Computer modelling of DNA structures involved in chromosome maintenance

Todd T.Eckdahl and John N.Anderson

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Received June 22, 1987; Revised and Accepted September 14, 1987

ABSTRACT

Sequence-dependent DNA bending of synthetic and natural molecules was studied by computer analysis. Modelling of synthetic oligonucleotides and of 107 kb of natural sequences gave results which closely resembled published electrophoretic data, demonstrating the powerful predictive capacity of the procedure. The analysis was extended to the study of DNA structures involved in chromosome maintenance. Centromeric DNAs from yeast were found to have sequences in their functional elements which cause them to be unusually straight. Autonomous replicating sequences were found to have two structural domains, one consisting of unusually straight sequences surrounding the consensus and the other of bending elements in flanking DNA. In addition to a structural homology, centromeric and autonomous replicating sequences share common sequence elements. These observations show that computer modelling of natural sequences is a viable approach to the study of the biological implications of alternative DNA structures.

INTRODUCTION

The three dimensional structure of DNA can vary in a sequence-dependent manner (1-4). One of the most pronounced variations is bent DNA, which results from curvature of the local axis of the double helix (5-7). The method most often used to detect DNA bending relies on the decreased mobility of bent DNA fragments during polyacrylamide gel electrophoresis (5-8). This technique has been used to identify bending sequences in a variety of prokaryotic and eukaryotic systems (8-11) and has been applied to the study of synthetic DNA oligonucleotides in order to identify the major sequence determinants of DNA curvature (6,7). These studies have led to the development of two models for DNA bending. The helix junction model (7) states that tracts of oligo(dA) longer than two nucleotides adopt an altered conformation related to the structure of poly(dA) (12). Increased tilt between successive base pairs at junctions between these stretches and neighboring B-DNA results in a bend in the local axis of the double helix. Macroscopic curvature of a molecule results from the summation of successive bends which are positioned with a periodicity near the helical repeat of DNA. The dinucleotide wedge model (13) proposes that ApA and TpT dinucleotides

cause changes in the direction of the local axis which may sum in the same fashion to create macroscopic curvature.

These models were incorporated into two computer programs for use in examining sequence-dependent structures in DNA. Computer analysis of synthetic oligonucleotides was used to refine features of the models and to demonstrate that the programs gave results which were consistent with the electrophoretic data (6,7). Modelling of natural DNA also gave results which were in excellent agreement with electrophoretic measurements (8). The analysis was then extended to the study of DNA structures involved in chromosome maintenance arising from sequences whose electrophoretic properties have not yet been reported. Centromeric DNAs from yeast were found to have sequences in their functional elements which cause them to be preferentially straight. Sequences capable of conferring autonomous replicating ability on plasmids in yeast were found to have two structural domains, one consisting of unusually straight sequences and the other of bent DNA.

METHODS

Computer analysis of sequence-dependent DNA structures involved construction of a helix junction program and a dinucleotide wedge program, written in BASIC for use on an IBM XT personal computer. The wedge program used the model of Ulanovsky and Trifonov (13) in which the angle of bending of ApA dinucleotides is divided into roll and tilt components of 8.4 and 2.3 degrees, respectively. These values were based on an overall wedge angle of 8.7 degrees as estimated by Ulanovsky et al. (15). For the helix junction program, this corresponded to a value of 14 degrees for the average junction bending angle. Varying the bending angle over a wide range did not affect the relative results from either program (data not shown). Since bending has been shown to be greater at the 3' than the 5' junction (7), the junction program used values of 15.4 and 12.6 for the 3' and 5' junctions, respectively. Because of the results of Hagerman (14), bending of the junctions of A or T tracts with B-DNA was assumed to be in the direction of tilt, as was the bending of the junction of A's with a 3' tract of T's, but bending of the junction of T's with a 3' tract of A's was assumed to be in the direction of roll.

Phasing of successive bends greatly affects macroscopic curvature (7). We found that the use of sequence-dependent dinucleotide twist angles gave better results than values based on the average pitch of a molecule. Therefore, twist angles for the ten possible dinucleotides as reported by Kabsch et al. (16) were used in both programs.

