Autonomous expansions of trinucleotide repeats with the general structure 5′-d(CNG)ₙ-3′ are associated with several human genetic diseases. We have characterized nuclear proteins binding to the unstable 5′-d(CGG)ₙ-3′ repeat. Its expansion in the human FMR1 gene leads to the fragile X syndrome, one of the most frequent causes of mental retardation in human males. Electrophoretic mobility shift assays using nuclear extracts from several human and other mammalian cell lines and from primary human cells demonstrated specific binding to double-stranded DNA fragments containing only a 5′-d(CGG)₁₇-3′ repeat or the repeat and flanking genomic sequences of the human FMR1 gene. Protein binding was inhibited by complete methylation of the trinucleotide repeat. The complex formed with crude nuclear extract apparently did not contain the human transcription factor Sp1 that binds to a characteristic GC-rich sequence. A 20-kDa protein involved in specific binding to the double-stranded 5′-d(CGG)₁₇-3′ repeat was purified from HeLa nuclear extracts by DNA affinity chromatography.

The autonomous, mechanistically still unexplained expansion of naturally occurring trinucleotide tandem repeats in the human genome has been recognized to be related to a number of serious human diseases: the fragile X syndrome (FRAXA locus), myotonic dystrophy, spinal and bulbar muscular atrophy, Huntington disease, mental retardation associated with myotonic dystrophy, and on the autosomes 11 (FRA11B; Ref. 10) and 16 (FRA16A; Ref. 11). All fragile sites identified so far have been found to be associated with amplifications of the simple unsta-

In the fragile X syndrome, the expanded tandem repeat 5′-d(CGG)ₙ-3′ is located in the 5′-untranslated region (UTR)¹ of the FMR1 gene in the human chromosomal location Xq27.3 (13). The number of repeat units varies between 6 and 54 in normal individuals, whereas more than 200 to up to 2000 repeat units can be found in affected individuals. Expansion of the repeat is accompanied by extensive methylation of the 5′-dCG-3′ dinucleotides in the repeat (14–16) and is associated with transcriptional silencing of the FMR1 gene (17–19). The function of the FMR1 protein is not yet known. The de novo methylation of the expanded trinucleotide repeat can be interpreted as a cellular defense against the invasion of foreign DNA or against unusual DNA structures (20, 21).

The cellular mechanism of triplet repeat amplification is not understood. Interestingly, procaryotic DNA polymerases are capable of expanding short synthetic oligodeoxyribonucleotides containing simple tandem repeat sequences to DNA stretches of several thousand nucleotides in lengths even in the absence of template DNA (21, 22). This finding suggests a slippage mechanism (23, 24) for the expansion of trinucleotide repeats presumably involving specific DNA-binding proteins. In transgenic mice for instance, a 5′-d(CAG)ₙ-3′ repeat in the andro-
gene receptor gene is stable upon transmission in the mouse, whereas it is expanding upon transmission in humans (25). The authors suggest the involvement of sequence-specific, probably species-specific, DNA-binding proteins in the amplification process. Experiments with crude nuclear extracts from human HeLa cells indeed have shown binding of proteins to tandem repeat sequences (26). In addition, an amplified 5′-d(CTG)-3′ repeat is a preferential target for nucleosome assembly (27, 28).

We have initiated experiments to characterize and purify human nuclear proteins that bind specifically to the double-stranded 5′-d(CGG)ₙ-3′ repeat. Such proteins are present in a variety of human and other mammalian cell lines, as well as in primary cells.

