on light exposures, and polySer at the top of polyacrylamide gels in SDS (Fig. 1B) or 8 M urea (Fig. S2A). As expected, these proteins were degraded by proteasome K, were not affected by RNase I or DNase I, and were not detected with addition of cycloheximide (Fig. 1B). The rel-


This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1B.G. and N.S.D. contributed equally to this work.

2Present address: Department of Molecular Genetics and Microbiology, University of Florida College of Medicine; Gainesville, FL 32610.

3To whom correspondence should be addressed. E-mail: ranum@ufl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013343108/-/DCSupplemental.


260–265
relative levels in individual cells can vary dramatically (Fig. 1D and Fig. S2B). Consistent with previous reports, the polyGln protein is localized primarily within the nuclear aggregates (3), the polyAla protein is primarily diffuse when in the cytoplasm and aggregated when nuclear (5, 10), and the polySer protein forms both nuclear and cytoplasmic aggregates (9). Additionally, an ATG start codon in the polyGln frame variably resulted in an additional higher molecular weight band, suggesting that translational initiation occurs at the ATG and one or more additional sites in some sequence contexts [compare polyGln for A8(‘KKQEXP’)-endo and A8(‘KKQEXP’)-3Tf1 (Fig. 1A and B)].

To determine if this repeat-associated non-ATG (RAN) translation is affected by sequence context, we modified the A8(‘KKQEXP’)-3Tf1 construct by removing 90 bp of the ATXN8 sequence so that the 6X STOP cassette was almost adjacent to the CAGEXP and by adding a seventh STOP immediately upstream of the polyGln, polyAla, or polySer frames (Fig. S3). These constructs also expressed polyGln and polyAla but only low levels of polySer, with the exception that a TAG stop immediately preceding the glutamine frame prevented translation of polyGln but not of polyAla or polySer.

Because these results were completely unexpected, we used several approaches to establish the identity of these homopolymeric proteins. First, each protein was detected with one or more C-terminal epitope tags (myc, His, and HA for Gln; HA for Ala; and HA and Flag for Ser), and the polyGln protein was detected with a monoclonal antibody (IC2) specific to polyGln expansion tracts (Fig. 1) (11). Second, [3H]-Gln, [3H]-Ala, and [3H]-Ser were preferentially incorporated into proteins immunoprecipitated with tags in the polyGln, polyAla, and polySer frames, respectively (Fig. 2A). HEK293T cells transfected with triple-tagged constructs containing the HA-tag in the Ala [A8(‘KKQEXP’)-3Tf1], Gln [A8(‘KKQEXP’)-3Tf2], or Ser [A8(‘KKQEXP’)-3Tf3] frames were grown with [3H]-Gln, [3H]-Ala, or [3H]-Ser amino acids, respectively. Immunoprecipitations were performed using α-HA antibody, separated by PAGE on duplicate gels and detected by either immunoblot or fluorography. Fig. 2A Upper shows that all three proteins in each set were pulled down by immunoprecipitation, and the corresponding fluorograph (Fig. 2A Lower) shows that [3H]-Gln was preferentially incorporated into the ~40-kDa proteins with the HA tag in the polyGln frame. Similarly, [3H]-Ala, and [3H]-Ser were preferentially incorporated into proteins immunoprecipitated with tags in the polyAla and polySer reading frames, respectively.

Third, we used MS to confirm that RAN translation results in the expression of a polyAla expansion protein. PolyAla was selected for MS because (i) polyAla-specific antibodies are not available, and (ii) the putative polyAla protein is expressed at sufficiently high levels required for MS in transfected cells. An arginine residue was introduced into the recombinant protein so that tryptic digestion of the N terminus would generate smaller peptide fragments of suitable size for MS (Fig. 2B). Intracellular (GCAEXP)-3T lysates were separated by PAGE, and MS was performed on proteins isolated from gel slices A1–A7 (Fig. 2C). Associated mass spectra were submitted for database searching against a human protein database and a list of all possible polyAla proteins in which translation could begin before or within the repeat and for which initiation would allow the possible inclusion of an N-terminal methionine residue. MS/MS identified a series of N-terminal peptides with varying numbers of alanines [(A)9–17R and AAADLEITR] (Fig. 2D and E). No peptides containing N-terminal methionine were detected, suggesting that translation initiation in cells occurs without incorporating an N-terminal methionine or that it is removed rapidly by methionine aminopeptidase or endopeptidase activity (12). Additionally, the predicted C-terminal digestion fragment (TTTTSYPYDVDPYDA) was identified (Fig. S4). In summary, these results demonstrate that RAN translation results in polyAla expression in transfected cells and that these proteins run as a broad smear on SDS-PAGE.

Zu et al.
To test if transcripts undergoing RAN translation are modified or edited, resulting in the introduction of AUG initiation codons, we isolated and characterized mRNA from actively translating polyribosomes isolated from cells transfected with (CAG_EXP)-3T constructs with or without an ATG initiation codon (Fig. 3A). Northern analysis showed that transcripts expressed from both the (+)ATG and (−)ATG constructs migrated at the predicted size (700 nt) and coassembled with light polyribosomal fractions (Fig. 3B). Further characterization of ribosome-bound CAG_EXP transcripts from fraction six by 5′ RACE and RT-PCR show that (i) transcription initiated within a few bp of the predicted transcription start site and (ii) the sequence predicted by the DNA was found in the corresponding transcripts and that no upstream AUG initiation codons were introduced by RNA splicing or editing among 140 independently isolated clones (Fig. 3C). To rule out independently the possibility that +ATG transcripts might be generated from these plasmids by a cryptic promoter or alternative splicing, RNA transfections were performed. We transcribed RNA in vitro from the ATT(CAG_EXP) construct and two additional linearized non-ATG constructs and transfected the RNA into cells. Consistent with previous results, RAN translation of polyGln protein also occurred with cell-free products. Following RNA transfections, we used to characterize (+)AUG and (−)AUG transcripts from polysome fraction 6 (Lower) Table summarizing sequencing results from cloned RT-PCR products. In vitro transcription and RNA transfection into HEK293T cells. (Upper) Constructs used to produce capped, polyadenylated mRNAs that extend from the T7 promoter to the poly(A) tail. (Lower) Immunoblots following RNA transfections.

Fig. 2. Protein labeling and MS. (A) Protein blot (Upper) and fluorograph (Lower) of [3H]-Gln, [3H]-Ala, or [3H]-Ser after immunoprecipitation with α-HA of cells transfected with (CAG_EXP)-3T construct with and without glycine. (Lower) Predicted sequence of polyAla protein with arginine interruption. (C) Preparative deep-purple gel showing slices digested with trypsin for MS. (D) N-terminal polyAla peptides identified by MS with varying numbers of alanine. (E) Representative spectrum of N-terminal peptide AAAAAAAAR with matched b-ions (red) and y-ions (blue).

To test if transcripts undergoing RAN translation are modified or edited, resulting in the introduction of AUG initiation codons, we isolated and characterized mRNA from actively translating polyribosomes isolated from cells transfected with (CAG_EXP)-3T constructs with or without an ATG initiation codon (Fig. 3A).
42–107 CAGs express polyGln by RAN translation, but constructs with 15 or 20 CAGs did not (Fig. 4A and Fig. S5A and B). PolyAla was robustly detected with 105 and 107 CAGs, moderately with 73 and 78 CAGs, and not with 42 and 58 CAGs. PolySer was detected with 58–107 repeats but not with 42 repeats (Fig. 4A). Thus, RAN translation is length dependent, and longer repeat tracts are associated with the simultaneous expression of multiple proteins.

