

# RNA structure of trinucleotide repeats associated with human neurological diseases

Krzysztof Sobczak, Mateusz de Mezer, Gracjan Michlewski, Jacek Krol and Włodzimierz J. Krzyzosiak\*

Laboratory of Cancer Genetics, Institute of Bioorganic Chemistry, Polish Academy of Science, Noskowskiego 12/14, 61-704 Poznan, Poland

Received June 11, 2003; Revised and Accepted August 8, 2003

## ABSTRACT

**The tandem repeats of trinucleotide sequences are present in many human genes and their expansion in specific genes causes a number of hereditary neurological disorders. The normal function of triplet repeats in transcripts is barely known and the role of expanded RNA repeats in the pathogenesis of Triplet Repeat Expansion Diseases needs to be more fully elucidated. Here we have described the structures formed by transcripts composed of AAG, CAG, CCG, CGG and CUG repeats, which were determined by chemical and enzymatic structure probing. With the exception of the repeated AAG motif, all studied repeats form hairpin structures and these hairpins show several alternative alignments. We have determined the molecular architectures of these co-existing hairpin structures by using transcripts with GC-clamps which imposed single alignments of hairpins. We have provided experimental evidence that CCUG repeats implicated in myotonic dystrophy type 2 also form hairpin structures with properties similar to that composed of the CUG repeats.**

## INTRODUCTION

Trinucleotide repeats are the most abundant type of simple sequence repeats in the coding sequences of all known eukaryotic genomes (1). In introns and intergenic regions they are significantly less frequent than dinucleotide and tetranucleotide repeats (2). The frequency of specific types of trinucleotide repeats and their localization in genes vary significantly between genomes reflecting the important role of the repeats in genome evolution (1). The high mutation rate of trinucleotide repeats makes them a rich source of quantitative genetic variation (3). They easily undergo slipped-strand mispairing during DNA replication which appears to be the primary mechanism leading to the natural polymorphism of the repeats number (3). In some human genes, the repeats may undergo pathogenic expansions that cause severe neurological, mostly neurodegenerative and neuromuscular

disorders (4) known as Triplet Repeat Expansion Diseases (TREDs).

A portion of the trinucleotide repeats which are present in transcribed sequences is retained in mature mRNAs. About 2% of mRNAs contain various types of trinucleotides repeated tandemly at least six times as revealed by our survey of the human transcriptome (5). The normal physiological function of triplet repeats in RNA is poorly understood. They are believed to play some regulatory roles mediated by their interactions with specific repeat binding proteins (6–10). In the nucleus they may be involved in the regulation of RNA splicing, maturation and transport (11–19). In the cytoplasm they may regulate mRNA stability and translation (20). It is now commonly accepted that the RNA repeats play a direct, causative role in the pathogenesis of myotonic dystrophy (15,21,22). It is also believed that the mechanism of RNA pathogenesis contributes to other diseases caused by expansions of triplet repeats in non-coding sequences (23,24).

Five different types of trinucleotide repeats are present in the transcripts of 16 genes known to be associated with TREDs. They include the (CUG)<sub>n</sub>, (CAG)<sub>n</sub>, (CGG)<sub>n</sub>, (CCG)<sub>n</sub> and (AAG)<sub>n</sub> repeats that occur either in translated or non-translated sequences (reviewed in 24–28). The (CGG)<sub>n</sub> and (CCG)<sub>n</sub> repeats implicated in the fragile X syndrome of the FRAXA and FRAXE types, respectively, occur in the 5'-UTR. The (AAG)<sub>n</sub> repeats involved in Friedreich's ataxia are present in an intron. The (CAG)<sub>n</sub> repeats associated with Huntington's disease, Kennedy's disease and several spinocerebellar ataxias (SCAs) occur in the ORF. The exception is SCA12 with the repeats located in the 5'-UTR. Finally, the (CUG)<sub>n</sub> repeats implicated in myotonic dystrophy type 1 (DM1) and SCA8 are located in the 3'-UTR. Tetranucleotide (CCUG)<sub>n</sub> repeats involved in myotonic dystrophy type 2 (DM2) occur in an intron. The normal variation in the repeat number in these transcripts is usually between 5 and 30, and pathogenic mutations begin in most cases from about 40 repeats (26).

Tandem repeats of the CUG motif form hairpin structures if they are of sufficient length (8,9,29,30). The hairpin stems are composed of regular base pairs interrupted by periodic U/U mismatches. In the transcripts of some genes involved in TREDs the (CNG)<sub>n</sub> hairpins may be extended and stabilized by double-stranded RNA regions formed by specific sequences that flank the repeats (5). In others such an effect is

\*To whom correspondence should be addressed. Tel: +48 61 8528503; Fax: +48 61 8520532; Email: wlojdkrzy@ibch.poznan.pl

**Table 1.** Oligodeoxynucleotides used and synthesized transcripts

Oligonucleotide symbol	Sequence (5'–3') <sup>a</sup>	Transcript symbol
cug17	(cag) <sub>17</sub> <u>cctatagtgagtcgtatta</u>	CUG17
cug17_cl2	g(cag) <sub>17</sub> <u>cctatagtgagtcgtatta</u>	CUG17_cl2
cug17_cl4	ggc(cag) <sub>17</sub> <u>gcctatagtgagtcgtatta</u>	CUG17_cl4
cug17_cl	ggccc(cag) <sub>17</sub> <u>ggcctatagtgagtcgtatta</u>	CUG17_cl
cug16_cl	ggccc(cag) <sub>16</sub> <u>ggcctatagtgagtcgtatta</u>	CUG16_cl
cag17	(ctg) <sub>17</sub> <u>cctatagtgagtcgtatta</u>	CAG17
cag17_cl	ggccc(ctg) <sub>17</sub> <u>ggcctatagtgagtcgtatta</u>	CAG17_cl
cag16_cl	ggccc(ctg) <sub>16</sub> <u>ggcctatagtgagtcgtatta</u>	CAG16_cl
cgg17	(cgg) <sub>17</sub> <u>cctatagtgagtcgtatta</u>	CGG17
cgg17_cl	ggccc(cgg) <sub>17</sub> <u>ggcctatagtgagtcgtatta</u>	CGG17_cl
cgg16_cl	ggccc(cgg) <sub>16</sub> <u>ggcctatagtgagtcgtatta</u>	CGG16_cl
ccg17	(cgg) <sub>17</sub> <u>cctatagtgagtcgtatta</u>	CCG17
ccg17_cl	ggccc(cgg) <sub>17</sub> <u>ggcctatagtgagtcgtatta</u>	CCG17_cl
gcc17_cl	ggccc(ggc) <sub>17</sub> <u>ggcctatagtgagtcgtatta</u>	GCC17_cl
aag17	(ctt) <sub>17</sub> <u>cctatagtgagtcgtatta</u>	AAG17
aag17_cl	ggccc(ctt) <sub>17</sub> <u>ggcctatagtgagtcgtatta</u>	AAG17_cl
ccug17	(cagg) <sub>17</sub> <u>cctatagtgagtcgtatta</u>	CCUG17
ccug17_cl	ggccc(cagg) <sub>17</sub> <u>ggcctatagtgagtcgtatta</u>	CCUG17_cl
ccug14_cl	ggccc(cagg) <sub>14</sub> <u>ggcctatagtgagtcgtatta</u>	CCUG14_cl
T7GG	taatacgaactactatagg	

<sup>a</sup>Promoter sequence underlined.

unlikely to occur and hairpins are predicted to be composed of the repeated sequence only. Some of these predictions were already verified by chemical and enzymatic probing of the repeats present in the specific sequence context of their host transcripts (29,31). However, in these earlier studies some complications were encountered caused by the different kinds of structure heterogeneity of the *in vitro* transcripts.

This study was designed to determine and directly compare the structural features of different types of simple sequence repeats in two experimental settings: first, in which the repeats have a freedom of alignment, and second, in which a single alignment of the repeat is imposed. Accordingly, we have analyzed the RNA structures formed by five different types of trinucleotide repeat and one type of tetranucleotide repeat. All parameters of the transcript characteristics and conditions of structure probing reactions were kept constant.

## MATERIALS AND METHODS

### Preparation of DNA templates for *in vitro* transcription

RNAs used in this study were prepared by the *in vitro* transcription of synthetic DNA templates with T7 RNA polymerase. DNA oligonucleotides were synthesized at MWG Biotech, and purified by polyacrylamide gel electrophoresis. Each oligomer contained the sequence complementary to the *in vitro* transcript and to T7 RNA polymerase promoter at its 3'-end (Table 1). Before transcription the oligomers containing (CAG)<sub>n</sub>, (CTG)<sub>n</sub> and (CTT)<sub>n</sub> repeats (200 pmol DNA) were annealed with the 19mer T7GG containing T7 RNA polymerase promoter (400 pmol) in a buffer containing 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA, in order to obtain a double-stranded promoter region. The samples were heated at 95°C for 1 min and cooled on ice. Double-stranded DNA templates containing (GCC)<sub>n</sub> (CGG)<sub>n</sub> and (CAGG)<sub>n</sub> repeats were prepared using the primer extension procedure (200 pmol

template oligomer and 1 mmol T7GG oligomer, 200 μM each dNTPs, standard PCR buffer and 0.5 U *Taq* DNA polymerase per 100 μl reaction) in two-step PCR: 50 cycles at 94°C for 15 s and 45°C for 15 s, and purified using Microcon YM30 (Millipore) centrifugal filter devices.

### Transcription *in vitro*

The transcription reaction carried out in a 50 μl volume contained 20 pmol of DNA template (described above), 50 μM rNTPs, 3.3 mM guanosine, 40 U of ribonuclease inhibitor RNase Out (Invitrogen), and 400 U of T7 RNA polymerase (Ambion). The reaction was carried out at 37°C for 1 h and transcripts were purified in denaturing 10% polyacrylamide gel, excised, eluted from the gel (0.3 M sodium acetate pH 5.2, 0.5 mM EDTA and 0.1% SDS) and precipitated. All transcripts were 5'-end-labeled with T4 polynucleotide kinase and [<sup>32</sup>P]ATP (4500 Ci/mmol; ICN). The labeled RNAs were purified by electrophoresis in a denaturing 10% polyacrylamide gel, localized by autoradiography and recovered as described above.

