The AUUCU Repeats Responsible for Spinocerebellar Ataxia Type 10 Form Unusual RNA Hairpins*

Received for publication, March 30, 2005, and in revised form, May 31, 2005
Published, JBC Papers in Press, June 20, 2005, DOI 10.1074/jbc.M503495200

Vaiashali Handa‡§, Herman J. C. Yeh‡§, Peter McPhie¶, and Karen Usdin‡**
From the ‡Laboratory of Molecular and Cellular Biology, the ¶Laboratory of Biochemistry and Genetics, and the **Laboratory of Bioorganic Chemistry, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0830

Spino cerebellar ataxia type 10 (SCA10) is an autosomal dominant disorder caused by expansion of a pentameric repeat tract (ATTCT-AGAAT)n in intron 9 of the gene that encodes ataxin-10. We have analyzed duplex DNA containing the repeat, the individual DNA strands, and the RNA that would be generated by transcription of the repeat. Circular dichroism and UV absorbance measurements suggest that the previously reported tendency of the repeat to unpair when supercoiled is probably related simply to GC content rather than reflecting any unusual property of the duplex. DNA containing d(ATTCT)n forms a folded structure at relatively low temperatures, whereas the antisense strand, d(AGAAT)n, does not form a structure even at 0 °C. In contrast r(AUUCU)n forms a folded structure under physiologically reasonable conditions. S1 nuclease analysis reveals a single region of hypersensitivity in the middle of the repeat tract, whereas V1 digestion is consistent with a hydrogen bonded or well stacked structure. CD spectroscopy shows that the structure is unmolecular and hydrogen bonded and has a significant amount of A-form helix. NMR spectroscopy demonstrates that these hydrogen bonds comprise an equal number of A-U and U-U base pairs. Our data thus suggest that the repeat forms an unusual RNA hairpin. Thus the ability to form an RNA hairpin seems to be a common property of those Repeat Expansion Diseases that are not recessively inherited and are caused by repeats that are transcribed but not translated.

Recent data show that loss of ataxin-10 causes apoptosis of cerebellar neurons (4). However, this observation is not necessarily proof of a loss of function mechanism for SCA10. For example, null mutations in the huntingtin gene also cause apoptosis (5), yet it was well established that the expansion responsible for this disease causes a toxic gain of function of the huntingtin protein that is responsible for disease pathology (6). Preliminary data suggest that the levels of SCA10 mRNA in cells from affected individuals are not significantly different from those in unaffected individuals (2). In addition, RNA corresponding to the expanded repeat tract can be detected in the lymphoblasts from affected individuals (3). Although it remains to be seen whether such RNA is present in brains of patients, it is possible that RNA toxicity, a mechanism proposed for a number of other Repeat Expansion Diseases (7–9), may also be a viable mechanism for SCA10 disease pathology (3).

The repeats that are thought to produce toxic RNA all form RNA hairpins (10–13). RNA with double-stranded character can have a wide range of biological effects, including sequestering of proteins bound by the repeat, activation of enzymes such as the RNA-dependent protein kinase PKR, which can lead to apoptosis (14), and regulation of gene expression via the RNA interference pathway (see Ref. 15 for recent review). Previous work on the SCA10 repeat has focused on the DNA version of these repeats. These studies showed no evidence of the formation of hairpins (16). Using enzymatic digestion and CD spectroscopy we have examined the behavior of the DNA repeats in more detail. In addition, using enzymatic digestion as well as CD and NMR spectroscopy, we have looked at the structures formed by the repeats in RNA. Our findings may have significance for disease pathology in this dominant cerebellar disorder.

MATERIALS AND METHODS

DNA and RNA Preparation—All oligonucleotides were synthesized by Integrated DNA Technology (Coralville, IA) using standard phosphoramidite chemistry. SCA-F (5′-TCGAGTTAATCGAATCTAT-AGGGAGAATTCCTAG-3′) contains at its 5′-end 4 bases that correspond to the 5′-overhang generated by XhoI as well as a T7 RNA polymerase promoter. SCA-R (5′-AGACTTAGAATAGAAT(AGAAT)14 AGAATTCTC- CCTATACTAGTCATTTAT-3′) has at its 5′-end 4 bases corresponding to the 5′-overhang that would be generated by HindIII digestion. These oligonucleotides were PAGE purified. For CD spectroscopy, DNA oligonucleotides containing either d(ATTCT)n or d(AGAAT)n were synthesized chemically and high pressure liquid chromatography (HPLC)1 purified. For both CD spectroscopy and NMR studies an RNA oligonucleotide containing r(AUUCU)n was synthesized chemically and purified by RNase-free HPLC.

