CTG Triplet Repeats from Human Hereditary Diseases Are Dominant Genetic Expansion Products in Escherichia coli

(Received for publication, October 10, 1995, and in revised form, November 29, 1995)

Keiichi Ohshima, Seongman Kang, and Robert D. Wells

From the Institute of Biosciences and Technology, Center for Genome Research, Texas A & M University, Texas Medical Center, Houston, Texas 77030-3303

The relative ability of the 10 triplet repeat sequences to be expanded in Escherichia coli was determined. Surprisingly, CTG tracts are expanded at least 8 times more frequently than any of the other nine triplets. Low levels of expansion were found also for CGG, GTC, and GTC. Thus, the structure of the CTG repeats and/or their utilization by the DNA synthetic systems in vivo must be quite different from the other triplets. These data further validate this genetically defined system for elucidating molecular mechanisms of expansion and may explain why most triplet repeat hereditary neuromuscular and neurodegenerative disease genes contain CTG repeats.

The molecular basis of several human genetic disorders involving CTG triplet repeat expansions including myotonic dystrophy (1, 2), Kennedy's disease (3), spinocerebellar ataxia type 1 (4), Huntington's disease (5), dentatorubral-pallidoluysian atrophy/Haw River syndrome (6–8), and Machado-Joseph disease (9) has been partially established. The CTG triplet repeats may be within the genes or in the 3' untranslated regions. The CTG expansion is associated with anticipation whereby the penetrance of the disease is increased in successive generations. Thus, this expansion process, possibly due to the slippage inefficiency: 

\[ \text{slippage} = 1 \times 10^{-8} \text{transfers/µg} \]

The transformants were grown in LB media containing ampicillin (150 µg/ml) and plasmids were isolated by the alkaline lysis method (20). To amplify the expanded components, the above experiments were repeated. After transformation, the cells were spread on ampicillin plates. Colonies were picked from the plates and grown in LB media (containing 75 µg/ml ampicillin). The plasmids were isolated, and the population of triplet repeats was determined by restriction enzyme digestions and/or DNA sequencing.

E. coli Strains—HB101 strain is F', mcrA44, supE44 galK2 lacY1 proA2 rpsL20(smr) xylS xne mtl1.

1 K. Ohshima, M. Shimizu, R. Gellibolian, S. Kang, and R. D. Wells, manuscript in preparation.

2 M. Shimizu, R. Gellibolian, and R. D. Wells, manuscript in preparation.

3 K. Ohshima, S. Kang, J. E. Larson, and R. D. Wells, manuscript in preparation.

4 The abbreviation used is: bp, base pair(s).

*This work was supported by National Institutes of Health Grants GM52982, National Science Foundation Grant DMB-9103942, and a grant from the Robert A. Welch Foundation. 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: Laboratory of Genetic Disease Research, National Center of Human Genome Research, National Institutes of Health, Bldg. 49, Bethesda, MD 20892.

§ To whom correspondence should be addressed: Institute of Biosciences and Technology, Center for Genome Research, Texas A & M University, Texas Medical Center, 2121 W. Holcombe Blvd., Houston, TX 77030-3303. Tel.: 713-677-7651; Fax: 713-677-7689; E-mail: rwells@bt.tamu.edu.

© 1996 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.
TABLE I
Plasmids used in this study

The number of triplet repeats and the length of the SacI-HindIII fragments are listed. After digestion with SacI and HindIII, the fragments containing the triplet repeats were between 179 and 208 bp in length for the shorter triplet repeats and between 245 and 265 bp in length for the fragments with the longer triplet repeats.

