Nucleic Acids Research, Vol. 18, No. 6

Conserved DNA structures in origins of replication

Todd T.Eckdahl and John N.Anderson*
Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Received September 25, 1989; Revised and Accepted February 11, 1990

ABSTRACT

According to the model of Bramhill and Kornberg, initiation of DNA replication in prokaryotes involves binding of an initiator protein to origin DNA and subsequent duplex opening of adjacent direct repeat sequences. In this report, we have used computer analysis to examine the higher-order DNA structure of a variety of origins of replication from plasmids, phages, and bacteria in order to determine whether these sequences are localized in domains of altered structure. The results demonstrate that the primary sites of initiator protein binding lie in discrete domains of DNA bending, while the direct repeats lie within well-defined boundaries of an unusual anti-bent domain. The anti-bent structures arise from a periodicity of A_3 and T_3 tracts which avoids the 10–11 bp bending periodicity. Since DNA fragments which serve as replicators in yeast also contain these two conserved structural elements, the results provide new insight into the universal role of conserved DNA structures in DNA replication.

INTRODUCTION

Origins of DNA replication from plasmids, phages, and bacteria may share a common organization (1,2). One region is tightly bound by an initiator protein which is specific to a particular origin, and another contains 2–5 direct repeats of 6–16 bp in length which are found in A+T rich DNA. Initiation of replication is thought to begin by binding of initiator proteins to specific sites in the first region, and duplex opening at the direct repeats (1–4).

Understanding of eukaryotic replication has been aided by the isolation and characterization of DNA sequences which confer upon plasmids in yeast the ability to replicate extrachromasomally (5–8). These autonomously replicating sequences (ARSs) from yeast and higher eukaryotes share common features. The conserved 11 bp ARS consensus sequence A/T TTTAT A/G TTT A/T is indispensable for function, and has been assumed to be the site of initial specific protein interaction (5–8). Regions outside the consensus sequence are necessary for full ARS activity, although point mutations and small deletions in them have only minor effects on replicator function (9,10). These regions do not share direct sequence homology, but contain two domains of altered DNA structure (11,12). The first consists of bent DNA, arising from oligo(dA) tracts of 3–6 bp in length which deflect the local axis of the double helix from a straight line (13–16). Spacing of the tracts near the helical repeat of DNA (10–11 bp) causes their deflections to be in the same direction, resulting in macroscopic bending. The second domain surrounds the ARS consensus sequence and contains oligo(dA) tracts which lie in a 6–8 bp periodicity. This spacing causes successive deflections at the tracts to cancel, producing an A+T rich, anti-bent structure. The importance of these structural features to ARS function is supported by deletion analysis (9,10), and by the observation that synthetic bent DNA can substitute functionally for the natural bending element in yeast ARS1 (17). In this report, we demonstrate that the conserved structural features of eukaryotic replicators are also found in origins of replication from prokaryotes.

METHODS

The dinucleotide wedge model of Ulanovsky and Trifonov (16) was used to determine the three dimensional axes of DNA fragments on the basis of sequence. A measure of DNA bending, the ENDS ratio, is then calculated as the ratio of the contour length of a given axis to the shortest distance between its ends (11). In a survey of natural DNA, the computer analysis gave results which were in complete agreement with electrophoretic data on DNA bending (11). Changes in DNA structure along a sequence can be quantified and mapped by comparison to the average ENDS ratio of 1.066 ± .030 S.E.M. (for 60% A+T) derived from 107 kb of natural DNA (11). For Figure 2, positional autocorrelation analysis was performed as previously described (11,15,18).

The general equations for the probability of occurrence in random sequence DNA of the consensus sequence $N_1N_2...N_n$ are as follows, where P (a,b) is the probability of a matches in b nucleotides.

$$P(n, n) = \prod_{i=1}^{n} P_i$$

$$P(n-1, n) = \sum_{i=1}^{n} \left[(1-P_i)\prod_{j} P_j\right], j = 1, 2, \ldots n, \text{ but } j \neq i$$

$$P(n-2, n) = \sum_{i=1}^{n-1} \left\{(1-P_i) \sum_{j=i+1}^{n}\left[(1-P_j)\prod_{k} P_k\right]\right\}, k = 1, 2, \ldots n, \text{ but } k \neq i, j$$

The probability of occurrence p_i of base N_i is determined by the base composition of the sequence analyzed.