### Nuclease digestions and lead cleavages

Prior to structure probing reactions, the <sup>32</sup>P-labeled RNAs were subjected to a denaturation and renaturation procedure in a solution containing 20 mM Tris-HCl (pH 7.2), 80 mM NaCl, 20 mM MgCl<sub>2</sub>, by heating the sample at 80°C for 1 min and then slowly cooling to 37°C. Limited RNA digestion was initiated by mixing 5 μl of the RNA sample (50 000 c.p.m.) with 5 μl of a probe solution containing lead ions or nuclease S1 or ribonucleases T1, T2, V1. All reactions were performed at 37°C for 20 min and stopped by adding an equal volume of stop solution (7.5 M urea and 20 mM EDTA with dyes) and sample freezing.

### Analysis of reaction products

To determine the cleavage sites, the products of lead-induced hydrolysis and nuclease digestion were separated in 15% polyacrylamide gels containing 7.5 M urea, 90 mM Tris-borate buffer and 2 mM EDTA, along with the products of alkaline hydrolysis and limited T1 or Cl3 nuclease digestion of the same RNA molecule. The alkaline hydrolysis ladder was generated by the incubation of the labeled RNA in formamide containing 0.5 mM MgCl<sub>2</sub> at 100°C for 10 min. The partial T1 ribonuclease digestion of RNAs was performed under semi-denaturing conditions (10 mM sodium citrate pH 5.0; 3.5 M urea) with 0.2 U/μl of the enzyme during incubation at 55°C for 15 min. For RNAs containing the (CGG)<sub>n</sub> repeats limited digestion with ribonuclease Cl3 (0.1 U/μl) was performed in the same way in a buffer containing 5 mM Tris-HCl (pH 8.0) and 3.5 M urea. The cleavage sites characteristic for nucleases S1 and V1 were determined by comparison with the homologous S1 ladder (not shown). The S1 and V1 generated RNA fragments terminate with 3'-hydroxyls and migrate more slowly than the corresponding formamide and T1 fragments. Electrophoresis was performed at 1500 V and was followed by autoradiography at -80°C with an intensifying screen. The products of the structure probing reactions were also visualized and analyzed by PhosphorImaging (Typhoon; Molecular Dynamics).

### Electrophoresis in a non-denaturing condition

The RNA structure homogeneity was analyzed for each transcript by the electrophoresis of radiolabeled samples in 10% non-denaturing polyacrylamide gel (acrylamide/bisacrylamide, 29:1) buffered with 45 mM Tris–borate, at a fixed temperature of 37°C. Prior to gel electrophoresis, the <sup>32</sup>P-labeled transcripts were subjected to a denaturation and renaturation procedure in a solution containing 10 mM Tris–HCl (pH 7.2), 40 mM NaCl, 10 mM MgCl<sub>2</sub>, by heating the sample at 80°C for 1 min and slowly cooling it to 37°C, and mixed with an equal volume of 7% sucrose with dyes. Electrophoresis performed at 15 W was followed by autoradiography and PhosphorImager analysis.

### RNA secondary structure modeling

RNA secondary structure prediction was performed using the Mfold program version 3.1 (32,33). This program is designed to determine optimal and suboptimal secondary structures of RNA calculated for 1 M NaCl solution at 37°C, and to count free energy contributions for various secondary structure motifs.

## RESULTS

Altogether 19 transcripts were analyzed in this study. Sixteen were composed of either the CUG, CAG, CCG, CGG or AGG motifs repeated 16 or 17 times. Three transcripts contained the tetranucleotide CCUG motif repeated 14 and 17 times. The above repeat number range was chosen as the corresponding transcripts composed of CNG repeats were expected to form relatively stable hairpin structures. Besides the repeats all transcripts contained two G-residues introduced at their 5'-end to ensure the specificity and efficiency of *in vitro* transcription. Twelve transcripts contained several additional G- and/or C-residues located at one or both ends of the repeated sequence as specified in Table 1. They were designed to form base pairs between transcript ends. The 5'-end labeled transcripts were analyzed by polyacrylamide gel electrophoresis in denaturing conditions for their length homogeneity and in non-denaturing conditions for their structure homogeneity (31).

The RNA structure analysis was performed with an adequate set of well characterized enzymatic and chemical probes (34,35). Ribonuclease T1 cleaves RNA after G-nucleotides present in the single-stranded regions. Ribonuclease T2 has more relaxed specificity and cleaves the single-stranded regions of RNA with the following order of reactivity: A>G=U=C. Nuclease S1 also cleaves the single-stranded form of RNA but without base specificity. Ribonuclease V1 cleaves the double-stranded regions of RNA as well as structured single-stranded regions. Lead ions cleave preferentially flexible single-stranded RNA regions (36,37), but also some relaxed double-stranded structures (29).

### (CNG)<sub>n</sub> repeats form alternatively aligned hairpins

The results of non-denaturing polyacrylamide gel electrophoresis of five transcripts, CUG17, CAG17, CGG17, CCG17 and AAG17, show that the electrophoretic mobility of the AAG17 is much lower than that of the other transcripts

(Fig. 1A). This implies that all CNG17 transcripts may fold into hairpin structures. It is also apparent that bands corresponding to the CNG17 transcripts appear rather diffused in gel which may suggest the micro-heterogeneity of their structure (two or more very similar stable conformers co-existing in conditions of analysis). The features of the transcript structures proposed above were then verified by a more detailed structure characterization. Each of the transcripts was subjected to structure probing as shown for the CAG17 in Figure 1B. For the other transcripts only the S1 nuclease digestion patterns are presented in Figure 1C. It is evident that several central repeats in the CNG17 show enhanced reactivity with all the enzymatic probes used in this experiment, although the level of this enhancement is not the same for all transcripts. The strongest T1 cuts occur at the G7 and G8 (the G-residues from the CNG repeats 7 and 8). The enhanced S1 digestions occur at the G7, A8, G8 and C9 in the CAG17, at the U8 and G8 in CUG17, at the first G-residues of repeat 8 and 9 in CGG17 and less profound cuts at the G8 and G9 in CCG17. These results confirm the conclusions drawn from the analysis of intact transcripts in non-denaturing gel (Fig. 1A), and show that these transcripts indeed assume hairpin structures. The number of centrally located repeats with increased reactivity, which is higher than expected for the loops of a typical size, is best explained by several co-existing alternatively aligned hairpins. In these hairpin variants, different combinations of the neighboring central repeats are involved in the formation of a terminal loop giving the impression that it is of an apparently large size. Several 3'-terminal repeats of the same variants form protruding ends which show enhanced reactivity. On the contrary, the AAG17 does not form a hairpin structure, which also supports the conclusion drawn from electrophoresis performed in non-denaturing conditions. Neither the S1 nuclease (Fig. 1C) nor other probes show any significant differences in the reactivity of the specific AAG repeats along their tract.

### The GC-clamp reduces hairpin structure heterogeneity

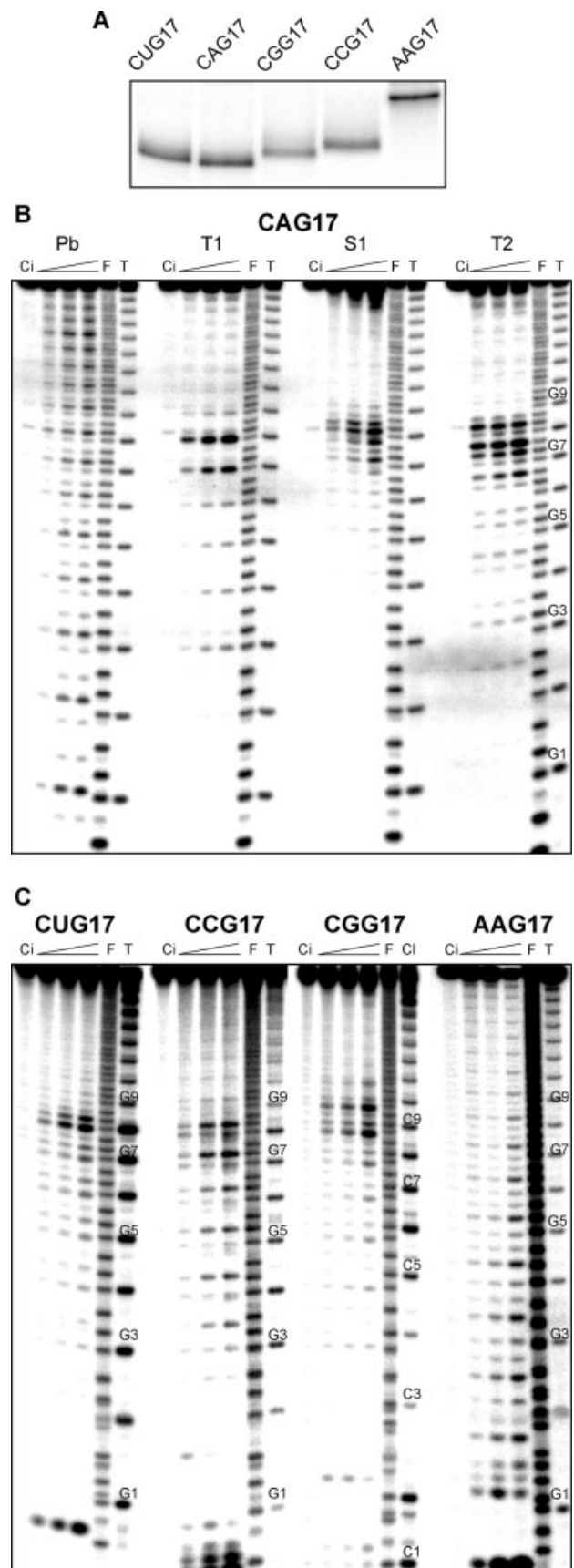
In order to gain a more detailed insight into the structures of different (CNG)<sub>n</sub> repeat hairpins, the elimination of their structure heterogeneity was required. One of the possible ways to achieve that was to add several G and C residues to the transcript ends. They form regular base pairs and fix the repeats in a single alignment. To find the strength of the GC-clamp which will be sufficient to overcome the tendency of the repeated sequence to form alternatively aligned hairpins, either two, four or six G-C and C-G base pairs were formed at the base of the CUG17 hairpin stem (Fig. 2A). The second, fourth or sixth base pair of this clamp is formed by the C and G residues from the first and the last repeat in the CUG17\_c12, CUG17\_c14 and CUG17\_c1 transcripts, respectively. The 5'-end labeled transcripts described above and CUG17 were subjected to the standard denaturation/renaturation procedure in a solution containing 10 mM MgCl<sub>2</sub> and 40 mM NaCl at pH 7.2, and were electrophoresed in non-denaturing gel. As revealed by a comparison of the PhosphorImager results obtained for all four transcripts (Fig. 2B), the CUG17\_c1 is structurally homogeneous. A single sharp peak characterizes its electrophoretic signal, as opposed to the two or more peaks of the CUG17\_c12 and at least three peaks of CUG17.