**EXPERIMENTAL PROCEDURES**

Cells and Cell Lines—Human HeLa cells were purchased from Gesellschaft für Biotechnologische Forschung, Braunschweig, Germany. Human KB and Jurkat cells, BHK21 hamster cells, and fat head minnow (FHM) fish cells (29) were propagated by standard meth-
ods. Primary human lymphocytes were prepared and grown as reported (30). Hamster cell line T637, an adenovirus type 12 (Ad12)-transformed BHK21 cell line, and the revertants of cell line T637, TR3, and TR12, with no detectable and about one genome equivalent of integrated Ad12 DNA, respectively (31), were all grown in Dulbecco's medium supple-
mented with 10% fetal calf serum. Cell lines 293 and HEK12, human embryonic kidney cells transformed with parts of adenovirus type 5 (Ad5) (32) and Ad12 (33), respectively, cell lines A549 (human lung cancer) and C4I (human cervix carcinoma), and a permanent cell line

¹ This research was supported by the Deutsche Forschungsgemeinschaft through SFB274-A1 and by the Fritz- Thyssen-Stiftung, Köln. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: Howard Hughes Medical Inst., CMM, University of California at San Diego, 950 Gilman Dr., La Jolla, CA 92039-0649.

§ To whom correspondence should be addressed: Institute of Genetics University of Cologne, Weyertal 121, D-50931 Köln, Germany. Tel.: 49-221-470-2386; Fax: 49-221-470-5163.

The abbreviations used are: UTR, untranslated region; FHM, fat head minnow; Ad12, adenovirus type 12; bp, base pair(s); EMSA, elec-
trophoretic mobility shift assays; CGGBP(s), CGG binding protein(s);

⁶C, 5′-methyldeoxycytidine.
isolate a human amnion tumor were gifts of the Institute of Cell Biology or of Molecular Biology (University of Essen, Medical School, Essen, Germany) as well as monkey Vero cells and the Ad12-transformed rat embryo fibroblast line REF 12.

Oligodeoxyribonucleotides and DNA Fragments—Oligodeoxyribonucleotides were synthesized in an Applied Biosystems 381A DNA synthesizer according to standard procedures. The oligodeoxyribonucleotides were carried out in a polymerase chain reaction thermal cycler (Perkin Elmer Cetus) under the following conditions: 10 min at 95°C, cooling to 70°C for 60 min, 60 min at 70°C, cooling to 58°C for 60 min, 60 min at 58°C, cooling to 17°C for 90 min, 60 min at 17°C. Oligodeoxyribonucleotides were subsequently purified by electrophoresis in agarose gels according to standard procedures.

Nuclear extracts from KB and BHK21 cells infected with Ad12 were prepared as described above, but smaller volumes of the buffers W 100 (1 ml) and W 150 (100 μl) were used. CGGBP(s) eluted in 20 μl of buffer E 750 to yield fraction IV. Only low activity remained after elution with buffer E 1000 (fraction IV).

Electrophoretic Mobility Shift Assay, Sodium Deoxycholate Treatment of DNA Protein Complexes, and Antibody Displacement/Super-shift Assay—End-labeled oligodeoxyribonucleotides or DNA fragments (30,000 cpm, equivalent to 2 fmol), unspecific DNA (poly(dA-dT), 1 μg) were incubated for 30 min at room temperature in 20 mM HEPES, 50-100 mM NaCl, 0.5 mM dithiothreitol, 10% glycerol, pH 7.9, with 0.5-2 μg of protein from crude nuclear extracts or 1 μl of fractions I to IV. DNA-protein complexes were separated by electrophoresis on polyacrylamide gels (7% for the separation of oligodeoxyribonucleotides, 10% for DNA fragments, 5% in 1 × TEB (89 mM Tris, 89 mM H2BO3, 2 mM EDTA, pH 8.4) without the addition of sodium dodecyl sulfate). The complexes were visualized by silver staining followed by densitometric scanning (37) followed by silver staining.

Electrophoretic Mobility Shift Assay, Sodium Deoxycholate Treatment of DNA Protein Complexes, and Antibody Displacement/Super-shift Assay—End-labeled oligodeoxyribonucleotides or DNA fragments (30,000 cpm, equivalent to 2 fmol), unspecific DNA (poly(dA-dT), 1 μg) were incubated for 30 min at room temperature in 20 mM HEPES, 50-100 mM NaCl, 0.5 mM dithiothreitol, 10% glycerol, pH 7.9, with 0.5-2 μg of protein from crude nuclear extracts or 1 μl of fractions I to IV. DNA-protein complexes were separated by electrophoresis on polyacrylamide gels (7% for the separation of oligodeoxyribonucleotides, 10% for DNA fragments, 5% in 1 × TEB (89 mM Tris, 89 mM H2BO3, 2 mM EDTA, pH 8.4) without the addition of sodium dodecyl sulfate). The complexes were visualized by silver staining followed by densitometric scanning (37) followed by silver staining.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.