* To whom correspondence should be addressed

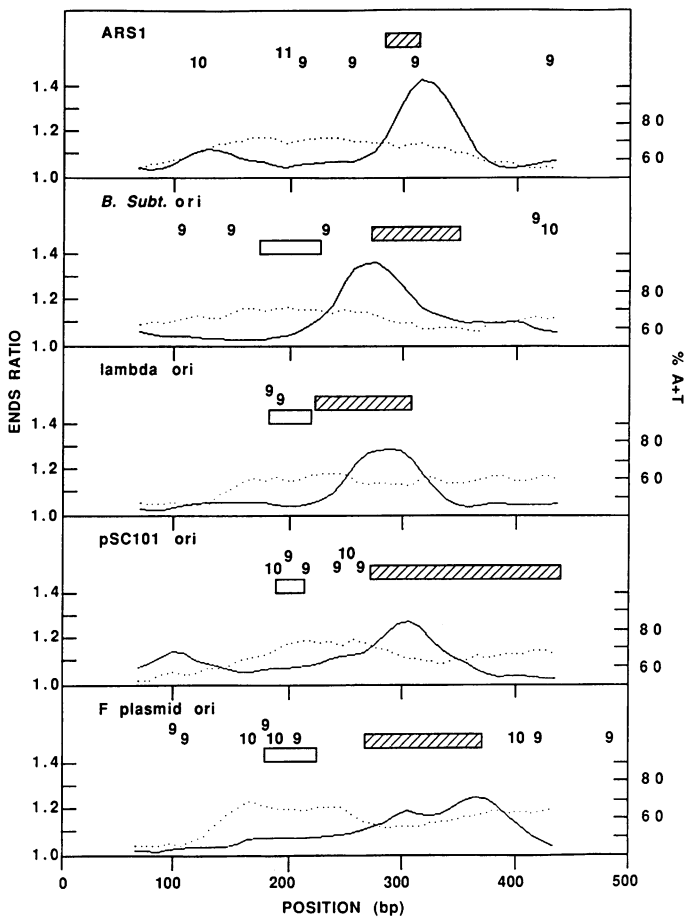

Figure 1. Higher-order DNA structure of origins of DNA replication from yeast and prokaryotes. Plots are shown of the ENDS ratio (solid lines) and A+T content (dotted lines) of 120 bp fragments as functions of their position in yeast ARS1 and prokaryotic origins of replication. The positions of matches to the 11 bp ARS consensus sequence are indicated by 9 (9/11), 10 (10/11), and 11 (11/11). Striped boxes represent binding sites for ABF1 and initiator proteins as identified by footprint analysis and open boxes show positions of direct repeat elements (1).

