helpful discussions. Supported by a grant from the International Human Frontier Science Program and by NSF grant DCB-9025115 (N.H.), Core grant CA14051 to the Center for Cancer Research at MIT, NIH grants HL-01855 (J.C.B.) and HD-20034 (T.F. and J.-K.Y.), NCI grant CA-58317 (T.F.), and a Stern Foundation grant (T.F.). Additional funding was contributed to N.H. by Mr. A. C. Merrill. S. Lin is supported by a fellowship from the American Cancer Society, Massachusetts Division. N. Gaiano was supported by a predoctoral training grant from the NIH.

7 April 1994; accepted 14 June 1994

Preferential Nucleosome Assembly at DNA Triplet Repeats from the Myotonic Dystrophy Gene

Yuh-Hwa Wang, Sorour Amirhaeri, Seongman Kang, Robert D. Wells, Jack D. Griffith*

The expansion of CTG repeats in DNA occurs in or near genes involved in several human diseases, including myotonic dystrophy and Huntington's disease. Nucleosomes, the basic structural element of chromosomes, consist of 146 base pairs of DNA coiled about an octamer of histone proteins and mediate general transcriptional repression. Electron microscopy was used to examine in vitro the nucleosome assembly of DNA containing repeating CTG triplets. The efficiency of nucleosome formation increased with expanded triplet blocks, suggesting that such blocks may repress transcription through the creation of stable nucleosomes.

Several human genetic diseases have been traced to the expansion of blocks of repeating nucleotide triplets. The expansion of a CGG repeat was first observed in the 5′-untranslated region of the *FMR-1* gene, responsible for fragile X syndrome (*1*). Subsequently, five more disorders were linked to expanded triplet blocks: a CTG (or CAG) repeat was found in the 3′-untranslated region of the myotonic dystrophy gene (*2*) and in the protein-coding regions of the genes for Kennedy's disease (*3*), Huntington's disease (*4*), spinocerebellar ataxia type 1 (*5*), and dentatorubralpallidoluysian atrophy (*6*). Although the repeat blocks occur at different locations relative to the coding sequences, a correlation between the length of the triplet block, the severity of the diseases, and the age of onset exists in all six disorders. Furthermore, a phenomenon called genetic anticipation has been observed in which the severity of the disease increases with succeeding generations (*4–7*). When the expression of a human myotonic dystrophy protein kinase gene containing an expanded CTG triplet block was examined in the absence of the normal allele, no mRNA from the kinase gene was detected (*8*). Furthermore, the threshold for the onset of the phenotype of the triplet diseases frequently approximates the amount of DNA (146 base pairs) in a nucleosome. These observations suggest that DNA that contains long triplet repeats might form stable nucleosomes that repress transcription.

To examine this hypothesis, we carried out in vitro nucleosome reconstitution experiments, starting from high salt buffer. This method has been used by researchers to examine DNA sequence–directed nucleosome positioning in vitro (*9–11*). We used electron microscopy (EM) in our analysis to create a map showing the relative affinity for nucleosome formation along plasmid DNAs containing CTG triplet blocks.

Numerous DNA structural transitions are facilitated by negative supercoiling or influenced by heating in the presence of Mg. A Bluescript-based plasmid, pSH2, isolated from *Escherichia coli*, with a negative superhelical density of −0.06 and containing 130 CTG repeats derived from an individual with myotonic dystrophy (*12*) was first treated with Mg and heat and then reconstituted with a limiting amount of histones. We then prepared the DNA for EM by cutting it with restriction enzymes and labeling one end with streptavidin (Fig. 1). This placed the CTG triplets in pSH2 between 32 and 43 map units along the DNA (*12*).