Because both repeat length and secondary hairpin structures are associated with CAG and several other disease-causing microsatellite expansions, we compared RAN translation of polyGln expansion proteins expressed from constructs containing hairpin-forming CAG and non–hairpin-forming CAA repeats. Cells transfected with CAG expansion constructs with or without ATG start codons expressed polyGln (Fig. 4B). In contrast, polyGln was expressed only from the CAA expansion constructs with an ATG start codon, suggesting that hairpin formation is required for RAN translation. All constructs were confirmed to express transcripts by RT-PCR (Fig. S5). Because CUGEXP transcripts also form hairpin structures, we tested CTG expansion constructs and show that RAN translation also occurs in all three frames (polyleucine, polyAla, and polycystine) (Fig. S5D).

Next, we addressed if RAN-translation products trigger apoptosis and therefore could be implicated in disease pathogenesis. Murine neuroblastoma (N2a) cells transfected with ATT(CAG105)-3T and ATT(CAG105)-3T, which express polyGln, polyAla, and polySer, showed significant increases in annexin-V staining (13), compared with control cells (Fig. 4C and Fig. S5E). These results indicate that the products of RAN translation can cause apoptosis.

Because most disease-causing CAG·CTG expansions are found in larger polyGln ORFs, we tested if RAN translation in the polyAla and polySer frames would still occur in the presence of an ATG-initiated polyGln ORF (Fig. 5A and B). PolyAla and polySer proteins were expressed irrespective of the polyGln ORF. The absence of the VS-tag on the polyGln from the polySer proteins was confirmed downstream of the VS tag and close to, or within, the repeat.

Immunoprecipitation using antibodies to 3′ epitopes followed by immunoblot with α-V5 showed only a small fraction of polyAla protein has undergone frame shifting from the ATG-initiated VS-polyGln frame (Fig. 5C). Although previously frame shifting has been suggested to result in hybrid polyGln-polyAla and polyGln-polySer proteins in spinocerebellar ataxia type 3 (SCA3) and Huntington disease (14, 15), our data demonstrate that frame shifting is rare, and an out-of-frame ATG initiation codon is not required for polyAla or polySer expression.

The potential for RAN translation in other disease-relevant sequence contexts was investigated using constructs with 20 bp of 5′ flanking sequence upstream of the CAG repeat from the HD, Huntington-like 2 (HDL2), SCA3, or DM1 loci (Fig. S6A). These constructs showed robust polyGln and polyAla and variable polySer expression with the highest non-ATG polySer translation occurring for A8(*KQEXP) and HDL2 (Fig. S6B). RT-PCR confirmed that each construct expressed unspliced transcripts with ATG-initiated ORFs only in the glutamine and serine frames for A8(*KQEXP) and DM1 constructs, respectively.

To understand better the conditions required for expression of these homopolymeric expansion proteins, we performed lentiviral transductions of HEK293T cells and mouse brain. Similar to the transfections described above, RAN translation of polyGln and polyAla proteins was detected, and polySer proteins were expressed irrespective of the polyGln ORF. The results indicate that the products of RAN translation can cause apoptosis.
polyAla also occurs in lentiviral-transduced cells and intact mouse brain (Fig. S6C–E). Taken together, these data demonstrate that RAN translation can occur when the transgene is integrated into the genome and that CAG expansions located in a variety of sequence contexts and under a variety of conditions can express homopolymeric expansion proteins in the absence of an ATG-start codon.

Next, rabbit reticulocyte lysates (RRLs) were used to test if non-ATG translation also occurs in a cell-free system. In contrast to cells, RAN translation in RRLs was limited. Only HDL2 produced polyGln without an ATG, none of the constructs generated detectable polyAla, and the highest levels of non-ATG–initiated polySer were from the HD and SCA3 constructs (Fig. S7A). Moreover, RAN translation in RRLs, but not in cells, is substantially affected by mutating previously reported alternative initiation codons (ATT and ATC) (16, 17) (Fig. S7B–D), indicating that sequence requirements for RAN translation in RRLs are less permissive than in cells. Next, we used constructs that undergo non-ATG translation in RRLs to test if N-terminal methionine incorporation still occurs in the absence of an AUG initiation codon. We showed that polyGln expressed from constructs lacking initiator and internal methionine codons incorporated S- methionine (Fig. S7E). Additionally, in vitro translation using S-p55-labelled Met-tRNAiMet intends that in RRLs non-ATG translation is initiated with tRNAiMet (Fig. S7F). The incorporation of an N-terminal methionine in RRLs in vivo is not surprising, because previously documented alternative initiation codons (ATT, ATC) are used. However, RAN translation in cells may use a different initiation mechanism, because we were not able to detect N-terminal methionine for any of the polyAla proteins using MS.

To determine if novel homopolymeric proteins are expressed in vivo, we developed peptide antibodies to putative RAN-translated SCA8-polyAla (SCA8GCA-Ala) and DM1 polyGln (DM1CAG-Gln) proteins (Fig. 6A and E). The specificity of both the α-SCA8GCA-Ala and α-DM1CAG-Gln antibodies was demonstrated by Western blot and IF detection in cells expressing recombinant SCA8GCA-Ala and DM1CAG-Gln proteins but not in control cells (Fig. 6B and F). Consistent with the possible role for RAN-translationed proteins in SCA8, we detected α-SCA8GCA-Ala immunostaining in Purkinje cell soma and dendrites throughout the cerebellum in an established SCA8 mouse model (7) (Fig. 6C). Similarly, α-SCA8GCA-Ala staining was found reproducibly in the remaining cerebellar Purkinje cells of postmortem samples from two patients with SCA8 (e.g., Fig 6D). For myotonic dystrophy, IF staining of DM1 mice (18) which express CUGEXP (Fig. S8A) and CAGEXP transcripts (Fig. S8B and C) show rare (2.32 ± 2.04%) but reproducible α-DM1CAG-Gln nuclear aggregates in cardiac myocytes (Fig. 6G and Fig. S9A and B) and more frequent α-DM1CAG-Gln staining in leukocytes (10.68 ± 3.66%) (Fig. S9C). The DM1 polyGln aggregates colocalize with caspase-8 (Fig. S9D), which is an early indicator of polyGln-induced apoptosis (19). Immunohistochemical staining of paraffin-embedded tissue with the polyGln-specific I2C antibody confirms staining in leukocytes in cardiac tissue from mice containing a CUG expansion of 55 repeats but not in control mice with 20 CTG repeats (Fig. S9E).

