Nucleic Acids Research, 1994, Vol. 22, No. 13 2479–2489

Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA

Drena L.Dobbs[1,2,*], Wen-Ling Shaiu and Robert M.Benbow[1,2,3]
Department of Zoology and Genetics, [1]Interdepartmental Genetics Major, [2]Molecular, Cellular, and Developmental Biology Program and [3]Nucleic Acid Research Facility, Iowa State University, Ames, IA 50011, USA

Received April 19, 1994; Revised and Accepted May 23, 1994

ABSTRACT

We have postulated that chromosomal replication origin regions in eukaryotes have in common clusters of certain modular sequence elements (Benbow, Zhao, and Larson, BioEssays 14, 661–670, 1992). In this study, computer analyses of DNA sequences from six origin regions showed that each contained one or more potential initiation regions consisting of a putative DUE (DNA unwinding element) aligned with clusters of SAR (scaffold associated region), and ARS (autonomously replicating sequence) consensus sequences, and pyrimidine tracts. The replication origins analyzed were from the following loci: *Tetrahymena thermophila* macronuclear rDNA gene, Chinese hamster ovary dihydrofolate reductase amplicon, human c-*myc* proto-oncogene, chicken histone H5 gene, *Drosophila melanogaster* chorion gene cluster on the third chromosome, and Chinese hamster ovary rhodopsin gene. The locations of putative initiation regions identified by the computer analyses were compared with published data obtained using diverse methods to map initiation sites. For at least four loci, the potential initiation regions identified by sequence analysis aligned with previously mapped initiation events. A consensus DNA sequence, WAWTTDDWW-WDHWGWHMAWTT, was found within the potential initiation regions in every case. An additional 35 kb of combined flanking sequences from the six loci were also analyzed, but no additional copies of this consensus sequence were found.

INTRODUCTION

Twenty-five years ago, Huberman and Riggs postulated the existence of fixed sites of initiation of DNA replication (origins) in eukaryotic chromosomal DNA based on results of DNA fiber autoradiography experiments performed in mammalian cells (1). Chromosomal DNA replication was posited to begin at the origins, and to proceed outward bidirectionally by means of fork-like growing points—branched structures in which DNA unwinding and DNA synthesis were tightly coupled. Although the paradigm of bidirectional replication forks emanating from a fixed origin has dominated interpretation of most subsequent studies, data proving the existence of specific origin sequences in higher eukaryotes are lacking.

The evidence in favor of specific origins of replication in mammalian and other eukaryotic organisms has recently been reviewed (2–8). As an alternative, we have proposed that origins of chromosomal replication in eukaryotes differ significantly from the well-characterized fixed sites of initiation used by prokaryotes such as *E.coli* [*ori* C] and mammalian tumor viruses such as SV40 [*ori* sequence] (9). We postulated that localization of initiation events in eukaryotes is determined by statistical probabilities (10, 11). In this view, any single initiation event in a chromosomal origin region depends on the joint probabilities of unwinding the duplex DNA and initiation of strand synthesis within the unwound region (9). Multiple sites of initiation are possible, but a few sites will be strongly preferred and thus may appear to be specific origin sequences (12).

Initiation at multiple sites rather than at one specific site has been observed in several eukaryotic organisms, including *Drosophila melanogaster* (13–15), *Xenopus laevis* (16, 17), and mammalian cells (3, 4, 18–20). However, other studies in the same organisms report initiation at fixed sites (21–26). The concept of statistically preferred start sites resolves the paradox posed by these apparently contradictory observations.

We have used a series of computer programs to analyze the DNA sequences from six eukaryotic chromosomal origin regions (and their flanking sequences). Each of these had been identified, characterized, and sequenced previously by other research groups (Table 1). The sequences analyzed are from a protozoan, a protostome, and various deuterostomes, and have widely divergent AT content. We tested the hypothesis that eukaryotic origin regions entail clusters of common modular sequence elements (9, 27, 28). Our analyses showed that each contained one or more potential initiation regions consisting of a putative DUE (DNA unwinding element) aligned with clusters of SAR (scaffold associated region) and ARS (autonomously replicating sequence) consensus sequences, and pyrimidine tracts. A

*To whom correspondence should be addressed at: 2114 Molecular Biology Building, Iowa State University, Ames, IA 50011, USA

consensus sequence was identified that was unique to the potential initiation regions. In addition, we have elaborated a proposed alternative mechanism for initiation at chromosomal origins in eukaryotes (9).

MATERIALS AND METHODS

Computer analyses of chromosomal DNA sequences

Six chromosomal DNA sequences retrieved from the GenBank/EMBL Genetic Data Base (Table 1) were analyzed using Oligo 4.0 on a Macintosh IIci (National Biosciences, Wolfgang Rychlik, 1992), the Thermodyn program (kindly provided by Dr David Kowalski) on an IBM 386 (29), the GentlBen program (kindly provided by Dr John N.Anderson) on an IBM 386 (30), and the University of Wisconsin Genetics Computer Group (GCG) Sequence Analysis Software Package (Version 7.3-UNIX) on a DEC 3100 workstation (31).

The primer stability analysis program of Oligo 4.0 (National Biosciences, Wolfgang Rychlik, 1992) was used to identify putative DNA unwinding elements (DUE-like elements), which are inherently unstable duplex DNA segments characterized in part by their hypersensitivity to single-strand-specific nucleases (32, 33). The program was used to identify the least stable 100 nucleotide long primers (defined as those with negative free energy values [ΔG in kcal/mole] closest to zero). Duplex stability calculations were based on nearest-neighbor dinucleotide free energy values determined by Breslauer and colleagues (34). The commercially available Oligo 4.0 program contained additional software modifications kindly provided by Dr Wolfgang Rychlik to permit graphic analysis using Microsoft Excel or Delta Graph. Preliminary analyses indicated that the optimal segment or window size was 100 nucleotides. Each data point in the DUE-like element plots in Figures 1B–4B, therefore, corresponds to the negative free energy of a theoretical primer-template duplex consisting of the 100 downstream nucleotides. This represents the magnitude of free energy required to unwind a 100 bp duplex of the analyzed sequence.

A major DUE-like element was defined as at least 200 contiguous nucleotides with an average ΔG at least 15 kcal/mole less negative than the mean for the sequence analyzed. A minor DUE-like element was defined as at least 400 contiguous nucleotides with an average ΔG at least 10 kcal/mole less negative than the mean for the sequence analyzed. These definitions differ slightly from our previous definitions (9), but identify the same elements.

The Thermodyn program of Kowalski and colleagues (29) can also be used to calculate the free energy difference (ΔG) between duplex and single-stranded states for multiple overlapping segments of DNA sequence. Free energy differences calculated using Thermodyn were identical to those calculated by Oligo 4.0 (the plots can be overlaid), but absolute free energy values (ΔG) differed by about 70 kcal/mole because of assumptions regarding temperature and ionic strength.

Putative bent DNA sequences were identified from ENDS ratio plots generated by the GentlBen program (30). A window size of 120 bp and a step size of 10 bp were used. The ENDS ratio is a measure of DNA bending and is calculated as the ratio of the contour length along a given axis to the shortest distance between the ends (35).

Consensus sequences were identified using the GCG program Findpatterns with 0, 1, or 2 mismatches as indicated. Results were graphed using Mapplot. Additional modular elements not shown in Table 2 were also analyzed. A complete list of all elements searched for may be obtained by contacting the corresponding author (D.L.D.). A cluster was designated when consensus sequences were found within a region (usually a putative DUE) at least 5-fold more frequently than in the total remaining sequence.

The putative initiation region consensus sequence was first detected by visual inspection of an alignment of putative initiation region sequences generated by the GCG progra Pileup. The consensus sequence was refined by manual alignment of sequences with the Pileup consensus, followed by trial and error replacement of consensus nucleotides at each position. The locations of the optimized initiation region consensus sequence were identified using Findpatterns (Table 3).

The statistical significance of the analyses in this study was validated using shuffle techniques (36). DNA sequences were randomized using the GCG program Shuffle. No clusters of major DNA unwinding elements, modular elements, and bent DNA regions were observed in multiple shuffled versions of any of the six sequences analyzed. The initiation region consensus sequence was not found in shuffled versions of any of the six complete sequences analyzed, and its natural frequency of occurrence in the rDNA 5′NTS was easily the extreme value [making it statistically significant, see (36)] relative to 20 shuffled versions of the 5′NTS (which is the most AT-rich initiation region of the sequences analyzed).

The statistical significance of the initiation region consensus sequence was also analyzed by calculating the expected number of occurrences in the sequences searched. These calculations were suggested and carried out by Dr David Landsman at NIH.