Analysis of the location of 100 single nucleosomes (Fig. 2A) revealed that 48% of all the nucleosomes were present in the region between 30 and 45 map units. In contrast, the data for the vector alone (Fig. 2B) showed a uniform distribution of nucleosomes over the length of the DNA with only 10% of the nucleosomes present in the region between 30 and 45 map units. When linear (Fig. 2C) or supercoiled (Fig. 2D) pSH2 DNA that had not first been treated with heat or Mg was reconstituted, in both cases 29% of all the nucleosomes were localized to the triplet repeat. The data suggest that primary DNA structure is the major factor favoring nucleosome positioning over the triplets and that supercoiling or exposure to heat and Mg is less effective.

In all triplet repeat–related diseases, the severity of the disease increases with the length of the triplet block. To explore the relation between the length of the triplet block and the efficiency of nucleosome assembly, we used six pUC19-based plasmids (treated with heat and Mg) containing 26 to 250 contiguous CTG triplet repeats [$(CTG)_{26}$ to $(CTG)_{250}$] (*12*) as templates. Nucleosome assembly was found to be two- to fivefold more likely over the triplet blocks than over the adjacent flanking regions (Fig. 3).

When the values were normalized (Fig. 4), we observed that as the number of the CTG repeats increased, the efficiency of nucleosome formation at the repeats also increased (Fig. 4), up to a repeat size of 180. For the DNAs with a $(CTG)_{250}$ repeat, two side-by-side nucleosomes were occasionally observed (Fig. 1C). Plasmids pSH2 and pRW1981 (*12*) contain the same $(CTG)_{130}$ repeat cloned into different vectors. Nonetheless, they showed similar efficiencies of nucleosome formation (Figs. 2 and 3), which suggests that the efficiency of nucleosome assembly is independent of the nature of the flanking sequences.

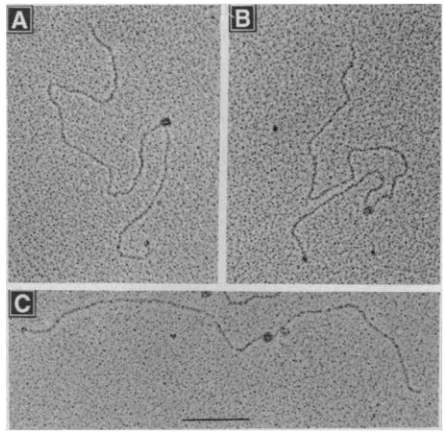

Fig. 1. Visualization of nucleosomes assembled on streptavadin-labeled pSH2 DNA containing a $(CTG)_{130}$ repeat (**A** and **B**) and pRW3222 containing a $(CTG)_{250}$ repeat (**C**), derived from individuals with myotonic dystrophy. The DNAs (curved filaments) contain streptavadin protein bound to one end [small particles at bottom in (A) and (B), and at left in (C)] and one or two nucleosomes assembled near the center of each DNA. In the absence of added histone, no nucleosome-like objects were observed. Reconstitution of the DNA with histones was as described (*16*) and preparation for EM, including rotary shadowcast with tungsten, was also as described (*17*). Bar, 100 nm.

Y.-H. Wang and J. D. Griffith, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA.
S. Amirhaeri, S. Kang, R. D. Wells, Institute of Biosciences and Technology, Texas A&M University, Houston, TX 77030, USA.
*To whom correspondence should be addressed.