In samples from patients with myotonic dystrophy, α-DM1CAG-Gln inclusions were found at low frequency in myoblasts (50–70

---

Fig. 6. In vivo evidence for RAN-translated SCA8-polyAla and DM1-polyGln. (A) ATXN8 ATG-polyGln ORF and putative non-ATG SCA8GCA-Ala protein. Underlined peptides used for antibody generation. (B) Antibody validation: α-SCA8GCA-Ala detection of recombinant protein in A8(*KMQEXP)-endo transfected cells by protein blot (Left) and IF (Right). (C) Immunohistochemical staining of SCA8 and control mouse cerebellum (FVB) using α-SCA8GCA-Ala. (D) In SCA8 human samples, α-SCA8GCA-Ala antibody shows consistent and specific staining (red-cy3) of surviving human SCA8 cells but not control Purkinje cells. Co-labeling with α-PKCγ antibody (cy5, yellow) independently stains Purkinje cells and confirms their presence in both samples. (E) Upper) DMT1 CAG anti-sense transcript. (Lower) Predicted non-ATG–initiated polyGln protein. Underlined peptide used for antibody generation. (F) Antibody validation. α-DM1CAG-Gln detects recombinant protein with endogenous DM1 polyGln C-terminus (CAGEXP-DM1-3') by protein blot (Left) and IF (Right). (G) IF using α-DM1CAG-Gln (cy3, red) detects DM1CAG-Gln protein in cardiomyocytes of DM1 mice with 55 (DM55), and >1,000 (DMSXL) CTGs but not in cardiomyocytes of control (WT) mice. (H) Staining with α-DM1CAG-Gln (cy3, red) in DM1 but not in control (WT) myoblasts.
CTG–CAG repeats) (Fig. 6H) and skeletal muscle and at higher frequency in blood (Fig. S10).

Discussion

Our understanding of the molecular basis of disease has been built on studying the expected effects of mutations on the functions of their corresponding genes. For microsatellite expansion disorders, cell culture and animal models have been developed to test specific hypotheses based on the prediction that CAG_{exp} mutations located in polyGln ORFs express protein only in the polyGln frame and that expansions located in noncoding regions do not encode proteins (2, 3). We demonstrate that these canonical rules of translation do not apply for CAG–CTG expansions and that in the absence of an ATG codon expanded CAG and CTG trinucleotide repeats often express homopolymeric expansion proteins in all three frames. RAN translation occurs in transfected and transduced HEK293T and N2a cells. In contrast, non-ATG translation is less frequent in RRLs, suggesting that RRLs may not recapitulate what is happening inside cells. The production of polyGln protein after RNA transfections in cells indicates that cellular factors and not promoter issues affect RAN translation.

While initiation at specific alternative codons has been reported previously (16, 20), our findings show translational initiation at different CAG–CTG expansion sites in cells is highly permissive. Data showing that hairpin-forming CAG and CUG repeats undergo RAN translation are consistent with previous reports that hairpin structures affect translational initiation (20). Hairpin structures are thought to allow translational initiation at suboptimal sites by delaying the 40S ribosomal subunit long enough to allow efficient interaction between the Met-tRNA_{i} anticodon and the AUG or non-AUG start site (21). Additionally, hairpins can recruit initiation factors and ribosomal subunits to internal ribosome entry sites (IRES) (1). Some IRESs, such as the one in the cricket paralysis virus, can facilitate translation initiation without eIFs or tRNA{Met} (1). These IRES hairpins function as a tRNA{Met} to initiate translation at non-AUG codons including GCA (22). This function could explain why translation in the polyAla (GCA) frame appears to initiate at multiple sites within the repeat. IRES hairpins form and are stabilized with the help of IRES translation-associated factors (ITAFs) (1), and at least one ITAF (CUGBP1) is known to bind to CAG and CUG repeats (23).

The apparent requirement for hairpins, the initiation from non-AUG codons, and the association of repeat transcripts with a known ITAF suggests that RAN translation may involve an IRES-like mechanism. Differences in repeat length required for the accumulation of polyGln, polyAla, and polySer proteins may reflect differences in protein stability and/or repeat length required for efficient initiation in each frame.

The discovery of RAN translation raises the possibility that polyAla and polySer proteins contribute to the pathogenesis of some of CAG polyGln diseases and that homopolymeric proteins contribute to diseases previously thought to involve primarily RNA gain of function (e.g., DM1). In SCA8, specific targeting of the SCAG{GCA-Ala} expansion protein is found in cerebellar Purkinje cells. In DM1, staining for the DMI{CAG-Gln} expansion protein is found in heart, skeletal muscle, and myoblasts. Further investigation will be required to determine which microsatellite expansions undergo RAN translation and which RAN-translated proteins contribute to disease. Our results indicate polyAla and polySer proteins are more likely to be expressed from CAG–CTG expansions exceeding 70 repeats, suggesting the possibility that RAN-translated proteins may contribute to the anticipation, the earlier onset, and increased disease severity associated with longer repeat lengths.

Recently, much of the genome (24) and a growing number of expansion mutations (25) have been shown to be transcribed bidirectionally. Given that CAG_{exp} and CUG_{exp} transcripts can express proteins without an ATG, and that these transcripts are reported to cause RNA gain-of-function effects (6, 25), the molecular pathology of microsatellite disorders may be far more complex than currently appreciated. Additionally, these results raise the possibility that other repetitive sequences in the genome also undergo RAN translation and contribute to proteome diversity.

Materials and Methods

Details of cloning, custom antibodies, and molecular techniques are available in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank patient families for participating; G. Nelles-tuen, L. Anderson, L. A. Higgens, C. Pearson, H. Orr, and N. Sonenberg for helpful discussions; J. Day (University of Minnesota, Minneapolis; AR057220) for providing tissue (P30 AR057220); J. Henriksen and the University of Minnesotas BioNet Core for providing tissue (P30 CA77598 and PS0 CA101955) and digital imaging; the National Institute of Neurological Disorders and Stroke for providing tissue (P01NS058901 and R01NS040389); and the Muscular Dystrophy Association for funding.

The ATXN8 minigene (A8(*KMQ)primer (5′-CTCGGTCTCAGTTACCGGGA-3′) and 5′-ATCTCGGTCTCAGTTACCGGGA-3′) containing a V5 tag at the 5′ end of the ATT (CAGEXP)-3T construct. The QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene) was used to change the ATG in front of the V5 tag to an ATC to generate the ATC-V5 (CAG_{105})-3T construct which contains no ORFs.

To generate the CAEXP constructs, a CA repeat was amplified by PCR using the ACA23 and TTG15 primers. PCR products varied in size. A gel slice containing 200–550 bp fragments (67–183 repeats) was purified, and the resulting fragments were cloned into the pSc-A-amp/kan vector using Stratagene PCR Cloning Kit (Stratagene). Clones were sequenced, and desirable CA repeats were excised and subcloned into pcDNA3.1/6Stops-3T. The resulting constructs were sequenced and CAA_{125} (−ATG), CAA_{90} (−ATG), and CAA_{38} (−ATG) constructs were obtained. Modified versions of these constructs containing an ATG in the polyGln frame [CAA_{125} (+ATG), CAA_{90} (+ATG), and CAA_{38} (+ATG)] were created using site-directed mutagenesis (Stratagene).

To generate CTGEXP (Cys-myc/His), CTGEXP (Ala-myc/His), and CTGEXP (leucine-myc/His) constructs, a fragment of expanded CTG repeats was subcloned into pcDNA3.1/myc-His (A, B, and C respectively), and each of the three reading frames was C-terminally tagged. In the three resulting constructs, there is no ORF in each of three frames, and polyC, polyA, and poly-leucine are tagged individually in frame with an myc-His tag.

