2654–2660 Nucleic Acids Research, 1995, Vol. 23, No. 14

Single-stranded DNA binding proteins isolated from mouse brain recognize specific trinucleotide repeat sequences *in vitro*

Hiroko Yano-Yanagisawa, Yin Li, Hua Wang and Yoshinori Kohwi*

La Jolla Cancer Research Foundation, La Jolla, CA 92037, USA

Received April 11, 1995; Revised and Accepted June 13, 1995

ABSTRACT

Expansion of trinucleotide repeats $(CAG)_n$ and $(CGG)_n$ is found in genes responsible for certain human hereditary neurodegenerative diseases. By gel-mobility shift assay, we detected a single-stranded $(AGC)_n$ repeat-binding activity primarily in mouse brain extracts and very low or undetectable activity in other tissue extracts. Two $(AGC)_n$-repeat binding proteins, with apparent molecular weights of 44 and 40 kDa, have been purified from mouse adult brain by a DNA affinity column and fast protein liquid chromatography. UV-cross linking of radiolabeled $(AGC)_n$ repeats with crude brain extracts and with purified two proteins of 44 and 40 kDa produced identical doublet bands, indicating that these proteins are in fact responsible for the $(AGC)_n$-binding activity in brain extracts. We designated these two proteins TRIP-1 for the 44 kDa protein and TRIP-2 for the 40 kDa protein, where TRIP represents trinucleotide repeat-binding protein. TRIP-1 and TRIP-2 bind to a specific subset of trinucleotide repeat sequences including $(AGC)_n$, $(AGT)_n$, $(GGC)_n$, and $(GGT)_n$ repeats but not to various other trinucleotide repeats. A minimum of eight (AGC) trinucleotide repeating units is required for TRIP-1 and -2 recognition and binding. The $(AGC)_n$ repeat-binding activity increases in the brain after birth and reaches a plateau within 3 weeks. In the brain, TRIP-1 and TRIP-2 may alter the function of the genes containing the expanded-trinucleotide repeats.

INTRODUCTION

The expansion of trinucleotide repeats, a novel form of mutation, has recently been found in the genes responsible for several genetic disorders (for review see 1). These include the trinucleotide-repeat expansion of $(CGG)_n$ in the mutant gene responsible for the inherited mental-retardation disorder fragile X syndrome (2–4), $(CAG)_n$ in the gene responsible for myotonic dystrophy (dystrophia myotonica, DM), an autosomal dominant neuromuscular disease (5–8) and also $(CAG)_n$ in the androgen receptor gene of X-linked spinal and bulbar muscular atrophy, a rare motor neuron disorder (SBMA or Kennedy's disease; 9,10). The discovery of the gene responsible for Huntington's disease (HD) also revealed that $(CAG)_n$ is expanded on HD chromosomes (11–14). Other expanded trinucleotide repeats were identified in the genes responsible for spinocerebellar ataxia type1 (SCA1; 15), hereditary dentatorubral-pallidoluysian atrophy (DRPLA; 16,17), Machado–Joseph disease (18) and the fragile site FRXE (19). In many of these conditions severity and/or the age of onset of the disease is determined by the number of repeating units. Although expansion of trinucleotide repeats has been associated with genetic disorders, it is not known how trinucleotide repeats expand or how this expansion causes disease.

We previously tested a hypothesis that trinucleotide repeats have an intrinsic propensity to adopt a non-B DNA structure which might interfere with processes such as DNA replication and transcription. Based on this idea, we found that $(CAG)_n$ repeats possess an unusual structural property: $(CAG)_n$ repeats consist of clusters of non-B DNA structural units formed by each AGC trinucleotide repeating unit {5′···(C AG)(C AG)(C AG)(C AG)···3′}. The hallmark of the non-B structure is that cytosines are specifically base-unpaired (20). The intrinsic propensity of trinucleotide repeats to adopt an unusual DNA structure may be involved in their abnormal behavior in cells (21,22).

To investigate the mechanism of $(AGC)_n$ repeats expansion seen in hereditary genetic disorders, we searched for an $(AGC)_n$-repeats-binding protein. In the present paper, we will employ the nomenclature $(AGC)_n$ repeats instead of $(CAG)_n$ repeats because (AGC) repeating units represent the DNA structure unit. Richards *et al.* (23) have reported a double-stranded (ds)-$(CCG/CGG)_n$ repeats-binding protein in HeLa cell nuclear extracts (CCG-BP1; 23). The CCG-BP1 binds ds-$(CCG/CGG)_n$ repeats but not methylated-ds-$(CCG/CGG)_n$ repeats that are observed in fragile X syndrome. The binding specificity of CCG-BP1 may be important to understand the mechanisms underlying fragile X syndrome. In addition to the CCG-BP1, several other double-stranded-simple tandem repeat-binding proteins have been detected in HeLa nuclear extracts as well as single-stranded (ss)-$(CCG)_n$ repeats and ss-$(CGG)_n$ repeats binding proteins. However, detailed characterization of those trinucleotide repeat-binding proteins have not been reported.

We detected and purified two proteins responsible for ss-$(AGC)_n$ repeats binding activity in mouse brain, which is absent in HeLa

* To whom correspondence should be addressed

cells. These two proteins bind a specific subset of single-stranded-trinucleotide repeat sequences, including ss-$(CGG)_n$ repeats. The proteins were expressed primarily in the brain and the ss-$(AGC)_n$ repeats-binding activity progressively increases after birth.

