Published in final edited form as:

Biochemistry. 2013 November 19; 52(46): 8219–8225. doi:10.1021/bi401129r.

The role of context in RNA structure: flanking sequences reconfigure CAG motif folding in *huntingtin* exon 1 transcripts

Steven Busan and **Kevin M. Weeks***

Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290

Abstract

The length of the CAG repeat region in the *huntingtin* messenger RNA is predictive of Huntington's disease. Structural studies of CAG repeat-containing RNAs suggest that these sequences form simple hairpin structures; however, in the context of the full-length *huntingtin* mRNA, CAG repeats may form complex structures that could be targeted for therapeutic intervention. We examined the structures of transcripts spanning the first exon of the *huntingtin* mRNA with both healthy and disease-prone repeat lengths. In transcripts with 17 to 70 repeats, the CAG sequences base paired extensively with bases in the 5′ UTR and with a conserved region downstream of the CCG repeat region. In *huntingtin* transcripts with healthy numbers of repeats, the previously observed CAG hairpin was either absent or short. In contrast, in transcripts with disease-associated numbers of repeats, a CAG hairpin was present and extended from a three-helix junction. Our findings demonstrate the profound importance of sequence context in RNA folding and identify specific structural differences between healthy and disease-inducing *huntingtin* alleles that may be targets for therapeutic intervention.

Huntington's disease (HD) is a devastating, ultimately fatal neurodegenerative disorder. In healthy individuals the first exon of each of the two alleles of the *huntingtin* gene contains a relatively short region of CAG triplet repeats that encode polyglutamine; the most common allele has 17 repeats. In HD patients, one *huntingtin* allele is abnormally expanded to contain between 36 and 70 CAG repeats, although patient alleles with shorter or significantly longer repeat regions have also been reported.[1,2] The length of this HD-expanded CAG repeat region is inversely correlated with patient age at the onset of symptoms, which include involuntary movements and dementia.[2] Pathogenesis is due to the polyglutamine peptides translated from the disease allele, and the expanded CAG repeat-containing RNA transcripts may also be toxic.[3,4]

This study was motivated by the potential for allele-selective therapeutic targeting of the *huntingtin* mRNA that might result if the RNA structure could be modeled with confidence. Huntingtin is nearly universally expressed and appears to be especially important for correct functioning of the adult nervous system.[5–8] An ideal therapeutic would therefore specifically destroy the disease-expanded *huntingtin* transcript or block its translation while preserving

*correspondence, ; Email: weeks@unc.edu, 919-962-7486

SUPPORTING INFORMATION

Includes figures S1 (RNase cleavage), S2 (additional secondary structure models), and S3 (alternative structures). This material is available free of charge via the Internet at http://pubs.acs.org.

the function of the healthy length transcript. Recent efforts to selectively target the expanded *huntingtin* transcript have focused either on targeting single-nucleotide polymorphisms associated with disease alleles[9,10] or on targeting the CAG repeats, taking advantage of the greater number of effective binding sites in the expanded transcript.[11] Allele-specific structures within the *huntingtin* mRNA could provide additional, and more precise, targets for therapeutic development.

Biochemical studies have consistently demonstrated that RNA transcripts containing CAG repeats fold into duplex helices and hairpins.[12–15] CAG-containing duplexes have been examined by NMR[14] and X-ray crystallography.[16] Recent studies have also shown that flanking sequences can modulate triplet repeat folding. The addition of even a short region of flanking *huntingtin* sequence to CAG repeats results in the formation of more complex structures.[17] We therefore sought to determine the folded structures of *huntingtin* transcripts with varying CAG repeat lengths in the context of the sequence of longer transcripts, more closely resembling those found in cells.

We designed five transcripts covering the entire first exon of the *huntingtin* mRNA. These exon 1 transcripts spanned the 5′ untranslated region (UTR), contained from 17 to 70 CAG repeats, and included the downstream region encoding polyproline repeats (mostly CCG). A combination of SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension), RNase T1 cleavage, and targeted antisense oligonucleotide binding was used to investigate the folded structures of these transcripts. We found that the sequence context had profound effects on the folded structure of the transcript because CAG repeats pair extensively with flanking *huntingtin* mRNA sequences. A CAG hairpin was absent or short in *huntingtin* transcripts with repeat lengths typical of healthy individuals (17 and 23 repeats) but was present in transcripts with disease-associated numbers of repeats (36, 41, and 70 repeats). Our data suggest that there are structural differences between healthy and disease-inducing alleles that may be promising targets for therapeutic intervention.