RESULTS AND DISCUSSION

Rationale

We selected for analysis six chromosomal DNA sequences representing putative replication origin regions in eukaryotes that satisfied the following criteria: (i) Preferred DNA replication start sites or localized initiation events were associated with a defined gene locus; (ii) High quality annotated DNA sequence was available in the GenBank/EMBL Genetic Data Base; (iii) The DNA sequence could be aligned with physical, genetic, or biochemical data used to map initiation events; (iv) The region sequenced was small (< 12 kb) and, therefore, less likely to contain multiple origins; (v) The same locus [for example, ribosomal genes (rDNA)] from different organisms was not included to avoid biasing for common elements unrelated to origin regions; (vi) Diverse species were represented, although there remains a bias in favor of mammalian origin regions. Origin regions in the yeasts, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, are analyzed elsewhere (Dobbs and Benbow, in preparation).

The six chromosomal DNA sequences (Table 1) were searched for the presence of the modular sequence elements listed in Table 2. Each of these elements might play a role in initiation, either by facilitating unwinding of the origin region, or by mediating the initiation of strand synthesis. The rationale for selection of these elements has been discussed previously (9). In the analyses that follow, we show that a putative DNA unwinding element (a DUE-like element) and a cluster of modular elements are aligned, within the limits of resolution of the localization techniques, with independently mapped initiation events. The modular elements that appear to align most frequently are SAR

Table 1. DNA sequences of chromosomal replication origin regions

GenBank accession number	Locus	Organism	Length (nt)	A−T (%)	References
X54512	macromolecular rDNA gene	*Tetrahymena thermophila*	10,315	64	(93)
X52034	dihydrofolate reductase (dhfr) amplicon	*Cricetulus griseus*	6,157	59	(56)
K01910	c-*myc* proto-oncogene	*Homo sapiens*	2,500	48	(58)
X00169	histone H5 gene	*Gallus gallus*	3,868	42	(65, 94)
X06257	chorion s18-1 and s15-1 gene cluster	*Drosophila melanogaster*	3,850	56	(95)
X61084	rhodopsin gene (opsin)	*Cricetulus griseus*	11,931	47	(28)

Table 2. Common modular sequence elements

Abbrev.	Element	DNA sequence or structure	Computer program	References
DUE	Putative DNA Unwinding Element	Inherently low helix stability: $-\Delta G$ (kcal/mole)	Oligo 4.0	(29, 32, 96, 97)
Bent DNA	Putative bent DNA predicted from ENDS ratio	dA_{2-6} repeated with a 10 to 11 nt periodicity	GentlBen	(30, 35, 98)
SAR A1	Scaffold Associated Sequence	AATAAAYAAA	Findpatterns	(99)
SAR A2	Scaffold Associated Sequence	WADAWAYAWW	Findpatterns	(99)
SAR T1	Scaffold Associated Sequence	TTWTWTTWTT	Findpatterns	(99)
SAR T2	Scaffold Associated Sequence	TWWTDTTWWW	Findpatterns	(99)
MAR	Matrix Attachment Sequence	AATATTTTT	Findpatterns	(100, 101)
PYR	Pyrimidine Tract	YYYYYYYYYYY	Findpatterns	(86, 87, 88, 89)
GAGA	Putative triple helices	GAGAGAGAGAGA	Findpatterns	(56, 102, 103, 104)
ARS-Sc	ARS (*S.cerevisiae*)	WTTTAYRTTTW	Findpatterns	(79, 105)
TopoII	DNA Topoisomerase binding site	IIGTNWAYATTNATNNG	Findpatterns	(106, 107)
Purl	*Pur* binding sequence	GGNNGAGGGAGARRRR	Findpatterns	(108, 109)

The sequence symbols used are R = A or G, W = A or T, Y = C or T, D = A or G or T, N = A or C or G or T.

and ARS consensus sequences and pyrimidine tracts. We call each region of alignment a potential initiation region.

Based on these alignments, we propose that the following algorithm (see Materials and Methods for details) can be used to identify potential initiation regions. First, each major and minor DUE-like element is identified. If a DUE-like element contains a cluster of SAR and ARS consensus sequences and is aligned with or flanked by pyrimidine tracts, it is designated as a potential initiation region. Potential initiation regions are shown as gray areas in Figures 1−4. Note that each DUE-like element extends 100 base pairs beyond its apparent 3′ end because of the search window used.

Tetrahymena thermophila macronuclear rRNA genes

The amplified rRNA genes (rDNA) in *Tetrahymena thermophila* are present in the macronucleus as 21 kb linear palindromic molecules (Figure 1), each containing two copies of the rRNA gene in inverted orientation (37). An origin of replication has been mapped within the 5′ non-transcribed spacer region (5′NTS) by electron microscopy (38). Restriction fragments that encompass the 5′NTS support autonomous replication in the macronucleus (39). We analyzed a 10,315 base pair sequence corresponding to one half of the palindrome, encompassing the 5′NTS, rRNA coding region and 3′NTS sequences (Figure 1).

The rDNA sequence contains a major DUE-like element in the 5′NTS extending from nucleotide position 1 (center of palindrome) to 1,400 (Figure 1B). Two DUE-like elements in the 3′NTS extend from 8,500 to 9,150 (minor) and from 9,750 to the end of the sequence (major) [only one telomeric GGGGTT repeat is included in the analyzed sequence]. Numerous regions of strongly bent DNA are predicted in the 5′NTS: two of these have ENDS ratios >1.4, two have ENDS ratios >1.3 (one of which is at the center of the palindrome), and several others have ENDS ratios >1.2 (Figure 1C). In contrast, the coding region has no predicted bent DNA segments. The 3′NTS is predicted to contain one strongly bent region with an ENDS ratio >1.4.

Scaffold associated region (SAR) consensus sequences are abundant in both the 5′NTS and 3′NTS, but rare in the coding region (Figure 1D). Pyrimidine tracts were found in both the 5′NTS and 3′NTS, and are more or less uniformly spaced along the entire sequence. No $(GA)_n$ and one Pur motif were identified. Five perfect ARS consensus sequences were found in the AT-rich 5′NTS and 3′NTS regions. Allowing two mismatches to the ARS consensus resulted in identification of over 200 ARS elements; these are concentrated in the NTS regions, presumably as a result of the AT-rich sequences localized there. DNA topoisomerase II cleavage site consensus sequences are rare in the 5′NTS, but otherwise are scattered throughout the entire sequence. In this study, a 21 bp consensus sequence unique to the potential initiation regions was identified (see Materials and Methods). Six of these initiation region consensus sequences were found in the rDNA: three in the 5′NTS and three in the 3′NTS (two originating in opposite orientations from nucleotide 10,100).

The origin region mapped by electron microscopy lies 650 ± 300 bp from the center of the 21kb palindrome (38) (Figure 1).