MATERIALS AND METHODS

Purification of native $(AGC)_n$-binding protein from the brain

Tissue extracts were prepared from brains of 4–6 week-old mice. A single mouse brain (~0.5 g) was homogenized using a dounce homogenizer (50 strokes, pestle A) in 1.5 ml of extraction buffer (0.4 M KCl; 10 mM sodium phosphate buffer pH 7.4; 10% glycerol; 0.5 mM DTT; 0.1 mM phenylmethylsulfonyl fluoride (PMSF); 10 μg/ml leupeptin and 10 μg/ml aprotinin). The extract was centrifuged for 1 h in a Sorvall centrifuge, T865.1 rotor at 38 000 r.p.m., and the supernatant (~3–5 mg protein/ml) was either subjected directly to gel-mobility shift assay, or used for affinity purification. TRIP-1 and -2 were purified by DNA affinity column chromatography (24). Briefly, synthetic oligonucleotide 5′-$(AGC)_{13}$-3′ were coupled to cyanogen-bromide activated Sepharose 6B (Pharmacia), according to the manufacturer's instructions. Prior to loading on the column, the cell extract was diluted 4-fold with buffer H (25 mM Hepes pH 7.4, 5 mM EDTA, 0.1% NP-40, 1 mM DTT, 20% glycerol). This diluted extract was loaded directly on the affinity column, which was then washed with buffer H containing 0.1 M NaCl. Bound protein was eluted with 0.5 ml/fraction of 0.5 M NaCl in buffer H. The DNA affinity-column purified proteins were renatured from gel sections excised from sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) gel and tested $(AGC)_{13}$-binding activity by the method of Ossipow (25).

The proteins in this DNA affinity column purified-fraction (10 mouse brains) were further purified by fast protein liquid chromatography (FPLC; Pharmacia). The 0.5 M NaCl fraction from the DNA-affinity column was equilibrated to 0.1 M NaCl in buffer-S (20 mM Hepes pH 7.4, 20% ethylene glycol, 0.01% Triton X-100, 1 mM DTT) with the use of a Centriprep-30 column (Amicon) and the proteins were applied to an FPLC-Mono-S column (cation exchange) pre-equilibrated with 0.1 M NaCl in buffer-S. The TRIP-1 and -2 were eluted from the column with a 0.1–1 M NaCl linear gradient in buffer-S (flow rate: 0.5 ml/min, collected in 60 fractions by 0.5 ml/fraction). The fractions with $(AGC)_{13}$-binding activity were pooled, equilibrated to 0.1 M NaCl in buffer-Q (20 mM Tris–HCl, pH 7.6, 20% ethylene glycol, 0.01% Triton X-100, 1 mM DTT) and loaded onto a FPLC Mono-Q column (anion exchange) pre-equilibrated with 0.1 M NaCl in buffer-Q. The TRIP-1 and -2 were eluted from this column using the same procedure employed for the Mono-S column. Fractions were analyzed for $(AGC)_{13}$-binding activity and the active fractions were pooled and concentrated in 400 μl of buffer H with 0.1 M NaCl.

Gel-mobility shift assays

The DNA probes used for gel mobility shift assay were prepared as follows; synthetic oligomers were first purified by oligonucleotide purification column (Applied Biosystem Inc.) and then by a denaturing 15% polyacrylamide gel electrophoresis. The twice purified oligomers were EtOH precipitated and the OD_{260} was measured to calculate the concentration of ssDNA. One μg of the purified oligomers were labeled by kinase and [γ-^{32}P]ATP as described (26).

Gel mobility shift assay was employed to study the DNA binding activity of proteins in tissue extracts and the purified fractions as described previously (27). Briefly, binding of proteins with radiolabeled oligomer probes was allowed in 20 μl total volume containing 10 mM Hepes pH 7.9, 1 mM DTT, 50 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.05 mg/ml poly(dI–dC)·poly(dI–dC) and 0.2–10 μg protein of the tissue extracts or the affinity-purified fractions (1–5 μl). Samples were pre-incubated at room temperature for 5 min before radiolabeled DNA probe (0.1 ng) was added. The binding reaction was allowed to proceed at room temperature for 15 min. Aliquots (4–10 μl) of each sample were electrophoresed and the gels were dried and exposed to XAR X-ray film. Using $(AGC)_{13}$ as a probe, 1 μg of protein of adult brain extracts (a total of 5–7 mg protein from one mouse brain) and 2 μl of FPLC purified-TRIP-1 and -2 fraction (the final volume of 400 μl from 10 brains) show 50–60% of gel shift activity.

UV cross-linking assay

Reaction mixtures (20 μl) of the protein (200 ng/μl of tissue extracts) and radiolabeled oligomer probes (0.5 ng) in reaction buffer (10 mM Hepes buffer pH 7.9, 50 mM KCl, 5 mM $MgCl_2$, 0.05 mg/ml dI–dC) were incubated at room temperature for 15 min and irradiated with a UV light (UV Stratalinker, Stratagene, 254 nm of 240 mW/cm^2/min) at 5 cm from the surface of the lamp for 5 min. SDS–PAGE sample buffer was added, boiled for 3 min and loaded on an SDS–10% polyacrylamide gel. After electrophoresis, the gel was stained and exposed to XAR X-ray film.

