Biochemical and Biophysical Research Communications **264,** 76–80 (1999)
Article ID bbrc.1999.1435, available online at http://www.idealibrary.com on IDEAL®

An Expanded CTG Trinucleotide Repeat Causes *trans* RNA Interference: A New Hypothesis for the Pathogenesis of Myotonic Dystrophy

Noboru Sasagawa,*,[1] Nobuhiro Takahashi,*,† Koichi Suzuki,† and Shoichi Ishiura*

**Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Tokyo 153-8902, Japan; and*
†Institute of Molecular and Cellular Biosciences, University of Tokyo, Tokyo 113-0032, Japan

Received August 14, 1999

Here we report a novel mechanism for the pathogenesis of myotonic dystrophy (DM). The DMPK mRNA with expanded CTG trinucleotide repeats interacts with other transcripts having expanded CAG repeats. This "*trans* RNA interference" occurs *in vitro* only when the number of CTG repeats is over 140 and the number of target CAG repeats exceeds 35. The *trans* RNA interference can explain all the phenomena previously reported about DM.

One pervasive mystery about DM is that the disease shows dominant inheritance although the mutation of the CTG triplet repeat expansion is in the 3′-untranslated region (3′-UTR) of the gene (1–3). We expect that the causative gene product, DMPK, has no amino acid alterations in DM patients. But this fact does not correlate with DM pathogenesis. DMPK knockout mice (4, 5) or overexpressing mice (6, 7) show less severe phenotype than DM patients. This fact indicates that the level of DMPK expression is not critical for the expression of the disease.

On the other hand, various CTG repeat expansion effects are observed in the cell. Krahe *et al.* reported that the levels of processed mRNA from the DM allele of expanded CTG repeats are reduced in comparison to normal controls (8). Wang *et al.* reported that the level of the mature poly(A)$^+$ RNA of DMPK is decreased in DM patients (9). Nuclear retention of DMPK transcripts with expanded CTG repeats has also been observed in cultured cell (10). These findings, however, are sporadic observations and their mechanisms are not totally explained.

Abbreviations used: EtBr, ethidium bromide; DM, myotonic dystrophy; DMPK, myotonic dystrophy protein kinase; ORF, open reading frame; PCR, polymerase chain reaction; UTR, untranslated region; TFIID, transcription factor IID; TBP, TATA binding protein.

[1] To whom correspondence should be addressed.

We report here that expanded CTG repeats in the DMPK transcript have a potential antisense effect on other genes with CAG triplet repeats. We previously obtained DMPK cDNA (11) and applied a method to produce artificial expanded CTG repeats (12) to obtain a DMPK cDNA with 140 CTG triplet repeats. We transcribed the mRNA from DMPK cDNAs by an *in vitro* transcription method and found that expanded CUG repeats in DMPK mRNA interact with long CAG repeats in certain mRNAs such as TFIID.

MATERIALS AND METHODS

Human TBP/TFIID and androgen receptor cDNAs were obtained by PCR cloning. Primers used for PCR were as follows; TFIID, 5′-ctggtttgccaagaagaaagtg-3′ and 5′-caacaccaccatttaaaggtacc-3′; androgen receptor, 5′-taagggaagtaggtggaagattca-3′ and 5′-aactccttggcgttgtcagaaat-3′. Primers for TBP/TFIID were designed to clone a 1139 bp fragment containing the full length ORF, and those of androgen receptor were designed to amplify a partial ORF of a 743 bp fragment in the vicinity of the CAG repeats. Our cloned human DMPK has 5 CTG repeats (11). We previously reported obtaining DMPK with CTG46 (13), CTG140, CAG140 (12), and DMPK with no CTG repeats (13). These cDNAs were ligated into pGEM7 vector (Promega). RNA transcription was achieved by an mCAP RNA capping kit (Stratagene) using T7 RNA polymerase. The transcription reaction was stopped by adding loading buffer, and then 5 μl of transcribed RNAs were immediately subjected to 1% agarose gel electrophoresis. The bands were stained with EtBr and visualized with an UV transilluminator. Polyacrylamide gel electrophoresis and silver staining were carried out using a PhastGel gel electrophoresis system (Pharmacia). Band density was analysed with an Imagemaster 2 (Pharmacia).