The programs calculate the three-dimensional coordinates of each position along the axis of a molecule, and these coordinates are used to draw two-dimensional projections of the axis. Figure 1 shows projections for the synthetic oligonucleotides (A5N5) 20 (A8N2) 20 (7), kinetoplast DNA (10), human ARS2 (17), and yeast centromeric element CEN3 (18). The programs also calculate the ratio of the contour length of the axis of a given molecule to the shortest distance between its ends. This is referred to as the ENDS ratio and is used to correlate the shape of various molecules with their degree of electrophoretic retardation. The ENDS ratio is related to another measure of DNA nonlinearity, Dmax, which has also been used to quantitate bending (19).

RESULTS

Comparison with Electrophoretic Data. Koo et al. (7) and Hagerman (14) measured bending of synthetic oligonucleotides by electrophoretic retardation. Their data were expressed as relative retardation as a function of length for oligonucleotides of up to 200 bp (see Figures 2-5 in Reference 7 and Figure 2 in Reference 14). As a measure of bending in the computer analysis, the ratio of the contour length of the axis of a molecule to the shortest distance between its ends was used (see Methods and Figure 1). This is referred to as the ENDS ratio and is plotted as a function of length of the synthetic oligonucleotides for comparison with the electrophoretic data.

Hagerman demonstrated that direct repeats of the sequence GA_4T_4C gave rise to bending while GT_4A_4C did not (14). The plots of ENDS as a function of length shown in Figure 2, Panel A, for the two oligonucleotides are in good agreement with this observation. Koo et al. demonstrated that bending according to the junction model was greater at the 3' junction of A tracts than at the 5' one, and that maximum bending occurred when successive tracts of oligo(dA) were arranged with a period near the helical repeat (7). Koo et al. also studied the electrophoretic properties of repeats of the sequence A_jN_{10-j} for $j=3$ to 10 (7). Figure 2, Panels B through D shows the ENDS ratios as a function of length for these sequences and demonstrates that both models yield curves which closely resemble the electrophoretic data in the relative maxima and overall shapes of the curves. At the locus of bending in a number of naturally occurring sequences, the tracts of oligo(dA) are flanked by C at the 5' end and T at the 3' end (8,10), and studies with synthetic oligonucleotides demonstrated that this arrangement caused maximal bending (7). Figure 2, Panel E shows that this result is reproduced by both models, simply as a consequence of sequence-dependent alterations in the dinucleotide twist angles. Indeed, of all the possible combinations of bases flanking the A tract in oligonucleotides of this type, the greatest bending as measured by modelling occurred when the 5' and 3' flanking bases were C and T, respectively.

We have reported an electrophoretic analysis of 107 kb of DNA from pBR322, phage lambda, SV40, adenovirus, yeast, and the chicken ovalbumin gene (8). These data were summarized in a plot of electrophoretic retardation as a function of A+T content which showed that a 10-20% increase in A+T content was accompanied by a 10% increase in electrophoretic anomaly (see Figure 6 in Reference 8). A small fraction of the fragments examined displayed

Figure 1. Two-dimensional projections of the double-helical axes of the molecules are shown as calculated by the wedge model program. Starting from the top in each panel, the projections are rotated about the axis at the 5' end in a counter clockwise direction by 0, 72, and 144 degrees. (A) synthetic oligonucleotide $(A_5N_5)_{20}$, (B) synthetic oligonucleotide $(A_8N_2)_{20}$, (C) 442 bp kinetoplast fragment, (D) 700 bp fragment of human ARS2 starting at position 800 bp, (E) 120 bp fragment of CEN3 containing functional elements.

unusually high electrophoretic retardation for their base composition. Here we report an independent study of the same 107 kb of DNA using computer analysis. Sequences were scanned to calculate ENDS ratios for 120 bp fragments starting every 40 bp. The staggers were to insure that the effects of positioning bending elements near the centers of fragments could be measured. A plot of ENDS ratios for the 120 bp molecules as a function of base composition is shown in Figure 3 for the wedge model. For this and other analyses involving natural sequences, only the wedge model results are presented due to the uncertainty of the junctions between consecutive tracts of A's and T's in the junction model. For simplicity, we calculated a least squares fitted line for the data, although the scattering in the plot does not allow us to preclude a nonlinear relationship. The mean ENDS ratio of the data was 1.05 with a standard error of $\pm .03$, indicating that the majority of the fragments displayed at least some bending. The positive slope of the line indicates greater bending with increased A+T content, as was seen by