RESULTS

Detection of an $(AGC)_n$ repeat sequence binding protein

To search for an $(AGC)_n$-binding protein(s), we initially examined an $(AGC)_n$-binding activity in mouse brain and liver extracts by gel-mobility shift assay. First, double-stranded (ds) $(AGC/GCT)_{20}$ was tested for binding. Although brain extracts showed a weak binding to ds$(AGC/GCT)_{20}$, the binding was non-specific because it was not competed with non-labeled ds$(AGC)_{20}$ (data not shown). Next, we examined single-stranded-(ss)-trinucleotide repeats-binding activity using synthetic oligomers, ss$(AGC)_{13}$ and its complementary sequence, ss$(GCT)_{13}$. As shown in Figure 1, migration of the ss$(AGC)_{13}$ oligomer was retarded by pre-incubation with brain extracts (Fig. 1A, lane 2), but not in liver extracts (data not shown), and the binding was specifically competed with non-labeled ss$(AGC)_{13}$ (Fig. 1A, lanes 3–6) but not with non-labeled ss$(ACG)_{13}$ and ds$(AGC)_{20}$ (data not shown). Using similar amounts of proteins in the brain extracts, the radiolabeled ss$(GCT)_{13}$ oligomer showed no binding. And this was true when a higher amount of proteins was used (Fig. 1B, lanes 2–4). This data suggest that there exists a protein(s) in mouse brain that specifically binds to a single-stranded $(AGC)_n$ repeat DNA, but not to single-stranded $(GCT)_n$ repeats or double-stranded $(AGC/GCT)_n$ repeats.

The $(AGC)_n$ binding activity was found primarily in the brain

The ss$(AGC)_n$ repeat sequence binding activity [$(AGC)_n$-binding activity] was examined by gel mobility shift assay in various adult mouse tissue extracts including brain, liver, lung, heart, skeletal muscle, kidney, testis and thymus. Brain extracts showed strong

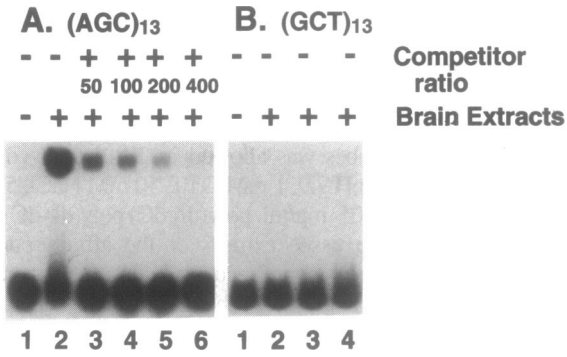


Figure 1. Single-stranded (AGC) repeat sequence binding activity detected in a crude brain extract. (**A**) Gel-mobility shift analysis with tissue extracts from mouse brain (4 μg of protein) was performed using ^{32}P-end-labeled $(AGC)_{13}$ oligomer as a probe. 50–400-fold molar excess of non-labeled $(AGC)_{13}$ oligomer was used as specific competitor as indicated above the figure. The – or + signs above the lanes indicate the absence or presence of protein or competitor. (**B**) Gel-mobility shift analysis with end-labeled $(GCT)_{13}$ oligomer as a probe with 0 μg (lane 1), 2 μg (lane 2), 4 μg (lane 3) and 8 μg (lane 4) of proteins from brain extracts.

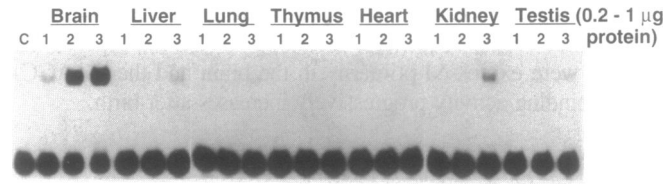


Figure 2. $(AGC)_n$-repeats binding activity predominantly expressed in mouse brain. Gel-mobility shift analysis with no extract (control lane c), brain extracts, liver extracts, lung extracts, thymus extracts, heart extracts, kidney extracts and testis extracts was performed using end-labeled ss$(AGC)_{13}$ as a probe. Lanes 1–3: 0.25, 0.5 and 1 μg protein of tissue extracts, respectively.

binding activity to ss$(AGC)_{13}$ repeat sequence, but this binding activity was very weak or virtually undetectable in other tissue extracts (Fig. 2). For brain extracts, 1 μg of protein in the extract showed more than 50% shift of the $(AGC)_{13}$ probe (Fig. 2, lane 2 in brain), whereas for testis extracts at least 20 times more protein was required to show comparable DNA binding activity (Fig. 2, testis). To evaluate the quality of the brain and other tissue extracts used in this study, we tested their binding activity to single-stranded poly(dC) sequences. The poly(dC) sequence binding protein, hnRNP-K, is known to be ubiquitously expressed (28). We detected similar levels of the poly(dC) sequence binding activity in all tissue extracts which we studied confirming the proteins were active in these extracts (data not shown). To identify proteins responsible for ss$(AGC)_{13}$ binding, we have purified $(AGC)_n$-binding proteins from mouse adult brain.