RESULTS AND DISCUSSION

Figure 1 shows the constructs used in these experiments. First, we selected DMPK(CTG 5) and DMPK(CTG 140) independently as templates for *in vitro* transcription. The transcribed RNAs were then analysed by agarose gel electrophoresis (1%, TAE). This revealed that the amount of transcribed mRNA did not change with the number of CTG repeats. The

0006-291X/99 $30.00

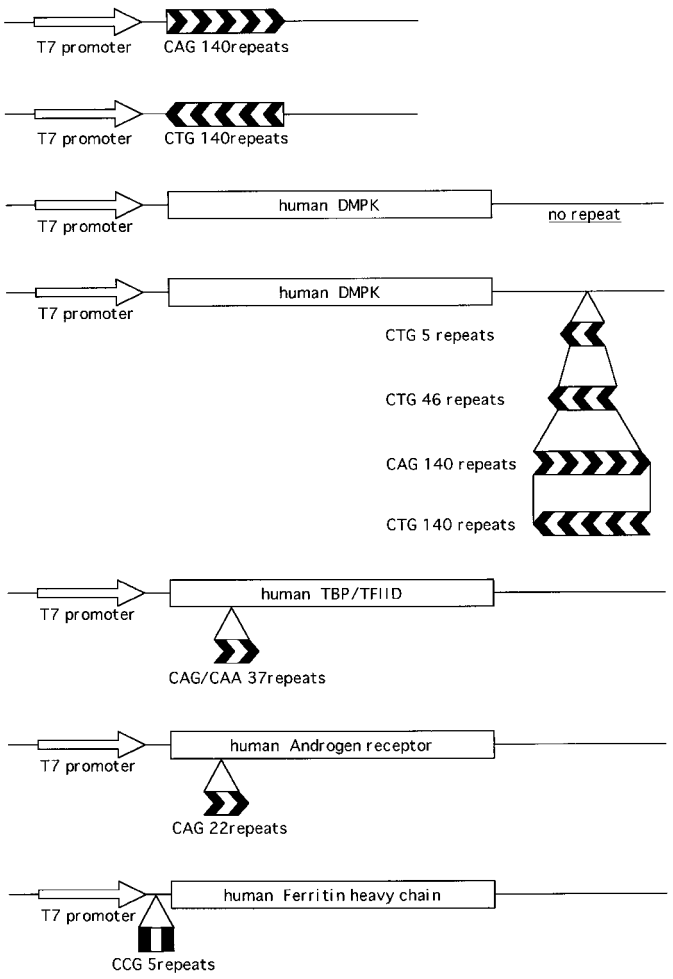

FIG. 1. Constructs used in these experiments. All constructs were ligated into pGEM7. The positions and lengths of the CAG/CTG repeats were analysed by sequence analysis at both the 5′- and 3′-sides and the fragment rate of flow on agarose electrophoresis.

transcribed RNA/template DNA ratios were 2.21 ± 0.37 (band density, CTG5) and 2.17 ± 0.36 (CTG140) (n = 17, data not shown). The identity was over 99% by Student's t test. The results correspond to those in our previous report (13).

We then cotranscribed mRNAs using DMPK(CTG 140) and CAG 140 as templates in the same test tube. We found a band shift on agarose gel electrophoresis, indicating that the RNAs anneal to one another causing a gel mobility shift. The cotranscription of DMPK(CTG140) with CTG140 did not produce a band shift, indicating that the annealing of the RNAs is in the CTG/CAG region (Fig. 2). In this experiment, CAG140 mRNA was not visualised by EtBr, possibly due to the structure of this single-strand RNA, because a band is visible on the silver-stained acrylamide gel (Fig. 2b).

