The EMBO Journal vol.6 no.2 pp.401–407, 1987

Endonuclease G: a $(dG)_n \cdot (dC)_n$-specific DNase from higher eukaryotes

Adolfo Ruiz-Carrillo and Jean Renaud

Cancer Research Center and Department of Biochemistry, Laval University Medical School, 11 Cote du Palais, Quebec G1R 2J6, Canada

Communicated by M.Beato

An endonuclease activity (termed endonuclease G) that selectively cleaves DNA at $(dG)_n \cdot (dC)_n$ tracts has been partially purified from immature chicken erythrocyte nuclei. Sites where $n \geq 9$ are cleaved in a manner that resembles types II and III restriction nucleases. The nicking rate of the G-strand is 4- to 10-fold higher than that of the C-strand depending on the length of the $(dG)_n \cdot (dC)_n$ tract and/or nucleotide composition of the flanking sequences. Endonuclease G hydrolyzes $(dG)_{24} \cdot (dC)_{24}$ of supercoiled DNA in a bimodal way every 9–11 nucleotides, the maxima in one strand corresponding to minima in the opposite, suggesting that it binds preferentially to one side of the double helix. The nuclease produces 5′ phosphomonoester ends and its activity is dependent on Mg^{2+} or Mn^{2+}. The wide distribution and high relative activity of endonuclease G in a variety of tissues and species argues for a general role of the enzyme. The striking correlation between genetic instability and poly(dG)·poly(dC) tracts in DNA suggests that these sequences and endonuclease G are involved in recombination processes.

Key words: genetic instability/'hot spot'/poly(dG)·poly(dC)/recombination/site-specific endonuclease

Introduction

Long (dG)·(dC) sequences have structural properties markedly different from those of DNA in the B conformation (Arnott and Selsing, 1974; McCall *et al.*, 1985). *In vitro*, they cannot be wrapped into nucleosomes (Simpson and Künzler, 1979), and under topological constraint they adopt an altered conformation (Nickol and Felsenfeld, 1983; Schon *et al.*, 1983).

One such homopolymer, $(G)_{18}$ [starting at −203 (Dolan *et al.*, 1983); site of transcription initiation = +1], occurs in a nuclease hypersensitive region (HS) of the chicken β^A-globin gene in erythrocyte nuclei (McGhee *et al.*, 1981). We have also recently observed that a HS upstream of the chicken histone H5 gene (HS3; Renaud and Ruiz-Carrillo, 1986), contains the sequence 5′-$(C)_{10}$TA$(C)_6$-3′ (starting at −1270; D.Bernier and A.Ruiz-Carrillo, unpublished data). In both cases, the hypersensitivity of these regions appears to be cell lineage-specific. As opposed to the family of shorter (dPu)·(dPy) sequence motifs occurring in the HS of several genes (Renaud and Ruiz-Carrillo, 1986), $(dG)_n \cdot (dC)_n$ tracts ($n \geq 9$) are much less abundant, although on a purely statistical basis they occur with a frequency 4.5 times higher than expected.

Since the homopolymers are found at quite variable distances from genes as well as in repeated DNA elements (see Discussion for references), we reasoned that their function might be related to mechanisms other than transcription. Nuclear processes including recombination, DNA replication, etc., could make use of the distinct structural features of these or other homo(co)polymer sequences. Although these functions are likely to leave some sort of footprint on chromatin, the structural alterations may escape detection when the genes are not in the pre-active or active unfolded conformation but packed in the less accessible 30 nm chromatin fiber.

In a search for nucleases specifically recognizing homopolymer sequences we have discovered a site-specific endonuclease activity, present in cells from different origins, that selectively cleaves DNA at $(dG)_n \cdot (dC)_n$ sites. On account of its striking specificity and preference for the G-strand, the activity has been termed endonuclease G. In this report we describe some aspects of its mechanism of recognition and catalysis, and discuss its possible involvement in recombination.

Results

Identification of endonuclease G activity in nuclear extracts

The strategy for detection of the nuclease activity used circular DNA substrates containing two $(dG)_n \cdot (dC)_n$ sequences. An endonuclease activity specific for these homopolymer tracts would produce two unique DNA fragments. In the initial experiments we used plasmid p554 DNA (see Figure 1B) containing a histone H5 cDNA (Ruiz-Vazquez and Ruiz-Carrillo, 1982) cloned at the *Pst*I site of pBR322 by oligo(dG)/(dC) tailing (Lobban *et al.*, 1973). The length of the homopolymer at the 5′ and 3′ ends of the cDNA is 24 and 17 respectively.