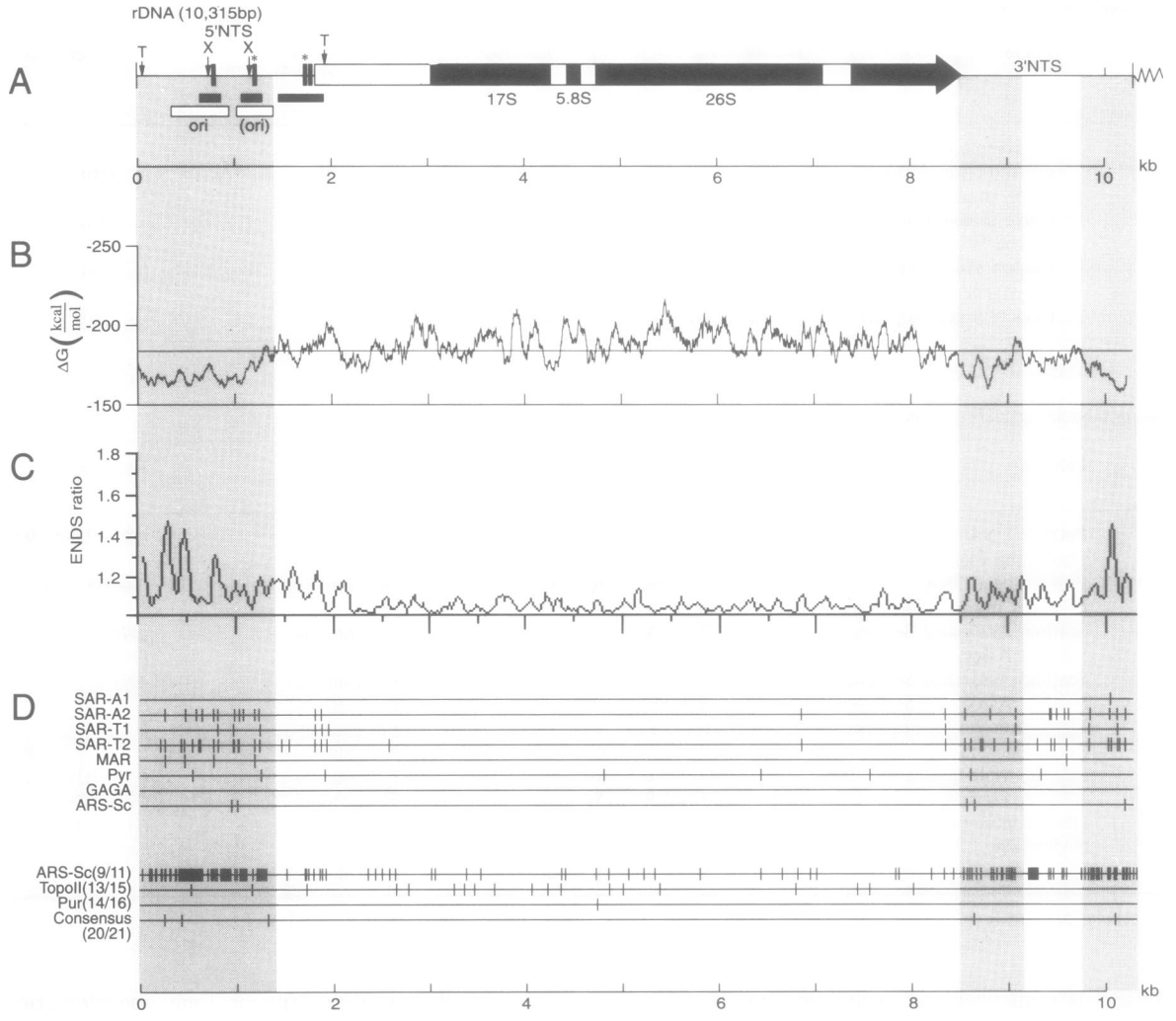

Figure 1. Computer analyses of the *Tetrahymena* rRNA gene region. (**A**) Schematic diagram of one half of the 21 kb palindromic rDNA. The center of the palindrome is shown at 0 kb. Solid boxes indicate exons; open boxes, introns; zig-zag line represents telomeric repeat sequences. The arrowhead indicates the direction of transcription. Putative replication origin(s) mapped by electron microscopy are indicated as open bars (38, 110). Nuclease hypersensitive regions in chromatin (111) are indicated as solid horizontal bars. Type I repeat sequence elements conserved among several ciliate species (112, 113) are represented as solid vertical bars. Restriction sites for XbaI (X) and TaqI (T) that are relevant to these analyses are shown. (**B**) Localization of predicted DUE-like elements. Negative free energy (ΔG) required to unwind 100 bp windows of duplex DNA was calculated and graphed as a function of nucleotide position. Valleys represent inherently unstable duplex DNA segments. (**C**) Predicted regions of intrinsically bent DNA. ENDS ratios were determined for 120 bp windows and graphed as a function of nucleotide position. Peaks represent regions predicted to display most pronounced DNA bending. (**D**) Distribution of consensus sequences. Consensus sequences for the elements shown in Table 2 were identified. Only perfect matches to consensus sequences are shown in the top 8 lines; elements with up to 2 mismatches are plotted in the lower 4 lines. Consensus refers to the putative initiation region consensus sequence identified in this study (see text and Table 3). Potential initiation regions (see text) based on the analyses in (B–D) are shown in gray.

This region is free of nucleosomes and hypersensitive to nuclease digestion in chromatin (40, 41), a feature common to viral and prokaryotic replication origins, as well as to the yeast ARS1 origin (42, 43). Approximately 400 bp of this origin region and its nuclease hypersensitivity are duplicated in the 5′NTS, generating a second potential origin (Figure 1A). Several mutations that affect replication and/or maintenance of the rDNA in the macronucleus (44, 45) are also located in the 5′NTS (asterisks in Figure 1A). All of these data are consistent with initiation events occurring within the potential initiation region (the leftmost gray area in Figure 1), as well as with the lack of clusters of modular sequence elements within the transcribed region. The existence in the 5′NTS of a DUE (based on sensitivity to mung bean nuclease digestion), regions of strongly bent DNA, and a restriction fragment preferentially associated with the nuclear matrix have been verified experimentally (46). Our computer analyses suggest that initiation events might also occur in the 3′NTS adjacent to the telomere.

Chinese hamster ovary dihydrofolate reductase locus

The origin region in the dhfr amplicon (Figure 2, left) is the most heavily studied of any metazoan origin region and has been reviewed extensively elsewhere (3, 4, 6, 8, 12, 47). Apparently specific initiation events in the *ori-β* region have been mapped by detection of the earliest labeled DNA fragments (48–51), by a replication origin trap (52), by nascent DNA fragment lengths (26), by the distribution of Okazaki fragment lengths (21), and by imbalanced DNA synthesis (53, 54). The apparent site of

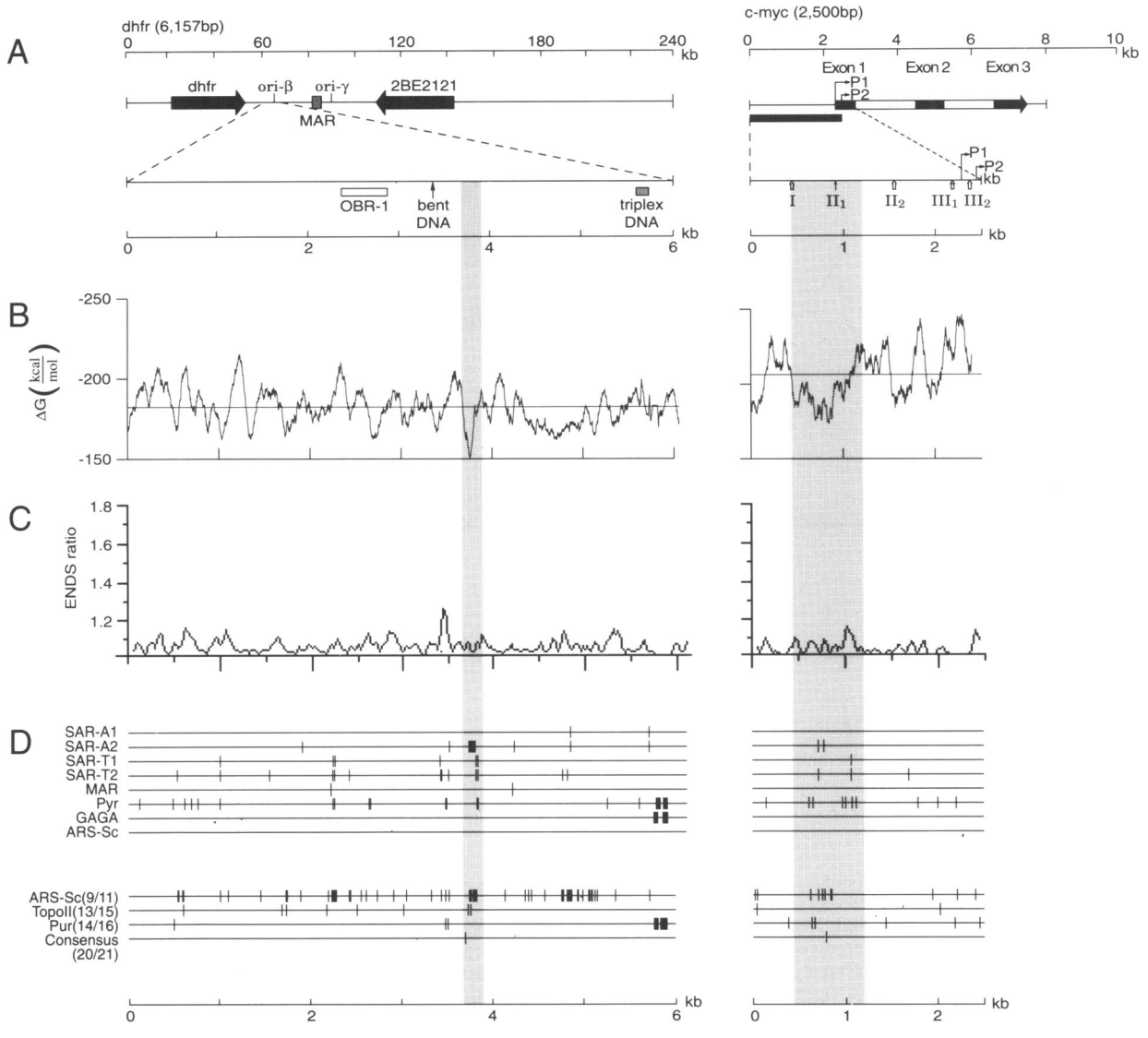

Figure 2. (**left**): Computer analyses of the Chinese hamster ovary dhfr locus. (**A**) Map of the 240 kb Chinese hamster ovary dihydrofolate reductase amplicon. Symbols are as in Figure 1. ori-β and ori-γ indicate origin regions mapped in (114). MAR, matrix attachment region (115). Inset: map of the region analyzed in (B–D). Locations of OBR-1 (origin of bidirectional replication), bent DNA, and triplex DNA are from (56). (**B–D**) are as in Figure 1. (**right**): Computer analyses of the human *c-myc* locus. (**A**) Map of the c-*myc* gene and 5′ flanking region. P_1 and P_2 are alternative promoter start sites for the *c-myc* mRNA. The dashed lines indicate the location of the sequence analyzed in (B–D). The nuclease hypersensitive region in chromatin is indicated by a solid horizontal bar. Inset: map of 5′ flanking region analyzed in (B–D). Arrows indicate nuclease hypersensitive sites; the size of the arrows is a qualitative representation of the intensities (58). (**B–D**) are as in Figure 1.

initiation has been designated OBR-1 (origin of bidirectional replication) (21). In contrast, initiation events have also been detected throughout a 30 to 70 kb initiation zone using neutral–neutral two-dimensional gel electrophoresis (18, 19, 55). The apparent contradiction between assays that detect specific initiation events localized to a region as small as 450 nucleotides, and assays that detect initiation events throughout the 70 kb region has generated considerable controversy (7). We analyzed a 6,157 base pair sequence containing *ori-β* and the OBR-1 (Figure 2, left).