METHODS

Sequences, primers, and antisense oligonucleotides

Huntingtin mRNA exon 1 transcripts (**n** =17, 23, 36, 41, 70): GCUGCCGGGA CGGGUCCAAG AUGGACGGCC GCUCAGGUUC UGCUUUUACC UGCGGCCCAG AGCCCCAUUC AUUGCCCCGG UGCUGAGCGG CGCCGCGAGU CGGCCCGAGG CCUCCGGGGA CUGCCGUGCC GGGCGGGAGA CCGCCAUGGC GACCCUGGAA AAGCUGAUGA AGGCCUUCGA GUCCCUCAAG UCCUUC $\mathbf{(CAG)_n}$ CAACAGCCGC CACCGCCGCC GCCGCCGCCG CCGCCUCCUC AGCUUCCUCA GCCGCCGCCG CAGGCACAGC CGCUGCUGCC UCAGCCGCAG CCGCCCCCGC CGCCGCCCCC GCCGCCACCC GGCCCGGCUG UGGCUGAGGA GCCGCUGCAC CGACC; reverse transcription primer: GGTCGGTGCAGCG; antisense oligonucleotides, listed by the 5′-most target nucleotide in the 70-CAG *huntingtin* transcript (* indicates a “locked” nucleotide[18]): (1) *TCC*CGG*CAG*C, (159) *ATC*AGC*TTT*T, (431) *AGG*AGG*CG*GCG*GCG*G, (464) *GTG*CCT*GCG*G, (475) *TGA*GGC*AG*CAG*CGG*C.

Transcript production and purification

Plasmids contained *huntingtin* sequences, a T7 promoter on the 5′ end, and a Bts I restriction site on the 3′ end, and were obtained by de novo synthesis (Blue Heron Biotechnology). Cells (SURE 2, Agilent Technologies) were transfected with plasmid, and 500 mL cultures prepared. Plasmids were extracted, and constructs verified by sequencing. Plasmids were linearized with Bts I (New England Biolabs), and linearization was confirmed by agarose gel electrophoresis. Linearized template sequences were transcribed using T7 RNA polymerase, and products were separated by polyacrylamide gel electrophoresis, excised from the gel, and recovered by precipitation with ethanol.[19] Transcripts were resuspended at 0.25 μM in 1/2× TE buffer, aliquoted for single use, and stored at −20 °C.

In vitro transcript folding, SHAPE, and RNase T1 probing

Transcripts were denatured at 95 °C for 2 minutes, snap-cooled on ice for 2 min, and refolded at 37 °C for 30 min in 50 mM Tris-HCl (pH 8), 75 mM KCl, and 3 mM $MgCl_2$. SHAPE probing was performed using 5–8 mM final concentration 1-methyl-7-nitroisatoic anhydride (1M7) for 5 min at 37 °C.[20] Enzymatic cleavage was carried out using RNase T1 (Ambion) at a final concentration of 0.2 U/μL for 5 min at 37 °C. Transcripts were recovered by ethanol precipitation. SuperScript III reverse transcriptase (Invitrogen) was used to extend the fluorescently labeled reverse transcription primer (above) for 1 hr at 37 °C. Fluorescent cDNA fragments were quantified using capillary electrophoresis.[21]

Structure disruption using antisense oligonucleotides

Transcripts were combined with five pooled antisense oligonucleotides (ASOs), containing “locked” nucleotides (Exiqon) to enhance RNA binding, at a four-fold excess of each ASO over RNA. Transcripts were then denatured, snap-cooled, folded, and modified as described above. To reduce the concentration of ASOs prior to reverse transcription, transcripts were incubated with three DNA oligonucleotides complementary to ASOs 431, 464, and 475 at a high concentration (200 times that of the RNA) at 95 °C for 2 minutes. Three serial rounds of binding, washing, and elution (RNeasy MinElute columns, Qiagen) were then performed to remove the ASOs and their complements. Structure analysis by reverse transcription was performed as outlined above.

Electropherogram analysis and structure prediction

Electropherograms were analyzed with QuShape.[21] SHAPE data were analyzed as follows: Nucleotides with no-reagent signals above the 99th percentile in any trial were excluded from analysis in all transcript datasets. SHAPE reactivity profiles were normalized as described,[22] except that the CAG repeat region was excluded from the normalization calculation to maintain a consistent SHAPE reactivity distribution across all transcripts. RNase T1 data were analyzed as follows: Nucleotides with background signals in the top 3% were excluded. Background and plus-RNase signals were normalized to the median of the plus-RNase signal. After background subtraction, guanosine residues showing normalized intensity values between 1 and 2 were designated low cleavage, between 2 and 4 medium cleavage, and above 4 high cleavage.

Secondary structures were modeled using the Fold module of RNAstructure,[23] version 5.4, using the latest parameters for incorporating SHAPE data.[24,25] Since the *huntingtin* mRNA likely forms many non-canonical base pairs and contains multiple regions of repeated sequence, structure modeling was challenging. Without constraining secondary structure models with SHAPE data, RNAstructure predicted a large number of alternative structures of similar energy. SHAPE constraints brought these predictions into agreement with experimental data and significantly reduced the number of plausible structures. Given the overall similarities in nucleotide reactivities across the five transcripts (Figs. 1 and 3, Suppl. Fig. 1), the lowest predicted free energy structure for the shortest transcript was used as a template to select the most likely structure for each of the CAG-expanded transcripts. In addition, we selected those structural models that showed reactive nucleotides in the CAG repeat region within two triplets of a CAG hairpin terminus.