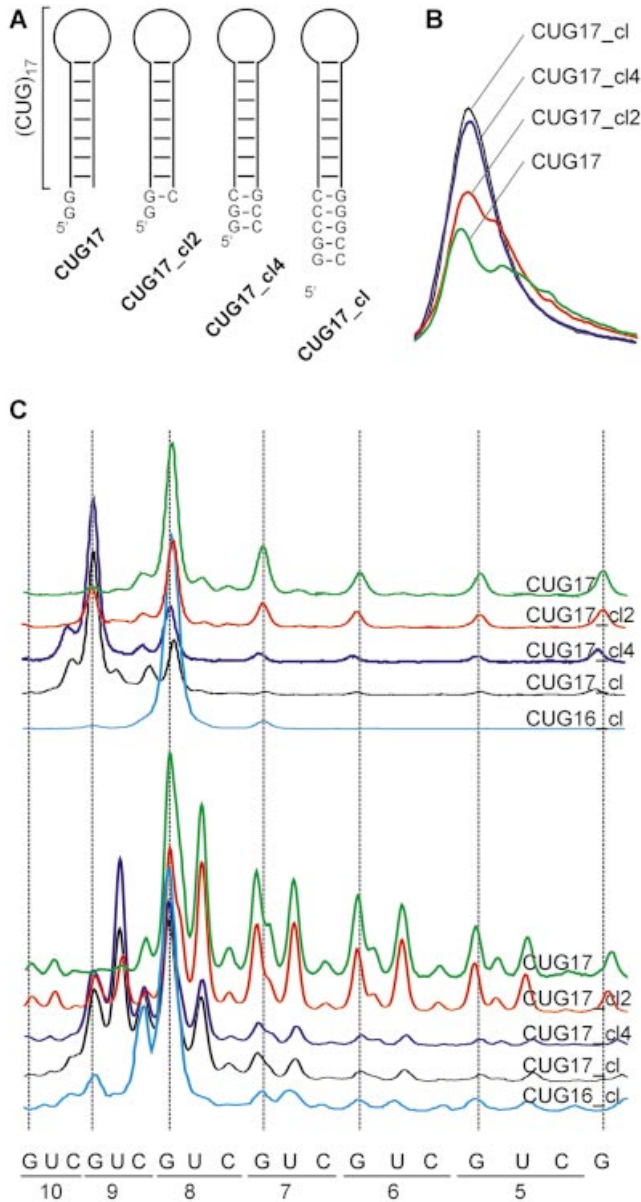
The structure of the CUG17 and that of its clamped variants was probed as described above. As expected, the number of centrally located repeats that showed enhanced reactivity with the single-strand specific probes was gradually reduced in going from CUG17 to CUG17\_cl. This trend is clearly seen in the quantitative representation of the T1 and S1 nuclease digestion patterns (Fig. 2C), as well as in the patterns generated by nucleases T2 and V1 (not shown). Note that the T1 digestions at the phosphodiester bonds between the G4 and G7 are significant in the CUG17 and CUG17\_cl2, still detectable in CUG17\_cl4 and practically absent in CUG17\_cl. In the homogeneous structure of the CUG17\_cl the highly reactive G8 and G9 form a terminal loop. Note also that the terminal loop is smaller and contains only a single reactive G8 residue, in clamped CUG16\_cl presented in Figure 2C for comparison. A similar picture emerges from the inspection of the S1 nuclease digestion patterns. These results show that the GC-clamp composed of six base pairs ensures the single alignment of the CUG17\_cl hairpin. A clamp of this strength was used also for transcripts containing other types of repeated motifs.

#### Structure probing results of the uniformly aligned hairpins are easier to interpret

The clamped transcripts CAG17\_cl, CGG17\_cl, CCG17\_cl and AAG17\_cl were analyzed in polyacrylamide gel in non-denaturing conditions and peaks representing their PhosphorImager signals were compared with those of the non-clamped transcripts (Fig. 3A). Also in this experiment, the clamped AAG17\_cl migrated much more slowly than the CNG17\_cl transcripts (not shown). The sharper peaks of clamped CNG17\_cl transcripts suggest the homogeneity of their hairpin structures. As revealed by the structure probing results obtained for the CAG17\_cl (Fig. 3B), CUG17\_cl, CCG17\_cl and CGG17\_cl (Fig. 3C) and by their more detailed analysis (see next sections), the GC-clamp is indeed effective in eliminating the structure heterogeneity of these hairpins. The strongest ribonuclease T1 cuts occur at the G8 and G9 in the CAG17\_cl, and in other CNG17\_cl transcripts. Strong S1 nuclease cuts appear between the G7 and G9 in the CAG17\_cl (Fig. 3B), between the G8 and G9 in the CUG17\_cl and CGG17\_cl, and between the C8 and G9 in CCG17\_cl. Weak cuts in all hairpin stems are usually stronger at their 5'-side.

**Figure 1.** Structure analysis of transcripts containing different types of triplet repeats. (A) Non-denaturing 10% polyacrylamide gel electrophoresis of 5'-end labeled CUG17, CAG17, CGG17, CCG17 and AAG17 transcripts dissolved in the structure probing buffer, heated to 80°C (denaturation) and cooled slowly to 37°C (renaturation). (B) Cleavage patterns obtained for 5'-end labeled CAG17 transcript treated with: Pb, lead ions at an increasing concentration (0.25, 0.5, 1 mM); T1, T1 ribonuclease (0.5, 1, 1.5 U/μl); S1, S1 nuclease (0.5, 1, 2 U/μl; 1 mM ZnCl<sub>2</sub> was present in each reaction mixture); T2, T2 ribonuclease (0.25, 0.5, 1 U/μl); lane Ci, incubation control (no probe); lane F, formamide ladder; lane T1, guanine specific ladder. Electrophoresis was performed in a 15% polyacrylamide gel under denaturing conditions. The positions of selected G-residues present in CAG17 are shown (G1 indicates the G residue from the first CAG repeat, etc.). (C) Patterns of cleavages generated by S1 nuclease in the 5'-end labeled CUG17, CCG17, CGG17 and AAG17 transcripts. In the case of CCG17 an additional lane Cl-limited C13 ribonuclease digest under semi-denaturing conditions and the positions of chosen C-residues are shown. Other conditions of analysis and abbreviations are as in (B).

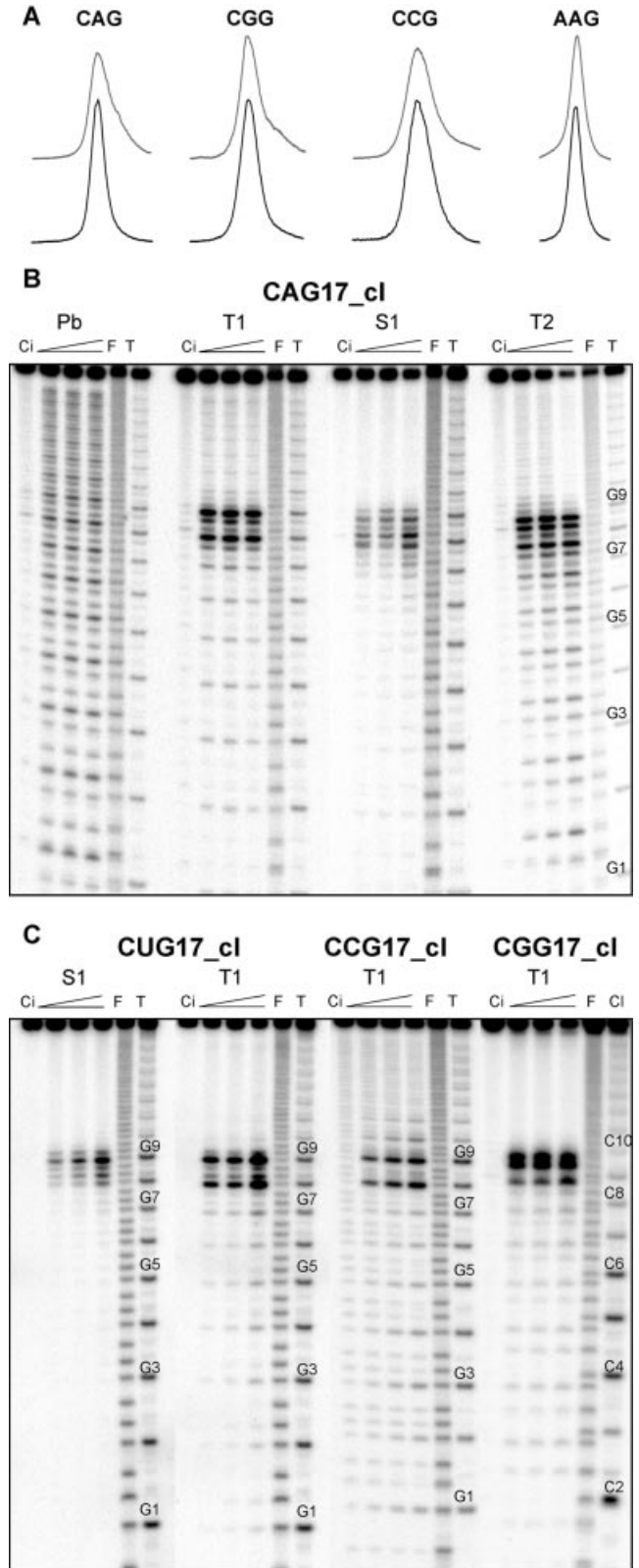




**Figure 2.** Structure analysis of (CUG)<sub>n</sub> repeat transcripts having the GC-clamp. (A) Schematic secondary structures of four transcripts, CUG17, CUG17<sub>cl2</sub>, CUG17<sub>cl4</sub> and CUG17<sub>cl</sub>, used in this analysis. (B) Qualitative comparison of 5'-end labeled transcripts shown in (A) containing a different number of base pairs forming a GC-clamp, obtained from PhosphorImaging. All transcripts were analyzed in the same polyacrylamide gel in non-denaturing conditions. (C) Cleavage patterns obtained for 5'-end labeled transcripts described above using T1 ribonuclease (upper panel) and S1 nuclease (lower panel). Results obtained for the CUG16<sub>cl</sub> transcript were added to show its structural similarity to non-clamped transcripts.