Purification of the trinucleotide $(AGC)_n$-binding proteins, TRIP-1 and -2

To identify the protein(s) responsible for $(AGC)_n$-binding activity, we prepared whole cell extracts from mouse brain tissues (4–9 weeks old) and applied them to a DNA affinity column containing single-stranded AGC oligomers as described in Materials and Methods. Protein with this binding activity was eluted in a 0.5 M NaCl fraction. Three proteins, with estimated molecular weight of 47, 44 and 40 kDa, were detected on SDS–PAGE (Fig. 3A, lane 2). The whole SDS–PAGE gel was sliced every 2 mm and each gel slice was subjected to protein extraction, renaturation and gel-mobility shift assay. We confirmed that the 44 kDa protein (TRIP-1) and 40 kDa protein (TRIP-2) showed $(AGC)_n$-binding activity (Fig. 3B). We further purified these proteins by FPLC Mono-S column (cation exchange column) followed by Mono-Q column (anion exchange column). The non-specific DNA-binding protein with 47 kDa was easily separated on a Mono-S column (in the flow through fraction) from TRIP-1 and TRIP-2. However, TRIP-1 and TRIP-2 could not be separated from each other by either Mono-S and Mono-Q columns as shown in Figure 3A, lane 3 and Figure 3C. No other proteins were detected in this purified-fraction by SDS–PAGE followed by silver staining. This indicates that their chemical properties are similar and suggests that these two proteins are closely related.

To determine whether TRIP-1 and -2 are responsible for $(AGC)_n$-binding activity predominantly found in brain extracts, we performed UV-cross linking studies of crude brain extracts and purified TRIP-1 and -2 using radio-labeled $(AGC)_{13}$ oligomer. Side-by-side gel electrophoresis of the UV-cross linked products showed a doublet that migrated at the same positions centered around 60 kDa (Fig. 3D). This data identified TRIP-1 and TRIP-2 to be the key proteins exhibiting the $(AGC)_n$-binding activity detected in the brain.

Trinucleotide repeat target selectivity of the purified TRIP-1 and TRIP-2

Using the final FPLC fraction containing both TRIP-1 and TRIP-2, we first screened various trinucleotide repeat sequences for binding by gel-mobility shift assay (Fig. 4A–D, and data not shown). For those DNAs that bound, UV-cross linking analysis was performed to determine which protein was responsible for the binding (Fig. 4E). The first analysis employing the gel-mobility shift assay used 13 units of various trinucleotide repeat sequence oligomers and the FPLC purified-fraction under protein excess. This series of experiments showed that another 'triplet disease' element, $(GGC)_{13}$ repeats, is also a preferred target for TRIP-1 and -2 (Fig. 4C) with an affinity similar to $(AGC)_{13}$ (Fig. 4B). However, the TRIP-1 and -2 fraction did not bind $(GAC)_{13}$ repeats, that differs in sequence but has the same composition as $(AGC)_{13}$ repeats (Fig. 4A). Other purine·purine·pyrimidine repeats, $(AAC)_{13}$ and $(AAT)_{13}$ also failed to bind the TRIP-1 and -2 fraction. In addition to $(AGC)_{13}$ and $(GGC)_{13}$, TRIP-1 and -2 strongly bind to $(AGT)_{13}$ repeats with an almost 8-fold greater affinity than to $(AGC)_{13}$ (Fig. 4D). TRIP-1 and -2 also bind $(GGT)_{13}$ repeats as strongly as $(AGT)_{13}$ (data not shown). Most trinucleotide repeats that are purine·pyrimidine·pyrimidine (GCT, GCC, GTC, ACT, ATC, ATT, ACC) show no affinity to the TRIP-1 and -2 fractions. This shows that both proteins in the TRIP-1 and -2 fraction select a specific subset of trinucleotide sequences for binding. It should also be noted that $(GTT)_{13}$ is an exceptional purine.pyrimidine.pyrimidine repeat that shows a significant level of binding.

UV-cross linking analysis employing the $(AGC)_{13}$, $(GGC)_{13}$, $(AGT)_{13}$ repeats revealed that both TRIP-1 and TRIP-2 recognize and bind this same set of trinucleotide sequences. All UV-cross linked protein–DNA complexes migrated as doublet at ~60 kDa.