Longer RNA tracts were synthesized enzymatically as follows. The oligonucleotides SCA-F and SCA-R were mixed and diluted to a final concentration of 10 μM in 1× annealing buffer (Invitrogen), heated to

1 The abbreviation used is: HPLC, high pressure liquid chromatography.
The SCA10 Repeat Forms Unusual RNA Hairpins

Nuclease Digestion—The chemically synthesized DNA and the enzymatically synthesized RNA were end labeled with T4 polynucleotide kinase and [γ-32P]GTP according to standard procedures. For S1 nuclease treatment at 20 °C, samples were heated in 10 μl of TE for 3 min at 95 °C and then diluted into a final volume of 100 μl of standard S1 buffer containing 2 μl of Supersafe I (Ambion, Austin, TX). 25-μl aliquots of this mixture were treated with 0, 1, 3, and 5 units of S1 nuclease for 20 min. For S1 treatment at 37 °C, samples were heated in 10 μl of TE for 3 min at 95 °C and then diluted into a final volume of 100 μl of standard S1 buffer containing 2 μl of Supersafe I. 20-μl aliquots were digested for 20 min with 0, 0.1, 0.25, 0.5, and 1 unit of S1. The reactions were then mixed with 180 μl of TE and extracted with phenol:chloroform:isoamyl alcohol, precipitated with ethanol, and dissolved in 8 μl of Stop buffer.

For RNase V1 analysis, RNA was heated in TE for 3 min at 90 °C. RNase V1 (Ambion) digestion was carried out at 20 °C for 15 min using 25 μl of buffer containing 10 mM Tris-HCl, pH 7.0, 100 mM KCl, 10 mM MgCl₂, and 0.01, and 0.1 units of enzyme. Reactions were terminated by addition of phenol:chloroform:isoamyl alcohol, and the RNA was precipitated with ethanol and dissolved in 8 μl of Stop buffer.

The S1 nuclease and the RNase V1-treated samples were heated at 95 °C for 3 min and resolved by electrophoresis on an 8% denaturing polyacrylamide gel. The RNAs show a low level of baseline cleavage at each A residue in the repeat tract even in the absence of added enzyme. This cleavage does not increase when enzyme is added, and because the original RNA used in the reaction was purified as a single product on a gel with similar resolution, it is probably due to acid-induced depurination that occurred during sample denaturation prior to gel loading. Although it does not reflect nuclease-specific cleavage, it serves as a useful internal molecular weight marker for identifying the nuclease-specific cleavages. The bands corresponding to residues A1-A14 were assigned by reference to a sequencing ladder.

Circular Dichroism (CD) Spectroscopy—CD and absorbance spectra were measured in a Jasco J-715 spectropolarimeter on solutions in a 1-cm or 1-mm path length quartz cuvette held in a cell holder thermostatted by a Neslab RTE-111 circulating water bath. Spectra were scanned four times, from 320 to 205 nm, and averaged (speed = 50 nm/min, time constant = 1 s.). Melting curves were determined for DNA between 0 and 95 °C and for RNA between 5 and 50 °C, at 260 nm, heating at 1 degree/min. After the water bath had reached the required temperature, solutions were recooled at 1 degree/min and spectra measured after 10 min of equilibration. After baseline correction, the measured ellipticities (in millidegrees) were converted into mean residue ellipticities using the formula \[ \text{MRE} = \frac{100}{c l} \left( \frac{1}{T} - 1 \right) \]

**Fig. 1.** Absorbance melting profiles of SCA10 DNA. The absorbance at 260 nm was measured as a function of temperature for d(AT-TCT)₉ and d(AGAAT)₉ as well as a mixture of both oligodeoxyribonucleotides. The triangles indicate the data for d(AT-TCT)₉, the squares the measurements obtained for d(AGAAT)₉, and the circles the results obtained for the mixture of the 2 oligodeoxyribonucleotides. The data are plotted as relative absorbance with the high temperature value set to 1.