RESULTS

SHAPE and RNase probing of *huntingtin* exon 1

We used SHAPE[26,27] chemical probing to analyze the structure of five RNA transcripts containing shorter CAG repeat lengths (17 and 23 repeats) typical of healthy alleles and longer, disease-associated, numbers of repeats (36, 41, and 70 repeats). Little degradation of RNAs was observed as judged by low peak intensities in reverse transcription products from the no-reagent controls as analyzed by capillary electrophoresis. SHAPE reactivity profiles for each of the transcripts are shown split in the center of the CAG repeat region and aligned at the 5′ and 3′ ends (Fig. 1). Overall, SHAPE reactivity profiles for the five transcripts are highly similar, suggesting that the global secondary structure is not affected by expanded CAG repeats (Fig. 1). Within the CAG repeat region in each transcript, most nucleotides were unreactive, consistent with formation of stable base pairing.[26,27] In addition, within each CAG repeat region, there was a short region with more reactive (conformationally flexible) nucleotides; this region was not centered in the CAG repeat region but, instead, was offset in the 3′ direction (Fig. 1, emphasized with solid arrows). This asymmetry in the CAG repeat regions was also observed by RNase T1 enzyme probing (Supplementary Fig. 1). The group of SHAPE-reactive nucleotides was consistently located six triplets 3′ of the center of the poly-CAG repeat.

Structural models of *huntingtin* transcripts

We used the SHAPE data to develop experimentally-supported[24,25] models for thermodynamically accessible states for each of our *huntingtin* RNA transcripts. The 5′ UTR and 3′ regions of the RNAs are predicted to form similar or identical structures, independent of CAG repeat length (Fig. 2). In general, these structural models are well defined (Supplementary Fig. 2). These models, based on RNA transcripts with long flanking sequences, likely capture features relevant to *huntingtin* mRNA structure *in vivo*. The 5′ end corresponds to the transcription start site − 145 nucleotides from the translation start, although transcripts starting at − 135 may also be present *in vivo*.[28,29] Some end effects are possible due to truncation of the studied transcripts at the 3′ exon boundary (155 nucleotides from the CAG repeat region).

Strikingly, the CAG repeat region forms extensive base pairing interactions with nucleotides outside the repeat region (Fig. 2, Suppl. Fig. 2, CAG repeat sequences are highlighted in orange). The 5′ end of the UTR, the CCG repeat region immediately downstream of the CAG repeat region, and an 11-nucleotide region with the sequence GCCGCUGCUGC (perfectly complementary to CAG repeats apart from one A:C mismatch) are all predicted to base pair with CAG repeat nucleotides. The remarkable result of this base pairing is that a hairpin formed only of CAG repeat nucleotides is entirely absent from the model of the healthy *huntingtin* transcript that contains 17 CAG repeats (Fig. 2, left). Moreover, the CAG-repeat hairpin and the three-helix junction from which it extends represent allele-specific structures that occur preferentially in the longer disease-associated alleles.

CAG hairpin induction

If base pairing between CAG repeats and flanking sequences prevents CAG hairpin formation in healthy-length *huntingtin* transcripts, disrupting this base pairing should allow the RNA to refold and form extended hairpins (Fig. 3, left). We folded all five *huntingtin* transcript RNAs in the presence of five antisense oligonucleotides designed to bind sequences flanking the CAG repeats and to compete for base pairing with these non-CAG sequences. Under these conditions, SHAPE-reactive nucleotides occurred at or near the center of the CAG repeat element in all transcripts, both healthy length and disease expanded (Fig. 3, right, site of hairpin loop is emphasized with open arrow). Thus, CAG repeat elements can be forced to form a simple hairpin structure by inhibiting pairing to flanking sequences present in the native transcript.

DISCUSSION

Our work provides the first empirical examination of *huntingtin* mRNA structure in the context of extended, native flanking sequences, in this case the entire first exon. Given the GC-rich nature of the *huntingtin* mRNA, it is not surprising that the transcripts are highly structured (Fig. 2). The CAG repeat regions adopt distinct structures that depended on repeat length and on the flanking sequence context (Fig. 4). In the absence of interacting flanking sequences, poly-CAG transcripts, which are found in several disease-related contexts,[30,31] fold back on themselves to base-pair into simple hairpins.[12,13] In *huntingtin* exon 1 mRNA sequences, CAG repeats are followed by poly-CCN sequences and a complementary GCCGCUGCUGC sequence, and our analysis indicates that these flanking sequences pair stably with the poly-CAG element. Because CAG repeats base pair with flanking sequences, a CAG hairpin was not observed in the transcript containing the 17 CAG repeats typical of a healthy individual.