The dhfr sequence contains a major DUE extending from nucleotide position 3,700 to 3,900 (Figure 2B, left). A minor DUE-like element extends from 4,350 to 5,100. One region of bent DNA with an ENDS ratio >1.2 is predicted at nucleotide position 3,450 (Figure 2C).

SAR consensus sequences are clustered within the major DUE (Figure 2D, left). Pyrimidine tracts were found within the major DUE and are also spaced along the remaining sequence. Two previously identified $(GA)_n$ tracts (with associated complementary strand pyrimidine tracts) were found (56). Pur motifs were identified associated with the $(GA)_n$ tracts. No perfect ARS consensus sequences were found, but allowing two mismatches to the ARS consensus resulted in clusters in the major DUE and OBR-1 region. DNA topoisomerase II cleavage site consensus sequences were found in the major DUE and the OBR-1 region, and were infrequent in the remainder of the sequence. One initiation region consensus sequence was identified in the major DUE.

Initiation events in the dhfr ori-β region have been mapped within a 450 bp region called OBR-1 (Figure 2, left). However,

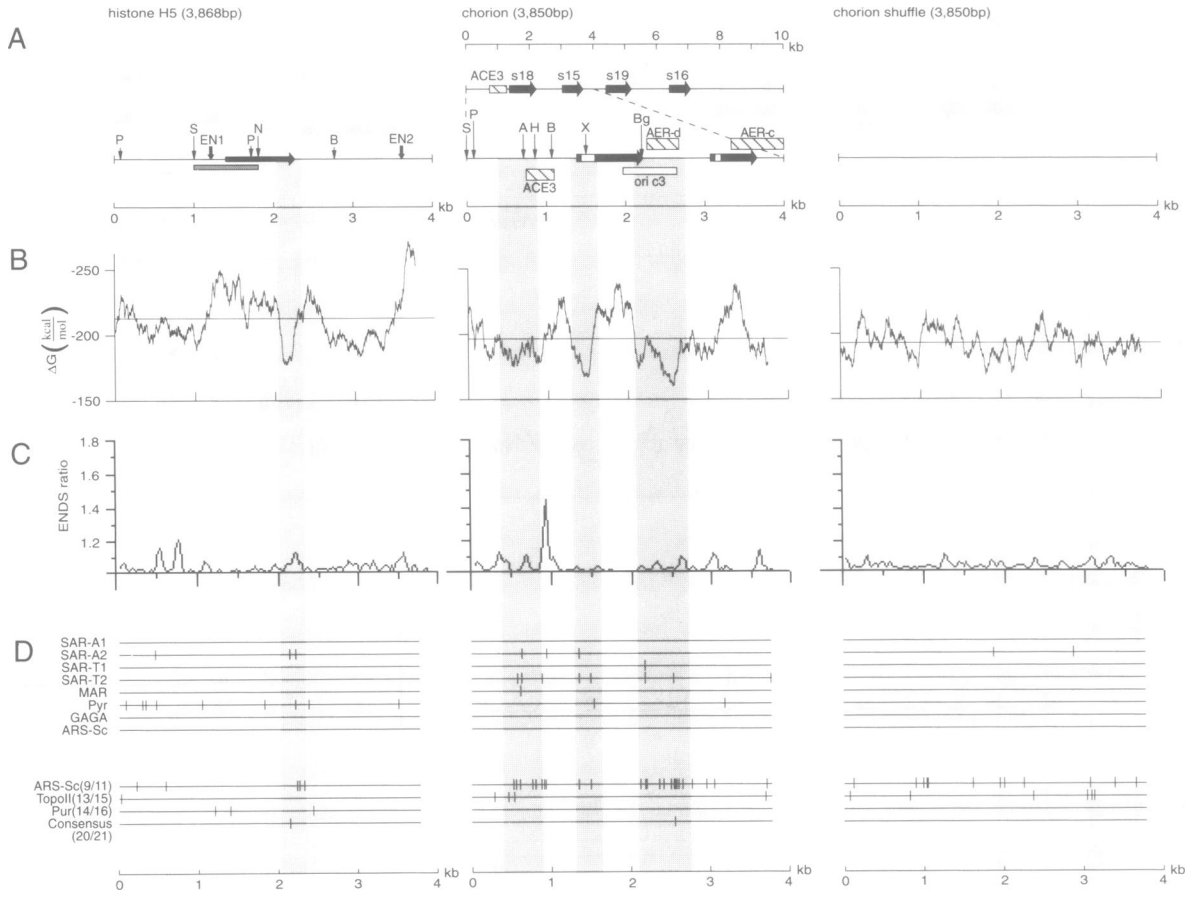

Figure 3. (**left**): Computer analyses of the chicken histone H5 locus. (**A**) Map of the chicken histone H5 gene transcript and flanking regions. Restriction enzyme sites (thin arrows) and endogenous nuclease cleavage sites (thick arrows) used in *in vitro* nuclear runoff DNA replication assays are indicated (66). Relevant restriction enzyme cleavage sites for PstI (P), SacI (S), NaeI (N), BamHI (B) are shown; EN1, EN2 are endogenous nuclease cleavage sites. (**B–D**) are as in Figure 1. (**center**): Computer analyses of the *D.melanogaster* chorion gene cluster. (**A**) Map of the chorion gene cluster on the third chromosome. Inset: map of the region (containing s18 and s15 chorion genes and flanking regions) analyzed in (**B–D**). Amplification control element 3 (ACE 3) or amplification enhancing regions (AER-d, AER-c) are indicated by cross-hatched horizontal bars. Open bar indicates ori C3 determined in (15). Relevant restriction enzyme cleavage sites for SalI (S), PvuII (P), Asp 718 (A), HindIII (H), BamHI (B), XmnI (X), BglI (Bg) are shown. (**B–D**) are as in Figure 1. (**right**): Computer analyses of a shuffled *Drosophila melanogaster* chorion gene cluster sequence. The 3.8 kb sequence analyzed in Figure 3 center, was randomized using the Shuffle program as described in Materials and Methods. (**B–D**) are as in Figure 1.

our computer analyses predict that unwinding is more likely to occur in a region about 1.2 to 1.5 kb downstream from OBR-1, rather than at OBR-1. Caddle and colleagues have shown that the OBR-1 is relatively refractory to unwinding (56). The potential initiation region (the gray area in Figure 2, left) contains a major DUE, a very substantial cluster of modular elements, and the initiation region consensus sequence. A possible explanation for the appearance of multiple initiation sites throughout a 30 to 70 kb region is provided by the strand separation model: unwinding of an entire chromosomal domain of duplex DNA would expose numerous sites for independent initiation of strand synthesis throughout the unwound region (11, 57).

Several features of the dhfr initiation region identified by our computer analyses have been verified experimentally. The major DUE was the preferred mung bean nuclease cleavage site in every context tested (56) confirming its identification as a DUE. The predicted bent DNA region between nucleotides 3410 to 3460 is within a restriction fragment that migrates anomalously on polyacrylamide gels, suggesting that it contains a bent region (56). Finally, pyrimidine tracts, which appear to serve as preferred start sites for initiation of strand synthesis (see Table 2), flank OBR-1 and provide a plausible explanation for the strand switching observed in this region (21).

Human c-*myc* locus

The c-*myc* gene is the normal cellular homologue of a transforming gene associated with tumors such as Burkitt lymphomas (58). A putative origin region was mapped to the 5′ flanking region 1.5 ± 2.0 kb upstream of the first exon of the c-*myc* gene by three independent methods (see below). We analyzed a 2,500 base pair sequence containing the 5′ flanking region and the initiation site of the c-*myc* primary transcript (Figure 2A, right).

The c-*myc* sequence contains a major DUE-like element extending from nucleotide position 550 to 1,100 (Figure 2B, right). Another major DUE-like element extends from 1 to 150, but is not associated with any SAR consensus sequences. There are no predicted regions of bent DNA with ENDS ratios >1.2, but a weakly bent region with an ENDS ratio >1.15 is predicted at nucleotide position 1,030 (Figure 2C, right).

SAR consensus sequences are moderately clustered in the major

DUE-like element at 550 to 1,100 (Figure 2C, right). Pyrimidine tracts were found within this major DUE-like element, and were also spaced along the remaining sequence. There were no $(GA)_n$ tracts and no perfect ARS sequences. Clusters of Pur motifs and ARS sequences allowing two mismatches were identified in the major DUE-like element, with a few scattered elsewhere. Only two DNA topoisomerase II cleavage site consensus sequences were found, one of which lies within a DUE-like element. One initiation region consensus sequence was identified within the major DUE-like element.