**Figure 3.** Structure probing of transcripts containing the GC-clamp. (A) PhosphorImager peaks representing electrophoretic bands of the intact 5'-end labeled transcripts analyzed in non-denaturing polyacrylamide gels. Upper and lower peaks represent transcripts not containing and containing the GC-clamp, respectively. (B) Cleavage patterns obtained for 5'-end labeled CAG17<sub>cl</sub> transcript treated with lead ions (Pb) and T1, S1 and T2 nucleases. (C) Patterns of cleavages generated by S1 nuclease in the 5'-end labeled CUG17<sub>cl</sub> transcript and by T1 ribonuclease in CUG17<sub>cl</sub>, CCG17<sub>cl</sub> and CGG17<sub>cl</sub> transcripts. Reaction conditions were as described in the legend to Figure 1.

Lead cleavages induced in the CAG17<sub>cl</sub> show some enhancement in central repeats. This effect was not observed in the non-clamped transcript. The comparison of nuclease digestion patterns in different repeat-containing transcripts



with those generated by lead ions clearly shows the different specificity of these probes. The repeats present in the single-stranded terminal loops of hairpins and in their relaxed stem structures are better differentiated by the nucleases which show a much stronger discrimination against the stems. For much smaller lead ions, the critical factor determining their reactivity is the flexibility of the phosphodiester bonds in RNA allowing the formation of a reactive intermediate (38). With the exception of the CUG17\_cl and CUG16\_cl this flexibility appears to be not much different in the loop and stem nucleotides of the analyzed transcripts.

### **The terminal loops of clamped hairpins have different sizes**

The clamped hairpins with either even or odd numbers of the (CNG)*n* repeats were analyzed by nuclease digestion in order to find the details of their loop structures. The efficiencies of the loop cuts were quantified and the peaks of different colors represent the effects of digestion by different nucleases (Fig. 4). The digestion sites and strengths are also shown in the proposed hairpin loop structures. It turned out that all analyzed CNG16\_cl transcripts behave similarly and form the 4-nt terminal loop. Cleavage patterns in the CAG16\_cl, CUG16\_cl and CGG16\_cl are very similar—only the G8 is cleaved by the T1 ribonuclease, the N8 and G8 are cut by nuclease S1, and five consecutive phosphodiester bonds between C8 and N9 are digested by ribonuclease T2. In contrast, the terminal loops in clamped transcripts containing 17 repeats are different. The CAG17\_cl and CCG17\_cl have the 7-nt terminal loop. These transcripts were efficiently digested by T1, S1 and T2 nucleases between C8 and C10. On the other hand, the terminal loops of the CUG17\_cl and CGG17\_cl hairpin structures are cleaved by the nucleases between G8 and G9 only, which implies the existence of smaller 3-nt loops. Another analyzed transcript, GCC17\_cl, differs from the CCG17\_cl discussed above only by the frame in which the repeat was inserted between the same flanking sequences. Its cleavage patterns revealed the presence of the 4-nt loop similar to that which occurs in the CNG16\_cl hairpins. Taken together, the clamped (CNG)*n* repeat hairpins have terminal loops of different sizes which clearly depend on the repeat number and also on its frame.

### **Stem structures of clamped hairpins are similar**

The uniformly aligned repeated sequences in structurally homogeneous hairpins made it possible to reveal details of their stem architecture. The number of 16 and 17 repeats in transcripts was selected to assure both the sufficient stability and flexibility of their stem structure which should be recognized by the probes specific for single-stranded and double-stranded RNA. Cleavage intensities in stem repeats were quantified and they are shown in Figure 5. The relatively high reactivity of stem nucleotides with lead ions suggests the relaxed stem structure in hairpins composed of all types of the repeats. It is evident that the strongest lead cleavages occur at the CpN phosphodiester bonds in all types of the repeats and at NpG phosphodiester bonds in the (CUG)*n* and (CCG)*n* repeats. The GpC bond is cleaved with the lowest efficiency in stems of all (CNG)*n* hairpins. Similar differences in reactivity of the internucleotide bonds can be observed in the case of ribonuclease T2. The nuclease S1, likewise the

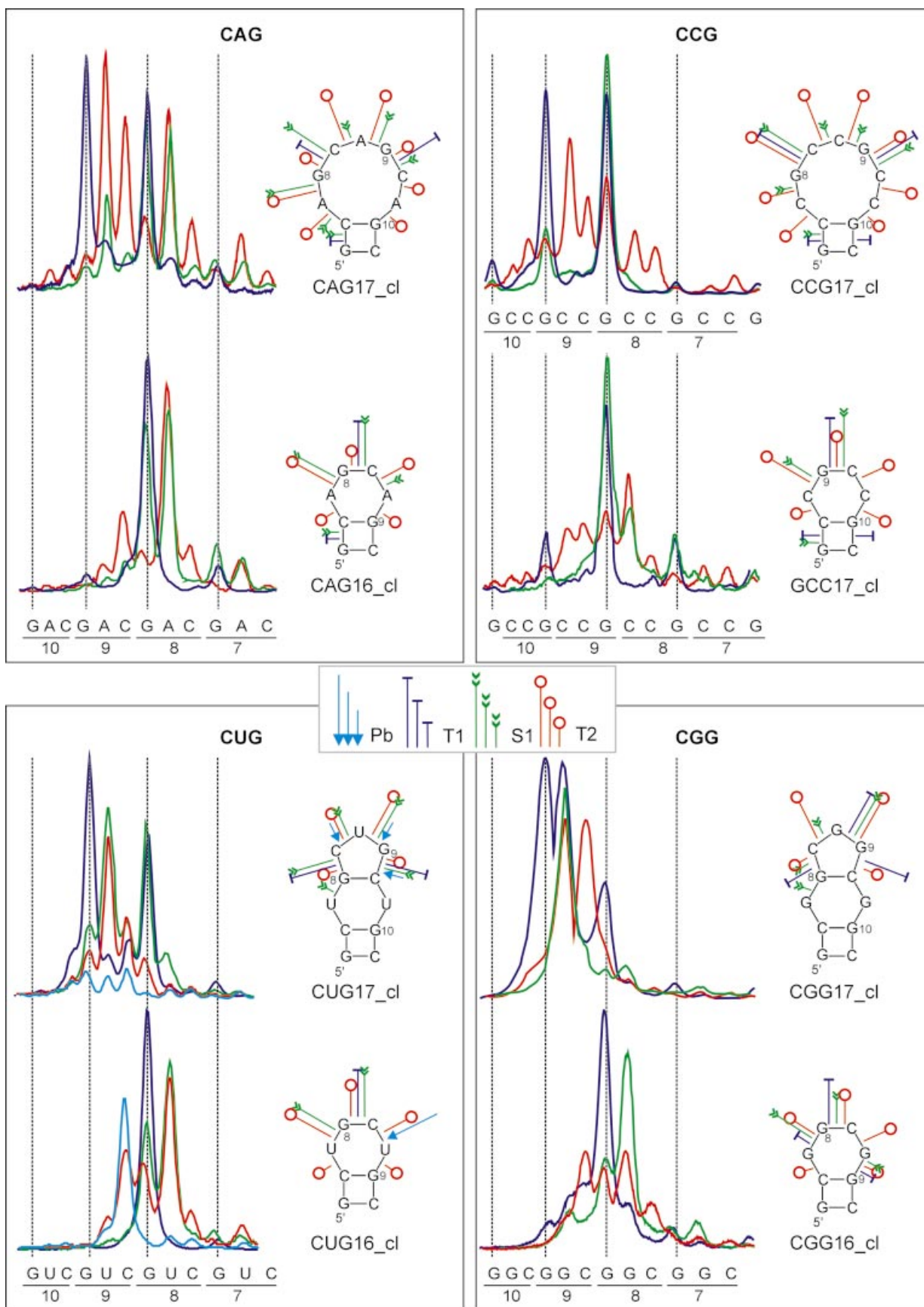
nuclease V1, cleaves the NpG bonds in the stem with the highest efficiency. In the hairpin stems formed by the (CAG)*n*, (CGG)*n* and (CCG)*n* repeats, the V1 cuts at the NpG are at least three times more efficient than those at other phosphodiester bonds. The cleavage characteristics shown above suggest a high similarity of stem structures composed of different types of repeats. It appears that in each analyzed stem structure, the same type of repeated motif is present: two consecutive base pairs 5'-G-C and C-G followed by the N/N mismatch. Although in hairpins composed of the (CGG)*n* and (CCG)*n* repeats other stem architectures are also theoretically possible they have not been observed in our experiments. Neither the specificities of the stem cleavages nor those of the terminal loop cuts give support to the existence of other arrangements of nucleotides in the (CNG)*n* hairpin stems. The experimentally determined stem structure is also favored by the most advanced RNA structure prediction tools (32).