The 3′ flanking sequence of DM1 in the CAG direction was amplified by PCR using 5′-CTCGGTCTCAGTTACCGGGACCCCGG-3′ and 5′-CTCGGTCTCAGTTACCGGGACCCCGG-3′ primers and a unique 5′ primer. All constructs were PCR-mediated mutagenesis was used to create several constructs in which the ATT or ATC alternative start codons were altered to ACT and ACC, respectively. All constructs were created using the 3′ primer (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′) and a unique 5′ primer. The ACT (CAG_{105})-3T primer (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′) was used to generate the ACT (CAGEXP)−3T construct from the ACT (CAGEXP)−3T template. The HDL2-3T: [ATT, ATC] construct was used as template to generate the HDL2-3T: [ATT, ACC], HDL2-3T: [ACT, ATC], and HDL2-3T: [ACT, ACC] constructs from the HDL2 [ATT, ACC] 5-1 (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′), HDL2 [ATT, ACC] 5-1 (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′), HDL2 [ACT, ACC] 5-1 (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′), and HDL2 [ACT, ACC] 5-1 (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′) primers, respectively. Likewise, the SCA3 [ACT] construct was generated from SCA3 template and the SCA3 [ACT] 5-1 (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′), HDL2 [ACT, ACC] 5-1 (5′-CTAAGGACAATGTTACCGGGACCCCGG-3′) primer was used along with the HDL3 [ATT, ATC] template to generate the HDL3 [ATT, ACT] construct.

The IntR (CAGEXP)-3T construct was made by inserting the duplex primer containing (CAG)_{n}CGCGCG into the EcoRI site of the ATT (CAG_{90})-3T construct. The extra nucleotides between (CAG)_{n}CGCGCG and (CAG)_{90} were removed by

Zu et al. 10.1073/pnas.1013343108

Supporting Information

Zu et al., www.pnas.org/cgi/content/short/1013343108
digested with EcoRI, followed by treatment with mung bean nuclease and DNA ligase.

All PCR reactions to generate the above constructs were performed with Pfx polymerase (#11708-013; Invitrogen) to mitigate PCR-induced mutations. The above PCR products were digested with HindIII and PmeI, gel purified, and cloned into pcDNA3.1 containing the 6X STOP cassette. The integrity of all constructs was confirmed by sequencing.

**Production of Polyclonal Antibodies.** The polyclonal antibodies were generated by New England Peptide. The α-SCA8GCA-Ala antisera were raised against a synthetic peptide corresponding to the C terminus of the predicted polyAla frame of SCA8 in the CAG direction (VKPGFLT). The α-DM1CAG-Glu antisera were raised against a synthetic peptide corresponding to the C terminus of a predicted glutamine frame of DM1 in the CAG direction (SPAARGRATITGLEG).

**Cell Culture and Transfection.** HEK293T cells were cultured in DMEM supplemented with 10% fetal bovine serum and incubated at 37 °C in a humid atmosphere containing 5% CO2. DNA transfections were performed using Lipofectamine 2000 Reagent (Invitrogen) according to the manufacturer's instructions.

Murine neuroblastoma (N2a) cells were cultured in unsupplemented DMEM and incubated at 37 °C in a humidified, 5% CO2 incubator. DNA transfections were performed using TransIT-Express (Mirus) according to the manufacturer’s instructions.

DM1 patient myoblasts with 50–70 CTG repeats, along with a normal control, were cultured in skeletal muscle growth medium (SGM; PromoCell) with Glutamax (Invitrogen), gentamicin (50 units/mL), and 2% decomplemented fetal calf serum, and the provided supplemental mix. Cells were grown to approximately 70% confluence on collagen-coated coverslips in six-well tissue-culture plates.

**RNA Transfections.** Plasmid DNA was linearized using PvuII. Transcription, capping, and polyadenylation was performed using 1 μg of DNA with the mScript mRNA Production System (Epigen). Transfections were performed in six-well plates using 3 μg of RNA and 10 μL Lipofectamine 2000 (Invitrogen) per well. Cell lysates were collected 18–24 h posttransfection, and immunoblotting was performed as described.

**Immunofluorescence.** The subcellular distribution of homopolymer proteins was assessed in transfected HEK293T cells by immunofluorescence. Cells were cultured on coverslips in six-well tissue-culture plates and transfected with plasmids the next day.

**Immunohistochemistry.** Myotonic dystrophy (DM) mutant and control mice were perfused in 10% formalin, and tissue was harvested and embedded in paraffin. Five-micrometer sections were deparaffinized in xylene and rehydrated through graded alcohol, incubated with 90% formic acid for 5 min, and washed with distilled water for 30 min. HIER was performed by steaming sections in citrate buffer, pH 6.0, at 90 °C for 20 min. For mouse and human tissues, 9-μm cryosections were fixed in 4% paraformaldehyde for 15 min. Heat-induced epitope retrieval (HIER) was employed by steaming sections in citrate buffer, pH 6.0, at 90 °C for 20 min. HIER was used in all immunofluorescence tissue experiments except for SCA8GCA-Ala and SCA8SCA-Ala mice and human experiments in which antigen retrieval was omitted altogether. A nonserum block (Biocare Medical) was applied to all tissues (except the SCA8 mouse tissue in which 10% NGS in a 0.3% Triton X-100 was used to block nonspecific immunoglobulin binding) and was allowed to incubate at room temperature for 1 h. The primary antibody (or antibodies, if double- or triple-labeled) was diluted either in a 1:5 solution of the nonserum block or in a solution of 5% NGS in PBS containing 0.3% Triton X-100 and was incubated at 4 °C overnight. Tissues were then incubated for 1 h in a 1:2,000 dilution of IgG–TRITC, in the dark at room temperature. If needed, a Sudan-black autofluorescence block was applied to the tissue for 1 h at room temperature in the dark (2). Staining was observed, and pictures were taken on an Olympus Fluoview 1000 1×2 inverted confocal microscope. All mutant and control images were adjusted in unison, to the same specifications, and in a linear fashion, for intensity and contrast when deemed necessary.

**Quantitative Imaging.** Cells were counted using Image-Pro Plus 6.0 (Media Cybernetics) using an automatic dark-object count method. Images first were digitally annotated by a pathologist to separate coagulated blood from cardiac myocytes into distinct regions of interest. The images were split into grayscale versions of their blue color channel (nuclear stain color) and red color channel (positive marker color). The grayscale images were analyzed in each region of interest with Image-Pro Plus using a dark-object count method with automatic thresholding for nuclear intensity.

For mouse and human tissues, 9-μm cryosections were fixed in 4% paraformaldehyde for 15 min. Heat-induced epitope retrieval (HIER) was employed by steaming sections in citrate buffer, pH 6.0, at 90 °C for 20 min. HIER was used in all immunofluorescence tissue experiments except for SCA8GCA-Ala and SCA8SCA-Ala mice and human experiments in which antigen retrieval was omitted altogether. A nonserum block (Biocare Medical) was applied to all tissues (except the SCA8 mouse tissue in which 10% NGS in a 0.3% Triton X-100 was used to block nonspecific immunoglobulin binding) and was allowed to incubate at room temperature for 1 h. The primary antibody (or antibodies, if double- or triple-labeled) was diluted either in a 1:5 solution of the nonserum block or in a solution of 5% NGS in PBS containing 0.3% Triton X-100 and was incubated at 4 °C overnight. Tissues were then incubated for 1 h in a 1:2,000 dilution of IgG–TRITC, in the dark at room temperature. If needed, a Sudan-black autofluorescence block was applied to the tissue for 1 h at room temperature in the dark (2). Staining was observed, and pictures were taken on an Olympus Fluoview 1000 1×2 inverted confocal microscope. All mutant and control images were adjusted in unison, to the same specifications, and in a linear fashion, for intensity and contrast when deemed necessary.