Cellular and animal models of HD indicate that disease symptoms correlate with several factors including repeat lengths, expression levels, localization of *huntingtin* transcripts, truncation of the *huntingtin* sequence, and stoichiometry of native and mutant sequences.[32–34] This work supports the hypothesis that the CAG repeat-containing transcript itself, and not just its ability to encode poly-glutamine, might be important for disease etiology. The two widely used mouse models of HD employ a yeast artificial chromosome (YAC128)[35] or a bacterial artificial chromosome (BACHD)[36]. Despite expressing similar mutant *huntingtin* mRNAs, BACHD mice do not show aggregate formation or display the transcriptional dysregulation present in YAC128 mice and HD patients.[37] An important distinction between these models is the use of nearly pure CAG repeats in YAC128 versus unnatural, mixed CAA/CAG repeats in BACHD. The presence of CAA triplets disrupts extended hairpin formation and favors branched secondary structures.[38] In addition, CAA sequences will not base pair strongly with CCG sequences and other flanking regions present within the authentic *huntingtin* transcript sequence. The allele used in the BACHD model will almost certainly lack the striking CAG repeat length-dependent hairpin formation found in this study; therefore, some of the phenotypic differences that distinguish pure CAG from mixed-codon HD models may reflect differences in RNA structure.

The secondary structure models developed in this work also suggest specific roles for *huntingtin* mRNA structure in splicing and translation. First, expanded CAG repeats within *huntingtin* transcripts contribute to misregulation of splicing. These defects include sequestration of the splicing factor Muscleblind-like protein 1[17,39] and mis-splicing of the *huntingtin* transcript, possibly due to recruitment of the splicing factor SRSF6.[40] We hypothesize that base pairing by healthy-length CAG repeats to flanking sequences reduces deleterious recognition by splicing factors. Second, the *huntingtin* 5′ UTR and the region surrounding the primary translation start site form stable RNA structures (Fig. 2); in general, structured UTRs reduce translation initiation.[41] Taken together with a putative active upstream open reading frame in *huntingtin*,[29] this work suggests that regulation of *huntingtin* translation may be complex and involve the interplay of the general translation initiation machinery, contributions of strong local structure at the translation initiation site, and the possible presence of multiple initiation sites.

The absence of a CAG hairpin in short, healthy-length *huntingtin* transcripts and its presence in transcripts with increased numbers of repeats suggests that allele-specific targeting of *huntingtin* mRNA structures will be possible. SHAPE-directed structure models suggest that CAG hairpins occur in disease-associated alleles but not in alleles with fewer repeats characteristic of healthy individuals. Molecules that bind specifically to CAG hairpins, especially if they discriminate against duplexes in which CAG sequences pair with CCG repeat sequences (Fig. 2), are likely to be very selective for disease-causing alleles. The three-helix junction from which the CAG hairpin extends represents another novel RNA target with the potential for both gene and allele selectivity. Broadly, our findings highlight the importance of flanking sequence in RNA folding and hint at the insights to be gained by conducting large-scale RNA structure analyses. Examinations of the effects of context on RNA structure are likely to identify new therapeutic targets in repeat-expansion diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are indebted to the CHDI Foundation for support of this work (E-0409 to K.M.W.).