Figure 2. Plots of ENDS as a function of length for synthetic oligonucleotides as calculated by the wedge model and helix junction model program. Sequences are repeats of the following oligonucleotides (7,14). (A) A4T4 is GAAAATTTTC and T4A4 is GTTTTAAAC, (B) A5-8 is CCAAAAACGGGCAAAAAA and A8-5 is CCAAAAACGGGCAAAAAA, (C) A3N7 is GGCCAAACCG, A4N6 is GGCCAAACCG, A5N5 is GGCAAAACCG, A6N4 is GGCAAAACCG, A8N2 is CCAAAAAA, and A9N1 is CAAAAA, (D) A5N5 is GGCAAAACCG, A5N6 is GGCCAAACCG, A5N7 is GGCCAAACCG, and A5N4 is CAAAAACCG, (E) FCT is GGCAAAATG, A5N5 is GGCAAAACCG, and FGG is CCGAAAAAGG.

Figure 3. ENDS as a function of base composition for 107 kb of natural DNA. Ratios are for 120 bp fragments as described in the text. The solid line is a least squares fit to the data and the dashed lines are for integral standard errors above the solid line. Sequences are labelled alphabetically from left to right and are centered at the positions indicated. (A) 35,060 bp in adenovirus, (B) 1925 bp in SV40, (C) 240 bp in adenovirus, (D) 750 bp in TRP1-ARS1, (E) 20,660 bp in phage lambda, (F) 220 bp in adenovirus, (G) 39,060 in lambda, (H) 500 bp in SV40, (I) 1280 bp in lambda, and (J) 540 in SV40.

electrophoresis. This result is consistent with our previous suggestion that increased anomaly with higher A+T content is due to bending at tracts of A's and T's (8). This trend was also observed for a similar analysis of 4000 molecules of the 120 bp length which were randomly generated with respect to base composition by the computer (data not shown). For both the random and natural DNA plots, more than 95% of the sequences had ENDS ratios lower than two standard errors above the mean, and none of the randomly produced molecules had ENDS ratios above five standard errors. In contrast, 10 of the fragments derived from natural sequences had ENDS ratios above this level (labelled A-J in Figure 3). These represented only eight unique sequences since two of the bending loci were present in two different 120 bp fragments as a result of the staggering. Remarkably, all the sequences identified by the electrophoretic analysis were present in this group, and only one additional sequence appeared. This agreement also extended to the relative degree of anomaly exhibited by the fragments in the group. The sequence which appeared bent by computer analysis but not by electrophoresis is found at

position 500 bp in SV40 and is comprised of two bending elements. Because the elements are separated by 26 bp they cause bending in opposite directions, resulting in an "S" shaped computer modelled structure (data not shown). This structure would not be expected to exhibit marked electrophoretic retardation, in agreement with experimental results (8 and unpublished data).

Sequences Involved in Chromosome Maintenance. Plasmids in Saccharomyces cerevisiae which contain an autonomous replicating sequence (ARS) and a centromeric DNA element (CEN) are replicated and segregated in a manner similar to chromosomes (18,20-22). Although ARS1 from yeast showed substantial electrophoretic retardation, CEN3 and CEN11 migrated at rates faster than that predicted from their base composition (8). These observations have led us to examine the DNA structure of CEN and ARS elements by computer modelling.

The sequences of twelve of the sixteen S. cerevisiae CENs and two from the related yeast S. uvarum have been reported (23-30). These sequences were studied by computer analysis. The results of the wedge model analysis are shown in Figure 4, Panel A, where ENDS ratios are plotted as a function of A+T content for the fourteen CEN sequences. Eleven of these are below the least squares fitted line of Figure 3, indicating that most of the CEN structures are preferentially straight. This type of structure is shown for CEN3 in Figure 1, Panel E. As a means of studying the sequence arrangement which gives rise to these unusual structures, we used an autocorrelation analysis (see reference 10). Since the average length of A and T tracts in the CEN elements is 3.5 bp, a composite autocorrelation analysis of A and T tetranucleotides was performed and is presented in Figure 4, Panel B. Peaks at multiples of 7.5 bp indicate that A and T tetranucleotides occur with a periodicity of 7.5 bp. The same periodicity was found for a composite analysis of trinucleotides and a similar analysis of the 14 individual CEN sequences showed a periodicity of 7.2 ± 1.3 bp (data not shown). The probability of occurrence of bending elements in the CEN sequences at intervals of 10-11 bp is lower than would be expected at random (Figure 4). Since the helical repeat of the DNA is close to 10.5 bp, this arrangement causes cancelling of successive bends, resulting in straightness. Two of the fourteen CEN elements, CEN5 and CEN14, have a weak periodicity of 10.5 bp superimposed upon the 7.5 bp one (data not shown), which explains their elevated ENDS ratios (Figure 4, Panel A).