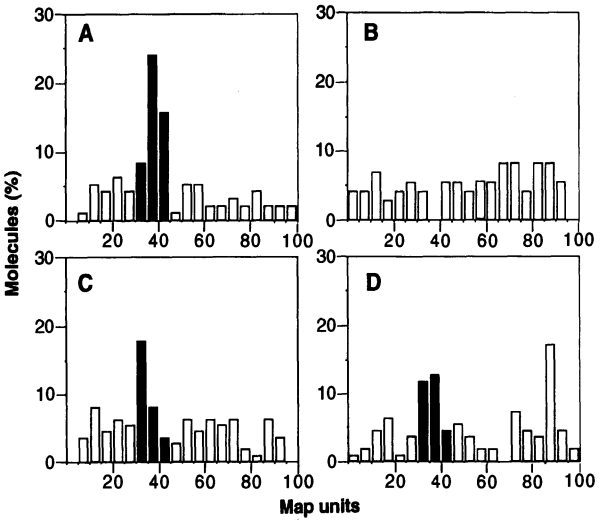

Fig. 2. Distribution of nucleosomes reconstituted on (**A**) closed, circular pSH2 DNA, which contains the $(CTG)_{130}$ repeat; (**B**) closed, circular Bluescript (Stratagene) vector without an insert; (**C**) linear pSH2 without Mg and heat treatment; and (**D**) closed, supercoiled pSH2 without Mg and heat treatment. Reconstitution and preparation for EM were performed as described (*16, 17*). At least 100 DNA molecules with single nucleosomes were photographed, and the position of each nucleosome from the streptavidin-labeled end was measured. We generated a histogram (with DNA length broken into 20 segments and each percentage of length equivalent to 1 map unit) showing the location of the nucleosomes along the DNA. The black bars indicate the positions of the CTG repeat sequences along the DNA. An additional peak observed between 85 and 90 map units in (D) maps to the 2012 to 2184 base pair region of the Bluescript vector, which contains a sequence-directed bend (*18*). Regions with no bars indicate that no nucleosomes mapped to that segment in this experiment.

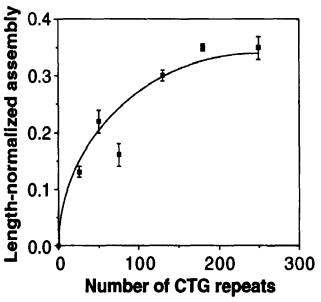

Fig. 4. Strength of sequence-directed nucleosome positioning correlates with the size of the repeated CTG block. Data from Fig. 3 were normalized as described (*20*).

We have shown that repeating CTG triplet blocks can create strong nucleosome positioning signals. Quantitative competitive reconstitution experiments (*13*) have shown that the $(CTG)_{130}$ triplet block is several times stronger than the *Xenopus borealis* 5S RNA gene, formerly the strongest known natural element. Studies of the 5S gene and other elements (*10*) have concluded that whereas bent DNA and DNA containing anisotropically flexible wedges preferentially assemble into nucleosomes, left handed Z-DNA and DNA containing polyadenine tracts are relatively resistant to nucleosome assembly. Although there can be a 100-fold difference in binding free energy between the strongest and weakest elements, the total energy difference measured in high salt buffers is relatively low (3 kcal/mol) and once assembled, all the various DNAs appear to adopt the same structure (*10, 11*). In the cell, at physiological ionic strength, however, these binding differences may be large enough to have biological effects.

The CTG inserts do not show electrophoretic retardations (*14*) or contain anisotropically flexible wedges. Thus, the reason why an extended CTG repeat would attract nucleosomes is unclear but may be related to long-range DNA structural effects (*15*). In vivo, primary DNA structure, histone H1, nucleosome assembly factors, and nonhistone proteins all participate in the positioning of nucleosomes. Thus, nucleosome positioning over repeating triplet blocks needs to be studied in vivo.

Frequently, the triplet-related disease symptoms become apparent when the triplet blocks expand from sizes that are a fraction of a nucleosome ($n \leq 30$) to sizes encompassing one or more nucleosomes ($n \geq 50$). The reason for this threshold has been unclear. The results reported here suggest that the generation of nucleosome positioning signals could have biological consequences, inducing an alteration of the local chromatin structure, inhibiting the passage of transcription complexes, or preventing the opening of the DNA before replication. These signals might also result in DNA polymerase pausing, slippage, or idling, leading to expansion of the triplet block.