**Fig. 2.** S1 sensitivity of the top strand of the SCA10 DNA repeat and the corresponding RNA. An oligodeoxyribonucleotide containing d(AT-TCT)₁₄ and RNAs containing r(AUUCU)₁₄, r(AUUCU)₁₇, and r(AUUCU)₁₁, were digested with S1 nuclease at the indicated temperatures and analyzed by electrophoresis as described under “Materials and Methods.” The first T of the 3 Ts in each DNA repeat is marked, with the numbering corresponding to the repeat number starting from the 5’ end. The As within the RNA repeat tracts are labeled with the number referring to the number of the repeat from the 5’ end of the repeat tract. The brackets demarcate the regions on the 3 RNAs that are S1 hypersensitive. The black brackets mark the hypersensitive region on r(AUUCU)₁₄, the gray brackets the S1-sensitive region on r(AUUCU)₁₇, and the dotted line the hypersensitive region on the RNA with an interrupted repeat tract. The solid gray line on the right hand side of the figure marks the pure r(AUUCU)₁₁, repeat tract in the r(AUUCU)₁₁, RNA. The dotted line delineates the interrupted portion of the repeat tract.
The melting curves were analyzed by assuming simple two-state equilibria between native and denatured states. At each temperature, \( T \) (degrees K), the equilibrium constant for unfolding was calculated as:

\[
K(T) = \frac{([\theta]_N - [\theta](T))([\theta](T) - [\theta]_D)}{[\theta](T)}
\]

where \([\theta]_N\) and \([\theta]_D\) are the extrapolated mean residue ellipticities of the native and denatured states, respectively, and \([\theta](T)\) is the measured mean residue ellipticity.

The temperature dependence of \( K(T) \) is given by:

\[
K(T) = \exp(-\Delta H (1 - \frac{T}{T_m}/R T_m))
\]

where \( \Delta H \) is the apparent enthalpy of denaturation and \( T_m \) is the melting temperature (\( K(T) = 1 \)). Curve fitting was performed using the PC-MLAB program (Civilised Software, Wheaton, MD).

**NMR Analysis**—Approximately 29 A_260 units (66 nmol) of RNA were resuspended in 500 \( \mu \)l of \( \mathrm{H}_2\mathrm{O}/\mathrm{D}_2\mathrm{O} \) buffer (containing 50 mM sodium phosphate, 0.1 mM EDTA, pH 7.2). NMR experiments were performed using a cryoprobe on a Bruker Avance 500 spectrometer equipped with pulsed-field gradient and shape pulse capabilities. Chemical shifts were referenced relative to the deuterated water peak at 4.72 ppm. The transmitter offsets were set on the water signal, which was suppressed using excitation sculpting and gradient pulse scheme as previously reported (17). A spectral width of 11,000 Hz was used in both \( f_1 \) and \( f_2 \) dimensions. Two-dimensional NOESY experiments were carried out with 250-ms mixing time and 2,048 points in \( t_2 \) dimension and 512 \( t_1 \) increments using the States-Haberkorn-Ruben method (18). All data were zero filled to typical matrix size of 4,096 \times 2,048 points, and shifted squared sine-bell weighting functions were used in the final Fourier transformation.

**RESULTS**

Pathological SCA10 alleles contain many hundreds of repeats. Unfortunately, repeat tracts of this size are extremely difficult to clone stably in bacteria or yeast and are far beyond the current limits of chemical synthesis. It is also extremely difficult to generate sufficient material for an analysis of this kind by amplification of the repeat from the genome of affected individuals. However, previous work on shorter versions of the repeat tracts responsible for the other Repeat Expansion Diseases has thrown light on the behavior of these repeats that has helped shape current thinking about the mechanisms of disease pathology (see Ref. 19 and references therein).