### **Alternatively aligned (CNG)*n* hairpin variants have protruding 3'-ends**

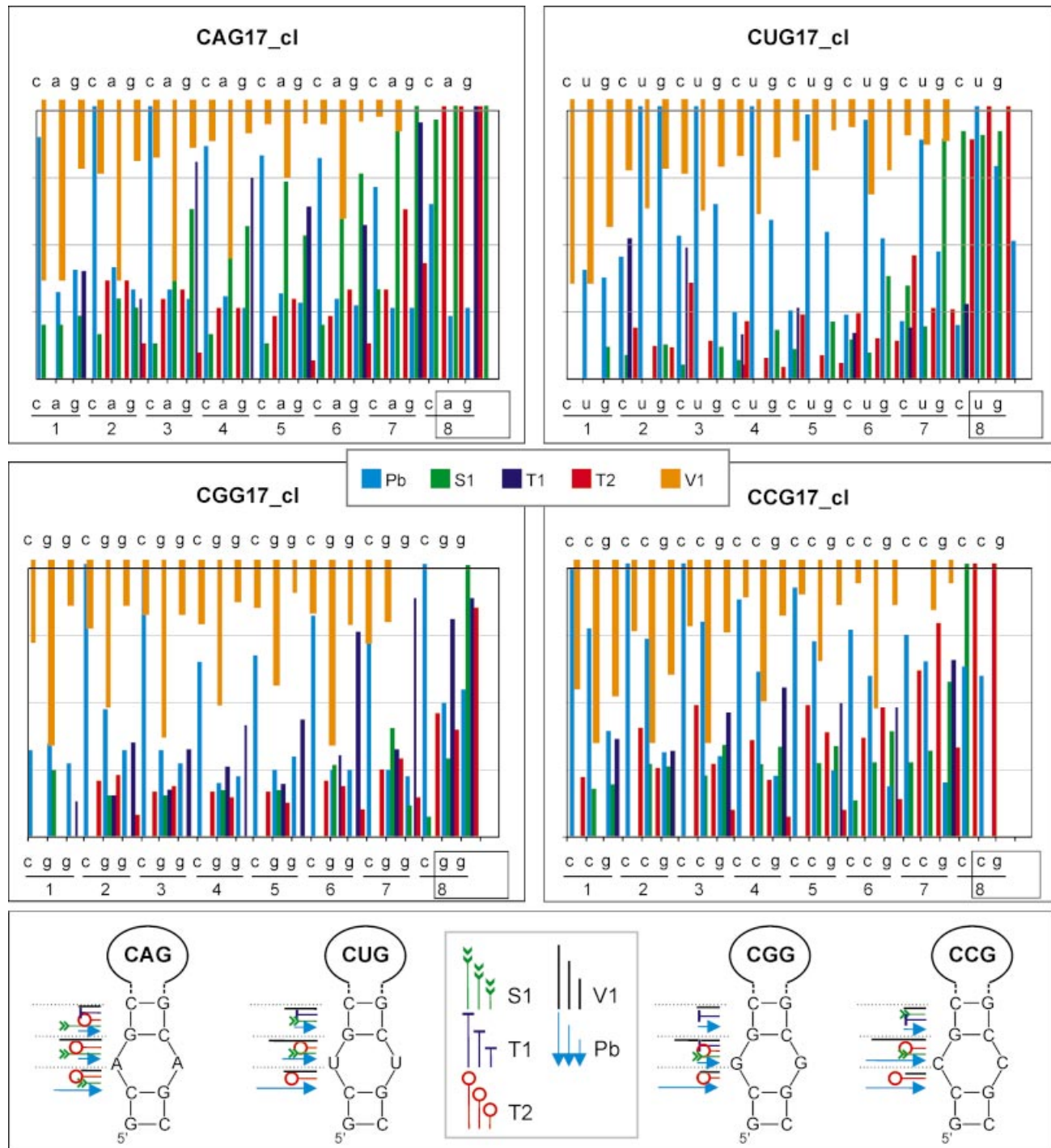
Having established the specific digestion patterns characteristic for terminal loops and the stems of clamped hairpins composed of either the odd or even number of the repeats it becomes possible to determine the structures of different variants of the non-clamped hairpins. The PhosphorImager quantifications of nuclease cuts at the loop nucleotides in the CAG17 are compared with those generated in clamped CAG16\_cl (Fig. 6A). It is evident that the digestion pattern observed in the CAG16\_cl is repeated in each of the CAG17 structural variants I–III (Fig. 6B), but the intensities of nuclease cleavages are gradually decreasing in accord with the decreased share of each variant in their mixture. All these variants contain a 4-nt terminal loop which is composed of the AG and CA of repeats 8 and 9, respectively (variant I), repeats 7 and 8 (variant II), and repeats 6 and 7 (variant III), as shown in Figure 6B. The same applies to the CUG17 hairpin variants (see cleavage patterns in Fig. 2C). Two or three 3'-terminal repeats, which are in some of these hairpin variants single-stranded, are partially digested by the S1 and T2 nucleases. By comparing the terminal loop digestions in the CAG16\_cl and CUG16\_cl with those generated in the non-clamped CAG17 and CUG17 hairpins, the loop cuts could be assigned to the individual hairpin variants I, II and III (Fig. 6A). The approximate contributions from these variants were determined from the averaged intensities of nucleases cuts. They are 75, 20 and 5%, and 80, 15 and 5% for variants I, II and III of the CAG17 and CUG17 hairpins, respectively. In the case of the CUG17\_cl2, which has a weak clamp capable of forming two base pairs only, the new hairpin variant containing a 7-nt loop makes a ~15% contribution to the CUG17 variants described above. This new variant is characterized by a T1 ribonuclease cut at the G9 and S1 nuclease cuts at the U9 and G10 (Fig. 2C).

The CGG17 structure has different properties. A comparison of the nuclease digestion patterns in the CGG17 terminal loop with those generated in CGG17\_cl and CGG16\_cl shows that the former pattern arises from the simple superposition of approximately equal contributions from the latter two (Fig. 6C and D). This suggests the co-existence of two conformers I and II. In contrast to the other (CNG)*n* repeat hairpins, we did not observe in this case 'slipped' hairpin variants. There were no





**Figure 4.** Structure of terminal loops in clamped (CNG)<sub>n</sub> hairpins. PhosphorImager peaks representing the CAG17\_cl, CAG16\_cl, CUG17\_cl, CUG16\_cl, CCG17\_cl, CGG16\_cl, CCG17\_cl and GCC17\_cl cleavage patterns. The cleavage sites and intensities are specified for each probe (blue, T1; red, T2; green, S1). The secondary structures of terminal loops are shown next to the PhosphorImages of cleavages.



**Figure 5.** Stem structures in clamped (CNG)<sub>n</sub> repeat hairpins. The quantitative representation of cleavage patterns generated in the sequence of repeats 1–8 of the CAG17<sub>cl</sub>, CUG17<sub>cl</sub>, CGG17<sub>cl</sub> and CCG17<sub>cl</sub> hairpin stems. Nucleotides involved in terminal loop formation are boxed. Intensities of cleavages generated by lead ions, and nucleases S1, T1 and T2 are shown in the same scale. A separate scale is used for stronger V1 nuclease cuts. The cleavage sites and intensities are also shown in the proposed stem structures.

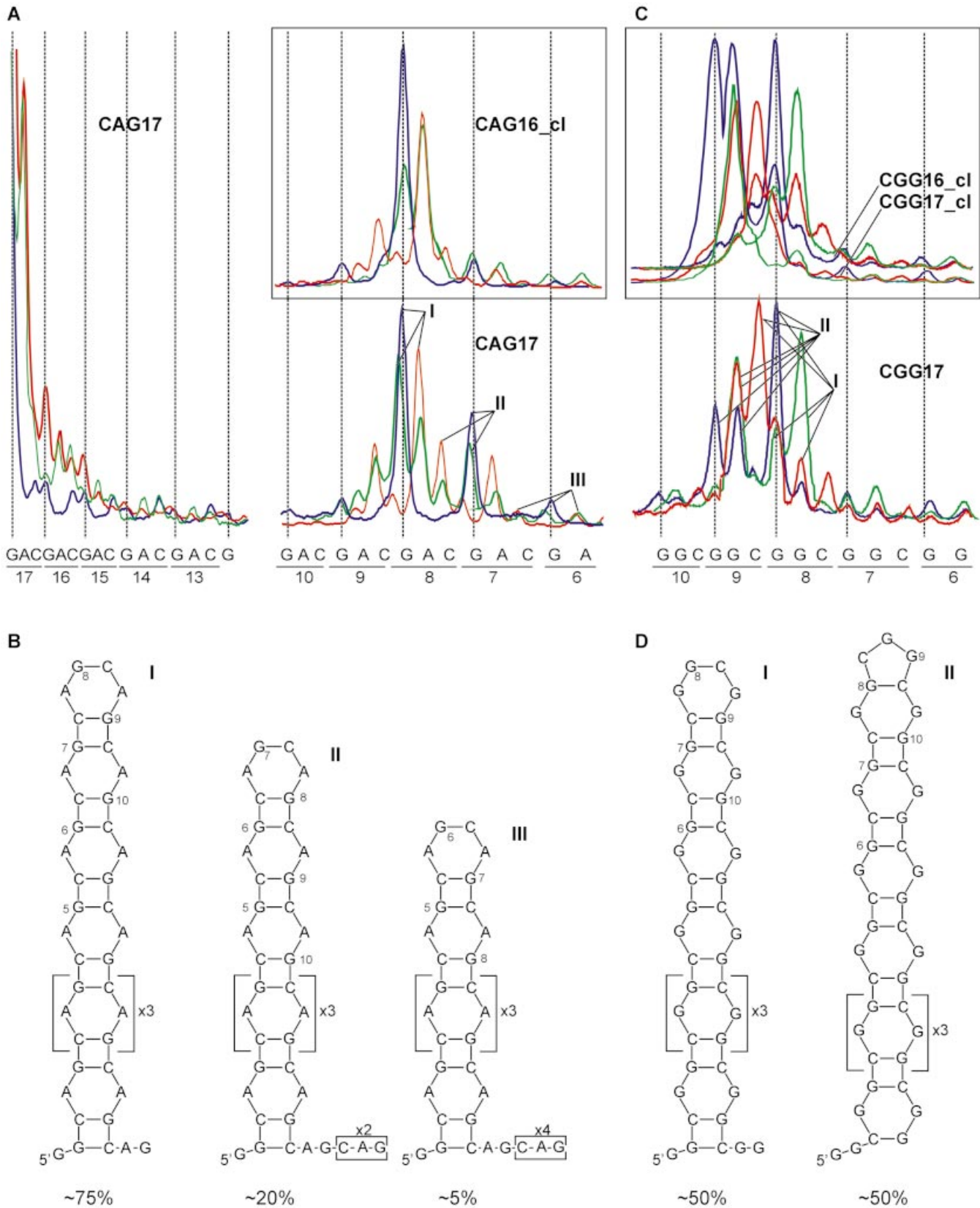
enhanced cleavages at the 3′-terminal repeats and no enhanced cuts at the 5′-side of the terminal loop.

