SUPPLEMENTARY MATERIAL

The presence of an intrinsic bend widens the curvature distribution

We consider DNA molecules that have a permanent or intrinsic bend of curvature C_i comprising a fraction f ($0< f <1$) of their total length and an absence of preferential bending direction in the remaining fraction (1-f). Since experiments are conducted at room temperature, the overall curvature distribution of the DNA molecules will include contributions from two distinct curvatures distribution, centered around C_i and 0, respectively. We describe both curvature distributions by Gaussians functions and associate a standard deviation σ_0 (σ_i) with the distribution centered at 0 (C_i). Because the DNA sample is adsorbed on a surface, the intrinsic bend can in principle have either positive or negative curvature. This degeneracy must be taken into account in the normalization constant. Consequently, the separate distributions of curvatures for the molecule are:

$$N_{-}(C,f)=\frac{f}{2}\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp^{-\frac{1}{2\sigma_i^2}(C+C_i)^2} \quad \text{S1}$$

$$N_{+}(C,f)=\frac{f}{2}\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp^{-\frac{1}{2\sigma_i^2}(C-C_i)^2} \quad \text{S2}$$

$$N_{0}(C,f)=(1-f)\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp^{-\frac{1}{2\sigma_0^2}(C)^2} \quad \text{S3}$$

The measured distribution is the sum of the individual curvature distributions $N_{-}(C,f)$, $N_{+}(C,f)$, and $N_0(C,f)$.

$$N(C,f)=(1-f)\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp^{-\frac{1}{2\sigma_0^2}(C)^2}+f\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp^{-\frac{1}{2\sigma_i^2}(C_i)^2}\exp^{-\frac{1}{2\sigma_i^2}(C)^2}\cosh\left(\frac{1}{\sigma_i^2}C_0C\right) \quad \text{S4}$$

If $f \neq 0$, the resulting curvature distribution is wider than the curvature distribution in the absence of intrinsic bends $N(C,0)$. Figure S1 shows an example of $N(C,f)$ using $f = 0.5$, $L = 5$ nm, C_i=0.05 nm^{-1}, $\sigma_i^2 = 0.008$ nm^{-2}, and $\sigma_0^2 = 0.004$ nm^{-2}.

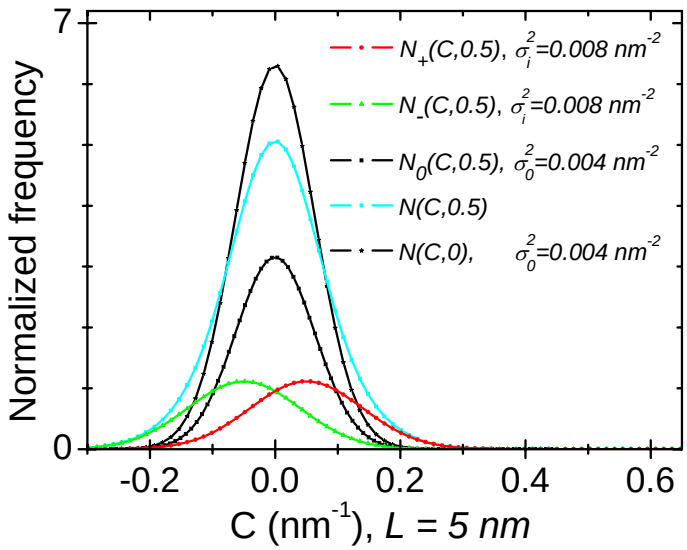

Figure S1. Example illustrating molecules with 50 % of their sequence with an intrinsic bend (red and green curves) and with 50 % without intrinsic bending (black, dots). The presence of an intrinsic bend results in a wider distribution of curvatures (light blue) compared to the distribution without intrinsic bends (black, stars).