| Generic sequence | Plasmid | Actual insert sequence | Length | Repeats |
|------------------|---------|------------------------|--------|---------|
| Shorter triplet repeats | CTG | pRW3293 | (CTG)$_{57}$ | 203 | 57 |
| | CGG | pRW3106 | (CGG)$_{11}$AGG(CGG)$_{8}$ | 197 | 62 |
| | GAT | pRW3167 | (GAT)$_{10}$GT(GAT)$_{8}$AT(GAT)$_{5}$ | 200 | 62 |
| | GTA | pRW3157 | (GTA)$_{10}$ATA(GTA)$_{8}$ | 206 | 60 |
| | GTG | pRW3421 | (GTG)$_{12}$ATG(GTG)$_{48}$ | 196 | 61 |
| | GTC | pRW3173 | (GTT)$_{11}$ATT(GTT)$_{14}$ | 189 | 57 |
| | GTT | pRW3415 | (GTC)$_{2}$GT(GTC)$_{34}$ | 208 | 61 |
| | AAG | pRW3439 | (AAG)$_{11}$AA(AAG)$_{2}$AGG(AAG)$_{19}$ | 193 | 57 |
| | CCT | pRW3195 | (CCT)$_{15}$ | 179 | 53 |
| | TTA | pRW3314 | (TTA)$_{9}$TTC(TTA)CTA(TTA)$_{2}$TAAT | 185 | 45 |

- **Longer triplet repeats**

| Repeat | Plasmid | Actual insert sequence | Length | Repeats |
|--------|---------|------------------------|--------|---------|
| CTG | pRW3292 | (CTG)$_{57}$ | 245 | 71 |
| CGG | pRW3311 | (CGG)$_{11}$AGG(CGG)$_{8}$ | 245 | 79 |
| GAT | pRW3441 | (GAT)$_{10}$GT(GAT)$_{8}$AT(GAT)$_{14}$ | 265 | 81 |
| GTA | pRW3451 | (GTA)$_{10}$ATA(GTA)$_{8}$ | 257 | 76 |
| GTG | pRW3428 | (GTG)$_{12}$ATG(GTG)$_{34}$ | 247 | 78 |
| TTA | pRW3418 | (TTA)$_{10}$TTC(TTA)CTA(TTA)$_{2}$TAAT | 254 | 68 |

- **a** Size (bp) of SacI and HindIII fragment.
- **b** Number of triplet repeats.

DH10B is F− mcrA Δ(mrr-hsdRMS-mcrBC) Δ80lacZAM15 ΔlacX74 deoR recA1 ara139 Δ(ara, leu)769 galU galK λ− rpsL endA1 nupG.

RESULTS AND DISCUSSION

For this study, a mixture of plasmids containing all 10 triplet repeats (Table I) were digested with SacI and HindIII and the products were analyzed on an agarose gel; all inserts were cloned into the polylinker of pUC19. In the case of the shorter triplet repeats (Table I), only fragments containing 45–62 repeats were observed along with the vector. However, areas of the gel that were larger (by approximately 50 bp) than this fragment that did not contain visually detectable DNA were eluted and the putative fragments were recloned back into pUC19. The sizes of the “recloned” inserts were determined by dideoxy sequencing and by restriction enzyme analysis.

When all 10 of the plasmids containing the shorter triplet repeats (Table I) were mixed in equimolar amounts and investigated by this method, we surprisingly found (Fig. 1A) that virtually all of the colonies contained CTG repeats (38 of the 43 colonies investigated, 88.4%). In addition, a smaller number of these colonies contained CGG (7%) and GTC (4.7%) repeats (Fig. 1A). Expansion of the other repeats was not observed.

As a control, we eluted the DNA from the major observable band on the gel at the expected repeat size (containing all 10 triplet sequences) and determined the relative ability of each of the 10 triplet repeat inserts to be recloned; the inset in Fig. 1A shows that all triplet repeats were recloned in approximately equal proportion indicating that the frequencies of ligation and/or transformation were not influenced by the triplet repeat sequences.