**The (CCUG)<sub>n</sub> repeats also form hairpin structure**

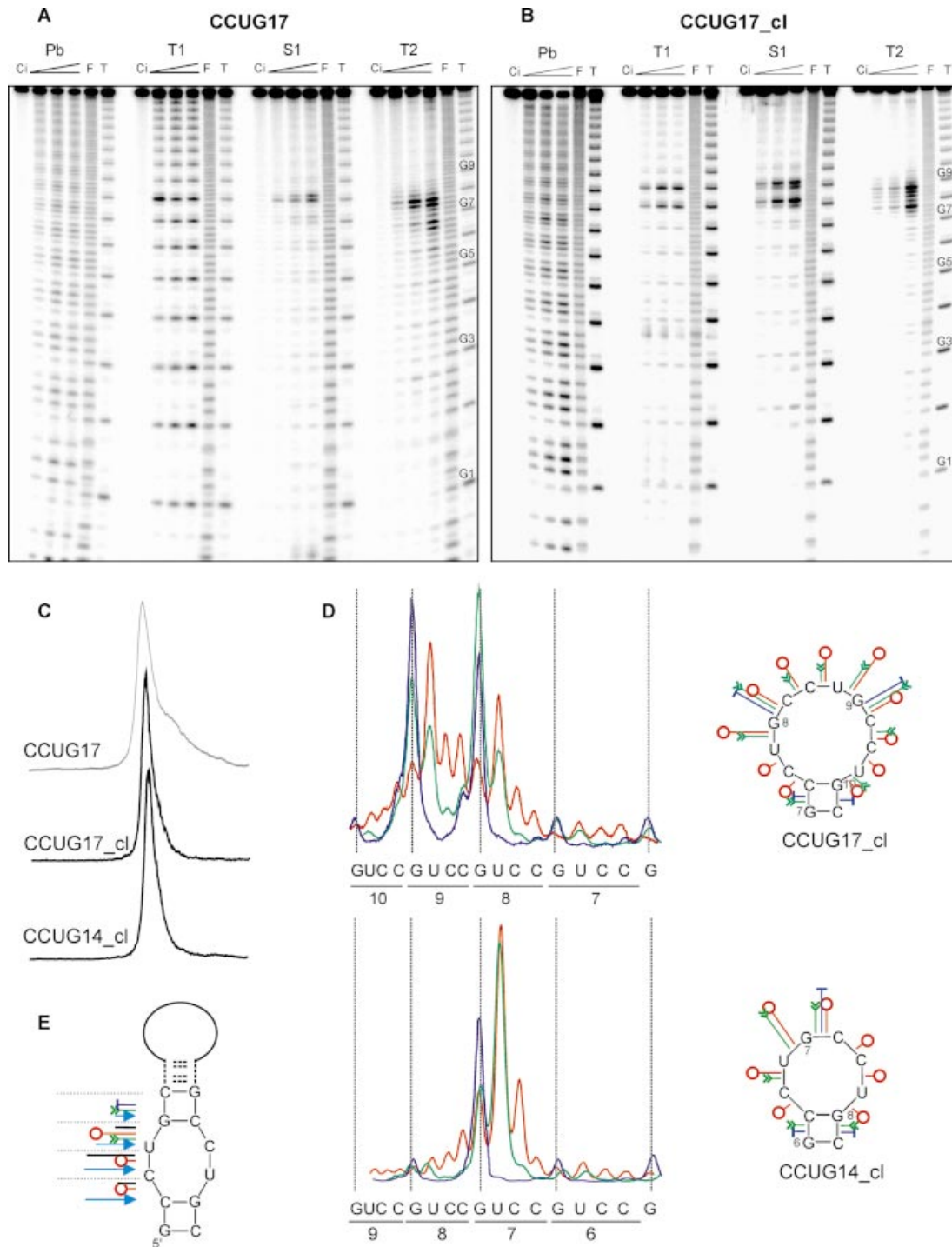
The structure probing data collected for the non-clamped and clamped (CCUG)<sub>n</sub> repeats containing transcripts (Fig. 7A and

B, respectively) look similar to those of the (CNG)<sub>n</sub> repeat transcripts. The cleavage patterns of the CCUG17 reflecting the heterogeneity of its structure become simpler in CCUG17<sub>cl</sub> in which the loop cleavages are more centered around repeat 8, and cuts at the neighboring phosphodiester bonds are suppressed. The reduction in the structure





**Figure 6.** Structural variants of the CAG17 and CGG17 hairpins. (A) PhosphorImager peaks representing terminal loop cleavages in the CAG17 and CAG16\_cl (inset). Cleavage sites and intensities are specified for each probe as described in the legend to Figure 4. Peaks corresponding to terminal loop cuts in variants I-III are indicated. (B) Secondary structures of the CAG17 variants and their relative contributions. (C) As in (A) but for the CGG17 and CGG17\_cl with CGG16\_cl (inset). (D) Secondary structures of two CGG17 variants.



**Figure 7.** Structure analysis of the (CCUG)*n* repeat containing transcripts. Denaturing polyacrylamide gel electrophoresis of 5'-end labeled CCUG17 (A) and CCUG17\_cl (B) treated with lead ions (lane Pb), ribonuclease T1, nuclease S1 and ribonuclease T2. The probe concentrations and lane indications are the same as described in the legend to Figure 1. Positions of selected G-residues in the CCUG repeat tracts are indicated. (C) PhosphorImager analysis of structure homogeneity of three transcripts CCUG17, CCUG17\_cl and CCUG14\_cl, separated in non-denaturing polyacrylamide gel. (D) Cleavage sites and intensities in the terminal loops of the (CCUG)*n* repeat hairpins: clamped CCUG17\_cl (upper) and CCUG14\_cl (lower), and the proposed terminal loop structures (symbols explained in Fig. 4). (E) Proposed stem structure of the (CCUG)*n* repeat hairpin including sites and intensities of cleavages generated by nucleases (symbols explained in Fig. 5).

heterogeneity is also observed in intact transcripts CCUG17\_cl and CCUG14\_cl electrophoresed in non-denaturing polyacrylamide gel (Fig. 7C). Details of the CCUG17\_cl and CCUG14\_cl loop structures are shown in

Figure 7D. It is apparent that these structures differ significantly in the number of reactive phosphodiester bonds. As many as 10 nt form a terminal loop in the CCUG17\_cl whereas 6 nt are looped out in CCUG14\_cl. In the non-clamped

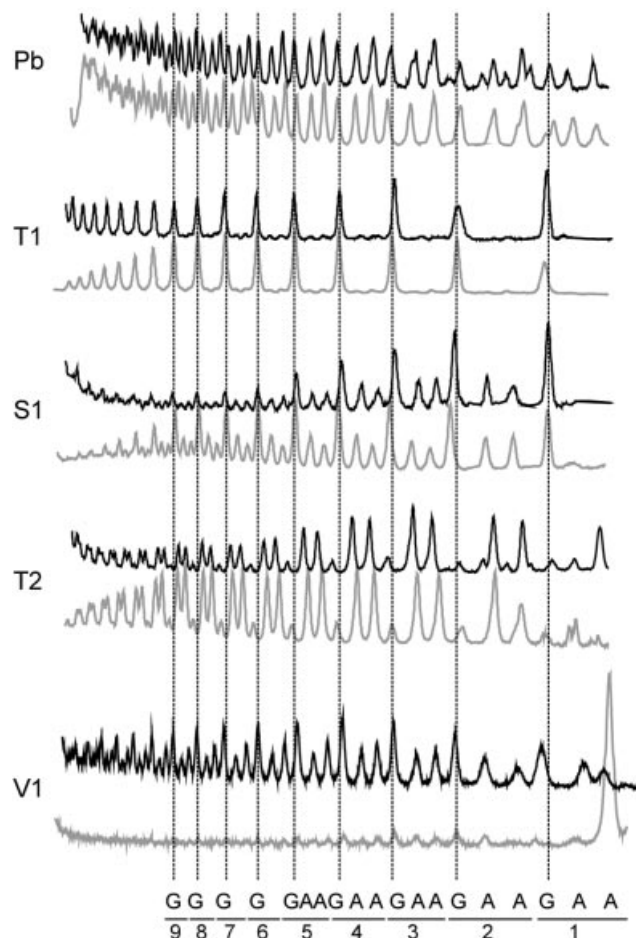
transcript CCUG17, three alternatively aligned conformers appear at the ratio 75, 20 and 5%, similar to that observed in CUG17 and CAG17. Each of these conformers contains the 6-nt terminal loop. Also the cleavages generated in the stems of the (CCUG)<sub>n</sub> repeat hairpins closely resemble those characteristic for the (CNG)<sub>n</sub> repeat hairpins (Fig. 7E). Ribonuclease T2 digests the CpC, CpU and most intensively the UpG phosphodiester bonds. Lead ions show a similar reactivity pattern. On the other hand, the S1 nuclease cuts are strong at the GpC bonds and weaker at the UpG bonds. The ribonuclease V1 cleaves the CpC, UpG and with the highest efficiency the CpU bonds in the stem. The periodic, symmetrical 4-nt internal loop present in the (CCUG)<sub>n</sub> repeat hairpin stem is well mapped by all enzymatic probes used in this study.

#### (AAG)<sub>n</sub> repeats do not form hairpin structures

Two AAG17 transcripts analyzed in this study are single-stranded as is shown by the results of structure probing and results of gel electrophoresis of intact transcripts performed in non-denaturing conditions. We compared the cleavage patterns generated by several structure probes in the non-clamped and clamped AAG17 transcripts. The latter were expected to form a large loop structure composed of about 50 repeat nucleotides and the GC-clamp. The structure probing results indicate that both AAG17 and AAG17\_cl are equally good substrates for lead ions, ribonucleases T1 and T2 and for nuclease S1 (Fig. 8). Lead ions show no cleavage preference within these repeats, the strongest S1 nuclease cuts occur after the G-residue of each AAG repeat, and the strong preference of ribonuclease T2 for digestions after each A-residue is also shown. Interestingly, the V1 ribonuclease cleaves efficiently the AAG17 but very poorly the clamped AAG17\_cl. This indicates the existence of an ordered single-stranded structure in the AAG17. The lack of this structure in AAG17\_cl is most likely due to the bending of the polynucleotide backbone forced by the GC-clamp formation. The existence of a functioning clamp is confirmed by the presence of the V1 ribonuclease cuts, and the absence of T1, T2 and S1 cuts in this region.

## DISCUSSION

In this study we have analyzed RNA structures formed by selected types of simple sequence repeats. These repeats are known to occur in the transcripts of genes associated with a number of human neurological diseases. The first question we asked was: how do the repeated sequences behave when allowed to fold without any restrictions imposed by the flanks? This question was important as the (CAG)<sub>n</sub> repeats in the *SCA3* and *DRPLA* transcripts as well as (CUG)<sub>n</sub> repeats in the *SCA8* and *DMPK* transcripts are predicted to form hairpin structures without flanking sequence participation (5). The answer was that the CAG17, CUG17, CCUG17 and CCG17 behave similarly and fold into several slipped hairpin variants differing by either the presence or the length of the single-stranded tail composed of the 3'-terminal repeats (Fig. 6C). The hairpin variant without the 3'-tail which has the longest stem predominates (70–80%) among those hairpins formed by each of the four repeat types, but contributions from the slipped variants are also significant. It is important to note that all the CAG17, CUG17 and CCG17 hairpin variants have 4-nt



**Figure 8.** Cleavage patterns obtained for the (AAG)<sub>17</sub> transcripts. Cleavage patterns of AAG17 (upper black lines) and AAG17\_cl (lower gray lines) obtained for different probes (Pb, 0.5 mM; T1 ribonuclease, 0.1 U/μl; S1 nuclease, 1.25 U/μl; T2 ribonuclease, 1.25 U/ml; V1 ribonuclease, 0.5 U/ml). Positions of selected residues are indicated.

terminal loops which are thermodynamically more stable than the 7-nt loops. In the stems of these hairpin variants, each of the two base pairs G-C and C-G are followed by either the U/U, A/A or C/C mismatches which have different contributions to stem stability: +0.4, +1.1 and +0.4 kcal/mol, respectively. The overall thermodynamic stabilities calculated for type I and II variants of the CAG17 hairpin (Fig. 6C) are, respectively, -17.1 and -14.8 kcal/mol. The type I variants of the CUG17 and CCG17 hairpins are more stable, -22 and -24.7 kcal/mol, respectively, and this increase in predicted hairpin stability correlates with the observed decreased contribution of slipped hairpin variants. Each of the slipped conformers has the single-stranded repeats protruding at the 3'-end. This alignment is more favored thermodynamically than the overhang of the 5'-repeats. The CCG17 also forms a hairpin structure which shows, however, different features to those discussed above. Two equally favored variants of the CCG17 hairpin have stem repeats in the same orientation with the first G-residues of the CGG repeats involved in mismatches. These variants differ in the size of the terminal loop and in the number of unpaired nucleotides which form small 5'- and 3'-overhangs. The fact that CCG17 behaves differently

may be explained by the helix stabilizing effect of the G/G mismatches:  $-1.4$  kcal/mol resulting in the highest overall stability of both CGG17 hairpin variants, about  $-35$  kcal/mol for each.