References

1. Kremer B, Goldberg P, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squitieri F, Lin B, Bassett A, Almqvist E, et al. A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. N Engl J Med. 1994; 330:1401–1406. [PubMed: 8159192]
2. Lee JM, Ramos EM, Lee JH, Gillis T, Mysore JS, Hayden MR, Warby SC, Morrison P, Nance M, Ross CA. CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. Neurology. 2012; 78:690–695. [PubMed: 22323755]
3. Hsu RJ, Hsiao KM, Lin MJ, Li CY, Wang LC, Chen LK, Pan H. Long tract of untranslated CAG repeats is deleterious in transgenic mice. PLoS ONE. 2011; 6:e16417. [PubMed: 21283659]
4. Shieh SY, Bonini NM. Genes and pathways affected by CAG-repeat RNA-based toxicity in Drosophila. Hum Mol Genet. 2011; 20:4810–4821. [PubMed: 21933837]
5. Velier J, Kim M, Schwarz C, Kim TW, Sapp E, Chase K, Aronin N, DiFiglia M. Wild-type and mutant huntingtin's function in vesicle trafficking in the secretory and endocytotic pathways. Exp Neurol. 1998; 152:34–40. [PubMed: 9682010]
6. Gauthier LR, Charrin BC, Borrell-Pagès M, Dompierre JP, Rangone H, Cordelières FP, De Mey J, MacDonald ME, Lessmann V, Humbert S, et al. Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. Cell. 2004; 118:127–138. [PubMed: 15242649]
7. Caviston JP, Holzbaur ELF. Huntingtin as an essential integrator of intracellular vesicular trafficking. Trends Cell Biol. 2009; 19:147–155. [PubMed: 19269181]
8. Zala D, Hinckelmann MV, Saudou F. Huntingtin's function in axonal transport is conserved in drosophila melanogaster. PLoS ONE. 2013; 8:e60162. [PubMed: 23555909]
9. Pfister EL, Kennington L, Straubhaar J, Wagh S, Liu W, DiFiglia M, Landwehrmeyer B, Vonsattel JP, Zamore PD, Aronin N. Five siRNAs targeting three SNPs may provide therapy for three-quarters of huntington's disease patients. Curr Biol. 2009; 19:774–778. [PubMed: 19361997]
10. Carroll JB, Warby SC, Southwell AL, Doty CN, Greenlee S, Skotte N, Hung G, Bennett CF, Freier SM, Hayden MR. Potent and selective antisense oligonucleotides targeting single-nucleotide polymorphisms in the Huntington disease gene/Allele-specific silencing of mutant huntingtin. Mol Ther. 2011; 19:2178–2185. [PubMed: 21971427]
11. Gagnon KT, Pendergraff HM, Deleavey GF, Swayze EE, Potier P, Randolph J, Roesch EB, Chattopadhyaya J, Damha MJ, Bennett CF, et al. Allele-selective inhibition of mutant huntingtin expression with antisense oligonucleotides targeting the expanded CAG repeat. Biochemistry. 2010; 49:10166–10178. [PubMed: 21028906]
12. Sobczak K, de Mezer M, Michiewski G, Kroi J, Krzyzosiak WJ. RNA structure of trinucleotide repeats associated with human neurological diseases. Nucleic Acids Res. 2003; 31:5469–5482. [PubMed: 14500809]
13. Michlewski G, Krzyzosiak WJ. Molecular architecture of CAG repeats in human disease related transcripts. J Mol Biol. 2004; 340:665–679. [PubMed: 15223312]
14. Broda M, Kierzek E, Gdaniec Z, Kulinski T, Kierzek R. Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. Biochemistry. 2005; 44:10873–10882. [PubMed: 16086590]
15. Yuan Y, Compton SA, Sobczak K, Stenberg MG, Thornton CA, Griffith JD, Swanson MS. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. Nucleic Acids Res. 2007; 35:5474–5486. [PubMed: 17702765]
16. Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. Nucleic Acids Res. 2010; 38:8370–8376. [PubMed: 20702420]
17. de Mezer M, Wojciechowska M, Napierala M, Sobczak K, Krzyzosiak WJ. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. Nucleic Acids Res. 2011; 39:3852–3863. [PubMed: 21247881]
18. Vester B, Wengel J. LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. Biochemistry. 2004; 43:13233–13241. [PubMed: 15491130]