**Immunohistochemistry.** Myotonic dystrophy (DM) mutant and control mice were perfused in 10% formalin, and tissue was harvested and embedded in paraffin. Five-micrometer sections were deparaffinized in xylene and rehydrated through graded alcohol, incubated with 90% formic acid for 5 min, and washed with distilled water for 30 min. HIER was performed by steaming sections in citrate buffer, pH 6.0, at 90 °C for 20 min. To block nonspecific avidin–biotin binding, the Avidin–Biotin blocking (Vector Laboratories, Inc.) was used as described. To block nonspecific immunoglobulin binding, a nonserum block (Biocare Medical) was applied for 30 min. Primary 1C2 antibody was applied at a dilution of 1/12,000 in nonserum block (Biocare Medical) and incubated overnight at 4 °C. Biotinylated secondary α-mouse IgG purified in goat (Vector Laboratories, Inc.) was applied at a dilution of 1:200 for 30 min at room temperature. ABC reagent (Vector Laboratories, Inc.) was used for detection with Chromagen SG (Vector Laboratories, Inc.) for 10 min and counterstained with nuclear fast red.

Leukocyte cell pellets were isolated from peripheral blood of DM1 patients and controls. The cell pellets were fixed in 10% neutral buffered formalin for 30 min, washed, encapsulated in HistoGel (Richard-Allen), and placed in 70% ethanol. The pellets then underwent a short, 2-h cycle in the tissue processor and were embedded in paraffin blocks. Five-micrometer sections were cut, deparaffinized, and hydrated to water. HIER was employed with steam and Reveal Decloaker (Biocare Medical). A nonserum block (Biocare Medical) was applied for 30 min to prevent nonspecific immunoglobulin binding. The nonserum block 1:10 in PBS was used to dilute the α-DM1CAG-Glu antibody to a concentration of 1:10,000. Slides were incubated overnight at 4 °C and washed three times for 5 min in PBS. The secondary antibody, DyLight 488-conjugated AffiniPure goat anti-rabbit (Jackson ImmunoResearch) was applied and incubated for 2 h in the dark at room temperature at a concentration of 1:1,000. Slides were washed three times for 5 min in PBS, mounted with
Cell Death Analysis. N2a cells were transfected with CAG<sub>105</sub> and CA<sub>AA90</sub> (control) constructs with or without ATG initiation codons and were grown in DMEM without FBS. After 48 h, cells were stained with annexin-V to detect phosphatidylserine (an early marker of apoptosis) on the plasma membrane and were sorted by flow cytometry. For flow cytometric annexin-V analysis, floating cells were collected and combined with trypsinized, adherent cells in ice-cold PBS supplemented with 2% FBS. After washing, cells were resuspended in annexin-binding buffer (BD Biosciences), vortexed, and stained with annexin-V–APC (BD Biosciences) according to manufacturer’s instructions. Cells were placed on ice and sorted immediately on a BD FACsCalibur flow cytometer. Thirty thousand total events were collected. Two independent experiments, each done in triplicate, were performed, and data were combined and normalized to the ATT(CAA<sub>90</sub>) average. Statistics were performed using a one-way ANOVA, F(3, 20) = 0.07, P = 0.0003, and post hoc P values were calculated with a one-tailed r-test.

Labeling and Immunoprecipitation of PolyGln, PolyAla, and PolySer Proteins with [<sup>3</sup>H]-Amino Acids. HEK293T cells were cultured in DMEM supplemented with 10% fetal bovine serum and were transfected with CAG expansion constructs. Twenty-four hours posttransfection, the DMEM-based medium was replaced with the glutamine-, alanine-, and serine-free MEM medium (Invitrogen) supplemented with 10% fetal bovine serum. Then [<sup>3</sup>H]-glutamine, [<sup>3</sup>H]-alanine, or [<sup>3</sup>H]-serine was added to the respective wells at 25 μCi/mL, and the cells were incubated for 16 h at 37°C. Cells in culture plates were rinsed with PBS and lysed in RIPA buffer [150 mM NaCl, 1% sodium deoxycholate, 1% Triton X-100, 50 mM Tris-HCl (pH 7.5), 1x protease inhibitors (Roche)] for 45 min on ice. The cell lysates were centrifuged at 16,000 × g for 15 min at 4°C, and the supernatant was collected. To immunoprecipitate [3H]-labeled protein, 500 μg of tissue lysate was incubated with the desired antibody at 4°C for 2 h and then with protein G-Sepharose at 4°C overnight. Protein G-Sepharose was washed three times with RIPA buffer. Bound proteins were eluted from the beads with 1x SDS sample buffer, incubated at 90°C for 10 min, and analyzed by protein gel electrophoresis.

Protein Labeling with [<sup>35</sup>S]-Methionine. A T7-coupled transcription and translation kit (Promega) was used with [<sup>35</sup>S]-methionine (MP Biomedical) for protein labeling. Labeled proteins were separated by PAGE on parallel two separate gels. One gel was dried for 1.5 h at 70°C and subsequently was used to generate a fluorograph. The other gel was transferred onto a nitrocellulose membrane and probed with the 1C2 monoclonal antibody specific to polyGln expansion tracts.

Preparation of N-Formyl-[<sup>35</sup>S]Methionyl-tRNA<sub>Met</sub>. Preparation of aminoacyl-tRNA synthetases and Met-tRNA<sub>Met</sub> transformylase from Escherichia coli MRE 600 (ATCC) was performed following Stanley’s method (3). The reaction mixture included the E. coli MRE 600 fraction (1.5 mg/mL), 0.1 mM each of 19 unlabeled amino acids (minus methionine), t-[<sup>35</sup>S]methionine (2 mCi/mL), 0.2 mM calcium leucovorin, unfracionated bovine liver tRNA (2 mg/mL), 10 mM ATP, 1 mM CTP, 15 mM MgCl<sub>2</sub>, and 50 mM sodium cacodylate, pH 7.4. The aminoacylation reaction was carried out at 37°C for 30 min. The reaction mixture was extracted with phenol, precipitated with ethanol, lyophilized, and suspended in water. N-formyl-[<sup>35</sup>S]methionyl-tRNA<sub>Met</sub> was added at a concentration of 100 μg/mL for in vitro translation using a T7-coupled transcription and translation kit. The reaction products were analyzed as described above.

Immunoprecipitation. The protein concentration of tissue lysates was determined using the protein assay dye reagent (Bio-Rad). To immunoprecipitate polyGln protein, 500 μg of tissue lysate was incubated with rabbit polyclonal anti-His antibody at 4°C for 2 h and then with protein G-Sepharose at 4°C for overnight. Protein G-Sepharose was washed three times with RIPA buffer. Bound proteins were eluted from the beads with 1x SDS sample buffer, incubated at 90°C for 10 min, and analyzed by immunoblotting.