When d(ATTCT)_9 is mixed with d(AGAAT)_9 at 10 °C a CD spectrum is produced at low temperature that is typical of a normal B-form double helix (data not shown). On heating, the UV absorbance shows a clear thermal transition characteristic of the melting of a normal Watson-Crick base-paired duplex (Fig. 1). The transition occurs at 72 °C in 50 mM Na⁺, which is somewhat higher than the \( T_m \) predicted by nearest neighbor calculations (63.9 °C). Thus the duplex does not demonstrate any unusual propensity to unpair beyond what would be expected of a sequence with a similar G+C content. If the mixture...
of oligonucleotides is then cooled to 10 °C, the duplex reanneals normally and the thermal transition is fully reversible. However, if the temperature of annealing is dropped to 0 °C no annealing of the duplex occurs. The bottom strand of the repeat shows no sharp thermal transition over the entire thermal range tested (Fig. 1). It is thus presumably unstructured. However, UV absorbance measurements indicate that an oligonucleotide containing the sequence d(ATTCT)₉ undergoes a sharp transition at neutral pH in 50 mM NaPO₄ over a 9-fold range from 0.023 to 0.26 mg/ml (data not shown), suggesting that it results from the presence of a folded (intra-strand) structure. The formation of this unimolecular structure might explain why reannealing of a mixture of the top and bottom strands does not occur when the annealing temperature drops below the Tₘ of the structure formed by the top strand.

S1 nuclease is an enzyme with a strong cleavage preference for single-stranded regions in RNA and DNA. As might be expected from our absorbance data, digestion of an oligonucleotide containing d(ATTCT)₁₄ with S1 nuclease at 20 °C shows no cleavage in the middle of the repeat tract that would be consistent with a hairpin (Fig. 2). There is a small amount of cleavage at the first T in each repeat. We attribute this cleavage to some sort of site preference of S1 nuclease that reflects some aspect of the backbone structure of the oligonucleotide. When RNA containing 14 AUUCU repeats is digested with S1 nuclease at 20 °C at pH 5.2, a different pattern of cleavage can be seen (Fig. 2). There is cleavage of the first U in each repeat, reminiscent of the cleavage at the first T in each repeat seen in DNA. This presumably reflects the same sort of site preference of S1 nuclease, perhaps because of some unusual conformation of this base in this sequence context. There is also a small amount of cleavage at A residues even in the absence of added enzyme. This cleavage is because of acid-induced depurination during sample preparation (see “Materials and Methods”). Finally, there is a cluster of bases in the middle of the repeat tract that is sensitive to S1 digestion. S1 hypersensitivity in the middle of the repeat tract is consistent with the formation of a hairpin. Similar results are seen with 17 repeats. Even a tract containing as few as 11 uninterrupted repeats is capable of forming a hairpin under these conditions.

RNase V1, which cleaves bases that are paired or well stacked (20), produces a periodic cleavage pattern at neutral pH that involves the 3 consecutive bases AUU (Fig. 3). Bases in the middle of the repeat tract hypersensitive to AUU are also digested by V1 nuclease. One way to reconcile this observation is if the bases in the hairpin loop are unpaired and thus vulnerable to S1 but relatively well stacked and thus also vulnerable to V1. This may not be surprising because RNA hairpin loops are frequently highly structured (21).

Circular dichroism spectroscopy of r(AUUCU)₉ at 5 °C shows a strong maximum at 262 nm, a small negative signal at ~240 nm, and a significant negative peak below ~210 nm at pH 7.5 (Fig. 4) or 5.2 (data not shown). These spectra are consistent with the presence of a highly structured RNA with a significant amount of A-form helix (22). A distinct change in the CD spectrum was seen when samples were heated to temperatures above 45 °C. A significant decrease in molar ellipticity is seen with a shift in the positive peak from ~262 to ~275 nm, typical of an unstructured single-stranded RNA. The thermal transition was relatively sharp, monophasic, and fully reversible (Fig. 5). The Tₘ was independent of strand concentration over an 11-fold concentration range from 0.023 to 0.26 mg/ml (data not shown). Thus our CD data suggest that the structure formed by the SCA10 RNA is an unimolecular one (23), most likely a hairpin. At neutral pH in 50 mM NaPO₄ the Tₘ of the structure was ~20.3 °C. Addition of 2 mM Zn²⁺ or Mg²⁺ resulted in a Tₘ of 37.7 °C. At pH 5.2 addition of divalent cation had a much smaller effect with the Tₘ of the structure without Zn²⁺ being ~35.4 °C and with Zn²⁺ being ~37.7 °C. The higher Tₘ at pH 5.2 indicates that the hairpin is stabilized by low pH. However, our data also show that physiologically reasonable levels of divalent cation can substitute for protonation at neutral pH.