19. Wyatt JR, Chastain M, Puglisi JD. Synthesis and purification of large amounts of RNA oligonucleotides. Biotechniques. 1991; 11:764–769. [PubMed: 1809333]
20. Mortimer SA, Weeks KM. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc. 2007; 129:4144–4145. [PubMed: 17367143]
21. Karabiber F, McGinnis JL, Favorov OV, Weeks KM. QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. RNA. 2012; 19:63–73. [PubMed: 23188808]
22. Wilkinson KA, Merino EJ, Weeks KM. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nature Protocols. 2006; 1:1610–1616. [PubMed: 17406453]
23. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics. 2010; 11:116. [PubMed: 20202215]
24. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci USA. 2009; 106:97–102. [PubMed: 19109441]
25. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc Natl Acad Sci USA. 2013; 110:5498–5503. [PubMed: 23503844]
26. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by Selective 2′-Hydroxyl Acylation and Primer Extension (SHAPE). J Am Chem Soc. 2005; 127:4223–4231. [PubMed: 15783204]
27. Weeks KM, Mauger DM. Exploring RNA structural codes with SHAPE chemistry. Acc Chem Res. 2011; 44:1280–1291. [PubMed: 21615079]
28. Lin B, Nasir J, Kalchman MA, McDonald H, Zeisler J, Goldberg YP, Hayden MR. Structural Analysis of the 5′ Region of Mouse and Human Huntington Disease Genes Reveals Conservation of Putative Promoter Region and Di- and Trinucleotide Polymorphisms. Genomics. 1995; 25:707–715. [PubMed: 7759106]
29. Lee J, Park EH, Couture G, Harvey I, Garneau P, Pelletier J. An upstream open reading frame impedes translation of the huntingtin gene. Nucleic Acids Res. 2002; 30:5110–5119. [PubMed: 12466534]
30. Nalavade R, Griesche N, Ryan DP, Hildebrand S, Krau S. Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death and Disease. 2013; 4:e752. [PubMed: 23907466]
31. Galka-Marciniak P, Urbanek MO, Krzyzosiak WJ. Triplet repeats in transcripts: structural insights into RNA toxicity. Biol Chem. 2012; 393:1299–1315. [PubMed: 23052488]
32. Hodgson JG, Smith DJ, McCutcheon K, Koide HB, Nishiyama K, Dinulos MB, Stevens ME, Bissada N, Nasir J, Kanazawa I, et al. Human huntingtin derived from YAC transgenes compensates for loss of murine huntingtin by rescue of the embryonic lethal phenotype. Hum Mol Genet. 1996; 5:1875–1885. [PubMed: 8968738]
33. Ehrnhoefer DE, Butland SL, Pouladi MA, Hayden MR. Mouse models of Huntington disease: variations on a theme. Dis Models Mech. 2009; 2:123–129.
34. Southwell AL, Warby SC, Carroll JB, Doty CN, Skotte NH, Zhang W, Villanueva EB, Kovalik V, Xie Y, Pouladi MA, et al. A fully humanized transgenic mouse model of Huntington disease. Hum Mol Genet. 2013; 22:18–34. [PubMed: 23001568]
35. Slow EJ, van Raamsdonk J, Rogers D, Coleman SH, Graham RK, Deng Y, Oh R, Bissada N, Hossain SM, Yang YZ, et al. Selective striatal neuronal loss in a YAC128 mouse model of Huntington disease. Hum Mol Genet. 2003; 12:1555–1567. [PubMed: 12812983]
36. Gray M, Shirasaki DI, Cepeda C, Andre VM, Wilburn B, Lu XH, Tao J, Yamazaki I, Li SH, Sun YE, et al. Full-length human mutant huntingtin with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. J Neurosci. 2008; 28:6182–6195. [PubMed: 18550760]
37. Pouladi MA, Stanek LM, Xie Y, Franciosi S, Southwell AL, Deng Y, Butland S, Zhang W, Cheng SH, Shihabuddin LS, et al. Marked differences in neurochemistry and aggregates despite similar behavioral and neuropathological features of Huntington disease in full-length BACHD and YAC128 mice. Hum Mol Genet. 2012; 21:2219–2232. [PubMed: 22328089]

38. Sobczak K, Krzyzosiak WJ. CAG repeats containing CAA interruptions form branched hairpin structures in spinocerebellar ataxia type 2 transcripts. J Biol Chem. 2005; 280:3898–3910. [PubMed: 15533937]

39. Mykowska A, Sobczak K, Wojciechowska M, Kozlowski P, Krzyzosiak WJ. CAG repeats mimic CUG repeats in the misregulation of alternative splicing. Nuc Acids Res. 2011; 39:8938–8951.

40. Sathasivam, et al. Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci USA. 2013; 110:2366–2370. [PubMed: 23341618]

41. Pickering BM, Willis AE. The implications of structured 5′ untranslated regions on translation and disease. Semin Cell Dev Biol. 2005; 16:39–47. [PubMed: 15659338]