RESULTS

Nuclear Proteins from Several Mammalian Cells Bind Specifically to the Double-stranded 5‘-d(CGG)n-3′ Repeat—We have examined nuclear extracts from various human and other mammalian cells by electrophoretic mobility shift assays (EMSAs) for the presence of proteins that bind to the double-stranded 5‘-d(CGG)n-3′ repeat located in the exon 1 of the human FMR1 gene (13). For this purpose, DNA fragments containing different tandem repeat sequences as competitors were preincubated with the specific DNA fragment as described above in the presence of the anti-Sp1 antibody (0.3-1 μg) for 60 min at room temperature. Complexes were separated by electrophoresis on polyacrylamide gels.
nucleotide (CGG)17ds. Formation of complex I could be competed by the oligodeoxyribonucleotide (CGG)17ds in at least 75-fold excess, but not by several oligodeoxyribonucleotides with different sequences (Fig. 3a). Additional DNA-protein complexes apparent in EMSAs shown in Fig. 2a were not specific as shown by competition experiments (Fig. 3a). Extracts from non-human cells like hamster BHK21 cells and rat embryo fibroblasts REF12 produced the same patterns as those from human cells (Fig. 2b). However, proteins from monkey Vero cells, from nonmammalian FHM fish cells, and from the insect cell line SF21 generated specific DNA-protein complexes (Fig. 2b), which were different from those with proteins from human cell lines.

Infection of the permissive human cell lines HeLa and KB with Ad12 did not abolish CGG-binding activity (Fig. 2a). However, the abortive infection of hamster BHK21 cells with Ad12 gave rise to two additional bands showing slightly higher mobility in EMSA (Fig. 2b). In contrast, extracts from the Ad12-transformed BHK21 cell line derivative T637 or from its revertants TR3 or TR12 showed the same patterns as proteins from extracts of the parental BHK21 cells. Interestingly, CGGBP(s) were not detectable in extracts isolated from BHK21 cells grown in suspension cultures (Fig. 2b).

The biological significance of these data had to be ascertained by repeating the binding experiments with authentic DNA fragments from the 5′-UTR of the FMR1 gene. Fragment 198ds gave rise to the DNA-protein complexes 1–4 (Fig. 3c, c1–c4) when nuclear extracts from human HeLa cells were used. Similar or identical patterns were found when extracts from other human or non-human cell lines were investigated.

**Table I**

Sequences of double-stranded oligodeoxyribonucleotides used in EMSAs

| Oligodeoxyribonucleotide | Sequence |
|--------------------------|---------|
| (CGG)8ds | 5′-(CGG CGG)8 CGG-3′ |
| (CGG)12ds | 5′-(CGG CGG)12 CGG-3′ |
| (CGG)17ds | 5′-(CGG CGG)17 CGG-3′ |
| (CAG)17ds | 5′-(CAG CAG)17 CAG-3′ |
| (CGA)17ds | 5′-(CGA CGA)17 CGA-3′ |
| (CAA)17ds | 5′-(CAA CAA)17 CAA-3′ |
| (TGG)17ds | 5′-(TGG TGG)17 TGG-3′ |
| CGGA8ds | 5′-(CGG AGG)8 CGG-3′ |
| CGGA7ds | 5′-(CGG TGG)7 CGG-3′ |
| (MGG)17ds | 5′-(MGG MGG)17 MGG-3′ |
| 8MCGGds | 5′ -(CGG)8 (CGG)8 (CGG)8 |
| 4MCGGds | 5′-(CGG)4 (CGG)4 (CGG)4 |