Immunoblotting. Cells in each well of a six-well tissue-culture plate were rinsed with PBS and lysed in 300 μL RIPA buffer for 45 min on ice. DNA was sheared by passage through a 21-gauge needle. The cell lysates were centrifuged at 16,000 × g for 15 min at 4°C, and the supernatant was collected. The protein concentration of the cell lysate was determined using the protein assay dye reagent (Bio-Rad). Twenty micrograms of protein were separated in a 4–12% or 10% NuPAGE Bis-Tris gel (Invitrogen) and transferred to a nitrocellulose membrane (Amersham). The membrane was blocked in 5% dry milk in PBS containing 0.05% Tween-20 and probed with the anti-His antibody (1:500) or 1C2 antibody (1:1,000) in blocking solution. After the membrane was incubated with anti-rabbit or anti-mouse HRP-conjugated secondary antibody (Amersham), bands were visualized by the ECL plus Western Blotting Detection System (Amersham).

The distinctive migration pattern of the polyAla may result from initiation at multiple sites and/or the formation of higher molecular weight protein complexes resistant to dissociation in SDS and 8 M urea. The polySer proteins might undergo beta elimination and addition from side chains of another polymer to create 3D-crosslinked polypeptides that would be extremely large (4).

MS. To immunoprecipitate polyAla protein for MS, transfected HEK293T cell lysate from five 150-mm dishes was incubated with mouse monoclonal antibody against the C-terminal HA tag at 4°C for 2 h and then with protein G-Sepharose at 4°C overnight. Protein G-Sepharose was washed three times with RIPA buffer. Bound proteins were eluted from the beads with 8 M urea. Samples were separated by parallel SDS-PAGE 4–15% Criterion Tris-HCl gels (Bio-RAD), one for MS and the other for immunoblotting. Protein bands of interest were excised manually after visualization with Imperial Protein Stain (Thermo Scientific). Specified bands were cut out and subjected to in-gel trypsin digestion using standard methods (5), and extracted peptides were cleaned up further using “stage” tips (6). Mass analysis was performed using an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific). Peptides derived from in-gel digestion were separated by reversed-phase chromatography with nanoHPLC. The gradient was 2–40% acetonitrile in H<sub>2</sub>O containing 0.1% formic acid for more than 60 min. Full MS scans were generated in the orbital trap at 60,000 resolution for 400 m/z. MS/MS scans were performed in a data-dependent manner using an inclusion list based on predicted tryptic peptides in the LTQ ion trap using collision induced dissociation (CID). Data were searched with SEQUEST V27 with semi-trypsin specification, cysteine carbamidomethylation as a fixed modification, and N-terminal methionine oxidation as a variable modification. The search was performed against the combined database consisting of the National Center for Biotechnology Information human database V200906 and its reversed complement and an additional list of all possible proteins that could be initiated anywhere in the polyAla frame of the IntR (CAG<sub>EXP</sub>)-3T construct with or without an N-terminal methionine, which totaled >76,000 entries. Identified proteins were organized using Scaffold (www.proteomesoftware.com), and peptide
probabilities were calculated within this program using Peptide Prophet (7). The identification output was filtered using a precursor mass tolerance at 7 ppm.

**In Vitro Translation.** In vitro translation was performed using coupled reticulocyte lysate systems (Promega). Coupled transcription/translation reactions (50 μL) contained 50% lysate, 1 μL T7 RNA polymerase, 20 μM amino acid mixtures, 1 μL (40 units) RNasin ribonuclease inhibitor (Promega), and 1 μg plasmid DNA. Incubation was performed at 30 °C for 90 min. Ten percent of each reaction was analyzed by Western blotting.

**Production and Purification of Lentiviral Vectors and Transduction of HEK293T Cells.** HEK293T cells were plated on 150-mm tissue culture dishes and transected the following day when cells were 80–90% confluent. Thirty micrograms of the transducing vector, 20 μg of the packaging vector ΔRRE, and 10 μg of the VSV envelope pMD.D were cotransfected by calcium phosphate-mediated transfection. The medium was changed the next day, and conditioned media were collected 48 and 72 h after transfection. Conditional medium then was cleared by filtering through a 0.45-μm filter. The viral particles were concentrated by ultracentrifugation at 50,000 × g for 80 min at 4 °C. The gradients were separated on a glyoxal gel, blotted to a nylon membrane, and probed with a [32P]ATP-labeled oligonucleotide (5′-TAAAGGCACAGCTGAGGAGGAGGTATGTTTGGTAAACCTCAAT-3′) complementary to the 3′ end of the CAG-containing transcripts. Blots subsequently were probed with a [32P]DATP-labeled GDPDH cDNA probe.

**RT-PCR.** For detection of CAG and CAA expansion transcripts, cells were transfected using Lipofectamine 2000 (Invitrogen) as described above. RNA and protein were harvested using Trizol (Invitrogen). Approximately 45 μg of RNA from each sample was resuspended in 50 μL diethylpyrocarbonate distilled water. The RNA sample was treated with an RNase-free DNase Set (Qiagen) and the RNeasy Plus Mini Kit (Qiagen) to remove DNA. A Superscript II Reverse Transcriptase System (Invitrogen) and the Myc Tag GSP Primer (5′-CAGATCTCCCTTCTGAGATGAGTTTTGTTCT-3′) were used to reverse transcribe the RNA, and PCR was performed using the 336 forward (5′-ACCCCAAAGCTGCTAATAGAC-3′) and 336 reverse (5′-TGTGTGTCGTGTCCTTTGTTA-3′) primers at 95 °C for 2 min, then 35 cycles of 94 °C for 45 s, 59.5 °C for 30 s, 72 °C for 45 s, and a 6-min extension at 72 °C. Control reactions were performed using the β-actin forward (5′-TCTGTGGTGACATTAAAGGAG-3′) and β-actin reverse (5′-GATCTTATTTGGCTGGTTGTT-3′) primers. PCR conditions were 95 °C for 2 min, then 35 cycles of 94 °C for 45 s, 59.5 °C for 30 s, 72 °C for 45 s, followed by a 6-min final extension at 72 °C. PCR products were separated on a 1% agarose gel. For detection of CAG expansion transcripts in DM humans and mice, total RNA was extracted from frozen tissues with Trizol (Invitrogen) following incubation with lysis buffer and 0.5 mg/mL proteinase K, as well as precipitation and DNase treatment. For strand-specific RT-PCR, a linker sequence (lk) was attached (5′-CGACTGGAGCAGGACACTAG-3′) to the 5′ end of primers specific for the antisense strand of dystrophy myotonica-protein kinase (DMPK): (i) 5′-CGCTTGACAGGACTCTTGAAGAGG-3′ or (ii) 5′-GACCTTTTCTTGGCCAGCAGGCAGG-3′. Three micrograms of RNA were reverse-transcribed with Superscript III (Invitrogen) at 55 °C. PCR against the antiB1, antiN3, and antiA2 regions was carried out using the CTCf1b (5′-CAGACTTCCCCGGCCACGACCACAGTT-3′) and anti3N (5′-GAGGCGCTGCTAGCAAAAG-3′) primers, respectively. The linker primer was used in all reactions. The PCR reactions were done using the following conditions: antiB1, 94 °C for 5 min, then 30 cycles of 94 °C for 30 s, 67 °C for 30 s, and 72 °C for 1 min followed by 10 min at 72 °C; antiN3, 94 °C for 5 min, then 30 cycles of 94 °C for 30 s, 63 °C for 30 s and 72 °C for 1 min followed by 10 min at 72 °C; antiA2, 94 °C for 5 min, then 40 cycles of 94 °C for 30 s, 57 °C for 30 s, and 72 °C for 1 min followed by 10 min at 72 °C. GAPDH was amplified using the GFw (5′-AGGTCGAGGTGAAGACGATGAG-3′) and GRev (5′-TGTGACACCTATGAGATGGTTCA-3′) primers at 94 °C for 5 min and then 24 cycles of 94 °C for 30 s, 65°C for 30 s, and 72 °C for 1 min followed by 10 min at 72 °C.