Leffak and colleagues (59–61) mapped the origin region to a location upstream of the c-*myc* gene by two separate methods. In an *in vitro* replication run-off assay, a transcribed c-*myc* gene was replicated from the 5′-upstream region in HeLa cells (60). The initiation region was close to, but upstream of, a cleavage site located between exon 1 and exon 2 (Figure 2, right). Both transient and long term replication assays indicate an initiation region localized to a 2.4 kb restriction fragment of the c-*myc* 5′ flanking DNA (61, 62). Bisection of this fragment indicated that the 1.2 kb immediately upstream from the P1 site is sufficient to confer autonomous replication. Polymerase chain reaction amplification of nascent DNA fragments indicated an initiation zone centered about 1.5 kb upstream of exon 1 (25). Both polymerase chain reaction amplification of nascent DNA fragments and strand switching analysis of 'Okazaki' fragments showed that there are multiple sites of initiation of strand synthesis within the region mapped in Figure 2, right (M.Leffak, personal communication). These data suggest that an initiation region is centered near the major DUE-like element, but that there are significant sites of initiation of strand synthesis on both sides as well as beyond the sequenced region. All of the above observations are consistent with the potential initiation region defined by our computer analyses (the gray area in Figure 2, right), and with the observed distribution of pyrimidine tracts.

Chicken histone H5 locus

The intronless chicken histone H5 gene (Figure 3, left) is a single copy gene that is not part of a histone gene cluster, and is not tightly linked to other histone genes (63–65). Nuclear run-off assays (66) showed that in avian embryonic erythrocytes where the H5 gene is transcribed, replication initiates from a zone located in the 5′ flanking region. In contrast, the inactive histone H5 gene of lymphoblastoid or chicken embryo fibroblast cell lines is replicated from a downstream origin (66). Transcription may facilitate use of the 5′ origin region by inducing alterations in chromatin structure (67). We analyzed a 3,868 base pair sequence containing the H5 gene and flanking 5′ and 3′ sequences (Figure 3A, left).

The histone H5 sequence contains a major DUE-like element extending from nucleotide position 2,050 to 2,350 (Figure 3B, left). Two additional DUE-like elements extend from 600 to 1,250 and from 2,600 to 3,600, but are not associated with clusters of modular sequence elements. There is one predicted region of bent DNA with an ENDS ratio >1.2 (Figure 3C, left).

Very few SAR consensus sequences were identified, two of which (SAR-A2s) were in the major DUE-like element at 2,050 to 2,350 (Figure 3D, left). Pyrimidine tracts were found in the major DUE-like element, and were also spaced along the remaining sequence. There are no $(GA)_n$ tracts and no perfect ARS sequences. Clusters of ARS sequences allowing two mismatches were identified in the major DUE-like element, as was one Pur sequence. Pur sequences were also scattered elsewhere. Only one DNA topoisomerase II consensus sequence was found, and it was not within a DUE-like element. One initiation region consensus sequence was identified in the major DUE-like element.

Although the data of Trempe *et al.* (66) do not allow a precise localization of initiation events in the 5′ and 3′ flanking regions of the histone H5 gene, our sequence analyses show that the 3′ flanking sequence contains a putative initiation region (the gray area in Figure 3, left) and the 5′ flanking sequence has a DUE-like element, pyrimidine tracts, and region of bent DNA.

Drosophila melanogaster chorion gene cluster on the third chromosome

Amplification of the chorion genes in *Drosophila melanogaster* is mediated by defined regulatory regions and initiates at specific origin regions within a cluster at 66D11-15 on the third chromosome (Figure 3A, center) (68). These origin regions have been mapped by genetics (13, 69), by neutral–neutral 2-D gel electrophoresis (13, 69), and by electron microscopy (70, 71). We analyzed a 3,850 base pair region containing the s18 and s15 chorion genes and upstream regions (Figure 3, center).

The chorion sequence contains two major DUE-like elements extending from nucleotide positions 1,300 to 1,650 and 2,100 to 2,750 (Figure 3B, center). A minor DUE-like element extends from 400 to 900. One region of bent DNA is predicted with an ENDS ratio >1.4 (Figure 3C, center) within the minor DUE-like element.

SAR consensus sequences (SAR-T1 and SAR-T2) were found within the DUE-like elements (Figure 3C, center). Pyrimidine tracts were found in or flanking the major DUE-like elements. There were no $(GA)_n$ tracts, no perfect ARS sequences, and no Pur elements. ARS consensus sequences with two mismatches were scattered everywhere, particularly in one of the major and in the minor DUE-like elements. DNA topoisomerase II consensus sequences were found at either end of the sequence analyzed. One initiation region consensus sequence was identified in the major DUE-like element at 2,100 to 2,750.

One region within the chorion gene cluster, ori C3, showed initiation activity when analyzed by Delidakis and Kafatos using neutral–neutral 2-D gel electrophoresis (15, 72). Ori C3 (Figure 3, center) is located by a BglII cleavage site, near the upstream end of AER-d (15). Another region bounded by a BamHI and a SalI site was also able to promote a low level of amplification, whereas the regions between the HindIII and SalI sites and PvuII and BglI sites did not function as initiation regions (Figure 3, center).

Heck and Spradling (13) also used neutral–neutral 2-D gel electrophoresis (72) to localize origins of amplification in the chorion cluster. The most frequently used origin mapped within AER-d and accounted for over 70% of initiation events. AER-d coincides with the major DUE-like element at 2,100 to 2,750, the proposed replication origin consensus sequence, and the cluster of modular elements located immediately downstream from the *BglII* cleavage site, and thus corresponds to the initiation region identified by our computer analyses (Figure 3, center). In apparent contradiction with the absence of initiation events in the ACE3 region reported by Delidakis and Kafatos (15), a second, less frequently used origin was localized within the ACE3 region by Heck and Spradling (13). This inconsistency has several possible resolutions: either the frequency of small replication bubbles was beneath their threshold of detection, or adjoining sequences are essential in addition to the ACE3 in order to obtain

replication origin activity. The ACE3 region is sufficient for amplification with correct developmental specificity (73), but the amount of amplification is low, even with many copies of ACE3 in tandem (74). Swimmer *et al.* (73) have suggested that amplification elements are functionally redundant throughout the amplified region, consistent with our redundant modular element hypothesis. Redundant transcriptional elements often have multiplicative rather than additive effects (see 15), so it is not

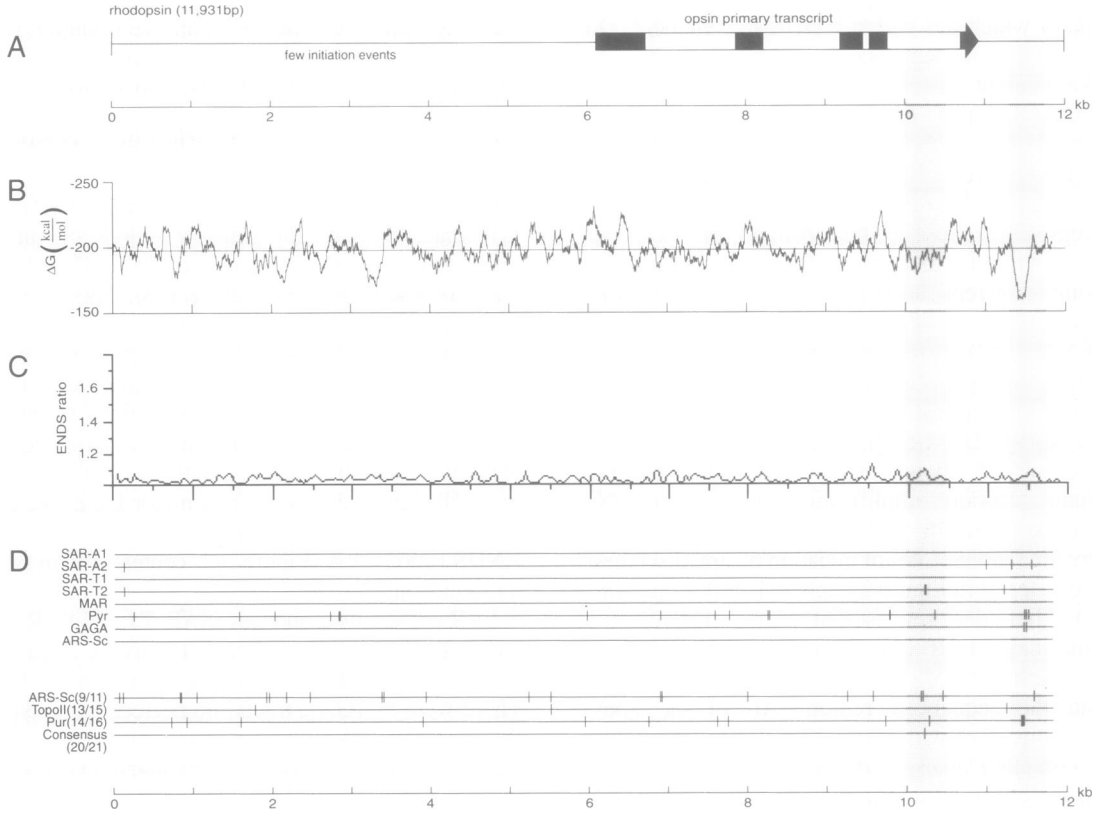

Figure 4. Computer analysis of the Chinese hamster ovary rhodopsin locus. (**A**) Map of rhodopsin gene and flanking regions. Region in which early replicating DNA was not detected is from (28). (**B–D**) are as in Figure 1.