Autonomous replicating sequences contain an 11 bp consensus sequence

Figure 4. ENDS as a function of base composition and composite autocorrelation for CEN elements. (A) In the top panel, the solid line is the least squares fit of the data in Figure 3 and the dashed lines are for integral standard errors above and below the solid line, also calculated using the data in Figure 3. ENDS ratios are for the following sequences, labelled alphabetically from left to right. (A) CEN14 (22), (B) CEN13 (23), (C) CEN5 (24), (D) CENB (23), (E) CENA (23), (F) CEN15 (25), (G) CEN1 (25), (H) CEN10 (25), (I) CEN11 (26), (J) CEN4 (27), (K) CEN7 (25), (L) CEN6 (28), (M) CEN3 (29), and (N) CEN16 (25). (B) Composite autocorrelation analysis of A and T tetranucleotides in the same 14 CEN sequences.

which is necessary for ARS activity although sequences adjacent to the consensus are required for maximal function (21,31-33). A plot of ENDS ratios as a function of base composition for the 120 bp region centered about the consensus sequence of 18 ARS elements appears in Figure 5, Panel A. Only those

Figure 5. ENDS as a function of base composition and composite autocorrelation analysis for 120 bp region surrounding the consensus of ARS elements. (A) In the top panel, the solid and dashed lines are from Figure 3 and the ARS sequences from left to right are (A) Yeast HML-L (34), (B) Yeast HO ARS (22), (C) Yeast HML-R (34), (D) Yeast HML-L (34), (E) Yeast 2 Micron ARS (35), (F) *N. tabaccum* ARS (36), (G) Yeast ARS2 (37), (H) Yeast TRP1-ARS1 (20), (I) Yeast copy-I H4 ARS (38), (J) *Drosophila* ARS-6 (39), (K) *Drosophila* satellite ARS (40), (L) *C. fasciculata* ARS 189 (41), (M) *C. acremonium* mitochondrial ARS (42), (N) Human ARS2 (17), (O) *Drosophila* ARS-2 (39), (P) Human ARS1 (17), (Q) Yeast HMR-R (34), and (R) *C. utilis* mitochondrial ARS (43). (B) Composite autocorrelation analysis of the 9 yeast ARSs listed above for the same 120 bp region.

ARS elements which contain a single functional consensus are included. Fourteen of the eighteen segments had ENDS ratios below the least squares fitted line of Figure 3, showing that the structures predicted are straighter than would be expected for their high A+T content. To our knowledge,

sequences are available for nine yeast ARSs which contain a single functional consensus. A composite autocorrelation analysis for the 120 bp region surrounding the consensus of these elements demonstrates that the probability of occurrence of A and T trinucleotides is highest at a periodicity of 9 bp (Figure 5, Panel B), and a similar analysis of the 9 individual ARSs showed a periodicity of 8.2 ± 1.1 bp (data not shown). This periodicity was also present in the 9 non-yeast ARSs analyzed in Figure 5, Panel A (data not shown). This sequence pattern results in an A+T rich but straight structure, as exemplified by the synthetic oligonucleotide A_5N_4 (Reference 7 and Figure 2, Panel D). The structure is localized within the 120 bp region centered at the consensus of the ARS elements since autocorrelation of sequences outside this region did not show the characteristic periodicity (data not shown). A search for the presence of A and T trinucleotides arranged in this periodicity yielded a value of 61 ± 31 bp for the length of the straight domain. The ARS consensus is at the approximate center of this domain in 15 of the 18 elements examined.