Other studies were conducted in the absence of the CTG triplet repeat (Fig. 1B) to more rigorously analyze the extent of expansion of the other nine triplet repeat sequences. Fig. 1B shows that CGG (32.3%), GTG (29.0%), and GTC (35.5%) were also expanded, whereas the other repeats were expanded to the extent of 3% or less. The reason for observing the expansion of GTG in this study but not in the experiment shown in Fig. 1A is uncertain. Again, a control experiment of recloning the principal observable band at the expected length of inserts (Fig. 1B, inset) showed that all of the nine triplet repeat sequences could be recloned, eliminating the potential artifact of problems in ligation, transformation, or other steps during the process.

Additionally, studies were conducted with the triplet repeats containing 68–81 repeats (Table II). Similar results were found as for the shorter triplet repeats (Fig. 1A). Although the expansion frequency is larger with increasing repeat numbers (19),
GAT, GTA, and TTA repeat expansions were still not observed.

To further evaluate the extraordinary finding that CTG is the dominant expansion product, we performed direct comparative analyses of this triplet repeat with its sequence isomer, GTC. Table III shows that both the 57-mer and the 71-mer of CTG were expanded much more readily than similar lengths of GTC. Both triplet repeats have the same base composition but are sequence isomeric. Hence, the type or stability of the structure of the CTG repeats or the manner in which this sequence is utilized by the DNA synthetic enzymes in vivo must be very different from the GTC repeat.

GTC repeats have different stacking free energies from CTG repeats, which result in less stable single-strand hairpin structures (22, 23). We proposed that expansion occurs in vivo during DNA replication due to the formation of a region of single-stranded CTG repeats (19). This hypothesis was based on the relative rate of expansion versus deletion as a function of the direction of DNA replication (19) as well as on oligonucleotide stability studies as determined by NMR spectroscopy (24, 25). CTG oligomers form antiparallel helical duplexes with the formation of T-T base pairs, whereas CAG oligomers form only metastable duplexes (25). Also, computer stability analyses (22, 24) infer that single-stranded CTG repeats as well as CAG repeats form antiparallel hairpin structures that are more stable than any of the other repeats. Our model (19) requires that, during the lagging strand replication, a stable hairpin may be formed in the newly synthesized DNA strand to achieve expansion, whereas a stable hairpin structure in the lagging strand template is likely to produce deletions.

A consideration of the effect of interruptions (polymorphisms) in the repeats is important since they have been statistically correlated with the stability of triplets; the loss of interruptions increases the genetic instability (26, 27). These interruptions consisted of AGG repeats within the CGG triplets for the FMR-1 gene and CAT polymorphisms in the CAG repeats for the SCA1 gene. The majority of the triplet repeats studied herein (Table I) did contain interruptions. We have evaluated the effect of interruptions on the frequency of expansion.  

### Table II

| Sequence | With CTG insert present | No CTG insert present |
|----------|------------------------|-----------------------|
|          | No. of clones | Frequency | No. of clones | Frequency |
| CTG      | 74           | 88.1       | NA           | NA        |
| CAG      | 0            | 0          | 0            | 0         |
| GAT      | 0            | 0          | 0            | 0         |
| GTA      | 0            | 0          | 0            | 0         |
| GTG      | 10           | 11.9       | 11           | 61.1      |
| TTA      | 0            | 0          | 0            | 0         |

### Table III

| Sequence | Plasmid  | Expanding size | Not expanding size |
|----------|----------|----------------|--------------------|
|          | No. of clones | Frequency | No. of clones | Frequency |
| 57 repeats |          |              |                  |
| (CTG)$_{57}$ | pRW3293 | 29          | 93.5            | 6       | 35.3 |
| (GTC)$_{57}$ | pRW3464 | 2           | 6.5             | 11      | 64.7 |
| 71 repeats |          |              |                  |
| (CTG)$_{71}$ | pRW3292 | 22          | 81.5            | 9       | 53.0 |
| (GTC)$_{71}$ | pRW3465 | 5           | 18.5            | 8       | 47.0 |

At present, the genetic functions responsible for the expansions and deletions are uncertain; the establishment of a genetic system$^6$ will be important for these studies.