Table 3. Localization of putative initiation region consensus cequences

Locus	Potential initiation regions	Start of consensus sequence	Consensus sequence WAWTTDDWWWDHWGWHMAWTT
rDNA	1–1,400	428	TATTTTTTAAGAgGTAAATTT
	1–1,400	1,301	AATTTTTAAAAATGAAAAAtT
	9,750–10,315	10,100	AATTTTTAATTAAaATAATTT
	1–1,400	Rev 257	AATTTGTAAATTTGTAAATTT
	8,500–9,150	Rev 8,653	TATTTGATAAAATGTAAAATc
	9,750–10,315	Rev 10,100	AAATTATTTTAATtAAAAATT
dhfr	3,700–3,900	3,698	AAATTAGTTTTAAGTCCAgTT
c-*myc*	450–1,200	778	TATaTTAAATATAGATCATTT
histone H5	2,050–2,350	2,162	AATTTTATAAGTTGATCtATT
chorion	400–900		
	1,300–1,650		
	2,100–2,750	2,544	AAATTATgTTTTTGTAAAATT
rhodopsin	10,050–10,550	10,241	TATTTGTAAAGCTGAAAAAgT
	11,350–11,750		

The sequence symbols used are M = A or C, W = A or T, D = A or G or T, and H = A or C or T.
Rev indicates that the consensus sequence was found on the complementary strand.

surprising that ACE3, the AERs, and adjoining sequences may interact synergistically to increase the frequency of initiation events.

Chinese hamster ovary rhodopsin locus

The rhodopsin gene from Chinese hamster ovary cell lines is a single-copy gene replicated very early [3 ± 2%] in the cell cycle (75), suggesting that it may lie close to an origin of DNA replication. The earliest replicating regions in this locus have been mapped to within the opsin primary transcript or in nearby 3′ flanking sequences (28). Sequence is available for 12 kb surrounding the rhodopsin gene, including 6 kb of 5′ flanking DNA in which no initiation events were detected (Figure 4A).

The rhodopsin sequence contains a major DUE-like element extending from nucleotide position 11,350 to 11,750 (Figure 4B). A minor DUE-like element also extends from 10,000 to 10,500. Another major DUE-like element extends from nucleotide position 3,100 to 3,500, and another minor DUE-like element from 1,800 to 2,350, both in the region of few initiation events. There are no predicted regions of bent DNA with an ENDS ratio >1.2 (Figure 4C).

A cluster of SAR elements was found in the major DUE-like element at 11,350 to 11, 750. SAR elements were also found in the minor DUE-like element at 10,000 to 10, 500 (Figure 4D). No clusters of sequence elements were found within the upstream DUE-like elements. Pyrimidine tracts are found in the downstream major and minor DUE-like elements and are also spaced along the remaining sequence. A previously identified $(GA)_{27}$ tract (with associated complementary strand pyrimidine tract) was found (28). Pur motifs were identified associated with the $(GA)_{27}$ tract near the major 3′ DUE-like element. No perfect ARS consensus sequences were found, but allowing two mismatches to the ARS consensus resulted in clusters in the minor downstream DUE-like element. One DNA topoisomerase II consensus sequence was found in the major downstream DUE-like element and two others elsewhere in the sequence. One initiation region consensus sequence was identified in the minor downstream DUE-like element.

Using cloned single copy DNA segments as hybridization probes to quantify replication of corresponding genomic segments, Gale and colleagues localized an origin region within a 6 kb sequence containing the opsin primary transcript and downstream region (28). Although the low resolution mapping methods used did not allow precise localization of initiation events within this 6 kb region, importantly for this study, no early replicating sequences were detected in the 6 kb region immediately upstream. We observed a cluster of common modular sequence elements, which could function as an initiation region, at the downstream end of the opsin transcript, within the 6 kb region that contained the earliest replicating DNA (Figure 4).

Within the six origin regions analyzed, some modular sequence elements appeared to be found in clusters associated with DUE-like elements. A different pattern were found when the 6 kb region upstream from opsin primary transcript was analyzed. One major and several minor DUE-like elements were observed, but modular sequence elements were rare in this region, and no clusters were found in association with the DUE-like elements. Thus, we suggest that potential initiation regions entail not only DUE-like elements, but also associated clusters of modular sequence elements.

Shuffled origin region sequences

To further investigate the significance of clusters of modular elements associated with DUE-like elements, each of the origin region sequences was randomized using the GCG Shuffle program and subjected to comparable computer analyses. In no case did a DUE-like element colocalize with a cluster of modular sequence elements: as an example, analyses of the shuffled chorion gene are shown in Figure 3 (right). No major DUE-like elements, regions of predicted strong bending or pyrimidine tracts were seen in the shuffled chorion sequence (Figure 3B, right). Fewer modular sequence elements were found (Figure 3D, right), with the exception of the DNA topoisomerase II consensus sequence. We conclude that clusters of modular sequence elements associated with the DUE-like elements do not arise randomly, and are statistically significant as defined in (36).

An initiation region consensus sequence

Based on our analyses, we designated the regions in which a DUE-like element coincided with clusters of common modular elements as potential initiation regions (the shaded gray areas in Figures 1−4). The Pileup program was then used to search for a consensus sequence within these regions. A 21 base pair sequence found in the potential initiation regions was identified as described in Materials and Methods (Table 3). This sequence, WAWTTDDWWWDHWGWHMAWTT, was not found (with one mismatch allowed) in any of the remaining 35 kb of sequences that did not contain potential initiation regions. The probability of finding the initiation consensus sequence in 40 kb of random DNA sequence is 0.0011: it was found 11 times in the six putative origin regions, always within a major DUE-like element and nowhere else. In a database search of GenBank (R) Release 70.0, October 15, 1993, 154 occurrences of the initiation consensus sequence were found in the 143,492 sequences analyzed (157,152,442 total bases). The expected number of occurrences in a random sequence of the same size was 4.45. This suggests there may be functional significance associated with the initiation consensus sequence.

A model for initiation in higher eukaryotes

We propose that initiation at chromosomal origins in eukaryotes involves a series of events that may occur with some probability anywhere, but which are more likely to occur within a cluster of redundant modular elements associated with a DUE-like element. DUEs have been implicated functionally at origin regions in *E.coli*, yeast, and SV40 (29, 32, 33, 76). The initial event in initiation in eukaryotes is probably unwinding of a relatively small (200 to 1,500 base pairs) major DUE-like element that is in close proximity to the nuclear matrix (77). Whether there are sequence-specific DNA binding proteins or complexes (78) that recognize the initiation region consensus sequence or otherwise participate in this process remains to be determined. Sequence-specific recognition *per se* is not an absolute requirement, at least for some eukaryotes, because specific origin sequences are not required in *Xenopus laevis*, sea urchin, or *Paramecium* (9). This is in contrast to a requirement for specific sequences in *Saccharomyces cerevisiae* (79), *Schizosaccharomyces pombe* (80), and *Physarum polycephalum* (Perron, personal communication).

The next event in initiation is likely to be an expansion of the unwound region, presumably by DNA helicases, to form clusters of microbubbles (10, 81−83) and 'initiation zones' (see Figure

3 in reference (9)). It seems likely that the initial unwinding of the major DUE facilitates unwinding within adjacent regions (84). Once an initiation zone is formed, initiation events at multiple sites may be an inevitable consequence. Conventional replication forks might be initiated at multiple sites throughout the initiation zone as proposed by the Jesuit model (47), or might be a consequence of the unwinding of the duplex throughout an extended region of the chromosomal domain as postulated by the strand separation model (10, 11, 85, 57).

Common modular elements in initiation zones

The most conspicuous feature of a potential initiation region is a DUE-like element. This is consistent with an absolute requirement to first unwind the duplex in order to obtain initiation. The alignment of bent regions is far less compelling: regions of strongly bent DNA are identified in some potential initiation regions, but not in others.

The individual SAR and MAR sequences do not reliably predict regions of matrix attachment. Thus, a matrix attachment region in the histone H5 gene region has been mapped to a SacI–NaeI restriction fragment (Figure 3, left) that overlaps the initiation region but does not contain any SAR or MAR consensus sequences. Similarly, a pyrimidine tract cannot be the only allowable DNA primase start site. Consensus sequences for initiation by DNA primase are beginning to emerge (86–89), however. Once these are more stringently defined, it should be possible to identify more accurately potential sites of initiation of strand synthesis.