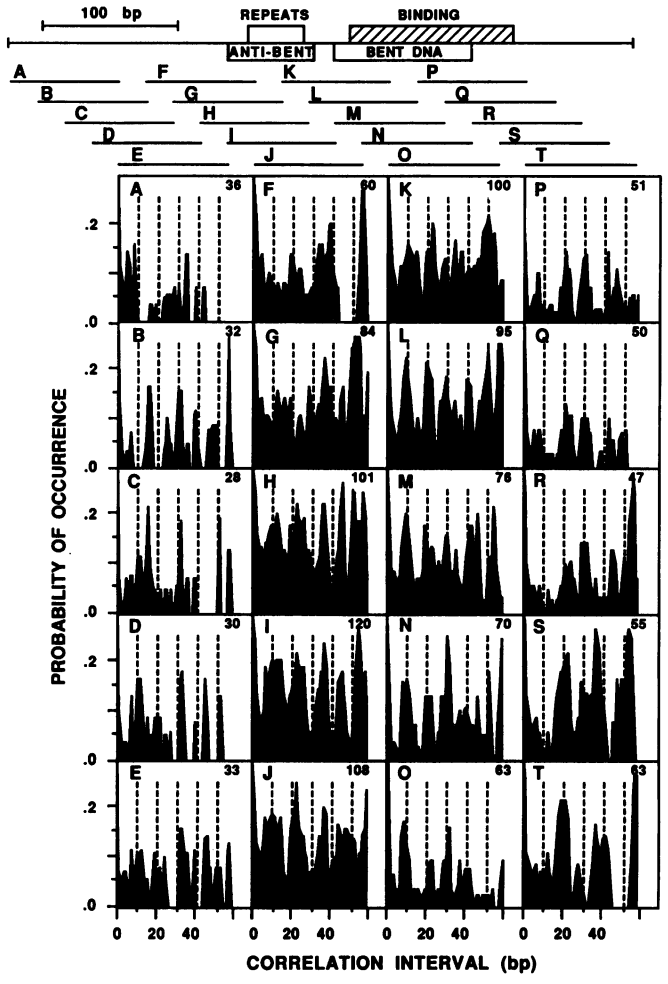

Figure 2. Positional autocorrelation analysis of A_3 and T_3 tracts in the twelve prokaryotic origins. As indicated in the map at the top of the Figure, sequences are aligned with the center of their direct repeats at position 200 bp with the sites of initiator protein binding downstream from them (as in ref. 1). The direct repeats span an average of 42 ± 12 S.E.M. bp and the binding sites are centered at 315 bp and are 119 ± 19 S.E.M. bp in length. Shown below the map are intervals of 80 bp staggered every 20 bp along the sequences which were analyzed. Autocorrelation plots appear in Panels A−T where the dotted lines represent multiples of a 10.5 bp periodicity. The total number of A_3 or T_3 tracts present in the sequences analyzed in each Panel is indicated in the upper right corner of the Panel. The positions of plots which display strong bending and anti-bending periodicities are indicated in the map.

RESULTS

Figure 1 shows plots of A+T content and the ENDS ratio (a measure of DNA bending) as functions of position in yeast ARS1. The largest peak in the ENDS ratio represents a bending element previously identified by electrophoresis (18,19) and the striped box marks the site of interaction for ARS binding factor 1 (ABF1), a presumptive replication protein (20). Figure 1 also shows plots for prokaryotic origins of replication from *B. subtilis*, phage lambda, pSC101, and the F plasmid. Strong bending elements are found in each origin. Bending as measured by electrophoresis has previously been reported for the origins in phage lambda and pSC101 (18,21,22). The computer analysis was performed for eight additional prokaryotic origins (*oriC*, P1, R1, R6K, phage82, phi80, RK2, and P4), which were the subject of a recent review (1). Of the twelve origins, ten contained bending elements with ENDS ratios greater than two standard errors above the mean for 107 kb of natural DNA, with an average of 1.219 ± .026 S.E.M. These bending elements overlap with regions of initiator protein binding, which are indicated by striped boxes in Figure 1.

The highest concentration of A_{3-4} / T_{3-4} tracts along the entire 1.4Kb TRP1-ARS1 fragment is found in the 80 bp region surrounding the perfect 11 bp ARS consensus sequence where 8 tracts are arranged in a pattern which avoids the 10−11 bp bending periodicity. This pattern produces an anti-bent structure with low ENDS ratios, and is a conserved feature of ARS elements (11, 12). A positional autocorrelation analysis was applied to the twelve prokaryotic origins as a group in order to detect this structure and to more precisely delineate the boundaries of the structural domains. In Figure 2, the twelve origins are aligned according to the positions of the direct repeats with the initiator protein binding sites downstream as in reference 1. The numbers in the boxes are the total number of A_3/T_3 tracts at 80 bp intervals and the tracings are the autocorrelation plots. The concentration of tracts varies over a 4-fold range with the highest density present in the segments containing the direct repeats. No clear patterns in the arrangement of the tracts are apparent upstream from the direct repeats (Panels A−G), while peaks in

Table 1. Matches to the ARS consensus sequence. Calculations for the expected (exp.) occurrences of 9/11, 10/11, and 11/11 matches to the 11 bp ARS consensus sequence A/T TTTAT A/G TTT A/T are compared in the table to the observed (obs.) occurrences for control DNA (SV40 genome and chicken ovalbumin gene), prokaryotic origins (listed in text, as in ref. 1), and yeast ARS elements (ARS1, ARS2, HMR-E, 2-micron, H4, A6C, C2G1, H9G, and J11D1, as in ref. 23). The 200 bp yeast ARS segments were those analyzed by Palzkill and Newlon while 500 bp segments of the same ARSs had the best match to the consensus at their centers.