We next tried to clarify whether the existing gene is the one that causes antisense annealing, because there are many genes in the human genome that have long CAG repeats (Table I). Many such genes are candidates as the causative factors in CAG triplet diseases, i.e., Huntington's disease, DRPLA, and so on. Interestingly, such genes contain many transcription factors. Among them, we cloned the TFIID/TATA binding protein because it has not been reported to be a triplet repeat disease gene and has an adequately long polyglutamine (CAG/CAA) codon. Our cloned TFIID/TATA binding protein has 37 polyglutamines, over 80% of them (31/37) CAG, and the rest CAA. We also cloned the human androgen receptor, which has 22 pure CAG repeats, but the length is shorter than the repeat region of TFIID.

We used DMPK(CTG140) and the TFIID/TATA binding protein as templates for RNA cotranscription. A band shift on agarose gel electrophoresis as seen in

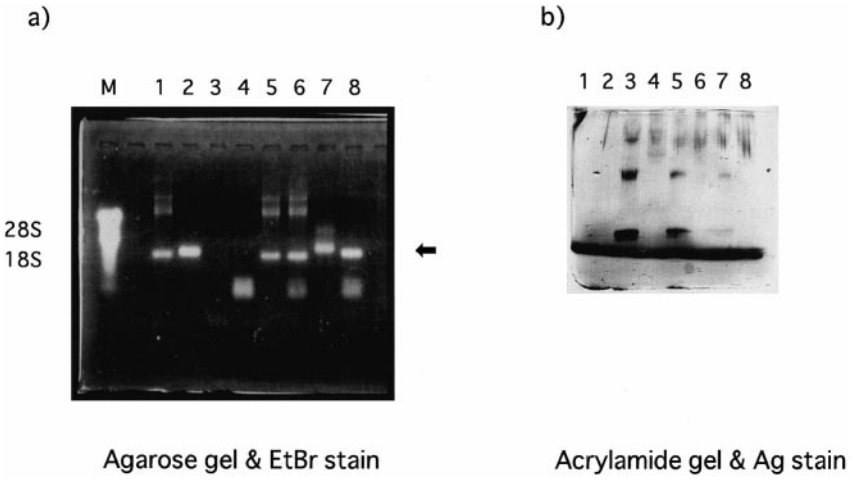

FIG. 2. RNA cotranscription with different templates. (a) agarose gel and EtBr stain, (b) acrylamide gel and silver stain. M: mouse ribosomal RNA, lane 1: DMPK(CTG5), lane 2: DMPK(CTG140), lane 3: CAG140, lane 4: CTG140, lane 5: DMPK(CTG5) + CAG140, lane 6: DMPK(CTG5) + CTG140, lane 7: DMPK(CTG140) + CAG140, lane 8: DMPK(CTG140) + CTG140. A band shift was observed in lane 7. The shifted band is indicated by the arrow.

TABLE I

Human Genes Containing Long CAG Repeats

ID	Gene	Repeat number	
gb:S82497	SCA1 = spinocerebellar ataxia type 1	44	A, B
gb:HUMMAJOPEA	Human Machado-Joseph disease (SCA3)	30	A, B
gb:HUMHUNTPEA	Human huntingtin	25	A, B
gb:HUMANREPEA	Human androgen receptor	26	A, B
gb:HSAAD20	Human mRNA for AAD20 protein (SCA2)	14	A, B
gb:HSTFIIDAA	Human transcription factor TFIID	38	A, C
gb:HSU15641	Human transcription factor E2F-4	15	C, D
gb:HUMASH1A	Human achaete scute homologous protein (ASH1)	14	A, C
gb:HSU91935	Human retina-derived POU-domain factor-1	11	A
gb:HSU38810	Human mab-21 cell fate-determining protein homolog	21	E
gb:HSMEF2	Human mRNA for myocyte-specific enhancer factor 2 (MEF2)	11	A, C
gb:HUMAF9X	Human AF-9 mRNA	44	D
gb:HSU79667	Human alpha1A-voltage-dependent calcium channel (SCA6)	12	A, B
gb:HSCANPX	Human mRNA for calpain-like protease CANPX	12	E
gb:HUMDRPLA	Human mRNA foro DRPLA protein	14	A, B
gb:HSU60325	Human DNA polymerase gamma	13	A
gb:HUMHSNF2A	Human mRNA for transcriptional activator hSNF2a	23	A, C
gb:HSHBRM	H. sapiens hbrm	23	F

Note. These genes were screened from BLAST search (National Institute for Biological Information), probing CAG10 repeats. A, polyglutamine in coding region; B, CAG triplet repeat disease; C, transcription factor; D, polyserine; E, CAG repeat at 5′-UTR; and F, DNA helicase.