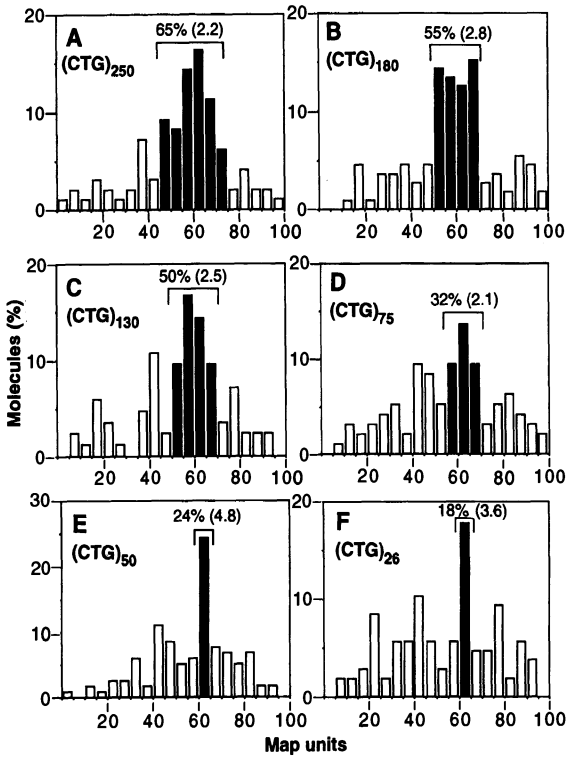

Fig. 3. Distribution of nucleosomes assembled onto six supercoiled plasmids containing CTG repeats of decreasing size (*12*). Closed circular plasmids were reconstituted with histones and analyzed by EM as in Fig. 2. The black bars indicate the positions of the CTG repeat sequences along the DNA. Values above the black bars represent the percent of total nucleosomes in the CTG block and the relative efficiency of nucleosome assembly for the repeat block (*19*).

REFERENCES AND NOTES

1. E. J. Kremer *et al.*, *Science* **252**, 1711 (1991); Y.-H. Fu *et al.*, *Cell* **67**, 1047 (1991); A. J. M. H. Verkerk *et al.*, *ibid.* **65**, 905 (1991).
2. M. Mahadevan *et al.*, *Science* **255**, 1253 (1992); Y.-H. Fu *et al.*, *ibid.*, p. 1256.
3. A. R. La Spada *et al.*, *Nature* **352**, 77 (1991).
4. The Huntington's Disease Collaborative Research Group, *Cell* **72**, 971 (1993).
5. H. T. Orr *et al.*, *Nat. Genet.* **4**, 221 (1993).
6. R. Koide *et al.*, *ibid.* **6**, 9 (1994).
7. C. T. Caskey *et al.*, *Science* **256**, 784 (1992); R. I. Richards and G. R. Sutherland, *Nat. Genet.* **1**, 7 (1992); R. R. Sinden and R. D. Wells, *Curr. Opin. Biotechnol.* **3**, 612 (1992).
8. P. Carango *et al.*, *Genomics* **18**, 340 (1993).
9. C.-H. Hsieh and J. D. Griffith, *Cell* **52**, 535 (1988).
10. T. E. Shrader and D. M. Crothers, *J. Mol. Biol.* **216**, 69 (1990); *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7418 (1989); J. J. Hayes *et al.*, *ibid.* **87**, 7405 (1990); J. J. Hayes *et al.*, *ibid.* **88**, 6829 (1991).
11. J. J. Hayes, J. Bashkin, T. D. Tullius, A. P. Wolfe, *Biochemistry* **30**, 8434 (1991).
12. The pSH1 and pSH2 plasmids are Bluescript-based plasmids and contain human complementary DNA (cDNA) derived from an individual with

myotonic dystrophy. The pUC19-based plasmids (pRW plasmids) used were derived from pSH1 and pSH2, containing 75 and 130 repeats of CTG, respectively (*14*). The DNA was cleaved with restriction endonucleases, and one end of the DNA was labeled with streptavidin. The length of the segment occupied by the triplet repeats, measured as a percentage of the distance from the labeled end for the $(CTG)_{250}$ insert in pRW3222 was 46 to 72%, 50 to 70% for the $(CTG)_{180}$ insert in pRW3219, 53 to 68% for the $(CTG)_{130}$ insert in pRW1981, 56 to 66% for the $(CTG)_{75}$ insert in pRW1980, 58 to 64% for the $(CTG)_{50}$ insert in pRW3213, 60 to 64% for the $(CTG)_{26}$ insert in pRW3212, and 32 to 43% for the $(CTG)_{130}$ insert in pSH2. In the figures, positions along the DNA are indicated as map units with each percentage point equal to 1 map unit.