DNA Source	9/11 obs.	9/11 exp.	10/11 obs.	10/11 exp.	11/11 obs.	11/11 exp.
SV40[1]	2.5 ±.52	2.3 ±.34	0.27 ±.15	0.25 ±.04	0.09 ±.09	0.01 ±.02
Ovalbumin Gene[1]	3.2 ±.62	3.2 ±.29	0.47 ±.24	0.37 ±.04	none ±.002	0.02
Prokaryotic Origins[2]	2.8 ±.80	1.8 ±.32	0.83 ±.34	0.20 ±.04	0.08 ±.09	0.01 ±.002
ARSs[2]	9.3 ±2.2	5.0 ±.93	1.78 ±.42	0.65 ±.15	0.67 ±.18	0.04 ±.01
ARSs[3]	2.8 ±.46	3.9 ±.81	1.11 ±.33	0.59 ±.15	0.67 ±.18	0.04 ±.01

[1] Mean ± S.E.M. number of matches in successive intervals of 500 base pairs.
[2] Mean ±S.E.M. number of matches per 500 base pair segment.
[3] Mean ±S.E.M. number of matches per 200 base pair segment

the probability of occurrence of the tracts in the region of the direct repeats are in a decidedly nonrandom periodicity (Panels H and I). This periodicity, which is responsible for the low ENDS ratios exhibited by this region (1.042 ± 0.006), is opposite to that which is characteristic of bent DNA since the probability of occurrence of tracts exhibits minima at 10−11 bp intervals. The strength of this pattern is comparable to the A+T rich anti-bent structure found in yeast ARS and CEN elements, the most pronounced such structures reported (11, 18). Panels J and K exhibit periodicities which are a composite of anti-bent and bent periodicities. Bent DNA is observed in initiator protein binding regions in Panels L and M, which display a 10.5 bp periodicity of tracts comparable to that of the locus of bending in kinetoplast DNA (15). This pattern persists downstream in the binding domain in Panels N and O, although the strength of the periodicity is weaker, an observation which is supported by a reduction in the bending periodicity of A_4 and T_4 tracts (data not shown). Since Panels P through T display no strong patterns in the arrangement of A_3 and T_3 tracts, the downstream boundary of the bending domain appears to be between Panels O and P. This is further evidenced by a decrease in the total number of tracts (Figure 2), a decrease in the fraction of tracts contributing to bending (Figure 2), and a decrease in the strength of the A_4 / T_4 bending periodicity as revealed by autocorrelation analysis (data not shown). The region downstream from the bending element, like the region of the direct repeats, exhibits ENDS ratios that are somewhat lower than the average observed for 107 Kb of DNA (see Figure 1 and ref. 11). However, the two regions are distinct in that the downstream DNA contains fewer tracts than the direct repeat segment and these tracts do not show a discernable periodicity. Based on this analysis, the distance between the centers of the anti-bent and bent structural domains is 90 bp, while the average distance between the centers of the direct repeat and initiator protein binding domains is 117 ± 14 bp. These results clearly show that the protein binding and direct repeat sequences in the prokaryotic origins are found in discrete domains of unusual higher-order DNA structure.

The 11 bp ARS consensus sequence is absolutely required for replicator function in yeast and Table 1 shows that the number of 11/11 matches to the consensus in yeast ARSs is 16-fold greater than the expected occurrence due to chance. Palzkill and Newlon reported that multiple weak matches of 9/11 or better to the consensus occur along yeast ARSs at a frequency which is higher than the expected occurrence due to chance, and this observation has been used to support models for ARS function based on reiterated copies of the consensus sequence (23). The presence of multiple copies of the ARS consensus in prokaryotic origins (Figure 1) raises the interesting possibility that this eukaryotic feature is important for prokaryotic replication. However, in contrast to the conclusions of Palzkill and Newlon, the results in Table 1 show that only the observed 10/11 and 11/11 matches to the consensus are significantly different from the expected values at the 95% confidence level. The 10/11 matches are not significantly different at this level if a 10/11 match is omitted for those ARSs which do not contain an 11/11 match. One of the 12 prokaryotic origins contains an 11/11 match and only half of them contain a 10/11 match. In addition, the observed presence of 9/11 or 10/11 matches in the origins is less than 2-fold greater than the expected occurrence. Palzkill and Newlon reported that the observed presence of matches of 9/11 or better to the ARS consensus in yeast ARSs ranged from 8 to 30-fold higher than the expected occurrence (23), while the correct calculations yield a total expected occurrence of 40 matches compared to the observed value of 41. The presence of multiple weak matches to the ARS consensus is therefore not an unusual feature of the prokaryotic or eukaryotic replicators.