Yeast ARS1 contains a region of bent DNA 83 bp from the consensus sequence (8,33). We used computer modelling to determine whether this structure is generally associated with ARS elements. Figure 6 shows ENDS as a function of percent A+T for 36 yeast and non-yeast ARSs including those which contain more than one consensus. All but two of the ARS sequences contain a segment whose ENDS ratio is above the least squares fitted line of Figure 3 and 39% of the ENDS ratios are over three standard errors above this line. This result is significant in that only 0.7% of the fragments derived from the 107 kb analyzed in Figure 3 had ENDS ratios above this level. The elevated ENDS ratios from ARSs arise from bending elements that are 49 ± 20 bp in length and which have 75% of their A tracts on the same DNA strand. Of the 17 ARS elements which do not contain segments with ENDS ratios above the second standard error line, 9 are smaller than 310 bp, while only one of the 19 elements containing fragments with ratios above this level is smaller than 310 bp. This indicates that lack of sequence information may have limited the number of ARS elements which were found to contain bending sequences. For 15 of the 18 sequences which exceed 460 bp in length, the structure may be described as a superhelix like the one shown in Figure 1, Panel D, which completes one superhelical turn in about 360 bp.

Each of the 18 sequences with a single functional consensus possesses a straight structural domain, and 7 of them also contain a bending domain with an ENDS ratio greater than three standard errors above the least squares line

Figure 6. ENDS as a function of base composition for ARS fragments. The solid and dashed lines are from the data in Figure 3 ENDS ratios are for 120 bp fragments from the sequences listed and centered at the positions indicated. Sequences are labelled alphabetically from left to right. (A) TRP1-ARS1 at 750 bp (20), (B) ARS from central rDNA of *T. thermophila* at 80 bp (44), (C) Yeast HMR-R at 160 bp (45), (D) ARS from terminal rDNA of *T. thermophila* at 80 bp (46), (E) ARS from *N. tabaccum* at 350 bp (47), (F) Human ARS2 at 1020 bp (17), (G) Yeast HMR-L at 320 bp (45), (H) Human ARS1 at 260 bp (17), (I) Yeast ARS2 at 120 bp (37), (J) *C. reinhardtii* ARS element 1 at 320 bp (48), (K) Drosophila ARS 406 at 220 bp (35) (L) Yeast HO ARS at 60 bp (22), (M) *C. reinhardtii* ARS element 2 at 370 bp (48), (N) *P. hybrida* chloroplast ARS element B at 220 bp (49), (O) Yeast 2 micron ARS at 3320 bp (35), (P) Yeast ARS A6C at 300 bp (50), (Q) *C. reinhardtii* ARS element 3 at 170 bp (48), (R) Yeast ARS J11D1 at 580 bp (50), (S) Yeast copy-I H4 ARS at 160 bp (38), (T) Yeast ARS H9G at 630 bp (50), (U) Yeast mitochondrial ARS at 160 bp (51), (V) t3-ARS from *N. tabaccum* at 780 (36), (W) Yeast HML-L at 70 bp (34), (X) ARS from Drosophila satellite DNA at 180 bp (40), (Y) Drosophila ARS 414 at 60 bp (35), (Z) Yeast ARS C2G1 at 180 bp (50), (AA) Drosophila ARS 312 at 200 bp (35), (BB) Drosophila ARS 410 at 60 bp (35), (CC) Yeast HML-R at 84 bp (34), (DD) Drosophila ARS 413 at 120 bp (35), (EE) Drosophila ARS-6 at 120 bp (39), (FF) *C. fasciculata* ARS 189 at 128 bp (41), (GG) *P. hybrida* chloroplast ARS element A at 280 bp (49), (HH) *C. acremonium* mitochondrial ARS at 560 bp (42), (II) Drosophila ARS-2 (39), and (II) *C. utilis* mitochondrial ARS at 100 bp (43).

of Figure 3. The two domains are positioned in either orientation with respect to one another and the distance between them ranges from 3 to 1026 bp, with an average of 251 bp. In the case of HO ARS, the two domains are present in the same 120 bp of DNA. For this reason, the ENDS ratio plotted in Figure 5 for HO ARS is higher than would be expected for a straight structure.