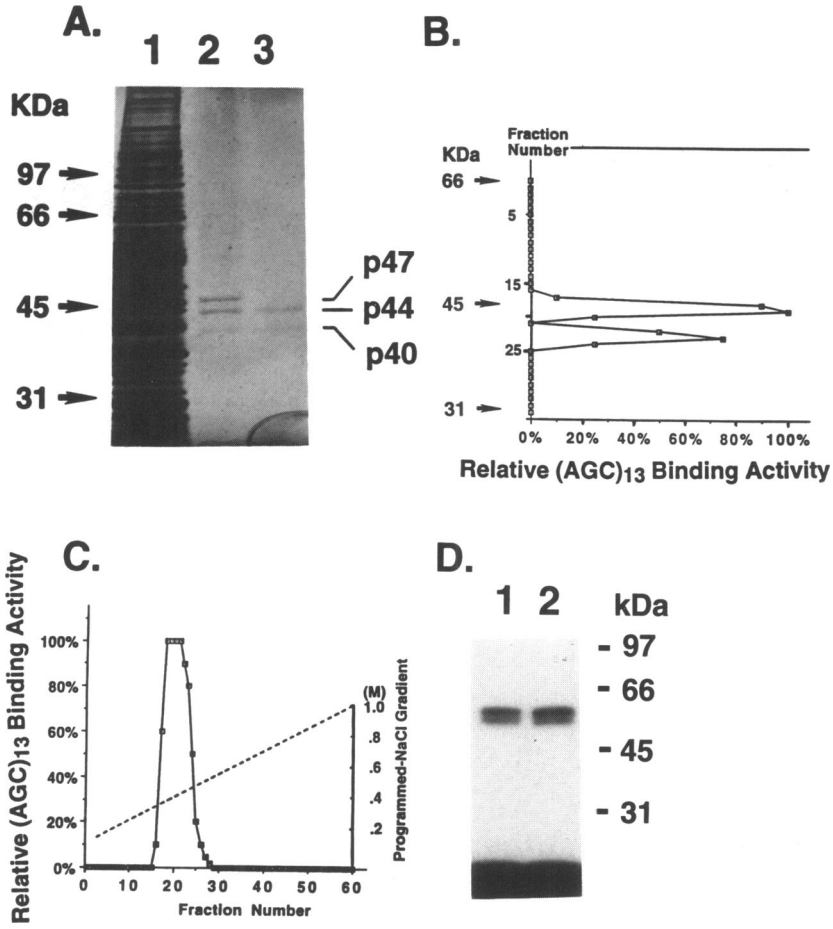


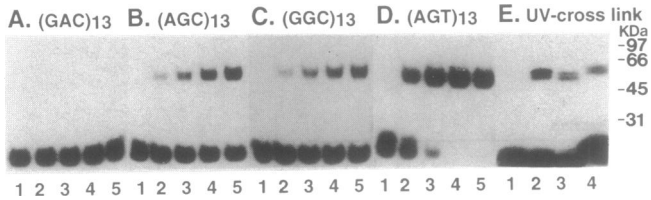
Figure 3. Purification of $(AGC)_n$ repeats-binding proteins, TRIP-1 and TRIP-2 from mouse brain. (**A**) Silver staining of a 10% SDS–polyacrylamide gel with crude brain extracts (lane 1), TRIP-1 and -2 purified by the affinity column (lane 2, 0.5 M NaCl elution) and TRIP-1 and -2 purified by affinity column and FPLC. Size markers are indicated in kDa. (**B**) Gel-mobility shift assay with renatured proteins from SDS–polyacrylamide gel of affinity column-purified TRIP-1 and -2. Fraction 17 corresponds to a 45 kDa molecular weight marker and fraction 34 corresponds to a 31 kDa molecular weight marker. Relative $(AGC)_{13}$ binding activity was shown by an intensity of shifted band of each fraction relative to the strongest shifted band in fraction 18. (**C**) Protein elution pattern from FPLC Mono-Q column detected by $(AGC)_{13}$ binding activity as described in Materials and Methods. Relative $(AGC)_{13}$ binding activity was shown by an intensity of the shifted band of each fraction relative to the strongest shifted band in fraction 18–21. (**D**) UV-cross linked products from 5 μg protein of brain extracts (lane 1) and 5 μl of affinity column and FPLC purified-TRIP-1 and -2 (lane 2) with radiolabeled $(AGC)_{13}$ in SDS–polyacrylamide gel.

Complexes containing $(AGC)_{13}$, $(GGC)_{13}$ and $(AGT)_{13}$ are shown in Figure 4E (lanes 2–5). No such doublet was seen with the negative control, $(GAC)_{13}$ (Fig. 4E, lane 1). The binding affinity of TRIP-1 and TRIP-2 to $(GGT)_{13}$ was similar to $(AGT)_{13}$ (data not shown). To ascertain that the doublet bands really contain the protein with 40 and 44 kDa cross linked to the target sequence, the TRIP-1 band and the TRIP-2 band were individually excised from an SDS–PAGE gel, the protein was renatured, and the renatured proteins were separately UV-cross-linked to the target sequences. The result confirmed that the doublet bands were derived from each of the two proteins in the TRIP-1 and -2 fraction (data not shown). The data suggest that TRIP-1 and TRIP-2 belong to the same family but possess slightly different target specificity. It is still possible that TRIP-2 (the 40 kDa protein) is a degradation product of TRIP-1 (the 44 kDa protein). The trinucleotide repeat sequences that did not bind TRIP-1 and -2 are listed in the middle panel of Figure 4.

Single-stranded DNA binding proteins are often found to be RNA binding proteins such as heterogeneous nuclear ribonucleoproteins isolated from mouse and human (hnRNPs; 29,30). We tested binding activity of the protein to both RNA and DNA using $d(AGC)_{13}$ DNA sequences and $r(AGC)_{13}$ RNA sequences. The TRIP-1 and -2 showed 20–30 times preferable binding to DNA over RNA (Fig. 4F). Therefore, TRIP-1 and -2 can be classified as single-stranded DNA binding proteins that bind specific trinucleotide repeats.