**Fig. 1.** Survey of DNA fragments used in EMSAs. DNA fragments were isolated from exon I of the human FMR1 gene cloned in the plasmid pE5.1. These fragments contained the double-stranded trinucleotide repeat 5′-d(CGG)n-3′ (gray boxes) flanked by genomic sequences of the 5′-untranslated region. The nucleotide numbers correspond to the sequence published in GenBank (accession number X61378). The start codon AUG is located approximately 70 bp downstream of the trinucleotide repeat.
Complex 3 was not always detectable. Complex 1 appeared to be specific for CGG binding, as its formation could be blocked by competition with the oligodeoxyribonucleotide (CGG)$_{17}$ds, but not with other oligodeoxyribonucleotides. The strong complex 4 seemed also to be formed by CGGBP(s), because its formation was partly competed by the oligodeoxyribonucleotide (CGG)$_{17}$ds (Fig. 3c) and also by 198ds. During the purification of CGGBP(s), complex 4 was the only detectable complex involving the 198ds fragment. Its formation could then be specifically competed by the oligodeoxyribonucleotide (CGG)$_{17}$ds and FMR1 promoter fragments 126ds, 198ds, and 248ds, but not by other oligodeoxyribonucleotides. Thus, complex 1 might contain additional factors that were probably associated with factors binding to flanking 3’-sequences. These additional factors could have been lost during purification and were no longer present in the CGGBP(s) in complex 4. Interestingly, the binding of proteins from nuclear extracts to the 126ds fragment with the same 5’-sequence as 198ds but a shorter 3’-end (Fig. 1) gave rise to only one complex and a pattern similar to that formed with the oligodeoxyribonucleotide (CGG)$_{17}$ds (data not shown). In contrast, binding of nuclear proteins to the 248ds fragment, which had the same 3’-sequence as 198ds but a longer 5’-sequence, produced the same pattern as the 198ds fragment.

Moreover, complex I was observed only with oligodeoxyribonucleotides (CGG)$_{17}$ds and (CGG)$_{12}$ds as binding probes, whereas (CGG)$_{8}$ds gave rise to a very faint complex (data not shown). The oligodeoxyribonucleotide FraxF isolated from the human FRAXF locus (9) did not serve as a specific binding probe for CGGBP(s) and did not compete for binding to (CGG)$_{17}$ds. The FraxF oligodeoxyribonucleotide contained eight 5’-d(CG)-3’ repeats and alternating 5’-d(CAGCGG)-3’-ds repeats (Table I). Hence, effective binding of CGGBP(s) to the recognition sequence required more than 8 repeat units.

Formulation of complex I was only partly competed by the synthetic oligodeoxyribonucleotide CGGT8ds (Fig. 3a), whereas no competition was observed with the oligodeoxyribonucleotide (TGG)$_{17}$ds (nucleotide sequences, see Table I). However, complex I formation was not competed by the addition of overlapping oligodeoxyribonucleotides with other triplet repeat sequences (Fig. 3a). Moreover, binding of nuclear proteins to the 5’-d(CAG)$_{17}$-3’ specific complex 1 (data not shown). When the authentic DNA fragments 198ds or 126ds were used as binding probes, the 5’-d(CGG)$_{17}$-3’-specific complexes 1, 3, and 4 were competed by the oligodeoxyribonucleotide CGGT8ds (Fig. 3c) but not with other oligodeoxyribonucleotides.

Complex 1 and complexes 1–4 were destroyed after the addition of the anionic detergent sodium deoxycholate (≈0.03%), whereas the nonionic detergents Triton X-100 or Tween 20 (≈2%) did not have any effects on complex formation (data not shown). Complex disruption by sodium deoxycholate was reversed in the presence of 0.6% Nonidet P-40. Although it cannot be ruled out that sodium deoxycholate as an anionic detergent affects protein-DNA interaction, the sodium deoxycholate sensitivity of the binding of CGGBP(s) to the 5’-d(CGG)$_{17}$-3’ repeat and the reversal by Nonidet P-40 suggest the involvement of protein-protein interactions in complex formation (38).