13. Y.-H. Wang and J. D. Griffith, in preparation.
14. S. Amirhaeri, Y.-H. Wang, R. D. Wells, J. D. Griffith, unpublished data.
15. DNA polymerase pausing studies and physical determinations on these plasmids reveal a base pair repeat periodicity of 12 + 9 (2 × 10.5; the helical repeat of B-form DNA is 10.5 base pairs) and longer range conformational effects that may enhance nucleosome formation (S. Kang, K. Ohshima, S. Amirhaeri, R. D. Wells, in preparation).
16. We incubated closed circular plasmids in 5 mM $MgCl_2$ at 55°C for 30 min before we mixed the DNA with purified calf thymus histone octamers (*9*) in a buffer containing 2 M NaCl. The amount of histone was experimentally determined to result in an average of one nucleosome per DNA strand. We slowly lowered the NaCl concentration by dilution to 0.6 M to form stable nucleosomes. The nucleosome-assembled DNA was fixed with 0.6% glutaraldehyde for 10 min at 21°C, chromatographed over 1 ml of Sephadex G-50 (Pharmacia), and then treated with Bsa HI and Sca I restriction enzymes (pSH2) or Bsr FI and Alw NI (pRW3222). We labeled only the Bsa HI (or Bsr FI) ends of the molecules with biotinylated deoxycytidine 5′-triphosphate using the large fragment of DNA polymerase; this end was then labeled with streptavidin.
17. DNA-histone complexes prepared as described (*16*) were mixed with a buffer containing 2 mM spermidine and applied to carbon foils treated with a high-voltage discharge in a mild vacuum to introduce charged groups on the surface to facilitate DNA binding, and complexes were processed for EM as described [J. D. Griffith and G. Christiansen, *Annu. Rev. Biophys. Bioeng.* **7**, 19 (1978)]. We analyzed only those DNAs with a single nucleosome to avoid errors resulting from the progressive compaction of DNAs with multiple nucleosomes.
18. N. C. Stellwagen, *Biochemistry* **22**, 6186 (1983); G. Muzard, B. Theveny, B. Revet, *EMBO J.* **9**, 1289 (1990).
19. The percent values above the black bars represent the fraction of nucleosomes present in the repeat blocks. Dividing this number by the average fraction of nucleosomes occupying a segment equal in size to the repeat block (from the flanking regions) gave a value (in parentheses) representing the efficiency of nucleosome assembly for the CTG repeat block, as compared with the flanking regions.
20. To compare the efficiency of nucleosome assembly for the plasmids in Fig. 3, we normalized the values from the black bars by subtracting the percentage background assembly. For example, 55% (0.55) of all the nucleosomes assembled on the plasmid containing a $(CTG)_{180}$ block (Fig. 3B) localized to the block $(CTG)_{180}$, which encompasses 20% (0.20) of the total plasmid length. Thus, the length-normalized assembly for the plasmid containing a $(CTG)_{180}$ block is 0.35 (0.55 to 0.20). These length-normalized assembly values were plotted against the CTG copy number. Bars represent the range of the values derived from dividing the DNA segment into 10, 20, 50, or 100 divisions. Curve was fit with a French curve. A DNA sequence containing no CTG repeats or other nucleosome positioning sequences has a length-normalized assembly value equal to zero.
21. We thank C. T. Caskey (Baylor College of Medicine) for the human myotonic dystrophy cDNA.

Supported by grants from NIH (GM31819 and GM42342) and the ACS (NP583) (J.D.G.) and by grants from NIH (GM 30822), NSF (86-07785), and the Robert A. Welch Foundation (R.D.W.).