Eight (AGC) trinucleotide repeating units are required for recognition by TRIP-1 and -2

$(AGC)_{13}$ repeats exhibit a single-shifted band of the same mobility on a gel mobility shift assay using varying amounts of TRIP-1 and -2 (Fig. 4B). Similar results were observed with $(AGC)_{20}$ repeats (data not shown). However, for $(AGC)_{26}$, an additional shifted band of higher mobility was detected (data not shown). These data suggest that up to $(AGC)_{20}$, there is a single binding site for TRIP-1 and -2. To determine the minimum length of $(AGC)_n$-repeats necessary to bind TRIP-1 and -2, an $(AGC)_{13}$



Trinucleotide repeats that do not bind TRIP-1 and -2

5'Pu.Pu.Py.3'	GAC=ACG=CGA	5'Pu.Py.Py.3'	GTC=TCG=CGT
	AAC=ACA=CAA		GCT=CTG=TGC
	AAT=ATA=TAA		GCC=CCG=CGC
5'Pu.Pu.Pu.3'	GGG=poly(G)		ACC=CCA=CAC
	GGA=GAG=AGG		ACT=CTA=TAC
	GAA=AAG=AGA		ATC=TCA=CAT
	AAA=poly(A)		ATT=TTA=TAT
5'Py.Py.Py.3'	TTT=poly(T)	CCC=poly(C)	
	CTT=TTC=TCT	CCT=CTC=TCC	

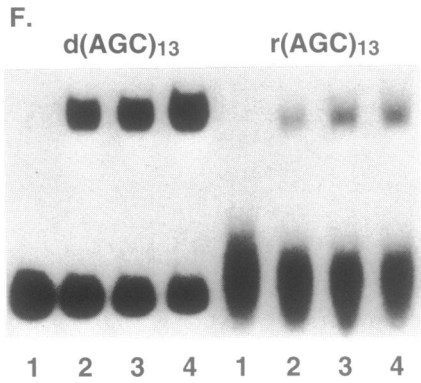


Figure 4. Target selectivity of TRIP-1 and -2. Gel-mobility shift assay with affinity column and FPLC purified-TRIP-1 and -2 was performed using various trinucleotide repeats $(NNN)_{13}$. (**A**) $(GAC)_{13}$, (**B**) $(AGC)_{13}$, (**C**) $(GGC)_{13}$, (**D**) $(AGT)_{13}$. (A–D) lane 1: probe alone, lanes 2–5: serial dilution of pooled solution of TRIP-1 and -2, 0.25, 0.5, 1 and 2 µl, respectively. (**E**) UV cross-linked TRIP-1 and -2 with ^{32}P-labeled probes in SDS–polyacrylamide gel, lane 1: $(GAC)_{13}$ and 5 µl of TRIP-1 and -2, lane 2: $(AGC)_{13}$ and 5 µl of TRIP-1 and -2, lane 3: $(GGC)_{13}$ and 5 µl of TRIP-1 and -2, lane 4: $(AGT)_{13}$ and 1 µl of TRIP-1 and -2. (**Table**) Trinucleotide DNA repeats that do not bind TRIP-1 and -2 as determined by gel-mobility shift assay are listed. (**F**) $d(AGC)_{13}$ DNA repeats and $r(AGC)_{13}$ RNA repeats were subjected to gel mobility shift assay using 1 µl (lane 2), 2 µl (lane 3) and 4 µl (lane 4) of purified-TRIP-1 and -2 fraction. lane 1: probe alone.

oligonucleotide of 39 bases in length was mutated such that different number of $(AGC)_n$ repeating units were flanked by $(AAC)_{13-n}$ repeats. The (AAC) repeats that do not bind TRIP-1 and -2 are included to make the length of the oligonucleotides constant. This series of oligonucleotides were tested for binding to TRIP-1 and -2 by gel mobility shift assay. The amount of TRIP-1 and -2 that shows 60% shift to the positive control $(AGC)_{13}$ was employed (Fig. 5, lane 2). In comparison with the $(AGC)_{13}$ oligomer as the binding target, TRIP-1 and -2 showed almost the same binding activity to 10 repeats of (AGC) trinucleotide repeating units, $(AAC)_3(AGC)_{10}$ (Fig. 5, lane 3), and ~20% less binding to eight repeats of (AGC) unit, $(AAC)_5(AGC)_8$ (Fig. 5, lane 4). However, it showed a marked reduction (80–90%) in binding to six repeats of (AGC) unit, $(AAC)_7(AGC)_6$ (Fig. 5, lane 5), and a negligible level of binding to four repeats of (AGC) unit,

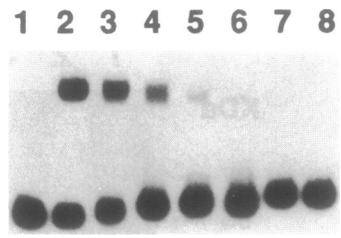


Figure 5. Minimal binding length of $(AGC)_n$ repeats for TRIP-1 and -2. Gel-mobility shift assay with affinity column purified-TRIP-1 and -2 was performed using varying length of $(AGC)_n$ repeats under the condition that gives 60% shifted band to $(AGC)_{13}$ (2 µl of TRIP-1 and -2 fraction, positive control, lane 2). lane 1: $(AGC)_{13}$; probe alone, lane 3; $(AAC)_3(AGC)_{10}$, lane 4; $(AAC)_5(AGC)_8$, lane 5; $(AAC)_7(AGC)_6$, lane 6; $(AAC)_9(AGC)_4$, lane 7; $(AAC)_{13}$ repeats (negative control), lane 8; $(AAC)_{13}$ probe alone.