For both DNA and RNA addition of salts increases the Tₘ by preferential binding of cations to the helical form. However, for double-stranded complexes such as acid poly(A) and acid poly(C), electrostatic interactions between the protonated bases and the negatively charged phosphates are important in stabilizing the helix. Increasing ionic strength masks these and results in decreases in Tₘ (24). The behavior of a structure containing both neutral and protonated base pairs can be expected to reflect both of these trends. At pH 5.2, they seem to balance out, with added cations producing little change in stability. The situation at neutral pH is more complex; the protonated bases may be maintained by pairing, but being well above their intrinsic pHₐₜ they will revert to the neutral form when the hairpin is absent. Therefore, at this pH divalent cations might be more important.
The NMR spectrum of the r(AUUCU)₉ oligoribonucleotide shows three low field resonances (9.8–14.0 ppm), typical of hydrogen-bonded imino protons (Fig. 6A). These resonances disappear above the Tₘ of the structure seen in the CD spectra and reappear when the sample is recooled. The resonance at 14.0 ppm is assigned to the imino proton of uracil in an A-U base pair, and the resonances at 9.8 and 10.8 are assigned to the two imino protons in a mismatch U-U base pair, as reported for other RNAs containing U-U mismatches (25, 26). Formation of a U-U base pair is evidenced by the strong nuclear Overhauser effect interaction between the two imino protons (9.8 and 10.8 ppm) (Fig. 6B). Each of the U-U imino protons also displays nuclear Overhauser effect interaction to the A-U-paired imino proton (14.0 ppm), indicating direct stacking of the U-U base pair to an adjacent A-U base pair. NMR integration data suggest that one U-U base pair is formed per every A-U base pair, and each repeat contributes bases toward the formation of 2 A-U base pairs and 2 U-U pairs. Fig. 7 shows all possible alignments of 2 strands of the r(AUUCU)₉ oligoribonucleotide for both the antiparallel and parallel orientations. Only one alignment (alignment C) has equal ratios of A-U and U-U pairs. In this alignment the 2 strands are antiparallel and the hairpin (or duplex) is composed of a series of stems, each containing 2 A-U base pairs flanked by 2 U-U base pairs. The stems are separated from one another by symmetrical loops containing a mismatch C-C pair.

The involvement of 4 bases in each repeat in a hydrogen bond at neutral pH is at seeming odds with the V1 nuclease data: V1 shows a periodicity of cleavage at 3 consecutive bases followed by 2 residues that are much less sensitive to cleavage. However, although V1 nuclease is known to cleave base-paired or well stacked regions, not all bases in this region are always cleaved; in some cases one base in the pair is resistant to cleavage while the other is sensitive (20). Our model suggests that the last U in each repeat is not a good substrate for V1 cleavage despite the fact that the base to which it is likely to be hydrogen bonded is readily cleaved. This might occur if the U-U base pairs are not symmetrical or if they cause significant distortion of the RNA helix (27).