DISCUSSION

Predicting DNA Structure. Computer programs using the helix junction and dinucleotide wedge models were used to study DNA bending of synthetic and natural sequences. The analysis involved modification and testing of the models using known electrophoretic data as well as prediction of structures for molecules which have not yet been examined for DNA bending. Modelling of the sequences of synthetic oligonucleotides studied by Hagerman (14) and Koo et al (7) showed excellent agreement with the electrophoretic data (Figure 2). The powerful predictive capacity of the procedure was further demonstrated by its ability to predict the electrophoretic properties of 107 kb of natural sequences (Figure 3). Since only A tract bending was considered in the computer analysis, the results show that this feature is the major determinant of electrophoretic anomaly in the sequences analyzed. The 107 kb of sequences examined are considered to be representative of natural DNA since they were derived from a variety of organisms.

The analysis of natural DNA presented in Figure 3 is based on the ENDS ratios of 120 bp fragments. This length was used because larger ones caused additional sequences to have elevated ENDS ratios which were not anomalous in electrophoretic mobility and smaller ones caused some bent fragments to have near normal ENDS ratios. This suggests that the optimal sequence window for predicting electrophoretic bending using the ENDS ratio is about 120 bp, which may be related to the maximum distance over which structural elements can influence one another in DNA (see Reference 52). The efficacy of the ENDS ratio in predicting relative electrophoretic results is also consistent with the view that electrophoretic mobility depends on the mean square end to end distance of DNA molecules (53,54). However, the absolute value of ENDS ratios did not always equal electrophoretic anomaly, especially for bending elements longer than 150 bp. This may be due to the limits of resolution in electrophoresis, as discussed in Ulanovsky and Trifonov (13). The ENDS ratio of 120 bp fragments is also insensitive to the relative position of a bending element along a restriction fragment although this positional effect can influence gel mobility (5).

Centromeric DNA Structure. DNAs cloned from yeast centromeres contain three functional elements (10). Elements I and III are 12 and 26 bp stretches which are sensitive to single nucleotide changes, while element II shows little sequence homology and consists of 78 to 84 bp of A+T rich sequence (10). The results of the wedge model analysis in Figure 4, Panel A, show that

the sequences of yeast CENs result in structures which are preferentially straight, especially for their high A+T content. The straightness of the CEN DNA structures arises from two types of sequence patterns. First, although there are 38 occurrences of the sequence motif oligo(dA) followed by oligo(dT) or vice versa, two-thirds of them have the T tract 5' to the A tract, which does not cause bending (13). Secondly, the sequences are non-randomly organized with respect to the distance separating blocks of A's and/or T's. As seen in Figure 4, Panel B, the highest probability of occurrence of A and T tetranucleotides is at a periodicity of 7.5 bp. This arrangement causes cancelling of successive bends and would be expected to produce a zig-zag structure. In the cases of CEN3 and CEN11, this structure is perceived by the electrophoretic method to be straight, since fragments containing these sequences have faster mobilities than predicted from their base composition (8). Studies on CEN3 (55) have shown that element II sequences may be replaced by linker DNA which is higher in G+C content with little loss of function, as long as a minimum distance between elements I and II is retained. We have analyzed these sequences by computer modelling and have found that the structural consensus of straightness is maintained in the mutants with near wild-type function (data not shown).

Two Structural Domains in ARS Elements. The presence of a consensus sequence is often noted for ARS fragments (34), although no further sequence homology between them has been found. However, A+T rich sequences directly adjacent to the consensus as well as regions further removed from it have been shown to be required for maximal ARS activity (21,31-33). The results of the present study permit us to advance a preliminary model for ARS elements based on structural homology. The model is applicable to yeast and non-yeast DNA fragments which have ARS activity in yeast, increasing the likelihood that non-yeast ARSs are functionally relevant. A typical ARS contains segments of A+T rich straight DNA on each side of the consensus sequence. The sequence pattern responsible for this structure is similar to that of the CEN DNAs in that A and T trinucleotides are distributed with a non-bending periodicity of 9 bp (Figure 5, Panel B). Bending elements are also associated with many ARSs, arising from A and T tracts spaced at intervals of 10-11 bp. Considerable variation exists among ARS elements in the length of each domain, the distance between them, and their orientation with respect to one another. The conserved structural domains are presumably involved in regulation of replication, since deletion of DNA containing the bending or straight domains from ARS elements reduces their activity (21,31-33). Moreover, the

replication deficiency observed after deletion of the natural bending element from ARS1 can be corrected by insertion of synthetically derived bent DNA (31).