$(AAC)_9(AGC)_4$ (Fig. 5, lane 6). The TRIP-1 and -2 did not show any binding to the negative control $(AAC)_{13}$ repeats (Fig. 5, lane 7). Although eight (AGC) trinucleotide repeating units are minimally required for recognition by TRIP-1 and -2, these units without any flanking sequences are not sufficient for TRIP-1 and -2 binding. Even $(AGC)_{11}$ repeats without flanking sequence did not show significant binding to TRIP-1 and -2 (data not shown). These results show that eight repeats of (AGC) trinucleotide repeating units are minimally required for recognition, but for TRIP-1 and -2 binding, an additional 15 bases are required. The added sequence $(AAC)_n$ is not recognized by TRIP-1 and -2, but it is there to provide the necessarily length for stable binding of the proteins to $(AGC)_8$.

The $(AGC)_n$-binding activity increases in the brain after birth

We next compared the $(AGC)_n$-binding activity in fetal brain and postnatal brain at different time points. By gel-mobility shift assay, $(AGC)_n$-binding activity was not detected in fetal brain. However, the binding activity was clearly detected as early as on day two, and this activity progressively increased and reached a plateau on day 21, as shown in Figure 6. We have also examined cerebrum and cerebellum separately and found similar $(AGC)_n$-binding activity in both tissues (data not shown).

Because $(AGC)_n$-binding activity is undetected in fetal brain and newly appears and increases after birth, it is unlikely that $(AGC)_n$-binding proteins are involved in neuronal determination, a process occurring primarily in embryonic brain. TRIP-1 and -2 may have an important role in brain development.

DISCUSSION

To obtain a clue for the mechanisms underlying certain neurodegenerative disorders in which responsible genes contain an expansion of trinucleotide repeats, we searched for an $(AGC)_n$-binding protein from mouse tissues. Single-stranded $(AGC)_n$-binding activity was found predominantly in the brain, and two DNA-binding proteins of 44 and 40 kDa, designated as TRIP-1 and TRIP-2, respectively, were purified. These proteins were confirmed to be responsible for the strong $(AGC)_n$-binding activity found in the brain. Both proteins strongly bound specific purine.purine.pyrimidine trinucleotide repeats $(AGT)_n$, $(GGT)_n$, $(AGC)_n$ and $(GGC)_n$, while neither bound $(GAC)_n$, $(AAC)_n$ or

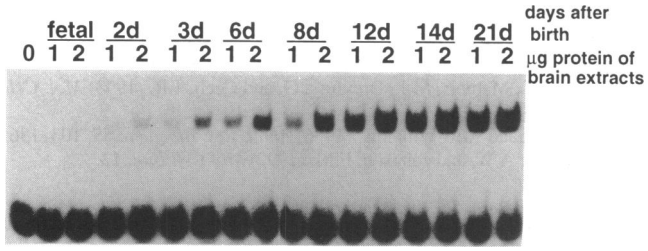


Figure 6. Increase in $(AGC)_n$-repeats binding activity in brain after birth. Gel-mobility shift analysis with fetal brain extracts and extracts prepared from brain of varying times after birth was performed using end-labeled $ss(AGC)_{13}$ as a probe. The amounts of proteins used are indicated above each lane (1 and 2 μg of protein of brain extracts). d: days/postnatal.

$(AAT)_n$ repeats. Among the purine.purine.pyrimidine trinucleotide repeats that bind TRIP-1 and -2, there are certain characteristically important requirements within each repeating unit. First, a guanine residue must exist and, secondly, a guanine, but not adenine, must be located 5′ of the pyrimidine residue. Therefore, TRIP-1 and -2 preferentially recognize ss(purine.guanine.pyrimidine)$_n$ repeats, such as $(AGC)_n$ over $(GAC)_n$ repeats, for example.

Several non-specific single-stranded DNA binding activities have been characterized (31), but little is known about the function of such proteins. Increasing evidence suggests that single-stranded purine-rich sequence binding proteins may have an active role in gene regulation and replication. A 28 kDa protein that recognizes a repeated, purine-rich element (PUR) present at putative origins of replication and in gene flanking regions in a variety of eukaryotes was isolated from HeLa nuclei extracts (32). A single-stranded $(GA)_n$ and $(GT)_n$ repeat binding protein (PGB with 33 kDa molecular weight) was also identified from human fibroblasts and it has been suggested that the binding of this protein to these repeats may enhance the transition from B- to non-B-DNA structure which may cause arrests of DNA replication and amplification (33). A chicken factor that binds regulatory regions of genes was identified and was found to recognize the intramolecular G–G base-pairing configuration of telomere repeats (34). Interestingly, the $(GGC)_n$ repeats, one of the TRIP-1 and -2 binding sequences, was demonstrated to form four-stranded structures *in vitro* (35). If such structures are formed *in vivo*, the structure may interfere with DNA replication and/or transcription. The $(AGC)_n$ repeats, another TRIP-1 and -2 binding sequence, form a novel non-B-DNA structure in supercoiled plasmid DNA (20). Whether TRIP-1 and -2 recognize target sequences when they form these unusual DNA structures awaits further experimentation.