CGGBP(s) Do Not Bind to the Fully Methylated Trinucleotide Repeat—The results of experiments with crude nuclear extracts from HeLa cells suggested methylation sensitivity of proteins binding to the 5’-d(CGG)$_{17}$-3’ repeat (26). In order to investigate this problem further, oligodeoxyribonucleotides, which contained partly or fully methylated trinucleotide repeats, were used as binding probes or in competition experi-
ments (Fig. 3, a and b). The completely methylated oligodeoxynucleotide (MGG)$_{17}$ds and the partly methylated oligodeoxynucleotides 8MCGGds and 4MCGGds (nucleotide sequences see Table I) were synthesized by incorporating 5-methyldeoxycytidine instead of C during chemical synthesis. Only weak competition for the formation of complex I (cI) was observed when the completely methylated double-stranded oligodeoxynucleotide (MGG)$_{17}$ds was added (Fig. 3a). Moreover, only proteins from crude nuclear extracts were capable of forming complexes with the methylated oligodeoxynucleotide (MGG)$_{17}$ds (Fig. 3b, lanes 5-8). Specific complexes M1 and MII were formed only with crude nuclear extract and (MGG)$_{17}$ds (lanes 5-8). Their formation was competed only by oligodeoxynucleotides of the general structure (CGGNGG)$_{8}$CGGds (with n = T or C). Complex 3 was not always detectable. Double-stranded competitor oligodeoxynucleotides were used at a 300-fold excess over the double-stranded binding fragment (2 fmol). Sequences of oligodeoxynucleotides and a summary of competition experiments were described in Tables I and II, respectively.

**Fig. 3.** Specific binding of nuclear proteins to the double-stranded trinucleotide repeat 5'-d(CGG)$_{17}$-3'. Binding of nuclear proteins from HeLa cells to the oligodeoxynucleotide (CGG)$_{17}$ds (a), to the fully methylated oligodeoxynucleotide (MGG)$_{17}$ds (b), or to the promoter-derived DNA fragment 198ds (c) led to the formation of several specific DNA-protein complexes. a, complex I (cI) with the oligodeoxynucleotide (CGG)$_{17}$ds could be competed only with oligodeoxynucleotides of the general structures (CGG)$_{n}$ds (n $\geq$ 12) and (CGGNGG)$_{8}$CGGds with n = T or 5-methyldeoxycytidine. Oligodeoxynucleotide (MGG)$_{17}$ds containing the fully methylated trinucleotide repeat 5'-d(mCGG)$_{17}$-3' did not function as a competitor. b, CGGBP(s) did not bind to the fully methylated repeat 5'-d(mCGG)$_{17}$-3' in the oligodeoxynucleotide (MGG)$_{17}$ds. Fraction I (see Fig. 4) was incubated with either (CGG)$_{17}$ds (lane 2) or (MGG)$_{17}$ds (lane 4); complex formation was only observed with (CGG)$_{17}$ds. Specific complexes M1 and MII were formed only with crude nuclear extract and (MGG)$_{17}$ds (lanes 5-8). c, DNA fragment 198ds contained the trinucleotide repeat 5'-d(CGG)$_{17}$-3' flanked by genomic sequences of the 5'-untranslated region from the human FMR1 gene. In binding experiments, it gave rise to the specific complexes 1, 3 and 4 (c1, c3, and c4). Their formation was competed only by oligodeoxynucleotides of the general structure (CGGNGG)$_{8}$CGGds (with n = T or C). Complex 3 was not always detectable. Double-stranded competitor oligodeoxynucleotides were used at a 300-fold excess over the double-stranded binding fragment (2 fmol). Sequences of oligodeoxynucleotides and a summary of competition experiments were described in Tables I and II, respectively.