Comparison of ARS and CEN Elements. It is interesting to note the striking similarity between CEN and ARS elements at the levels of sequence and higher-order DNA structure. The portion of CEN DNAs which is most crucial for their function is element III, the consensus for which is TGTTTATGNTTTCGAAANNNA (18). This sequence bears a strong resemblance to the ARS consensus of A/T TTTATPuTTT A/T (34). In fact, the CEN element III consensus consists of two near-perfect copies of the ARS consensus in opposite orientations. Each of the consensus sequences lies in a region of A+T rich DNA which is of a unique sequence arrangement which causes it to be preferentially straight. Whether this homology in structure and sequence relates to replicator function within CEN DNA (18), to segregation of ARS elements (21), or to some other common feature merits further investigation.

ACKNOWLEDGEMENTS

We wish to thank B. Battaile and P. Seiler for advice on the computer programming. We also appreciate critical comments on the manuscript by D. J. Asai and I. Tessman and helpful discussions with P. T. Gilham. This work was supported by Grant P01-CA36761-01A1 from the National Cancer Institute.

REFERENCES

1. Calladine, C.R. (1982) J. Mol. Biol. 161, 342-352.
2. Dickerson, R. (1983) J. Mol. Biol. 166, 419-441.
3. Nussinov, R. (1985) J. Theor. Biol. 115, 179-189.
4. Tung, C-S. and Harvey, S.C. (1987) Nucl. Acids Res. 15, 4973-4985.
5. Wu, H., and Crothers, D.M. (1984) Nature (London) 308, 509-513.
6. Hagerman, P.J. (1984) Proc. Nat. Acad. Sci. USA 81, 4632-4636.
7. Koo, H.S., Wu, H.M., and Crothers, D.M. (1986) Nature (London) 320, 501-506.
8. Anderson, J.N. (1986) Nucl. Acids Res. 14, 8513-8533.
9. Zahn, K., and Blattner, F.R. (1985) Nature (London) 317, 451-453.
10. Marini, J.C., Levene, S.D., Crothers, D.M., and Englund, P.T. (1982) Proc. Nat. Acad. Sci. USA 79, 7664-7668.
11. Marini, J.C., Effron, P.N., Goodman, T.C., Singleton, C.K., Wells, R.D., Wartell, R.M., and Englund, P.T. (1984) J. of Biol. Chem. 259, 8974-8979.
12. Arnott, S., Chandrasekaran, R., Hall, I.C., and Puigjaner, L.C. (1983) Nucl. Acids Res. 11, 4141-4155.
13. Ulanovsky, L., and Trifonov, E.N. (1987) Nature (London), in press.
14. Hagerman, P.J. (1986) Nature (London) 321, 449-450.
15. Ulanovsky, L., Bodner, M., Trifonov, E.N., and Choder, M. (1986) Proc. Nat. Acad. Sci. USA 83, 862-866.
16. Kabsch, W., Sander, C., and Trifonov, E.N. (1982) Nucl. Acids Res. 10, 1097-1104.