Although the function of TRIP-1 and TRIP-2 is unknown at present, the observation that their $ss(AGC)_n$-binding activity is detected predominantly in the brain but at very low levels in other tissue sources tested in mice suggests that these proteins may have an important tissue-specific function in the brain. The $(AGC)_n$-binding activity is absent in the fetal brain and increases postnatally. It can be detected as early as on day 2 after birth, reaches a plateau between postnatal days 14 and 21 (Fig. 6), and maintains the same level of activity to at least 7–9 months (unpublished data). In genes responsible for each neurodegenerative disorder, the trinucleotide repeats have been expanded in all patient tissues. However, the expanded gene may readily become the binding targets for TRIP-1 and -2 present predominantly in the brain. If this is the case, TRIP-1 and -2 binding to these genes may alter the expression of their gene products and may lead to functional damage in the brain. Whether TRIP-1 and TRIP-2 serve as an important clue in understanding the mechanisms behind neurodegenerative disorders must await further research. Characterization of cDNAs encoding TRIP-1 and TRIP-2 and the generation of antibodies specific to TRIP-1 and TRIP-2 will provide tools to test this hypothesis in the future.

ACKNOWLEDGEMENTS

This study was supported by the fund from the National Institutes of Health (RO1CA51377) and by the fund from the Hereditary Disease Foundation (Santa Monica, CA) to Y.K.

REFERENCES

1 Mandel, J.-L. (1994) *Nature Genet.* **7**, 453–455.
2 Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G. and Warren, S.T. (1991) *Cell* **67**, 1047–1058.
3 Kremer, E.J., Prichard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R. and Richards, R.I. (1991) *Science* **252**, 1711–1714.
4 Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F. and Zhang, F.P. (1991) *Cell* **65**, 905–914.
5 Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P. and Hudson, T. (1992) *Cell* **68**, 799–808.
6 Fu, Y.H., Pizzuti, A., Fenwick, R.G.Jr, King, J., Rajnarayan, S., Dunne, P.W., Dubel, J., Nasser, G.A., Ashizawa, T. and de Jong, P. (1992) *Science* **255**, 1256–1259.
7 Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barcelo, J. and O'Hoy, K. (1992) *Science* **255**, 1253–1255.
8 Imbert, G., Kretz, C., Johnson, K. and Mandel, J.-L. (1993) *Nature Genet.* **4**, 72–76.
9 La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E. and Fischbeck, K.H. (1991) *Nature* **352**, 77–79.
10 Mhatre, A.N., Trifiro, M.A., Kaufman, M., Kazemi-Esfarjani, P., Figlewicz, D., Rouleau, G. and Pinsky, L. (1993) *Nature Genet.* **5**, 184–187.
11 Huntington's Disease Collaborative Research Group. (1993) *Cell* **72**, 971–983.
12 Andrew, S.E., Goldberg, Y.P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B. and Kalchman, M.A. (1993) *Nature Genet.* **4**, 398–403.
13 Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M. and Abbott, M. (1993) *Nature Genet.* **4**, 387–392.
14 Snell, R.G., MacMillan, J.C., Cheadle, J.P., Fenton, I., Lazarou, L.P., Davies, P., MacDonald, M.E., Gusella, J.F., Harper, P.S. and Shaw, D.J. (1993) *Nature Genet.* **4**, 393–397.
15 Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, T.J.,Jr, Servadio, A, Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P. and Zoghbi, H.Y. (1993) *Nature Genet.* **4**, 221–226.
16 Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A. and Hayashi, T. (1994) *Nature Genet.* **6**, 9–13.
17 Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I. and Murayama, N. (1994) *Nature Genet.* **6**, 14–18.
18 Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M. and Akiguchi, I. (1994) *Nature Genet.* **8**, 221–228.
19 Knight, S.J., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R., Barnicoat, A. and Pembrey, M.E. (1993) *Cell* **74**, 127–134.
20 Kohwi, Y., Wang, H. and Kohwi-Shigematsu, T. (1993) *Nucleic Acids Res.* **21**, 5651–5655.
21 Kohwi, Y., and T. Kohwi-Shigematsu. (1991) *Genes Dev.* **5**, 2547–2554.

22 Kohwi, Y. and Panchenko,Y. (1993) *Genes Dev.* **7**, 1766–1778.
23 Richards, R.I., Holman, K., Yu, S. and Sutherland, G.R. (1993) *Hum. Mol. Genet.* **2**, 1429–1435.
24 Kadonaga J. T. and R. Tjian. (1986) *Proc. Natl. Acad. Sci.USA* **83**, 5889–5893.
25 Ossipow, V., Laemmli, U. K. and Schibler, U. (1993) *Nucleic Acids Res.* **25**, 6040–6041.
26 Sambrook, K.J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* 2nd edition Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27 Nakagomi, K., Kohwi, Y., Dickinson, L.A. and Kohwi-Shigematsu, T. (1994) *Mol. Cell. Biol.* **14**, 1852–1860.
28 Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1992) *Mol. Cell. Biol.* **12**, 164–171.
29 McKay, S.J. and Cooke, H. (1992). *Nucleic Acids Res.* **20**, 6461–6464.
30 Ishikawa, F., Matunis, M.J., Dreyfuss, G. and Cech, T.R. (1993) *Mol. Cell. Biol.* **13**, 4301–4310.
31 Chase, J.W. and K.R. Williams. (1986) *Annu. Rev. Biochem.* **55**, 103–136.
32 Bergemann, A.W. and Johnson, E.M. (1992) *Mol.Cell.Biol.* **12**, 1257–1265.
33 Aharoni, A., Baran, N. and Manor, H. (1993) *Nucleic Acids Res.* **21**, 5221–5228.
34 Gualberto, A., Patrick, R.M. and Walsh, K. (1992) *Genes Dev.* **6**, 815–824.
35 Fry, M. and Loeb, L.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4950–4954.