Fig. 2, was observed, indicating the RNAs annealed to one another to form a double strand (Fig. 3). This band shift was not observed when DMPK(CTG46), DMPK(CTG0), and androgen receptor were used as templates. Besides, the human ferritin heavy chain, which has CCG 5 repeats in the 5′-UTR, also did not interact with CTG repeats (Fig. 4). Thus, we conclude that DMPK(CTG140) and TFIID mRNAs anneal to one another at the CUG/CAG site, and the CTG repeat in DMPK has an antisense effect only when the number of repeats is over 140. Furthermore, the boundary number of CAG repeats in the target gene is about 35–40. Only when both the CTG and CAG repeats have adequate repeat numbers does this trans antisense effect occur.

In humans, the decreased level of DMPK protein caused by the expansion of the CTG repeat has been discussed (14, 15). However, this haplo insufficiency model has not been confirmed because DMPK knockout mice do not show any typical DM symptoms, and also because various levels of DMPK mRNA have been reported (14, 16, 17). Expanded CTG repeats may affect the expression of DMAHP (18), the neighbor gene on the downstream side of DMPK. But both decrease (19, 20) and the normal expression (21) of DMAHP has been reported. Moreover, these models do not explain the abnormal processing of the DMPK mRNA.

We report in this paper that expanded CUG repeats in the DMPK mRNA form abnormal conformations

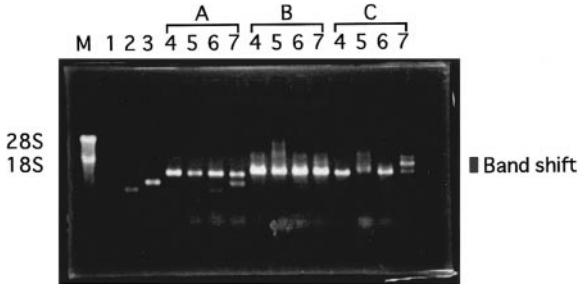

FIG. 3. Expanded CTG repeats have a trans-interference effect with existing human genes that have long CAG repeats. M: mouse ribosomal RNA, lane 1: CAG140, lane 2: androgen receptor, lane 3: TFIID, lane 4: DMPK only, lane 5: DMPK + CAG140, lane 6: DMPK + androgen receptor, lane 7: DMPK + TFIID. A, B, C means DMPK(CTG5), DMPK(CTG46), DMPK(CTG140), respectively. Shifted bands are observed in lanes C-5 and C-7.

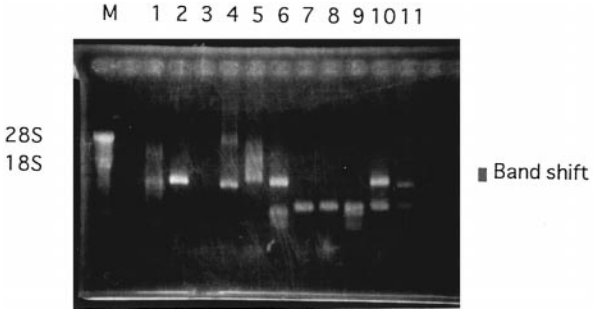

FIG. 4. CTG repeats do not interact with the CCG5 repeat in ferritin. M: mouse ribosomal RNA, lane 1: DMPK(CTG5), lane 2: DMPK(CTG140), lane 3: CAG140, lane 4: DMPK(CTG5) + CAG140, lane 5: DMPK(CTG140) + CAG140, lane 6: DMPK(CTG140) + CTG140, lane 7: ferritin, lane 8: ferritin + CAG140, lane 9: ferritin + CTG140, lane 10: ferritin + DMPK(CTG140), lane 11: ferritin + DMPK(CTG5). The band shift is observed only in lane 5.