DISCUSSION

We have shown that the DNA duplex formed by the complementary strands of the SCA10 repeat has a CD spectrum typical of normal B-form DNA. It does not seem to have a particular predisposition to unpairing other than that resulting from its relatively low G+C content. Although d(AGAAT)₉ from the bottom strand of the SCA10 repeat does not form a folded structure even at 0 °C, even relatively short d(ATTCT) tracts from the top strand of the SCA10 repeats form a hairpin albeit at very low temperatures. In plasmids containing SCA10 repeats, the unpaired “bubble” visible by atomic force microscopy in less highly supercoiled plasmids forms “locally condensed” or
SCA10 alleles have repeat tracts that are among the largest found in the Repeat Expansion Diseases. The repeat tract in affected alleles is 50 to >250 times longer than the ones we have examined here (2). It is thus possible that RNA from these alleles could form larger hairpins with increased stability. An alternative scenario can also be envisioned: alleles with 800–4000 repeats might produce transcripts that form a tangled network of hundreds of smaller, less stable hairpins. RNA hairpins may cause disease pathology in a number of ways. By analogy with what has been suggested for the hairpins formed by the repeats responsible for myotonic dystrophy type 1 and type 2, these hairpins may affect the export of the SCA10 transcript from the nucleus or facilitate the sequestering of proteins important for splicing or other processes (10, 13, 30). They may also activate enzymes like RNA-dependent protein kinase PKR that lead to apoptosis (31) or affect transcription and translation via the RNA interference pathway (11).

Hairpins, particularly those in A+T-rich regions, may also increase the likelihood of premature transcript release from the RNA polymerase complex (32), leading to decreased levels of full-length transcript, or interfere with the normal secondary structure in the intron important for splicing or silencer/enhancer function (33). It remains to be seen which, if any, of these effects contribute to the symptoms of SCA10.

### FIG. 7. Possible anti-parallel (A–E) and parallel (F–J) alignments for an RNA duplex (or hairpin) containing 9 repeats of AUUCU. The loop region for a hairpin or extension in a duplex strand is denoted by ~~~. The first base in each repeat is underlined. A-U and U-U base pairs are indicated by the shaded areas. The boxed alignment (C) is the only one consistent with the NMR data showing 2 A-U base pairs and 2 U-U mismatch base pairs/repeat.

| Possible anti-parallel hairpin (or duplex) alignments | Base pair ratio |
|------------------------------------------------------|-----------------|
| A | A-U-U-U-U-U-U-U-U | AU: UU = 2:1 |
| B | A-U-U-U-U-U-U-U-U | AU: UU = 0:2 |
| C | A-U-U-U-U-U-U-U-U | AU: UU = 2:2 |
| D | A-U-U-U-U-U-U-U-U | AU: UU = 2:1 |
| E | A-U-U-U-U-U-U-U-U | AU: UU = 0:3 |

| Possible parallel duplex alignments | Base pair ratio |
|-----------------------------------|-----------------|
| F | A-U-U-U-U-U-U-U-U | AU: UU = 0:3 |
| G | A-U-U-U-U-U-U-U-U | AU: UU = 2:1 |
| H | A-U-U-U-U-U-U-U-U | AU: UU = 1:2 |
| I | A-U-U-U-U-U-U-U-U | AU: UU = 1:2 |
| J | A-U-U-U-U-U-U-U-U | AU: UU = 2:1 |

“collapsed” structures in more highly supercoiled ones (16). Because under conditions of high superhelical stress more of the repeat region would be expected to be unpaired, it may be that the collapsed structure results, at least in part, from hairpin formation by the top strand of the repeat. Formation of this hairpin would leave the complementary strand free to interact as it does with small molecule probes (16). It also impedes reannealing of the complementary strands, at least at low temperatures.

The stability of the SCA10 DNA hairpin is reminiscent of that reported for the GAA repeat responsible for Friedreich ataxia (28). Thus our data are consistent with the idea that hairpin-forming ability is a common property of at least one strand of all disease-associated repeats (29). However, the SCA10 and FRDA repeats are considerably less stable than the other repeats in this category.

In contrast, RNA containing as few as 9 (AUUCU) repeats forms a folded structure that can be seen even at physiological temperatures. The structure formed is a hairpin in which the stem contains a mixture of A-U and U-U base pairs in a 1:1 ratio. U-U mismatches are the third most common mismatch seen in a data base of 101 small subunit rRNAs, 218 large subunit rRNAs, and 75 group I introns (25). U-U mismatches are indicated by the shaded areas underlined. A-U and U-U base pairs are indicated by the shaded areas. The boxed alignment (C) is the only one consistent with the NMR data showing 2 A-U base pairs and 2 U-U mismatch base pairs/repeat.