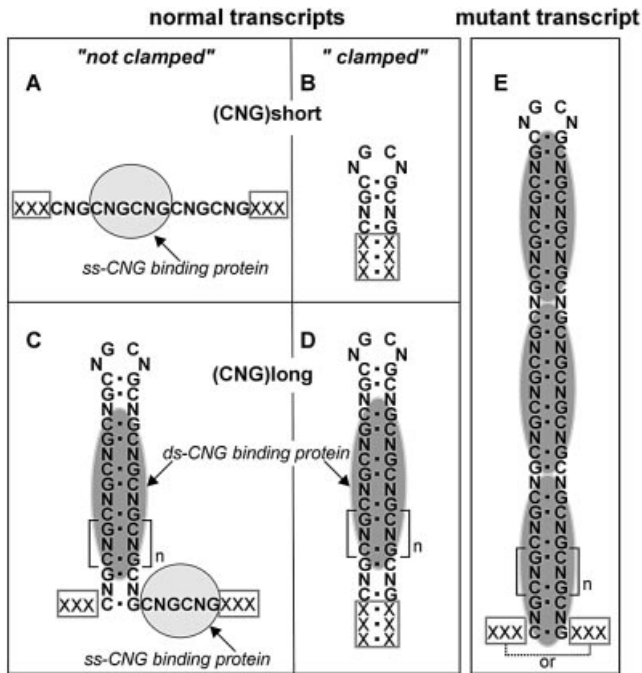
The next biologically relevant question we asked was: what is the structure of repeats having a single alignment defined by interacting flanking sequences? According to structure prediction, this situation occurs in the (CGG) $_n$  and (CCG) $_n$  repeats present in the *FMR-1* and *FMR-2* transcripts, respectively, as well as in (CAG) $_n$  repeats located in the *SCA6*, *SCA12* and *AR* transcripts (5). To facilitate comparisons of the structural features of different repeated sequences in hairpins, the same GC-clamp was used in each case instead of the specific flanking sequences present in their host transcripts. This clamp composed of six G-C and C-G pairs has a  $-16.6$  kcal/mol contribution to the overall hairpin stability. This is sufficient to nucleate the extended hairpin stem structure during RNA folding and force the 17 repeats of any CNG motif into a single alignment. The comparative analysis performed using these clamped transcripts revealed details of their hairpin loop and stem structures. It turned out that the clamped hairpins form terminal loops of different sizes, depending on the odd or even number of the repeats present in the hairpin and the type of the repeated sequence. The 4-nt loops are present in all (CNG) $_n$  hairpins with an even number of repeats, the 7-nt loop occurs in the (CAG) $_n$  and (CCG) $_n$  hairpins with an odd number of repeats and the 3-nt loops are present in the (CUG) $_n$  and (CGG) $_n$  hairpins with an odd number of repeats.

The stem structures of all clamped hairpins have a similar architecture in which the central nucleotides of the repeated motifs form different mismatches. This is an important observation especially for the (CCG) $_n$  and (CGG) $_n$  repeats for which different arrangements of the repeats in the stem could also be considered. The CGG17\_cl hairpin is the most stable:  $-45.9$  kcal/mol as compared to the  $-33.7$ ,  $-32.6$  and  $-28.3$  kcal/mol calculated for CUG17\_cl, CCG17\_cl and CAG17\_cl, respectively. The change of the repeat frame from (CCG) $_n$  to (GCC) $_n$  to form the GCC17\_cl transcript causes an important structural difference. In spite of the odd number of repeats, the 4-nt terminal loop is formed which is characteristic for clamped hairpins with an even number of repeats CNG16\_cl. The stem structure remains, however, unchanged. This information may be relevant to transcripts in which the terminal repeats base pair with specific flanking sequences in the predicted structures of the repeat regions. Such interactions may result in the formation of the repeat hairpins having terminal loop sizes different to those predicted solely on the basis of the odd or even number of the repeats in the transcript.

Another important finding of this study is related to the mechanism of RNA pathogenesis in two types of myotonic dystrophy DM1 and DM2. Similar to the expanded (CUG) $_n$  repeat containing transcript in DM1, the expanded (CCUG) $_n$  repeat containing RNA in DM2 forms nuclear foci and is not transported to the cytoplasm (22). The common feature of these two mutant RNAs is their ability to form hairpin structures which appear to be major factors in DM pathogenesis (15). They are thought to recruit the double-stranded (CUG) $_n$  and (CCUG) $_n$  repeat binding proteins by sequestering them from other ds(CUG) $_n$  and ds(CCUG) $_n$  repeat containing

transcripts and altering their functions (11,39). They are also known to enhance the expression of the CUG-BP (7) and other CELF proteins (13,14,40–42) resulting in the impaired splicing of other genes (12,16,19,43,44). The formation of the hairpin structure by the (CUG) $_n$  repeats was shown earlier (29), and was confirmed by other authors (8,9,30). The (CCUG) $_n$  repeat hairpin structure predicted earlier (15) has been shown experimentally for the first time in this study. The CCUG17 hairpin has a lower predicted thermodynamic stability ( $-17.3$  kcal/mol) as compared to that of CUG17 ( $-22.0$  kcal/mol). This difference results mainly from the stronger destabilizing effect:  $+1.4$  kcal/mol of tandem CU/CU mismatches which are present in the hairpin stem. The larger terminal loop of the CCUG17 hairpin also contributes to its lower stability.

In this study, we have shown that all types of the CNG repeats form similar hairpin structures. The convincing evidence for the presence of hairpin structures in the expanded CUG and CAG repeats comes from earlier *in vitro* studies (8,9,29,30). Although there is no direct evidence for the existence of CNG repeat hairpins *in vivo*, the CUG hairpin structure is strongly supported by the co-localization of the expanded transcript and the muscleblind proteins in cells (45–47). It was demonstrated that the CUG repeat-containing transcript specifically binds the muscleblind protein in the HeLa cell extract (30). This binding which was shown to be repeat length-dependent did not occur in the (CUG)11 transcript but occurred in the (CUG)20 and longer transcripts. These observations correspond well to the results of our *in vitro* studies on hairpin formation ability of transcripts containing CUG repeats of different length (29). The results of muscleblind protein binding experiments (30) mean that the hairpin structure composed of the 20 repeats is not destabilized by other proteins present in the cellular extract. On the other hand, shorter CUG repeats, which do not form hairpins, were shown to bind different proteins-CUG-BP. In this light, we postulate that not only will CUG repeats interact with their specific single-stranded and double-stranded RNA-binding proteins, but similar interactions may also occur between the CAG, CCG and CGG repeats and their putative binding proteins, as shown in the hypothetical model (Fig. 9). The different types of structures shown in this model are supported by our published (29) and unpublished experimental data and the results of RNA structure prediction. Whether this model shows only the possibility of similar physiological RNA-protein interactions for different hairpin-forming repeats or it presents a more general version of the RNA pathogenesis mechanism triggered by the repeats in transcripts remains to be established. In the postulated mechanism of RNA pathogenesis in myotonic dystrophy type 1, there are many different transcripts and at least two different types of proteins involved, muscleblind and CELF, that bind to double-stranded and single-stranded (CUG) $_n$ , respectively. The causative role is assigned to the expanded, mutant transcript which sequesters the dsCUG repeat binding proteins from other transcripts. There are many transcripts in cells that contain the repeat capable of forming hairpin structures. They could be co-regulated in cells taking advantage of this property. This study sheds more light on this group of transcripts by showing which repeat lengths are sufficient to form stable hairpins.



**Figure 9.** A model depicting the structures and protein binding properties of the CNG repeats in transcripts. Different types of structures formed by the repeats and their flanking sequences in transcripts as well as two different types of the CNG repeat binding proteins are schematically shown. The ability of the CNG repeats to form hairpin structures depends on two factors: the presence of a stable duplex structure (a 'clamp') formed by the natural sequences flanking the repeats in their host transcripts, and the repeat length. (A) Short non-clamped repeats do not form hairpin structures and bind the ssCNG repeat binding protein (gray oval). (B) Short clamped repeats form the hairpin structure which may not have a protein binding capacity. (C) Long normal (>20) non-clamped repeats form stable 'slippery' hairpin structures that may bind both the dsCNG repeat binding protein (dark gray oval) to the hairpin stem, and ssCNG repeat binding protein to the protruding repeat tail as demonstrated for the CUG repeat transcript and the CUG-BP (8). (D) Long normal clamped repeats form stable hairpin structures that bind the dsCNG repeat binding protein only. (E) An expanded repeat has the structure and protein binding specificity similar to that of the long normal repeats (C) or (D) but binds the dsCNG repeat binding protein in a length-dependent manner.

Finally, we have investigated the structure of the (AAG) $n$  repeats that are known to occur in the first intron of the X25 gene associated with Friedreich's ataxia (48). The formation of alternative triple-stranded DNA structures between the (AAG) $n$  repeats and their complementary (CTT) $n$  repeats is thought to inhibit the transcription and processing of the X25 gene resulting in the deficiency of its protein product frataxin (49–52). The ability of single-stranded (AAG) $n$  repeats to form hairpin structure in DNA is a matter of controversy (50–53). We have shown that the AAG17 does not form a hairpin structure in RNA, at least in the conditions of analysis used in this study. However, it forms an ordered single-stranded structure maintained most likely by the extensive stacking interactions. This structure present in the AAG17 is not present in AAG17\_cl. The natural environment of the (AAG) $n$  repeats in the X25 transcript resembles that of the AAG17 rather than AAG17\_cl. Thus, the rigid RNA structure of the (AAG) $n$  repeats may be required for their RNA functions including their proposed role as splicing activators mediated by interactions with Tra2 protein in humans (54).