DISCUSSION

The results of this study indicate that prokaryotic origins of replication, like eukaryotic ARS elements (11, 12), consist of two domains of unusual structure. The A+T rich 80bp segments surrounding the direct repeats can be compared to the 80 bp around the 11 bp yeast ARS consensus sequence. An average of 5.3 and 7.2 A_{3-6} tracts are present in these prokaryotic and eukaryotic segments, respectively, with 73% and 75% of their A tracts on the same DNA strand. In both cases, the tracts are arranged in a pattern which avoids the 10−11 bp periodicity producing straight structures with ENDS ratios that are lower than expected for their high A+T content. It has generally been viewed that the ARS consensus sequence in eukaryotes serves as a site for initial protein binding (5−8). However, the results

of the present analysis raise the possibility that an initiator protein binds first to the bending domain, by analogy to the prokaryotic origins. A candidate for such a protein is ABF1, which binds to non-conserved elements in yeast ARS1 and the HMR-E ARS (20), and these elements contain the sequence requirements for DNA bending (11,18,19). The ARS1 and HMR-E ARS bending elements have ENDS ratios of 1.43 and 1.20, respectively (ref. 11 and Figure 1) and overlap with the regions of ABF1 binding identified by footprint analysis (20).

Two observations suggest that the A+T rich, anti-bent structure is important for prokaryotic origin function. First, the A tracts are not limited to the direct repeats themselves. An average of 3.1 tracts of A_3 or greater was found in the direct repeats, which span an average of 41.5 bp, while the 80 bp regions surrounding the repeats had an average of 5.3 A tracts. Second, the straight structure is produced by two different ways of avoiding the 10−11 bp bending periodicity of A tracts; some origins have an 8 bp periodicity of A tracts while others have a 12 bp periodicity (data not shown). Initiator proteins may catalyze duplex opening in prokaryotic origins by stepwise melting of the direct repeats (1−4). The ARS consensus domain has also been shown to undergo unwinding in response to the torsional stress of superhelicity and this effect has been attributed to the A+T richness of the region and to additional features whose precise nature is unknown (24). Our results show that the A+T rich, anti-bent structure is found in the regions where duplex opening is likely to occur in the prokaryotic and eukaryotic replicators (Figures 1 and 2, ref. 11) and the structure may play some role in this process. Perhaps the non-B-DNA in the A tracts (25) on different sides of the helix creates an unstable conformation which can be relieved by unwinding. Considerations at this level of structure are appropriate since an analysis of nearest neighbor nucleotide frequencies in the replicators did not yield unusually low total free energies of base stacking interactions (data not shown). Alternatively, the straight DNA in origins may facilitate duplex melting by conferring resistance to binding of duplex stabilizing proteins such as histones for ARS elements and the histone-like proteins in prokaryotes, in a manner similar to the way in which the rigid poly(dA)-(dT) resists nucleosome formation (26).

The prokaryotic origins of replication studied in this report share common structural features with yeast replicators. Does this conservation in structure extend to function as well? The *S. aureus* plasmids pC194, pC221, and pE194 replicate autonomously in the yeast *S. cerevisiae*, as well as in *B. subtilis* and *E. coli* (27−29). No 11/11 matches to the ARS consensus sequence are present in the three origin regions and the total number of 9/11 and 10/11 matches is 60, compared to the expected occurrence of 47. Computer analysis of the bacterial origins of the plasmids revealed bending elements with ENDS ratios of 1.19, 1.15, and 1.16 in pC194, pC221, and pE194, respectively, and regions of A+T rich DNA with ENDS ratios which are less than one standard error below the mean for 107 kb of natural DNA. These observations lend support to the view that conserved DNA structures play an important and universal role in DNA replication.