**Table II**

Summary of the results obtained in EMSAs and in competition experiments

| Oligodeoxynucleotide used for competition | Does oligodeoxynucleotide compete for binding? |
|-----------------------------------------|-----------------------------------------------|
| (CGG)$_{17}$ds                          | Weakly ND*                                     |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | Yes Yes (bands 1, 3, 4)                       |
| (CGG)$_{17}$ds                          | Yes ND                                        |
| (CGG)$_{17}$ds                          | No ND                                         |
| (CGG)$_{17}$ds                          | No ND                                         |

* ND, not done.
proteins that interacted specifically with highly methylated DNA sequences (40, 41).

It is concluded that proteins in nuclear extracts from primary human cells, from established cell lines, and from several mammalian as well as from some nonmammalian cells form a specific complex with the synthetic double-stranded oligodeoxyribonucleotides (CGG)$_n$ ds, with 12 $\leq n \leq$ 17. The oligodeoxyribonucleotide (CGG)$_8$ ds suffices for weak complex formation. The authentic DNA fragments 248ds, 198ds, or 126ds from the 5'-UTR of the human FMR1 gene can also form at least one 5'-d(CGG)$_{16}$-3' ds-specific complex and additional, probably less specific complexes. Some of the more complicated EMSA patterns (Fig. 3c) might be accounted for by additional complex formation with nucleotide sequences that flank the 5'-d(CGG)$_{16}$-3' repeat. Modifications of the specific 5'-d(CGG)$_{17}$-3' ds sequence can be tolerated for its efficiency in competition experiments when exchanges of the C are limited to 8 and to the pyrimidines T or 5-methyldeoxycytidine. CGGBP(s) do not bind to the fully methylated trinucleotide repeat sequence. The ubiquitous expression of CGGBP(s) points to an important function of these proteins. This binding activity seems to be highly conserved, since similar proteins have been found in extracts from nonmammalian fish or insect cells.

Binding of Nuclear Proteins from Human Cells to the Single-stranded Oligodeoxyribonucleotides (CGG)$_{17}$ ss and (CCG)$_{17}$ ss Sequences Is Unspecific—Several reports suggested that single-stranded oligodeoxyribonucleotides 5'-d(CGG)$_n$-3' and 5'-d(CCG)$_n$-3' (n $\geq$ 4) might adopt unusual structures in vitro (39, 42). In fact, these oligodeoxyribonucleotides exhibited abnormally high electrophoretic mobility in polyacrylamide gels (data not shown). We therefore examined these oligodeoxyribonucleotides for their capacity to bind nuclear proteins from human cells. The oligodeoxyribonucleotide (CGG)$_{17}$ ss led to the formation of several complexes that could, however, be prevented by competition with single-stranded oligodeoxyribonucleotides of the general sequence 5'-d(CGGCCSK)$_n$-3' (S could be G or C and K could be G or T), but not with double-stranded oligodeoxyribonucleotides. The oligodeoxyribonucleotide (CGG)$_{17}$ ss did not give rise to any specific complex at all. It is therefore likely that the generation of complexes between nuclear proteins and the single-stranded repeat sequences is rather unspecific and probably due to a single-strand binding protein.

The Human GC Box Binding Transcription Factor Sp1 Is Not Part of the CGGBP(s)-(CGG)$_{17}$ ds Complex—A possible candidate protein for complex formation with the double-stranded 5'-d(CGG)$_{17}$-3' repeat was the transcription factor Sp1, which recognized the consensus sequence 5'-dGGCCGG-3' (43). Therefore, an oligodeoxyribonucleotide Sp1 ds containing the Sp1 binding sequence (Table I) was tested for its capacity to compete for protein binding to the 5'-d(CGG)$_{17}$-3' repeat. It failed to function as a specific competitor (Fig. 3a).

In addition, we tried to assess the participation of Sp1 in the formation of the CGGBP(s)-(CGG)$_{17}$ ds complex by testing the effect of an anti-Sp1 monoclonal antibody on complex formation. This antibody did not affect complex formation (data not shown).