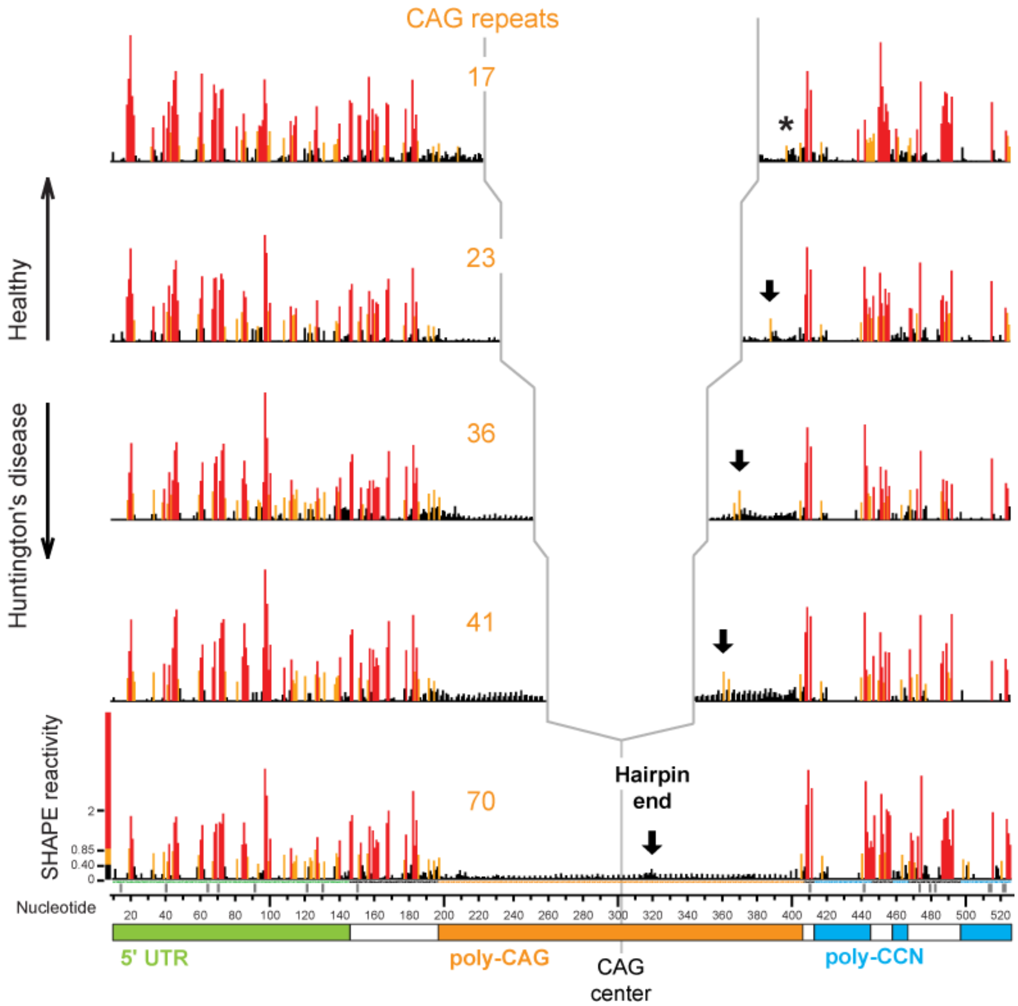

Figure 1. SHAPE profiles for *huntingtin* exon 1 transcripts as a function of CAG repeat length

Reactivity profiles are shown split in the center of the CAG repeat region and aligned at the 5′ and 3′ ends. The black, yellow, red scale indicates low, medium, and high SHAPE reactivities, respectively. The most SHAPE-reactive region within the CAG repeat consistently falls six CAG repeats 3′ of the CAG repeat region center, as emphasized with solid arrows. The region likely to form an internal loop in the 17-CAG repeat transcript is indicated with an asterisk (top panel). Data shown are the average of three independent experiments. The small number of nucleotides for which no data were obtained (due to strong electropherogram peaks in the no-reagent control, see Methods) are marked with gray boxes at the x-axis.

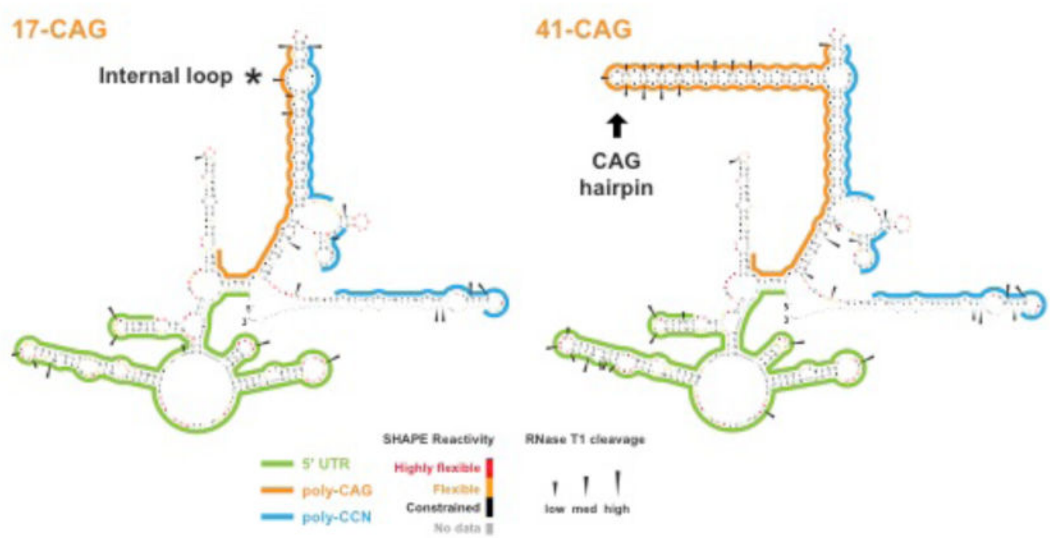

Figure 2. Structural models for representative normal and disease-associated *huntingtin* transcripts

Secondary structure models for the most common healthy length transcript (17-CAG) and for a strongly disease-associated (41-CAG) RNA are shown. SHAPE and T1 RNase probing are shown with colored nucleotides and arrowheads, respectively. Absence of a CAG hairpin in the 17-CAG repeat RNA is emphasized with an asterisk.