A

![Diagram of splicing in HEK293T cells](image)

Frame 1: RTHCLLAR...YDSL...GDPSWLKLGTELGSTNQRGCAGIPYLL...INS'VRDQYVEEWWNKI'QL'HQIKV'DNIFLKNAAA...

Frame 2: REPTAYWLEINTHYRETQAQ'LSLVPSLDPLVTAASVLEFALTCSR'ILK'EISSMRKYGKIVNYN'N'K'IYF'KMQQQ...

Frame 3: ENPLTLGSLRILGTRPKLAA'SAWYRARN'RPVPCWNSPLPAKLFSKRAV'AG'MSEMKYTIITSNKIR'YIFKKCSSS...

Frames of full-length A8(*KMqMq)-endo transcript

Frame 1: RTHCLLAR...YDSL...GDPSWLKLGTELGSTNQRGCAGIPYLL...INS'VRDQYVEEWWNKI'QL'HQIKV'DNIFLKNAAA...

Frame 2: REPTAYWLEINTHYRETQAQ'LSLVPSLDPLVTAASVLEFALTCSR'ILK'EISSMRKYGKIVNYN'N'K'IYF'KMQQQ...

Frame 3: ENPLTLGSLRILGTRPKLAA'SAWYRARN'RPVPCWNSPLPAKLFSKRAV'AG'MSEMKYTIITSNKIR'YIFKKCSSS...

Frames of spliced transcript

Frame 1: RTHCLLAR...YDSL...GDPSWLKLGTELGSTNQRGCAGIPYLL...INS'VRDQYVEEWWNKI'QL'HQIKV'DNIFLKNAAA...

Frame 2: REPTAYWLEINTHYRETQAQ'LSLVPSLDPLVTAASVLEFALTCSR'ILK'EISSMRKYGKIVNYN'N'K'IYF'KMQQQ...

Frame 3: ENPLTLGSLRILGTRPKLAA'SAWYRARN'RPVPCWNSPLPAKLFSKRAV'AG'MSEMKYTIITSNKIR'YIFKKCSSS...

B

![Diagram of splicing in HEK293T cells](image)

Frame 1: RTHCLLAR...YDSL...GDPSWLKLGTELGSTNQRGCAGIPYLL...INS'VRDQYVEEWWNKI'QL'HQIKV'DNIFLKNAAA...

Frame 2: REPTAYWLEINTHYRETQAQ'LSLVPSLDPLVTAASVLEFALTCSR'ILK'EISSMRKYGKIVNYN'N'K'IYF'KMQQQ...

Frame 3: ENPLTLGSLRILGTRPKLAA'SAWYRARN'RPVPCWNSPLPAKLFSKRAV'AG'MSEMKYTIITSNKIR'YIFKKCSSS...

Frames of full-length A8(*KQqKq)-3Tf1 transcript

Frame 1: RTHCLLAR...YDSL...GDPSWLKLGTELGSTNQRGCAGIPYLL...INS'VRDQYVEEWWNKI'QL'HQIKV'DNIFLKNAAA...

Frame 2: REPTAYWLEINTHYRETQAQ'LSLVPSLDPLVTAASVLEFALTCSR'ILK'EISSMRKYGKIVNYN'N'K'IYF'KMQQQ...

Frame 3: ENPLTLGSLRILGTRPKLAA'SAWYRARN'RPVPCWNSPLPAKLFSKRAV'AG'MSEMKYTIITSNKIR'YIFKKCSSS...

Frames of spliced transcript

Frame 1: RTHCLLAR...YDSL...GDPSWLKLGTELGSTNQRGCAGIPYLL...INS'VRDQYVEEWWNKI'QL'HQIKV'DNIFLKNAAA...

Frame 2: REPTAYWLEINTHYRETQAQ'LSLVPSLDPLVTAASVLEFALTCSR'ILK'EISSMRKYGKIVNYN'N'K'IYF'KMQQQ...

Frame 3: ENPLTLGSLRILGTRPKLAA'SAWYRARN'RPVPCWNSPLPAKLFSKRAV'AG'MSEMKYTIITSNKIR'YIFKKCSSS...

Zu et al. www.pnas.org/cgi/content/short/1013343108 5 of 13
Fig. S2. (A) Immunoblot of HEK293T lysates after transfection with constructs in Fig. 1B and separation on a polyacrylamide gel in 8 M urea. (B) Immunofluorescence of tagged polyGln (α-His/cy3), polyAla (α-HA/cy5) and polySer (α-FLAG/FITC) proteins in transfected HEK293T cells shows RAN-translated proteins are expressed.
Fig. S3. Immunofluorescence of tagged polyGln (α-His/cy3), polyAla (α-HA/cy5), and polySer (α-FLAG/FITC) proteins in transfected HEK293T cells shows repeat-associated non-ATG (RAN)-translated proteins are expressed. (A) Schematic diagram showing constructs with and without stop codons immediately preceding repeats in the CAG(polyGln), GCA(polyAla), and AGC(polySer) frames. These constructs expressed full-length unspliced transcripts in transfected cells. All constructs contain 3' epitope tags: myc-His (polyGln), HA (polyAla), and Flag (polySer). (B) Protein blots from transfected HEK293T cells probed with 1C2, α-HA, or α-Flag antibodies.

Fig. S4. Identification of C-terminal peptides of the polyAla protein by tandem MS. Representative identified spectrum of polyAla C-terminal peptide TTTTSSYPYDVPDYA. Matched b-ions are shown in red, and y-ions are shown in blue for the product ions of the associated precursor ion. Below each spectrum are fragmentation tables displaying matched product ions. The precursor ion was +2 charged with a mass error of 2.1 ppm. The SEQUEST Xcorr and ΔCN values were 2.33 and 0.39, respectively. Ten spectra with peptide probabilities at 95% were assigned to this peptide.