In conclusion, the results presented in this study have revealed the basic structural features of several types of simple sequence repeats in RNA. These results shed more light on the normal roles played by the repeated sequences in transcripts, and give a better understanding of the mechanisms of RNA pathogenesis in human neurological diseases in which the repeats are involved.

## ACKNOWLEDGEMENTS

This work was supported by the State Committee for Scientific Research, Grant No. 6P04B03118 and PBZ/KBN/040/P04/12 and the Foundation for Polish Science, Grant No. 117/96 and 8/2000.

## REFERENCES

- Toth,G., Gaspari,Z. and Jurka,J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.*, **10**, 967–981.
- Borstnik,B. and Pumpernik,D. (2002) Tandem repeats in protein coding regions of primate genes. *Genome Res.*, **12**, 909–915.
- Kashi,Y., King,D. and Soller,M. (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.*, **13**, 74–78.
- Wells,R.D. and Warren,S.T. (1998) *Genetic Instabilities and Neurological Diseases*. Academic Press, San Diego, CA.
- Jasinska,A., Michlewski,G., de Mezer,M., Sobczak,K., Kozlowski,P., Napierala,M. and Krzyzosiak,W.J. (2003) Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Res.*, **31**, 5463–5468.
- McLaughlin,B.A., Spencer,C. and Eberwine,J. (1996) CAG trinucleotide RNA repeats interact with RNA-binding proteins. *Am. J. Hum. Genet.*, **59**, 561–569.
- Timchenko,L.T., Timchenko,N.A., Caskey,C.T. and Roberts,R. (1996) Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum. Mol. Genet.*, **5**, 115–121.
- Michalowski,S., Miller,J.W., Urbinati,C.R., Paliouras,M., Swanson,M.S. and Griffith,J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res.*, **27**, 3534–3542.
- Tian,B., White,R.J., Xia,T., Welle,S., Turner,D.H., Mathews,M.B. and Thornton,C.A. (2000) Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*, **6**, 79–87.
- Peel,L., Rao,R.V., Cottrell,B.A., Hayden,M.R., Ellerby,L.M. and Bredesen,D.E. (2001) Double-stranded RNA-dependent protein kinase, PKR, binds preferentially to Huntington's disease (HD) transcripts and is activated in HD tissue. *Hum. Mol. Genet.*, **10**, 1531–1538.
- Timchenko,L.T. (1999) Myotonic dystrophy: the role of RNA CUG triplet repeats. *Am. J. Hum. Genet.*, **64**, 360–364.
- Mankodi,A., Takahashi,M.P., Jiang,H., Beck,C.L., Bowers,W.J., Moxley,R.T., Cannon,S.C. and Thornton,C.A. (2002) Expanded CUG repeats trigger aberrant splicing of CIC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell*, **10**, 35–44.
- Lu,X., Timchenko,N.A. and Timchenko,L.T. (1999) Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. *Hum. Mol. Genet.*, **8**, 53–60.
- Philips,A.V., Timchenko,L.T. and Cooper,T.A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science*, **280**, 737–741.
- Tapscoff,S.J. and Thornton,C.A. (2001) Biomedicine. Reconstructing myotonic dystrophy. *Science*, **293**, 816–817.
- Charlet,B.N., Savkur,R.S., Singh,G., Philips,A.V., Grice,E.A. and Cooper,T.A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell*, **10**, 45–53.
- Amack,J.D. and Mahadevan,M.S. (2001) The myotonic dystrophy expanded CUG repeat tract is necessary but not sufficient to disrupt C2C12 myoblast differentiation. *Hum. Mol. Genet.*, **10**, 1879–1887.



18. Amack, J.D., Reagan, S.R. and Mahadevan, M.S. (2002) Mutant DMPK 3'-UTR transcripts disrupt C2C12 myogenic differentiation by compromising MyoD. *J. Cell Biol.*, **159**, 419–429.
19. Savkur, R.S., Philips, A.V. and Cooper, T.A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nature Genet.*, **29**, 40–47.
20. Timchenko, N.A., Welm, A.L., Lu, X. and Timchenko, L.T. (1999B) CUG repeat binding protein (CUGBP1) interacts with the 5' region of C/EBPbeta mRNA and regulates translation of C/EBPbeta isoforms. *Nucleic Acids Res.*, **27**, 4517–4525.
21. Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R.C., Angelini, C. and Hoffman, E.P. (1995) Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum. Mol. Genet.*, **4**, 599–606.
22. Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W. and Ranum, L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**, 864–867.
23. Feng, Y., Zhang, F., Lokey, L.K., Chastain, J.L., Lakkis, L., Eberhart, D. and Warren, S.T. (1995) Translational suppression by trinucleotide repeat expansion at FMR1. *Science*, **268**, 731–734.
24. Ranum, L.P. and Day, J.W. (2002) Dominantly inherited, non-coding microsatellite expansion disorders. *Curr. Opin. Genet. Dev.*, **12**, 266–271.
25. Richards, R.I. and Sutherland, G.R. (1997) Dynamic mutation: possible mechanisms and significance in human disease. *Trends Biochem. Sci.*, **22**, 432–436.
26. Cummings, C.J. and Zoghbi, H.Y. (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.*, **9**, 909–916.
27. Bowater, R.P. and Wells, R.D. (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog. Nucleic Acid Res. Mol. Biol.*, **66**, 159–202.
28. Sinden, R.R. (2001) Neurodegenerative diseases. Origins of instability. *Nature*, **411**, 757–758.
29. Napierala, M. and Krzyzosiak, W.J. (1997) CUG repeats present in myotonin kinase RNA form metastable 'slippery' hairpins. *J. Biol. Chem.*, **272**, 31079–31085.
30. Miller, J.W., Urbinati, C.R., Teng-Ummuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A. and Swanson, M.S. (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.*, **19**, 4439–4448.
31. Krzyzosiak, W.J., Napierala, M. and Drozd, M. (1999) RNA structure modules with trinucleotide repeat motifs. In Barciszewski, J. and Clark, B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer, Dordrecht, The Netherlands, pp. 303–314.
32. Zuker, M., Mathews, D.H. and Turner, D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In Barciszewski, J. and Clark, B.F.C. (eds), *RNA Biochemistry and Biotechnology*. Kluwer, Dordrecht, The Netherlands, pp. 11–43.
33. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
34. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.P. and Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.*, **15**, 9109–9128.
35. Giege, R., Helm, M. and Florentz, C. (2001) Classical and novel chemical tools for RNA structure probing. In Sool, D., Nishimura, S. and Moor, P.B. (eds), *RNA*. Pergamon, Amsterdam, The Netherlands, pp. 71–90.
36. Krzyzosiak, W.J., Marciniak, T., Wiewiorowski, M., Romby, P., Ebel, J.P. and Giege, R. (1988) Characterization of the lead(II)-induced cleavages in tRNAs in solution and effect of the Y-base removal in yeast tRNA<sup>Phe</sup>. *Biochemistry*, **27**, 5771–5777.
37. Ciesiolka, J., Michalowski, D., Wrzesinski, J., Krajewski, J. and Krzyzosiak, W.J. (1998) Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J. Mol. Biol.*, **275**, 211–220.
38. Brown, R.S., Dewan, J.C. and Klug, A. (1985) Crystallographic and biochemical investigation of the lead(II)-catalyzed hydrolysis of yeast phenylalanine tRNA. *Biochemistry*, **24**, 4785–4801.
39. Mankodi, A. and Thornton, C.A. (2002) Myotonic syndromes. *Curr. Opin. Neurol.*, **15**, 545–552.
40. Good, P.J., Chen, Q., Warner, S.J. and Herring, D.C. (2000) A family of human RNA-binding proteins related to the *Drosophila* Bruno translational regulator. *J. Biol. Chem.*, **275**, 28583–28592.
41. Ladd, A.N., Charlet, N. and Cooper, T.A. (2001) The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol. Cell Biol.*, **21**, 1285–1296.
42. Suzuki, H., Jin, Y., Otani, H., Yasuda, K. and Inoue, K. (2002) Regulation of alternative splicing of alpha-actinin transcript by Bruno-like proteins. *Genes Cells*, **7**, 133–141.
43. Seznec, H., Agbulut, O., Sergeant, N., Savouret, C., Ghestem, A., Tabti, N., Willer, J.C., Ourth, L., Duros, C., Brisson, E. et al. (2001) Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum. Mol. Genet.*, **10**, 2717–2726.
44. Buj-Bello, A., Furling, D., Tronchere, H., Laporte, J., Lerouge, T., Butler-Browne, G.S. and Mandel, J.L. (2002) Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells. *Hum. Mol. Genet.*, **11**, 2297–2307.
45. Fardaei, M., Larkin, K., Brook, J.D. and Hamshire, M.G. (2001) *In vivo* co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.*, **29**, 2766–2771.
46. Mankodi, A., Urbinati, C.R., Yuan, Q.P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S. and Thornton, C.A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.*, **10**, 2165–2170.
47. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshire, M.G., Harper, P.S. and Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.*, **11**, 805–814.
48. Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A. et al. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423–1427.
49. Bidichandani, S.I., Ashizawa, T. and Patel, P.I. (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.*, **62**, 111–121.
50. Suen, I.S., Rhodes, J.N., Christy, M., McEwen, B., Gray, D.M. and Mitas, M. (1999) Structural properties of Friedreich's ataxia d(GAA) repeats. *Biochim. Biophys. Acta*, **1444**, 14–24.
51. Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M. and Wells, R.D. (2001) Sticky DNA, a self-associated complex formed at long GAA\*TTT repeats in intron 1 of the frataxin gene, inhibits transcription. *J. Biol. Chem.*, **276**, 27171–27177.
52. LeProust, E.M., Pearson, C.E., Sinden, R.R. and Gao, X. (2000) Unexpected formation of parallel duplex in GAA and TTC trinucleotide repeats of Friedreich's ataxia. *J. Mol. Biol.*, **302**, 1063–1080.
53. Heidenfelder, B.L., Makhov, A.M. and Topal, M.D. (2003) Hairpin formation in Friedreich's ataxia triplet repeat expansion. *J. Biol. Chem.*, **278**, 2425–2431.
54. Tacke, R., Tohyama, M., Ogawa, S. and Manley, J.L. (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93**, 139–148.