It is therefore concluded that the transcription factor Sp1 is not part of the CGGBP(s)-(CGG)$_{17}$ ds complex. In addition, putative Sp1 binding sites located in the 3'-flanking region of the genomic 5'-d(CGG)$_{18}$-3' repeat are not bound by Sp1, since the antibody against this factor did not affect the formation of any complex formed with the authentic 198ds fragment (data not shown).

Partial Purification of a Nuclear Protein (p20) Associated with the Binding to the Double-stranded 5'-d(CGG)$_{17}$-3' Repeat—CGGBP(s) participating in complex I formation were isolated from HeLa nuclear extracts by the purification scheme outlined in Fig. 4. Nuclear extracts were prepared from 2 $\times$ 10$^7$ HeLa cells, and the proteins were first fractionated by anion-exchange chromatography (Fig. 5a). Protein binding activity to the double-stranded oligodeoxyribonucleotide (CGG)$_{17}$ ds was recovered in the flow-through designated as fraction I (Figs. 4 and 5a). About 60% of unrelated proteins and nucleic acids from the nuclear extracts were eliminated in this purification step. Fraction I was then incubated in a batch procedure with the double-stranded oligodeoxyribonucleotides CGGBAd5 or (CAG)$_{17}$ ds coupled to Sepharose beads to remove proteins that bound unspecifically to DNA of similar structure (Fig. 4). CGGBP(s) were recovered almost quantitatively in the supernatant. This material was designated as fraction II. Fraction II was subsequently adsorbed to a (CGG)$_{17}$ ds-Sepharose matrix, and active fractions (fraction III) were eluted with >300 mM NaCl (Fig. 5b). After a second passage of fraction III over the (CGG)$_{17}$ ds-matrix, a major band of 20 kDa was detected in the active fraction IV (Fig. 5, b and c) by SDS-polyacrylamide gel electrophoresis followed by silver staining. The 20 kDa band was accompanied by an additional faint band of 120 kDa. In order to determine which of the two bands was responsible for specific (CGG)$_{17}$ ds binding, proteins of fraction I were bound to the mutated CGGBAd5s-matrix (Fig. 4, dashed line). The material was washed and eluted as described above. Fraction III'
eluting with buffer E750 did not show (CGG)$_{17}$-ds-binding activity (Fig. 5b). Analyses by SDS-polyacrylamide gel electrophoresis followed by silver staining revealed that this fraction contained several bands at 120, 70, and 55 kDa. However, a band around 20 kDa was not detected (Fig. 5c).

It is concluded that the protein p20 is involved in the formation of complex I and also of complex 4 established with the repetitive oligodeoxyribonucleotide (CGG)$_{17}$-ds and the authentic DNA fragment 198ds, respectively. However, participation of additional proteins in complex I and complex 4 cannot be ruled out, since their amounts might be below the detection limit of silver staining.

**DISCUSSION**

This research has been initiated on the premise that the size stability of trinucleotide repeats in the human genome and their controlled replication may be regulated by factors that are encoded at chromosomal sites far remote from the locus of the trinucleotide repeats, e.g. of the FRAXA location on Xq27.3 in the instance of the fragile X syndrome (13). Alterations in such regulatory proteins might be implicated in eliciting the repeat expansions that are causally related to a number of serious genetic diseases in humans. In addition, it needs be investigated whether the trinucleotide repeat itself might influence the regulation of the expression of adjacent genes.
Whatever the underlying mechanisms underlying these striking trinucleotide repeat amplifications or the function of the repeat itself may turn out to be, we have considered it interesting to study cellular proteins that can bind specifically to these sequences. The 5′-d(CGG)₉-3′ repeat in the 5′-untranslated region of the human FMR1 gene has been chosen as a system of considerable theoretical and medical importance.

We have partly purified a protein that is involved in specific binding to the double-stranded form of the synthetic 5′-d(CGG)₁₇-3′ repeat and its naturally occurring counterpart in the 5′-regulatory region of the human FMR1 gene. Further experiments will be focused on the isolation of a cDNA encoding this protein and on elucidating its function. Whether additional proteins are involved in complex I formation has to be investigated. However, the GC box binding protein Sp1 (43) does not participate in CGGBP(s)-(CGG)₁₇ds complex formation. This specific complex is sensitive to sodium deoxycholate treatment, and this sensitivity can be abrogated by sufficient concentrations of the nonionic detergent Nonidet P-40. This finding is indicative of a complex in which more than one protein is involved and which might be based in part on protein-protein interactions.