PolyA C-terminal peptide
TTTTSSYPYDVPDYA - Xcorr = 2.33, deltaCN = 0.39, Precursor charge +2, Precursor Error = 2.1 ppm

Fig. S5. Identification of C-terminal peptides of the polyAla protein by tandem MS. Representative identified spectrum of polyAla C-terminal peptide TTTTSSYPYDVPDYA. Matched b-ions are shown in red, and y-ions are shown in blue for the product ions of the associated precursor ion. Below each spectrum are fragmentation tables displaying matched product ions. The precursor ion was +2 charged with a mass error of 2.1 ppm. The SEQUEST Xcorr and ΔCN values were 2.33 and 0.39, respectively. Ten spectra with peptide probabilities at 95% were assigned to this peptide.
Fig. S5. (A) Immunoblots of lysates after transfection with ATT(CAG)$_{2}$-3T with varying repeat lengths. (B) Immunoblot of polyGln protein from cells transfected with constructs of varying repeat lengths show RAN-translation of polyGln proteins occurs in the presence of 107 CAG repeats but not 20 CAG repeats. The ATT(CAG)$_{22}$-3T construct is a positive control. (C) Semiquantitative RT-PCR of CAG and CAA transcripts. (Upper) Schematic diagram depicting the RT-PCR strategy. The Myc-RT primer was used in a first-strand synthesis reaction, and 336 forward (336 F) and 336 reverse (336 R) primers were used for subsequent amplification of the repeat. (Lower) RT-PCR results for the CAG and CAA repeat constructs and β-actin control in the presence (+) or absence (−) of reverse transcriptase. (D) Corresponding immunoblots. (E) Summary of data for individual annexin-V experiments.
Fig. S6. Non-ATG translation in various sequence contexts. (A) Constructs with 20 bp of 5′ endogenous human flanking sequence upstream of repeat (CAG direction) at the HD, HDL2, DM1, and SCA3 loci. (B) Corresponding protein blots. (C) Schematic diagram showing triply tagged lentiviral constructs used for infection of HEK293T cells and mouse brains. All lentiviral constructs are in the CSII lentiviral vector. (D) Protein blots of HEK293T cells after lentiviral vector (Lt) infection with Lt-GFP, Lt-A8(*KMQ_EXP)f1, Lt-HD, Lt-HDL2, Lt-SCA3, and Lt-DM1(M_S). Infected HEK293T cells show robust non-ATG translation of polyGln proteins for Lt-HDL2 and Lt-DM1. PolyAla but not polySer is expressed from all four constructs (Lt-HD, Lt-HDL2, Lt-SCA3, and Lt-DM1) without an ATG in the polyAla frame. (E) Protein blots of mouse cerebellar extracts after lentiviral vector infection and immunoprecipitation. The ∼40 kDa 1C2-positive protein was detected in cerebellar lysates injected with Lt-A8(*KMQ_EXP)f1, Lt-HDL2, and Lt-DM1(M_S), but not Lt-HD, Lt-SCA3, or Lt-GFP. Two FVB animals were injected with each of these viruses, and 4 wk postinjection, polyGln-tagged protein was immunoprecipitated with anti-His antibody and probed with 1C2. As shown, polyGln-tagged protein was immunoprecipitated from tissue infected with the ATG-positive control virus Lt-A8(*KMQ_EXP)f1 as well as from tissue infected with the Lt-DM1 and Lt-HDL2 lacking an ATG in the glutamine frame, although at a substantially lower level.
RAN translation in cell-free rabbit reticulocyte lysates (RRLs) is less permissive and requires alternative start codons. (Lower) Protein blot of the same in vitro translation products probed with the 1C2 antibody. (Upper) Fluorograph showing [\textsuperscript{35}S]-methionine incorporation. (Lower) Protein blot of the same in vitro translation products probed with the 1C2 antibody.
**Fig. S8.** CUG-containing ribonuclear inclusions and CAG DMPK antisense transcripts in DM mice. (A) FISH of CAG probe to detect CUG-containing RNA foci in cardiac sections from DMSXL and WT control animals. (B) Diagram showing the DMPK 3’ UTR and the location of antisense-specific primers for the CAG transcript. For strand-specific priming, a linker sequence (lk1 or lk-2) was attached to the DM1-specific primers for cDNA synthesis. PCR was performed using a primer complementary to the linker sequence and reverse primers anti1B, antiN3, or antiA2. The 3’ end of the DM1 CAG RNA is unknown. (C) Strand-specific RT-PCR of the human DMPK antisense strand in transgenic mice. Strand-specific reverse transcription and PCR were performed with RNA from a pool of 5-month-old mouse hearts (n = 3) and with RNA from DM1 and control human heart samples. Various lines of transgenic mice have been assessed: DM20 mice with 20 CTGs, DM55 mice with 55 CTGs, DM300 mice with ~600 CTGs, and DMSXL mice with >1,000 CTGs. Ctrl hs heart, human control heart; DM1 hs heart, DM1 human heart; M, 250-bp DNA ladder; wt ms, wild-type mouse. Asterisks to the right of corresponding lanes indicate PCR products with large repeats that amplified with low efficiency. Primers used for DNA synthesis and for PCR are indicated on the left. Gapdh indicates PCR with primers for the mouse Gapdh cDNA that self-primed during reverse transcription. Note that these primers also amplified endogenous human GAPDH cDNA, at lower efficiency.
Fig. S9. RAN-translated polyGln protein in DM1 mouse cardiomyocytes and leukocytes and DM1 human leukocytes. (A) Immunofluorescence staining of cardiomyocytes using α-DM1<sub>CAG-Gln</sub> (cy3, red) in DM 300 (~600 CTG repeats) and WT control mice show rare (2.32 ± 2.04%) but reproducible staining in expansion mice but not control animals based on quantitative digital imaging of six representative fields. (B) H&E staining of cardiac tissue comparable to that used in A shows typical cardiac histology including large, boxy, centrally located myocyte nuclei in both DM300 and WT samples. (C) (Upper) Leukocytes in coagulated blood within heart chambers show frequent positive staining (10.69 ± 3.67%) with α-DM1<sub>CAG-Gln</sub> for DM300 mice but not DM20 control animals based on four representative fields. (Lower) Comparable (nonserial) H&E sections. (D) Colocalization of α-DM1<sub>CAG-Gln</sub> (cy3, red) with caspase-8 (Alexa Fluor 488, green) in mouse leukocytes colocalized with caspase-8, an early indication of polyGln-induced apoptosis (1). (E) Conditions for the α-DM1<sub>CAG-Gln</sub> antibody were optimized using frozen tissue. The 1C2 antibody does not adequately detect polyGln inclusions in frozen samples using available methods (2). Therefore we used the 1C2 antibody on fixed tissue to demonstrate independently that DM1 expansion animals express the DM1<sub>CAG-Gln</sub> protein. The 1C2 antibody previously has been shown to recognize expanded polyGln repeat tracts (>40) specifically (3). HRP-labeled 1C2-positive cytoplasmic stain (blue) is seen in leukocytes of the DM55 mouse but not of the DM20 control mouse.

Fig. S10. DM1 polyGln protein in human tissue. (A) Immunofluorescence staining of skeletal muscle with α-DM1_cAG-Gln (cy3, red) and α-desmin (green) antibodies. Infrequent but reproducible α-DM1_cAG-Gln staining (cy3, red) was found in skeletal muscle from one DM1 autopsy case. Boxed regions are enlarged as insets. (B) DM1 human longitudinal skeletal muscle section showing coexpression of polyGln (red) and the early polyGln apoptotic marker caspase-8 (green). (C) Immunofluorescence α-DM1_cAG-Gln (cy3, red) staining of human DM1 leukocytes from peripheral blood. (D) Protein blots of human peripheral blood from a patient with 85 CTG•CAGs probed with 1C2 and α-DM1_cAG-Gln antibodies provide independent evidence that the DM1 cAG-Gln expansion protein is expressed in blood.