29 March 1994; accepted 8 June 1994

Pf1 Virus Structure: Helical Coat Protein and DNA with Paraxial Phosphates

David J. Liu and Loren A. Day

The helical path of the DNA in filamentous bacteriophage Pf1 was deduced from different kinds of existing structural information, including results from x-ray fiber diffraction. The DNA has the same pitch, 16 angstroms, as the surrounding helix of protein subunits; the rise and rotation per nucleotide are 6.1 angstroms and 132 degrees, respectively; and the phosphates are 2.5 angstroms from the axis. The DNA in Pf1 is, therefore, the most extended and twisted DNA structure known. On the basis of the DNA structure and extensive additional information about the protein, a model of the virion is proposed. In the model, the DNA bases reach out, into the protein, and the lysine and arginine side chains reach in, between the DNA bases, to stabilize the paraxial phosphate charges; the conformation of the protein subunit is a combination of α and 3_{10} helices.

The Pf1 virus is 2 μm long and contains a circular single-stranded DNA of 7349 nucleotides wrapped in a coat of protein subunits (*1–3*). This virus stands out among all filamentous bacteriophage (all species of the genus *Inovirus*) in having the highest axial distance, about 6 Å, between two nucleotides (*3, 4*) and an integer nucleotide/subunit ratio (n/s) of 1 (*5, 6*). This is the lowest n/s value of all known DNA-protein complexes. Structural data for Pf1 are available from spectroscopic methods (*5–15*), ion probes of its DNA (*16*), chemical modification of its protein (*17, 18*), x-ray fiber diffraction (*18, 19*), and neutron diffraction (*20*). Structural models for the major coat protein of Pf1 virus have been proposed (*20–22*), but heretofore, no model of the entire virion has been presented that contains DNA and systematically accounts for all relevant data, including the x-ray diffraction results for the DNA. The problem has been that the apparent structural repeat in x-ray fiber diffraction patterns for Pf1 calls for odd numbers of subunits (*5, 18, 23*), yet the unit stoichiometry (n/s = 1) calls for an even number. Specifically, the stoichiometry implies that each protein subunit directly contacts one nucleotide, either up-strand or down-strand; hence, there are two types of protein subunit and an even total number of protein subunits in a true repeat. In this paper, we provide a solution that establishes the symmetry of the DNA helix and places the phosphates close to the axis (paraxial). Working outward from the phosphates and making use of experimental (*4–19, 24*) and theoretical (*25–27*) results, we have developed models for the DNA and for the protein of Pf1 virus.

The establishment of the symmetry of any macromolecular assembly, such as a virus, is the critical step in solving its structure. In the case of a filamentous virus, it is the establishment of the helical symmetries of both its DNA and its protein coat. The x-ray fiber diffraction data for Pf1 have established the approximate helical symmetry of the coat, showing that the protein subunits are uniformly oriented with respect to the structural axis (*18, 19*). Information relating to DNA symmetry has been from the ^{31}P solid-state nuclear magnetic resonance (NMR) spectrum of oriented Pf1, from which it was concluded that the phosphates of both strands are "uniformly oriented" with respect to the structural axis (*10*). We have found that information on DNA symmetry is also in the x-ray diffraction data. On the basis of the protein coat symmetry, an electron density map for the entire structure was calculated (*19*). In the map, there are three spots of high density at radial positions below 10 Å. Our integration of the electron density above 0.6 $e/Å^3$ in these spots yielded 25 electrons, and the number of electrons in the largest spot approximates the sum for the two smaller spots. The maximum electron density in the spots exceeds 1.4 $e/Å^3$, which is too high for water but is appropriate for phosphate. There are 31 electrons in one O–P–O group (phosphorus and the two nonesterified oxygen atoms). The largest spot has the shape of a tilted O–P–O group. These spots are not from protein or solvent, so they must be from DNA.

Diffraction patterns contain information

Public Health Research Institute, 455 First Avenue, New York, NY 10016, USA.