To date, only two triplet repeat sequences, CTG (CAG) and CGG (CCG), have been associated with human genetic diseases due to their massive expansions (1–11, 17, 18). Herein, we have shown that GTG and GTC repeats also have the capacity to be expanded, at least in E. coli. Although other long triplet repeats have been found in the human genome (28, 29) in addition to CTG and CGG repeats, their expansion has not been correlated with human hereditary diseases. Since GTG repeats have similar base composition characteristics to CGG and CTG repeats (30), our determinations make this a likely candidate for disease involvement. The close relationship between our studies in E. coli (expansion-deletion data (19), DNA polymerase pausing (31), and mismatch repair (32)) and eucaryotic observations (see Refs. 19, 31, and 32) infers that the dominant expansion of CTG described herein may reveal significant molecular insights into human systems.

The surprising discovery that CTG triplet repeats are the dominant expansion products in E. coli, as found (1–11, 17, 18) in clinical samples from human hereditary diseases, suggests the importance of DNA structural properties. Other investigations have revealed that duplex CTG and CGG repeats have unorthodox properties, including nucleosome assembly (33), their capacity to cause DNA polymerases to pause within the repeat sequences (31), as well as conformational features as revealed by helical repeat and polyacrylamide gel migrations (34).$^7$ The further elucidation of the involvement of DNA conformational features in the etiology of human

---

$^5$ S. Kang, K. Ohshima, A. Jaworski, and R. D. Wells, manuscript in preparation.

$^6$ P. Parniewski, R. P. Bowater, R. R. Sinden, and R. D. Wells, manuscript in preparation.

$^7$ R. Gellibold, M. Shimizu, S. Amirhaeri, S. Kang, K. Ohshima, J. E. Larson, Y.-H. Fu, C. T. Caskey, B. A. Oostra, and R. D. Wells, manuscript in preparation.
hereditary neuromuscular and neurodegenerative diseases will be intriguing.

Acknowledgments—We thank Drs. Adam Jaworski and Richard P. Bowater for helpful discussions and Ela Klysik for technical assistance.