$(GA)_n$ tracts are found only in the rhodopsin and dhfr gene regions and are associated with a major DUE-like element only for rhodopsin. Pur and DNA topoisomerase II consensus sequences also do not appear to align with potential initiation regions except in mammalian cells in which Pur sequences were found near all initiation regions. Recognition sites for a variety of transcription factors were found within the potential initiation regions, but were too abundant to include in the analyses reported here. No energetically favorable, imperfect, inverted repeats ($\Delta G < -7$ kcal) were detected in the initiation regions by the LKB 2020 DNAsis sequence analysis software (90, 91) or the available GCG programs.

Our computer analyses identified clusters of certain modular elements associated with DUE-like elements in six origin regions examined. These initiation regions corresponded with previously mapped DNA replication initiation events. Adjoining regions did not show these clusters associated with DUE-like elements. The predictive validity of the algorithm developed in this study was then tested experimentally in another organism: computer analyses of the complete *Xenopus laevis* ribosomal gene repeat sequence predicted localization of early initiation events in a cloned *Xenopus laevis* rDNA gene repeat replicated in cell-free extracts of *Xenopus laevis* eggs (92). Although additional features and elements related to chromatin structure, matrix attachment, and other requirements for initiation will undoubtedly emerge, we suggest that the paradigm developed in this study can be used to identify potential initiation regions in chromosomal DNA of eukaryotes.

ACKNOWLEDGMENTS

We are grateful to Dave Kowalski for thoughtful discussions and providing the THERMODYN program, to John Anderson for introducing us to the analysis of bent DNA and providing the GentlBen program, and to Maria Zannis-Hadjopoulos and Robert Martin for discussions and computer analyses dealing with inverted repeats. We are indebted to Dr David Landsman for showing us how to calculate the statistical significance of the initiation consensus sequence, and for carrying out the calculations. We thank Alan Spradling, Michael Leffak, Mel DePamphilis, Joyce Hamlin, John Gale and Joe D'Anna for preliminary data and discussions regarding the genes analyzed in this study. We thank Alan Myers for assistance with computer analyses, and Zhen Hou and Wolfgang Rychlik for assisting with the graphic presentation of computer analyses. We thank Jiyong Zhao, John Mayfield, Charisse Buising, Pam Hines, Al Myers, Linda Martin, and Chang-Yeol Yeo for discussions regarding the analyses and comments on the manuscript, and Skyla Larsen for her invaluable assistance with the references and manuscript preparation. We thank members of the DNA Replication Society for their thoughtful insights. This study would never have been completed without our resurrection from the excesses of our computers by Tom Hiett and Rich Beachler of Media Resources at Iowa State University. This research was supported by grants from the National Institutes of Health, the National Science Foundation, the Iowa Soybean Promotion Board, and the Office of Biotechnology and Graduate College of Iowa State University. This is Journal Paper No. J-15625 of the Iowa Agriculture and Home Economics Experiment Station, Ames IA (Project 2974 and 0164).

REFERENCES

1. Huberman, J.A. and Riggs, A.D. (1968) J. Mol. Biol. 32, 327–341.
2. Hamlin, J.L., Vaughn, J.P., Dijkwel, P.A., Leu, T.-H. and Ma, C. (1991) Curr. Opin. Cell Biol. 3, 414–421.
3. Hamlin, J.L. (1992) BioEssays 14, 651–659.
4. Hamlin, J.L., Dijkwel, P.A. and Vaughn, J.P. (1992) Chromosoma 102, S17-S23.
5. Held, P.G. and Heintz, N.H. (1992) Biochim. Biophys. Acta. 1130, 235–246.
6. Heintz, N.H., Dailey, L., Held, D.P. and Heintz, N. (1992) Trends Genet. 8, 376–381.
7. Linskens, M.H.K. and Huberman, J.A. (1990) Cell 62, 845–847.
8. Vassilev, L.T. and DePamphilis, M.L. (1992) Crit. Rev. Biochem. Mol. Biol. 27, 445–472.
9. Benbow, R.M., Zhao, J. and Larson, D.D. (1992) BioEssays 14, 661–670.
10. Benbow, R.M. (1985) In C.B. Metz and A. Monroy (eds.), Biology of Fertilization. Academic Press, New York, pp. 299–345
11. Benbow, R.M., Gaudette, M.F., Hines, P.J. and Shioda, M. (1986) In A.L. Boynton and H.L.Leffert (eds.), Cell Proliferation: Recent Advances. Academic Press, pp. 449–483.
12. DePamphilis, M.L. (1993) J. Biol. Chem. 268, 1–4.
13. Heck, M.M.S. and Spradling, A.C. (1990) J. Cell Biol. 110, 903–914.
14. Shinomiya, T. and Ina, S. (1991) Nucleic Acids Res. 19, 3935–3941.
15. Delidakis, C. and Kafatos, F.C. (1989) EMBO J. 8, 891–901.
16. Mahbubani, H.M., Paull, T., Elder, J.K. and Blow, J.J. (1992) Nucleic Acids Res. 20, 1457–1462.
17. Hyrien, O. and Mechali, M. (1992) Nucleic Acids Res. 20, 1463–1469.
18. Dijkwel, P.A., Vaughn, J.P. and Hamlin, J.L. (1991) Mol. Cell. Biol. 11, 3850–3859.
19. Vaughn, J.P., Dijkwel, P.A. and Hamlin, J.L. (1990) Cell 61, 1075–1087.
20. Vaughn, J.P., Dijkwel, P.A., Mullenders, L.H.F. and Hamlin, J.L. (1990) Nucleic Acids Res. 18, 1965–1969.
21. Burhans, W.C., Vassilev, L.T., Caddle, M.S., Heintz, N.H. and DePamphilis, M.L. (1990) Cell 62, 955–965.
22. Botchan, P.M. and Dayton, A.I. (1982) Nature 299, 453–456.
23. Hines, P.J. and Benbow, R.M. (1982) Cell 30, 459–468.
24. Hiraga, S., Sudo, T., Yoshida, M., Kubota, H. and Ueyama, H. (1982) Proc. Natl. Acad. Sci. USA 79, 3697–3701.
25. Vassilev, L. and Johnson, E.M. (1990) Mol. Cell. Biol. 10, 4899–4904.
26. Vassilev, L.T., Burhans, W.C. and DePamphilis, M.L. (1990) Mol. Cell. Biol. 10, 4685–4689.