The protein-DNA complex investigated responds to specific 5′-d(CGG)-3′ methylation in the repeat sequences. This observation lends further credence to the biological significance of this complex formation since it has been demonstrated that in patients with the fragile X syndrome, the repeat sequence is hypermethylated (14–16). The biochemical functions of the protein(s) actually contained in the complex require further detailed analyses.

Acknowledgments—We thank Stephen T. Warren, Emory University Medical School, Atlanta, GA for a gift of the pES1 plasmid. We also thank Helmut Deissler and Hans-Christoph Kirch, Institute of Cell Biology of Molecular Biology, University of Essen, Medical School, Essen, Germany, respectively, for cell lines and valuable comments on the manuscript, and Irmgard Hölker for the synthesis of oligodeoxyribonucleotides.

Note Added in Proof—Southwestern blotting analyses with purified protein fractions III, IV, or III′ revealed that the p20 protein detected in fractions III or IV bound directly to the oligodeoxyribonucleotide (CGG)₁₇ds but not to the control oligodeoxyribonucleotide (CAG)₁₇ds. Proteins in fraction III′ exhibited only unspecific binding to several different oligodeoxyribonucleotides. These results confirmed the conclusions drawn in this report that p20 bound specifically to the trinucleotide repeat 5′-d(CGG)₁₇-3′.

REFERENCES
1. Caskey, C. T., Pizzuti, A., Fu, Y.-H., Fenwick, R. G. J., and Nelson, D. L. (1992) Science 256, 784–789
2. Richards, R. J., and Sutherland, G. R. (1992) Cell 70, 709–712
3. Riggins, G. J., Do, L., Chastain, J., Leiner, H. A., Sherman, S. L., Wilkinson, K. D., and Warren, S. T. (1992) Nature Genetics 2, 186–191.
4. Knight, S. J. L., Flannery, A. V., Hirst, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., Pointon, J., Middleton-Price, H. R., Barnicoat, A., Pembrey, M. E., Holland, J., Ostra, B. A., Babrow, M., and Davies, K. E. (1993) Cell 74, 127–134
5. Orr, H. T., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, K., Kondo, R., Ishikawa, A., Hayashi, T., Salta, M., Tomoda, A., Miki, T., Naito, H., Ikeda, F., and Tsujii, S. (1994) Nature Genet. 5, 9–13
6. Sutherland, G. R. (1979) Am. J. Hum. Genet. 31, 125–135
7. Sutherland, G. R. (1993) Hum. Mol. Genet. 2, 1659–1665
8. Sutherland, G. R. (1979) Science 208, 680–685
9. Sutherland, G. R. (1993) Am. J. Med. 95, 9–13
10. Sutherland, G. R., and Richards, R. I. (1994) Science 264, 5085–5092
11. Sutcliffe, J. S., Nelson, D. L., Ostra, B. A., and Doerfler, W. (1991) Hum. Mol. Genet. 2, 1476–1481
12. Nancarrow, J. K., Holman, K., Yu, S., and Sutherland, G. R. (1995) Hum. Mol. Genet. 4, 372–376
13. Berry, M. J., and Bird, A. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 905–914
Purification of Nuclear Proteins from Human HeLa Cells That Bind Specifically to the Unstable Tandem Repeat (CGG) in the Human FMR1 Gene
Heidrun Deissler, Annett Behn-Krapa and Walter Doerfler

J. Biol. Chem. 1996, 271:4327-4334.
doi: 10.1074/jbc.271.8.4327

Access the most updated version of this article at http://www.jbc.org/content/271/8/4327

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 43 references, 12 of which can be accessed free at http://www.jbc.org/content/271/8/4327.full.html#ref-list-1