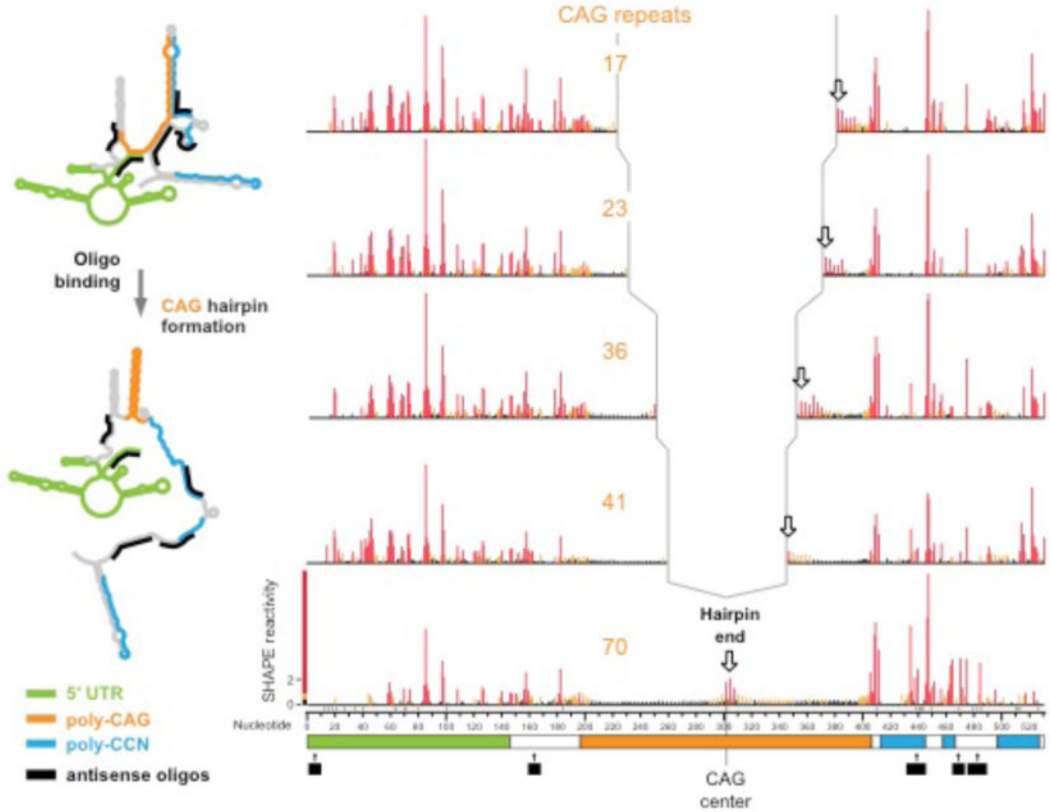

Figure 3. SHAPE analysis of *huntingtin* transcripts in the presence of antisense oligonucleotides designed to disrupt pairing between CAG sequences and flanking regions

Five antisense oligonucleotides were designed to bind specific, non-CAG, sequences in the *huntingtin* mRNA to disrupt base-pairing with the CAG repeat region and to promote formation of a CAG hairpin. Oligonucleotide binding sites are shown with black bars. The center (reactive) region of each CAG repeat element is emphasized with an open arrow and is consistent with simple hairpin formation by self-paired CAG sequences.

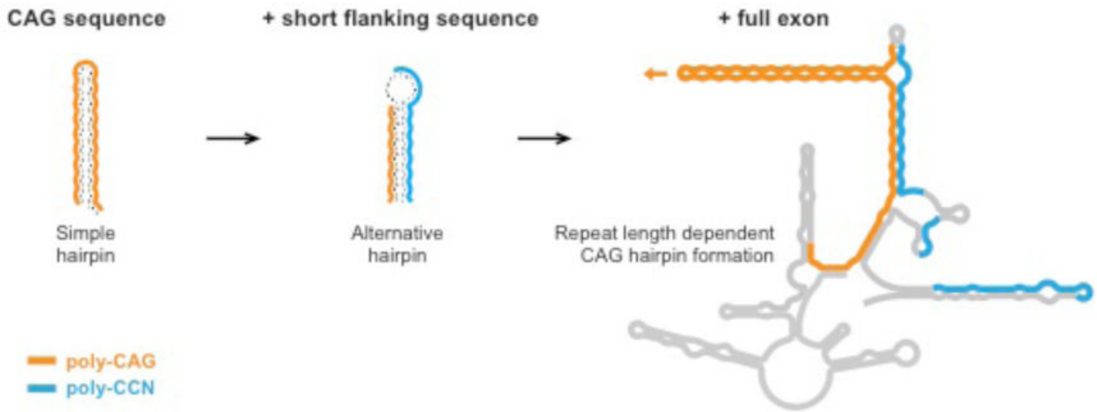

Figure 4. Role of flanking sequence in defining CAG repeat RNA structures

Shown are secondary structure models for the CAG repeat sequence,[12] for a CAG repeat with short flanking sequences,[17] and for the full-length *huntingtin* exon 1 sequence studied in this work. A CAG hairpin begins to form with intermediate-length repeat expansion and preferentially forms a long classical hairpin (shown) with disease-associated CAG expansions.