27. Heintz, N.H. (1992) Curr. Opin. Cell Biol. 4, 459–467.
28. Gale, J.M., Tobey, R.A. and D'Anna, J.A. (1992) J. Mol. Biol. 224, 343–358.
29. Natale, D.A., Schubert, A.E. and Kowalski, D. (1992) Proc. Natl. Acad. Sci. USA 89, 2654–2658.
30. VanWye, J.D., Bronson, E.C. and Anderson, J.N. (1991) Nucleic Acids Res. 19, 5253–5261.
31. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
32. Umek, R.M. and Kowalski, D. (1988) Cell 52, 559–567.
33. Kowalski, D. and Eddy, M.J. (1989) EMBO J. 8, 4335–4344.
34. Breslauer, K.J., Frank, R., Blocker, H. and Marky, L.A. (1986) Proc. Natl. Acad. Sci. USA 83, 3746–3750.
35. Eckdahl, T.T. and Anderson, J.N. (1987) Nucleic Acids Res. 15, 8531–8545.
36. Karlin, S. and Ghandour, G. (1985) Proc. Natl. Acad. Sci. USA 82, 5800–5804.
37. Yao, M.-C. (1986) In J.G. Gall (ed.), The Molecular Biology of Ciliated Protozoa. Academic Press, Inc., Orlando, Florida, pp. 179–201.
38. Cech, T.R. and Brehm, S.L. (1981) Nucleic Acids Res. 9, 3531–3543.
39. Yu, G.L., Hasson, M. and Blackburn, E.H. (1988) Proc. Natl. Acad. Sci. USA 85, 5151–5155.
40. Palen, T.E. and Cech, T.R. (1984) Cell 36, 933–942.
41. Bonven, B.J. and Westergaard, O. (1982) Nucleic Acids Res. 10, 7593–7608.
42. Thoma, C., Bergman, L.W. and Simpson, R.T. (1984) J. Mol. Biol. 177, 713–728
43. Brown, J.A., Holmes, S.G. and Smith, M. (1991) Mol. Cell. Biol. 11, 5301–5311.
44. Yaeger, P.C., Orias, E., Shaiu, W.-L., Larson, D.D. and Blackburn, E.H. (1989) Mol. Cell. Biol. 9, 452–460.
45. Larson, D.D., Blackburn, E.H., Yaeger, P.C. and Orias, E. (1986) Cell 47, 229–240.
46. Shaiu, W.-L., Du, C., Sanzgiri, R., Choi, J.-K., Hou, Z. and Larson, D.D. (1993) Nucleic Acids Res. submitted.
47. DePamphilis, M.L. (1993) Annu. Rev. Biochem. 62, 29–63.
48. Burhans, W.C., Selegue, J.E. and Heintz, N.H. (1986) Proc. Natl. Acad. Sci. USA 83, 7790–7794.
49. Burhans, W.C., Selegue, J.E. and Heintz, N.H. (1986) Biochemistry 25, 441–449.
50. Leu, T.-H. and Hamlin, J.L. (1992) Mol. Cell. Biol. 12, 2804–2812.
51. Heintz, N.H. and Hamlin, J.L. (1982) Proc. Natl. Acad. Sci. USA 79, 4083–4087.
52. Anachkova, B. and Hamlin, J.L. (1989) Mol. Cell. Biol. 9, 532–540.
53. Burhans, W.C., Vassilev, L.T., Wu, J., Sogo, J.M., Nallaseth, F. and DePamphilis, M.L. (1991) EMBO J. 10, 4351–4360.
54. Handeli, S., Klar, A., Meuth, M. and Cedar, H. (1989) Cell 57, 909–920.
55. Dijkwel, P.A. and Hamlin, J.L. (1992) Mol. Cell. Biol. 12, 3715–3722.
56. Caddle, M.S., Lussier, R.H. and Heintz, N.H. (1990) J. Mol. Biol. 211, 19–33.
57. Gaudette, M.F. and Benbow, R.M. (1986) Proc. Natl. Acad. Sci. USA 83, 5953–5957.
58. Siebenlist, U., Hennighausen, L., Battey, J. and Leder, P. (1984) Cell 37, 381–391.
59. Kumar, S. and Leffak, M. (1989) Nucleic Acids Res. 17, 2819–2833.
60. Leffak, M. and James, C.D. (1989) Mol. Cell. Biol. 9, 586–593.
61. McWhinney, C. and Leffak, M. (1990) Nucleic Acids Res. 18, 1233–1242.
62. Iguchi-Ariga, S.M.M., Okazaki, T., Itani, T., Ogata, M., Sato, Y. and Ariga, H. (1988) EMBO J 7, 3135–3142.
63. Ruiz-Carrillo, A., Affolter, M. and Renaud, J. (1983) J. Mol. Biol. 170, 843–859.
64. Krieg, P.A., Robins, A.J., D'Andrea, R. and Wells, J.R.E. (1983) Nucleic Acids Res. 11, 619–627.
65. Affolter, M. and Ruiz-Carrillo, A. (1986) J. Biol. Chem. 261, 11496–11502.
66. Trempe, J.P., Lindstrom, Y.L. and Leffak, M. (1988) Mol. Cell. Biol. 8, 1657–1663.
67. Adams, C.C. and Workman, J.L. (1993) Cell 72, 305–308.
68. Orr-Weaver, T.L. (1991) BioEssays 13, 97–105.
69. Delidakis, C. and Kafatos, F.C. (1987) J. Mol. Biol. 197, 11–26.
70. Osheim, Y.N., Miller, Jr., O.L. and Beyer, A.L. (1988) Mol. Cell. Biol. 8, 2811–2821.
71. Osheim, Y.N. and Miller Jr., O.L. (1983) Cell 33, 543–553.
72. Brewer, B.J. and Fangman, W.L. (1987) Cell 51, 463–471.
73. Swimmer, C., Delidakis, C. and Kafatos, F.C. (1989) Proc. Natl. Acad. Sci. USA 86, 8823–8827.
74. Carminati, J.L., Johnston, C.G. and Orr-Weaver, T.L. (1992) Mol. Cell. Biol. 12, 2444–2453.
75. D'Anna, J.A., Church, V.L., Gale, J.M. and Tobey, R.A. (1990) Anal. Biochem. 187, 1–9.
76. Lin, S. and Kowalski, D. J. Mol. Biol. in press.
77. Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C. and Kohwi-Shigematsu, T. (1992) Science 255, 195–197.
78. Bell, S.P. and Stillman, B. (1992) Nature 357, 128–134.
79. Campbell, J.L. and Newlon, C.S. (1991) In J.R. Broach, J.R. Pringle and E.W. Jones (eds.), The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis and Energetics. Cold Spring Harbor Press, Cold Spring Harbor, New York, pp. 41–146
80. Zhu, J., Newlon, C.S. and Huberman, J.A. (1992) Mol. Cell. Biol. 12, 4733–4741.
81. Baldari, C.T., Amaldi, F. and Buongiorno-Nardelli, M. (1978) Cell 15, 1095–1107.
82. Bozzoni, I., Baldari, C.T., Amaldi, F. and Buongiorno-Nardelli, M. (1981) Eur. J. Biochem. 118, 585–590.
83. Micheli, G., Baldari, C.T., Carri, T., Cello, G. and Buongiorno-Nardelli, M. (1982) Exp. Cell Res. 137, 127–140.
84. Krysan, P.J. and Calos, M.P. (1991) Mol. Cell. Biol. 11, 1464–1472.
85. Linskens, M.H.K. and Huberman, J.A. (1988) Mol. Cell. Biol. 8, 4927–4935.
86. Yamaguchi, M., Hendrickson, E.A. and DePamphilis, M.L. (1985) J. Biol. Chem. 260, 6254–6263.
87. Kaiserman, H.B., Marini, N.J., Poll, E., Hiriyanna, K.T. and Benbow, R.M. (1990) In P.R. Strauss and S.H.W. (eds.), The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies. Teleford Press, Caldwell, New Jersey, pp. 783–811.
88. Suzuki, M., Savoysky, E., Izuta, S., Tatebe, M., Okajimna, T. and Yoshida, S. (1993) Biochemistry **32**, 12782–12792.
89. Suzuki, M., Izuta, S., Savoysky, E., Sakurai, T., Simbulan, C., Tatebe, M., Kojima, K. and Yoshida, S. (1993) Biochemistry Mol. Biol. Int. 29, 645–652.
90. Rao, B.S., Zannis-Hadjopoulos, M., Price, G.B., Reitman, M. and Martin, R.G. (1990) Gene 87, 233–242.
91. Landry, S. and Zannis-Hadjopoulos, M. (1991) Biochim. Biophys. Acta 1088, 234–244.
92. Yeo, C.-Y., Zhao, J., Larson, D.D. and Benbow, R.M. (1993) submitted.
93. Engberg, J. and Nielsen, H. (1990) Nucleic Acids Res. 18, 6915–6919.
94. Ruiz, J.C. and Wahl, G.M. (1990) Mol. Cell. Biol. 10, 3056–3066.
95. Levine, J. and Spradling, A. (1985) Chromosoma 92, 136–142.
96. Kowalski, D., Natalie, D.A. and Eddy, M.J. (1988) Proc. Natl. Acad. Sci. USA 85, 9464–9468.
97. Umek, R.M. and Kowalski, D. (1990) Proc. Natl. Acad. Sci. USA 87, 2486–2490.
98. Williams, J.S., Eckdahl, T.T. and Anderson, J.N. (1988) Mol. Cell. Biol. 8, 2763–2769.
99. Gasser, S.M., Amati, B.B., Cardenas, M.E. and Hofmann, J.F.-X. (1989) Int. Rev. Cytol. 119, 57–96.
100. Cockerill, P.N. and Garrard, W.T. (1986) Cell 44, 273–282.
101. Cockerill, P.N., Yuen, M.H. and Garrard, W.T. (1987) J. Biol. Chem. 262, 5394–5397.
102. Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D. (1986) J. Biomol. Struct. and Dyn. 3, 667–669.
103. Kailey, L., Caddle, M.S., Heintz, N. and Heintz, N.H. (1990) Mol. Cell. Biol. 10, 6225–6235.
104. Wells, R.D. (1988) J. Biol. Chem. 263, 1095–1098.
105. Stinchcomb, D.T., Mann, C., Selker, E. and Davis, R.W. (1981) ICN-UCLA Symp. Mol. Cell. Biol. 22, 473–483.
106. Spitzner, J.R. and Muller, M.T. (1988) Nucleic Acids Res. 16, 5533–5566.
107. Sander, M. and Hsieh, T.-s. (1985) Nucleic Acids Res. 13, 1057–1072.
108. Bergemann, A.D., Ma, Z.-W. and Johnson, E.M. (1992) Mol. Cell. Biol. 12, 5673–5682.
109. Bergemann, A.D. and Johnson, E.M. (1992) Mol. Cell. Biol. 12, 1257–1265.
110. Cech, T.R. (1986) In J.G. Gall (ed.), The Molecular Biology of Ciliated Protozoa. Academic Press, Inc., Orlando, pp. 203–225.
111. Palen, T.E. and Cech, T.R. (1984) Cell 36, 933–942.
112. Niles, E.G., Cunningham, K. and Rakesh, J. (1981) J. Biol. Chem. 256, 12857–12860.
113. Challoner, P.B., Amin, A.A., Pearlman, R.E. and Blackburn, E.H. (1985) Nucleic Acids Res. 13, 2661–2680.
114. Leu, T. and Hamlin, J.L. (1989) Mol. Cell. Biol. 9, 523–531.
115. Dijkwel, P.A., Wenink, P.W. and Poddighe, J. (1986) Nucleic Acids Res. 14, 3241–3249.