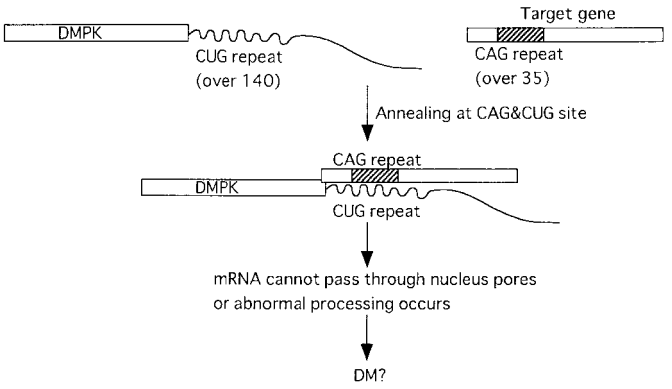

FIG. 5. Scheme for trans RNA interference deduced from these experiments. The mechanism works only when the number of CTG repeats in DMPK is over 140 and the number of CAG repeats is over 35.

with other genes with long CAG repeats. This abnormal CUG/CAG double strand RNA structure would explain the nuclear retention and altered processing of DMPK mRNA with an expanded CUG repeat: CUG/CAG double strand RNA could be recognized by a specific RNase and digested.

Koch *et al.* reported that the hairpin structure formed by long CTG repeats (n = >44) cannot pass through nucleic pores. Thus, the CUG/CAG double strand RNA with over 40 repeats found in this study may also be retained in the nucleus (22). Moreover, this trans-RNA interference can explain the multisystemic features of DM patients, because there can be multiple CAG repeat targets (vs. the CTG repeats of DMPK) in various organs and tissues.

Usuki *et al.* reported that the overexpression of DMPK with expanded CTG repeats retards muscle cell differentiation (23). Okoli *et al.* confirmed the same results that DMPK overexpression leads to the inhibition of cell differentiation (24). On the other hand, Sabourin *et al.* reported that the 3′-UTR of DMPK itself inhibits myoblast differentiation (25). So, the molecular pathogenesis of DM should occur at the transcription/translation stage, not simply at the protein level. Also, DMPK knockout (4, 5) and overexpressing mice (6, 7) do not show any of the typical symptoms seen in DM patients, indicating that these symptoms occur only in humans. Interestingly, as far as we know, the polyglutamine repeats in TBP are unique to humans. Although TBP is well-conserved from yeast to humans, only the human protein has polyglutamine repeats in it. This fact may explain why these genetically-engineered mouse models of DM show no DM-like symptoms.

In conclusion, we propose a new model of trans RNA interference by expanded CTG triplet repeats in DM patients (Fig. 5). The expanded CUG trinucleotide repeats in the DMPK mRNA may anneal to other mRNAs with sufficiently long stretches of CAG repeats, resulting in abnormal RNA processing and localization in the cell.

ACKNOWLEDGMENTS

This work was supported by research grants from the Ministry of Health and Welfare, Japan, the Naito Foundation, and the Human Science Foundation.