REFERENCES

1. Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G., Neville, C., Narang, M., Barceló, J., O'Hoy, K., Leblond, S., Earle-MacDonald, J., de Jong, P. J., Wieringa, B., and Korneluk, R. G. (1992) Science 255, 1253–1255
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. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., and Fischbeck, K. H. (1993) Nature 362, 77–79
4. Orr, H. T., Chung, M.-Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Yu, A., Dill, J., Kamp, T. J., Servadio, A., Beaudet, A. L., McCull, A. E., Duvick, L. A., Ranum, L. P. W., and Zoghbi, H. Y. (1993) Nat. Genet. 4, 221–226
5. The Huntington’s Disease Collaborative Research Group (1993) Cell 72, 971–983
6. Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., Salto, M., Tomoda, A., Mikes, T., Naoto, H., Ikuta, F., and Tsuji, S. (1994) Nat. Genet. 6, 9–13
7. Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, H., Ohsaki, E., Bundo, M., Kadoya, T., Sato, M., Tomoda, A., Mikes, T., Naoto, H., Ikuta, F., and Tsuji, S. (1994) Nat. Genet. 6, 13–20
8. Burke, J. R., Wingfield, M. S., Lewis, K. E., Huang, G., Lee, V. H., Haworth, I. S., and Mitas, M. (1995) Nucleic Acids Res. 23, 2604–2606
9. Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994) Nucleic Acids Res. 22, 2213–2218
10. Nelson, D. L. (1993) Genome Analysis: Genome Rearrangement and Stability (Petersen, G. M., Kinzler, K. W., Vogelstein, B., and delachapelle, A., eds) Vol. 7, pp. 1–24, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
11. Wieringa, B. (1994) Hum. Mol. Genet. 3, 1–7
12. Weber, J. L. (1990) Genomics 5, 252–258
13. Aaltonen, L. A., Peltomäki, P., Leach, F. S., Sistonen, P., Pykkänen, L., Mecklin, J.-P., Järvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Nykänen, G. M., Kizior, K. W., Vogelstein, B., and de la Chapelle, A. (1993) Science 260, 812–816
14. Thibodeau, S. N., Bren, G., and Schaid, D. (1993) Science 260, 816–819
15. Ionov, Y., Peinado, M. A., Makowsky, S., Shibata, D., and Perudo, M. (1993) Nature 363, 558–561
16. Wooster, R., Cleon-Janson, A.-M., Collins, N., Magoni, J., Cornelis, R. S., Cooper, C. S., Gusterson, B. A., Ponder, B. A. J., von Demling, A., Wietzler, O. D., Cornish, C. J., Devilee, P., and Stratton, M. R. (1994) Nature Genet. 6, 152–156
17. Verkerk, A. J. M. H., Pijlman, G., Sutcliffe, J. S., Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F., Eussen, B. E., van Ommeren, G. J. B., Blonden, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991) Cell 65, 905–914
18. Knight, S. J. L., Flannery, A. V., Hird, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., Pionton, J. L., Miller-Price, H. R., Barnicoat, A., Pembrey, M. E., Holland, J. L., Oostra, B. A., Bobrow, M., and Davies, K. E. (1993) Cell 74, 127–134
19. Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995) Nature Genet. 10, 213–218
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) Nucleic Acids Res. 16, 6127–6145
22. Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., and Haworth, I. S. (1995) Nucleic Acids Res. 23, 1050–1059
23. Yu, A., Dill, J., Wirth, S. S., Huang, G., Lee, V. H., Haworth, I. S., and Mitas, M. (1995) Nucleic Acids Res. 23, 2705–2714
24. Gao, A. M., Goelinger, G. J., uranic, N., Macura, S., and McMurray, C. T. (1995) Cell 81, 533–540
25. Smith, G. K., Jie, J. J., Fox, G. E., and Gao, X. (1995) Nucleic Acids Res. 23, 4303–4311
26. Chung, M.-Y., Ranum, L. P. W., Duvick, L. A., Servadio, A., Zoghbi, H. Y., and Orr, H. T. (1993) Nat. Genet. 5, 254–258
27. Eichler, E. E., Holden, J. J. A., Popovich, B. W., Reiss, A. L., Snow, K., Thibodeau, S. N., Richards, C. S., Ward, P. A., and Nelson, D. L. (1994) Nat. Genet. 8, 18–94
28. Bedford, M. J., and Weber, J. L. (1992) Genomics 12, 627–631
29. Lidblad, K., Zander, C., Schalling, M., and Hudson, T. (1994) Nat. Genet. 7, 124
30. Huang, H., Sui, C., Zhu, Z., Longshore, J. W., and Finley, W. H. (1994) Nucleic Acids Res. 22, 1735–1740
31. Kang, S., Ohshima, K., Shimizu, M., Anirhaeri, S., and Wells, R. D. (1995) J. Biol. Chem. 270, 27014–27021
32. Jaworski, A., Roche, W. A., Dill, J., Kamp, T. J., Jenkins, Z., Kizior, K. W., Vogelstein, B., and de la Chapelle, A. (1993) Science 260, 812–816
33. Wang, Y.-H., Anirhaeri, S., Kang, S., Wells, R. D., and Griffith, J. D. (1994) Science 265, 669–671
34. Chastain, P. D., II, Eichler, E. E., Kang, S., Nelson, D. L., Levene, S. D., and Sinden, R. R. (1995) Biochemistry 34, 16125–16131
CTG Triplet Repeats from Human Hereditary Diseases Are Dominant Genetic Expansion Products in *Escherichia coli*

Keiichi Ohshima, Seongman Kang and Robert D. Wells

*J. Biol. Chem.* 1996, 271:1853-1856.
doi: 10.1074/jbc.271.4.1853

Access the most updated version of this article at http://www.jbc.org/content/271/4/1853

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 33 references, 7 of which can be accessed free at http://www.jbc.org/content/271/4/1853.full.html#ref-list-1