The values of σ_i and σ_0 used in the example above can be correlated with values of persistence length P_0 and a P_i for the non-intrinsic bend fraction of the DNA and for the intrinsically bend fraction using the Worm-Like Chain model. In the WLC model, the standard deviation of the Gaussian curvature distribution is $\sigma = \sqrt{1/(P \cdot L)}$ where P is the persistence length, L is the length of the segment considered (S1). The example described above used P_0 = 50 nm and P_i = 25 nm.

In summary, the mere observation of a wider curvature distribution for a DNA molecule of a given sequence does not inherently imply increased flexibility. As we have shown, this effect can also be caused by the presence of an intrinsic bend. Additional approaches such as the absolute curvature distributions described in the main text are necessary to detect the presence of intrinsic bends. Together, they allow one to correctly

assay whether a wider curvature distribution should be attributed to the presence of sequences with increased flexibility, or to the presence of an intrinsic bend.

Graphical profiles of periodicity for DNAs used in this study.

The AA/TT periodicity of each DNA molecule analyzed was assessed using the 'PATC' algorithm developed and described in detail by Fire, Alcazar, and Tan (S2). For the purposes of analysis in this paper, we summarize the salient features of the algorithm briefly:

The PATC algorithm provides a quantitative measure of the enrichment of AA/TT dinucleotides on a single face of a helical DNA molecule. The algorithm starts by assigning an initial AA/TT score to each 5-nt window (pentamer) in the sequence (pentamers containing four AA/TT dinucleotides are assigned a score of 30, while those with three, two, one, and zero AA/TT dinucleotides are assigned scores of 20, 10, 0, and –5, respectively). Next, the algorithm attempts to link AA/TT-rich pentamer sequences that are located on adjacent faces of the helix (e.g., 10 nt center-to-center distance), giving each possible series of linked pentamers a score that is the sum of the individual pentamer scores. To accommodate some flexibility in helical structure, the algorithm allows for non-canonical helical spacings (9-12 bp), but assesses a penalty related to the deviation from a standard ~10.2 base pair helical density (16, 8, 16, and 32 points are assessed for proposed helical spacing of 9 bp, 10 bp, 11 bp, and 12 bp, respectively). For each base, a scalar PATC value is then calculated as the highest score assigned to any helically-linked series of pentamers covering the region. Assignment of PATC values involves sampling of very large numbers of proposed pentamer strings for any given DNA segment to determine which has the highest score. Scores from the PATC algorithm range from zero to several hundred, and we note that randomized segments with GC content comparable to *C. elegans* DNA generate scores with an average PATC score of 4.6, with 99.9% of scores in the range of 0-60 (2). Above this score, the random fraction of positives for a given PATC score decreases roughly exponentially with increasing PATC score (approximately 1 log10 for every 15-20 points in score).

Figure S2 shows the periodicity values (PATC scores) of the DNAs used in this study as a function of position along the DNA.