17. Montiel, J.F., Norbury, C.J., Tuite, M.F., Dobson, M.J., Mills, J.S., Kingsman, A.J., and Kingsman, S.M. (1984) *Nucl. Acids Res.* **12**, 1049-1068.
18. Clarke, L., and Carbon, J. (1985) *Ann. Rev. Genet.* **19**, 29-56.
19. Tung, C.-S. and Burks, C. (1987) *J. Biomolec. Struct. Dynamics* **4**, 553-559.
20. Tschumper, G., and Carbon, J. (1980) *Gene* **10**, 157-166.
21. Celniker, S.E., Sveder, K., Srienc, F., Bailey, J.E., and Campbell, J.L. (1984) *Mol. and Cell. Biol.* **4**, 2455-2466.
22. Kearsley, S. (1984) *Cell* **37**, 299-307.
23. Neitz, M., and Carbon, J. (1985) *Mol. Cell. Biol.* **5**, 2887-2893.
24. Huberman, J.A., Pridmore, R.D., Jager, D., Zonneveld, B., and Philippsen, P. (1986) *Chromosoma* **94**, 162-168.
25. Maine, G.T., Surosky, R.T., and Tye, B.-K. (1984) *Mol. Cell. Biol.* **4**, 86-91.
26. Heiter, P., Pridmore, D., Hegemann, J.H., Thomas, M., Davis, R.W., and Philippsen, P. (1985) *Cell* **42**, 913-921.
27. Fitzerald-Hayes, M., Buhler, J.-M., Cooper, T., and Carbon, J. (1982) *Mol. Cell. Biol.* **2**, 82-87.
28. Stinchcomb, D.T., Mann, C., and Davis, R.W. (1982) *J. Mol. Biol.* **158**, 157-179.
29. Panzeri, L., and Philippsen, P. (1982) *EMBO J.* **1**, 1605-1611.
30. Clarke, L., and Carbon, J. (1980) *Nature (London)* **287**, 504-509.
31. Williams, J.S., Eckdahl, T.T., and Anderson, J.N. (1987), Submitted.
32. Strich, R., Woontner, M., and Scott, J.F. (1986) *Yeast* **2**, 169-178.
33. Snyder, M., Buchman, A.R., and Davis, R.W. (1986) *Nature (London)* **324**, 87-89.
34. Broach, J.R., Li, Y.-Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A., and Hicks, J.B. (1983) *CSHSQB* **47**, 1165-1173.
35. Genbank (1986) BBN Laboratories Inc.
36. Ohtani, T., Kiyokawa, S., Ohawara, T., Harada, H., and Uchimiya, H. (1985) *Plant Mol. Biol.* **5**, 35-39.
37. Tschumper, G., and Carbon, J. (1982) *J. Mol. Biol.* **156**, 293-307.
38. Bouton, A.H., Stirling, V.B., and Smith, M.M. (1987) *Yeast* **3**, 107-115.
39. Mills, J.S., Kingsman, A.J., and Kingsman, S.M. (1986) *Nucl. Acids Res.* **14**, 6633-6648.
40. Marunouchi, T., and Hosoya, H. (1984) *Mol. Gen. Genet.* **196**, 258-265.
41. Kim, R., and Ray, D.S. (1984) *Gene* **29**, 103-112.
42. Penalva, M.A., Patino, C., and Rubio, V. (1986) *FEBS Letters* **198**, 92-98.
43. Tikhomirova, P., Kryukov, V.M., Strizhov, N.I., and Bayev, A.A. (1983) *Mol. Gen. Genet.* **189**, 479-484.
44. Kiss, G.B. and Pearlman, R.E. (1981) *Gene* **13**, 281-287.
45. Abraham, J., Nasmyth, K.A., Strathern, J.N., Klar, A.J.S., and Hicks, J.B. (1984) *J. Mol. Biol.* **176**, 307-331.
46. Kiss, G.B., Amin, A.A., and Pearlman, R.E. (1981) *Mol. Cell. Biol.* **1**, 535-543.
47. Ohtani, T., Uchimiya, H., Kato, A., Harada, H., Sugita, M., and Sugiura, M. (1984) *Mol. Gen. Genet.* **195**, 1-4.
48. Vallet, J.-M., Rahire, M., and Rochaix, J.-D. (1984) *EMBO J.* **3**, 415-421.
49. de Haas, J.M., Boot, K.J.M., Haring, M.A., Kool, A.J., and Nijkamp, H.J.J. (1986) *Mol. Gen. Genet.* **202**, 48-54.
50. Palzkill, T.G., Oliver, S.G., and Newlon, C.S. (1986) *Nucl. Acids Res.* **14**, 6247-6264.
51. Mabushi, T., Nishikawa, S., and Makabayashi, K. (1984) *J. Gen. Appl. Microbiol.* **30**, 469-478.
52. Jolly, D., and Eisenberg, H. (1976) *Biopolymers* **15**, 61-95.
53. Lumpkin, O.J., and Zimm, B.H. (1982) *Biopolymers* **21**, 2315-2316.
54. Lerman, L.S., and Prisch, J.L. (1982) *Biopolymers* **21**, 885-997.
55. Gaudet, A., and Fitzgerald-Hayes, M. (1987) *Mol. Cell. Biol.* **7**, 68-75.