ACKNOWLEDGEMENTS

We gratefully acknowledge critical reading of the manuscript by David J. Asai and helpful discussions with P. T. Gilham. Special thanks go to Irwin Tessman for help with the calculations of the probability of occurrence of consensus sequences. This work was supported by grant #RO1 GM41708-01 from the National Institutes of Health.

REFERENCES

1. Bramhill,D. and Kornberg,A. (1988) Cell *54*, 915−918.
2. Bramhill,D and Kornberg,A. (1988) Cell *52*, 743−755.
3. Schnos,M.,Zahn,K.,Inman,R.B., and Blattner,F. (1988) Cell *52*, 385−395.
4. Mukherjee,S.,Patel,I., and Bastia,D. (1985) Cell *43*, 189−197.
5. Hsiao,C. and Carbon,J. (1979) Proc. Natl. Acad. Sci. USA *76* 3829−3833.
6. Williamson,D.H. (1985) Yeast *1*, 1−14.
7. Huberman,J.A.,Spotila,L.D.,Nawotka,K.A.,El-Assouli,S.M., and Davis,L.R. (1987) Cell *51*, 473−481.
8. Broach,J.R.,Li,Y.,Feldman,J.,Hayaram,M.,Abraham,J.,Nasmyth,K., and Hicks,J. (1983) CSHSQB *47*, 1165−1173.
9. Celniker,S.E.,Sweder,K.,Srienc,F.,Bailey,J.E., and Campbell,J.L. (1984) Mol. Cell. Biol. *4*, 2455−2466.
10. Strich,R.,Wootner,M., and Scott,J.F. (1986) Yeast *2*, 169−178.
11. Eckdahl,T.T. and Anderson,J.N. (1987) Nucl. Acids Res. *15*, 8531−8545.
12. Eckdahl,T.T.,Bennetzen,J.L., and Anderson,J.N. (1989) Plant Mol. Biol. *12*, 507−516.
13. Koo,H.S.,Wu,H.M., and Crothers,D.M. (1986) Nature (London) *320*, 501−506.
14. Nadeau,J.G. and Crothers,D.M. (1989) Proc. Natl. Acad. Sci. USA *86*, 2622−2626.
15. Marini,J.C.,Effron,P.N.,Goddman,T.C.,Singleton,C.K.,Wells,R.D.,Wartell,R.M., and Englund,P.T. (1984) J. Biol. Chem. *259*, 8974−8979.
16. Ulanovsky,L. and Trifonov,E.N. (1987) Nature (London) *326*, 720−722.
17. Williams,J.S.,Eckdahl,T.T., and Anderson,J.N. (1988) Mol. Cell. Biol. *8*, 2763−2769.
18. Anderson,J.N. (1986) Nucl. Acids Res. *14*, 8513−8533.
19. Snyder,M.,Buchman,A.R., and Davis,R.W. (1986) Nature (London) *324*, 87−89.
20. Diffley,J.F.X. and Stillman,B. (1988), in Cancer Cells *6*, Eukaryotic DNA Replication, Cold Spring Harbour Laboratory 235−243.
21. Zahn,K. and Blattner,F.R. (1987) Science *236*, 416−422.
22. Stenzel,T.T.,Pankaj,P., and Bastia,D. (1987) Cell *49*, 709−717.
23. Palzkill,T.G. and Newlon,C.S. (1988) Cell *53*, 441−450.
24. Umek,R. and Kowalski,D. (1987) Nucl. Acids Res. *15*, 4467−4480.
25. Aymami,J.,Coll,M.,Frederick,C.A.,Wang,A.H.-J., and Rich,A. (1989) Nucl. Acids Res. *17*, 3229−3245.
26. Prunell,A. (1982) EMBO Jour. *1*, 173−180.
27. Goursot,R.,Goze,A.,Niaudet,B., and Ehrlich,S.D. (1982) Nature (London) *298*, 488−490.
28. Polumienko,A.L., and Grigor'eva, S.P. (1986) Mol. Biol. *20*, 183−186.
29. Horinouchi,S. and Weisblum,B. (1982) J. Bact. *150*, 804−814.