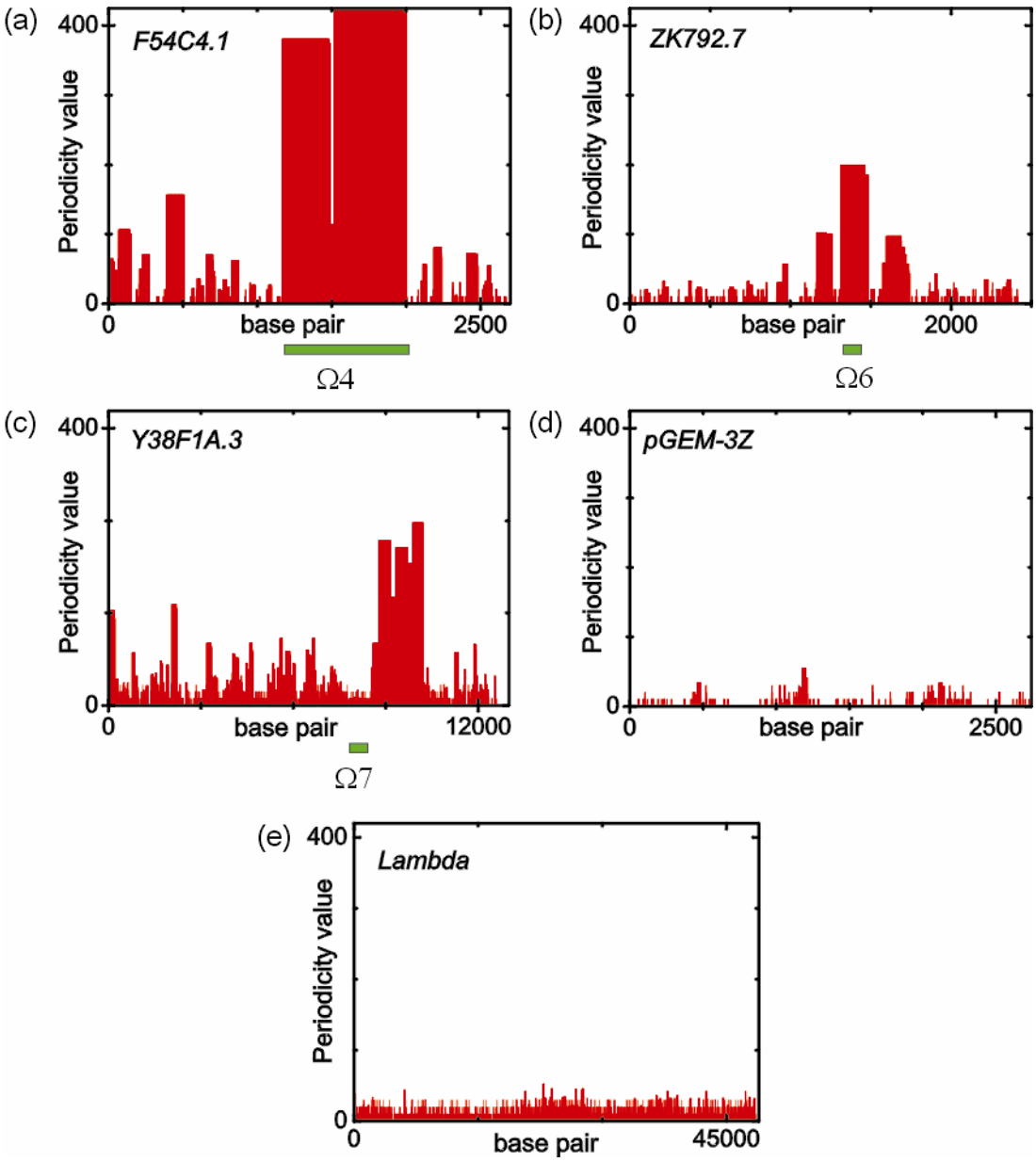

Figure S2 Each panel in the figure describes the periodicity values (PATC scores) of the DNAs used in this study as a function of position along the DNA molecule. For the *C. elegans* DNAs, we show the entire gene encompassing each fragment, with an indicator bar denoting the segment of interest. Gene F54C4.1 is the genomic segment encompassing fragment Ω4 (bases 1126 to 2048 of the segment shown) (**a**). A portion of

this segment (nt 1517 to 2000) gives PATC values, which are over 400 with a maximum value 573. Genes ZK792.7 (**b**) (nt 1280 to 1426) and Y38F1A.3 (**c**) (nt 8313 to 8462) encompass the nucleosome-length fragments Ω6 and Ω7, respectively. Equivalent profiles for pGem3C (**d**) and lambda (**e**), DNAs used as controls in this work, are also shown.

REFERENCES

S1. Rivetti, C., Guthold, M. and Bustamante, C. (1996) Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J Mol Biol*, **264,** 919-932.

S2. Fire, A., Alcazar, R. and Tan, F. (2006) Unusual DNA structures associated with germline genetic activity in Caenorhabditis elegans. *Genetics (in press)*.