REFERENCES

1. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P. S., Shaw, D. J., and Housman, D. E. (1992) *Cell* **68,** 799–808.
2. Fu, Y. H., Pizzuti, A., Fenwick, R. G Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., Wieringa, B., Korneluk, R., Perryman, M. B., Epstein, H. F., and Caskey, C. T. (1992) *Science* **255,** 1256–1258.
3. Harley, H. G., Brook, J. D., Rundle, S. A., Crow, S., Reardon, W., Buckler, A. J., Harper, P. S., Housman, D. E., and Shaw, D. J. (1992) *Nature* **355,** 545–546.
4. Jansen, G., Groenen, P. J., Bachner, D., Jap, P. H., Coerwinkel, M., Oerlemans, F., van den Broek, W., Gohlsch, B., Pette, D., Plomp, J. J., Molenaar, P. C., Nederhoff, M. G., van Echteld, C. J., Dekker, M., Berns, A., Hameister, H., and Wieringa, B. (1996) *Nat. Genet.* **13,** 316–324.
5. Reddy, S., Smith, D. B., Rich, M. M., Leferovich, J. M., Reilly, P., Davis, B. M., Tran, K., Rayburn, H., Bronson, R., Cros, D., Balice-Gordon, R. J., and Housman, D. (1996) *Nat. Genet.* **13,** 325–335.
6. Gourdon, G., Radvanyi, F., Lia, A. S., Duros, C., Blanche, M., Abitbo, M., Junien, C., and Hofmann-Radvanyi, H. (1997) *Nat. Genet.* **15,** 190–192.
7. Monckton, D. G., Coolbaugh, M. I., Ashizawa, K. T., Siciliano, M. J., and Caskey, C. T. (1997) *Nat. Genet.* **15,** 193–196.
8. Krahe, R., Ashizawa, T., Abbruzzese, C., Roeder, E., Carango, P., Giacanelli, M., Funanage, V. L., and Siciliano, M. J. (1995) *Genomics* **28,** 1–14.
9. Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R. C., Angelini, C., and Hoffman, E. P. (1995) *Hum. Mol. Genet.* **4,** 599–606.
10. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., and Housman, D. E. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 7388–7393.
11. Sasagawa, N., Sorimachi, H., Maruyama, K., Arahata, K., Ishiura, S., and Suzuki, K. (1994) *FEBS Lett.* **351,** 22–26.
12. Takahashi, N., Sasagawa, N., Suzuki, K., and Ishiura, S. (1999) *Neurosci. Lett.* **262, 45–48.**
13. Sasagawa, N., Saitoh, N., Shimokawa, M., Sorimachi, H., Maruyama, K., Arahata, K., Isiura, S., and Suzuki, K. (1996) *Biochim. Biophys. Acta* **1315, 112–116.**
14. Fu, Y. H., Friedman, D. L., Richards, S., Pearlman, J. A., Gibbs, R. A., Pizzuti, A., Ashizawa, T., Perryman, M. B., Scarlato, G., Fenwick, R. G., and Caskey, C. T. (1993) *Science* **260,** 235–238.
15. Koga, R., Nakao, Y., Kurano, Y., Tsukahara, T., Nakamura, A., Ishiura, S., Nonaka, I., and Arahata, K. (1994) *Biochem. Biophys. Res. Commun.* **202,** 577–585.
16. Carango, P., Noble, J. E., Marks, H. G., and Funanage, V. L. (1993) *Genomics* **18,** 340–348.
17. Sabouri, L. A., Mahadevan, M. S., Narang, M., Lee, D. S., Surh, L. C., and Korneluk, R. G. (1993) *Nat. Genet.* **4,** 233–238.
18. Boucher, C. A., King, S. K., Carey, N., Krahe, R., Winchester, C. L., Rahman, S., Creavin, T., Meghji, P., Bailey, M. E.,

Chartier, F. L., Brown, S. D., Siciliano, M. J., and Johnson, K. J. (1995) *Hum. Mol. Genet.* **4,** 1919–1925.

19. Klesert, T. R., Otten, A. D., Bird, T. D., and Tapscott, S. J. (1997) *Nat. Genet.* **16,** 402–406.
20. Thornton, C. A., Wymer, J. P, Simmons, Z., McClain, C., and Moxley, R. T. 3rd (1997) *Nat. Genet.* **16,** 407–409.
21. Hamshere, M. G., Newman, E. E., Alwazzan, M., Athwal, B. S., and Brook, J. D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 7394–7399.
22. Koch, K. S., and Leffert, H..L. (1998) *J. Theor. Biol.* **192,** 505–514.
23. Usuki, F., Ishiura, S., Saitoh, N., Sasagawa, N., Sorimachi, H., Kuzume, H., Maruyama, K., Terao, T., and Suzuki, K. (1997) *Neuroreport* **8,** 3749–3753.
24. Okoli, G., Carey, N., Johnson, K. J., and Watt, D. J. (1998) *Biochem. Biophys. Res. Commun.* **246,** 905–911.
25. Sabourin, L. A., Tamai, K., Narang, M. A., and Korneluk, R. G. (1997) *J. Biol. Chem.* **272,** 29626–29635.