When naturally supercoiled p554 DNA was incubated with increasing amounts of a 0.3 M KCl nuclear extract from purified immature chicken erythrocytes (IE-$E_{0.3}$, see Materials and methods), the DNA was first relaxed by topoisomerases (Figure 1A, lane b), then nicked (lanes b and c), and finally linearized (lane c). Incubation with larger amounts of the extract resulted in the production of shorter molecules among which fragments of 4.36 and 0.88 kb predominate (Figure 1A, lanes d and e). Similar results were obtained with 0.1 M KCl and 0.2 M KCl nuclear extracts. Since the lengths of the insert and the vector are respectively 870 and 4363 bp, the results suggested that a nuclease present in the extract was capable of introducing double-stranded breaks in the DNA at two major sites, probably at or near the homopolymer stretches. Part of the additional complexity in the reaction products is due to fragments originating from dimer and trimer plasmid DNA molecules (cf. Figures 4 and 5 where these are not present). The activity of the endonuclease was not affected by pre-digestion of the nuclear extract with micrococcal nuclease (in the presence of 2 mM $CaCl_2$) or RNase A, but it was inactivated by digestion with proteinase K in 0.2% SDS (not shown).

The presence of a topoisomerase in the extract, presumably of type I since its activity is independent of Mg^{2+} and ATP, and the introduction of two major cuts per DNA molecule indicate that the endonuclease can cleave the target sequences in the relaxed substrate. A difference in the thermostability of endonuclease G and topoisomerase was used to demonstrate that the two activities are unrelated (Figure 1C). This approach also allowed

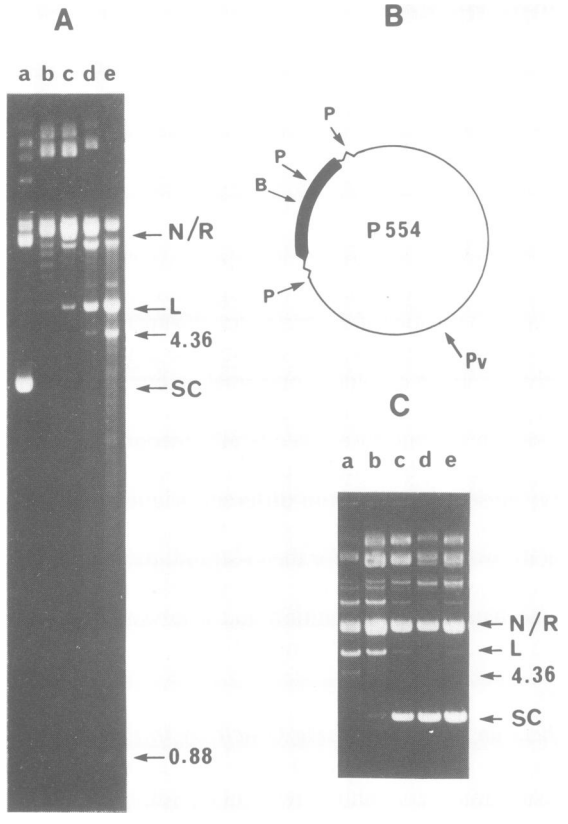

Fig. 1. Identification of endonuclease G activity in nuclear extracts. **(A)** Supercoiled p554 DNA (2 μg) was incubated for 30 min in standard buffer with **(a–e)**: 0, 1, 2, 3 and 6 μl (30 μg) of a non-heated 0.3 M KCl nuclear extract from immature erythrocytes (IE-$E_{0.3}$). **(B)** H5 cDNA clone used in the analysis. The pointed thick line indicates the H5 cDNA (from +54 to +883) inserted at the *Pst*I (P) site of pBR322 (thin line) by oligo(dG)·(dC) connectors (wavy lines): B, *Bst*EII; Pv, *Pvu*II. **(C)** Differential thermostability of endonuclease G and topoisomerases. Supercoiled p554 DNA was incubated for 30 min under standard conditions **(a, b)**, or in the presence of 10 mM EDTA **(c, d)** with 3 μl of IE-$E_{0.3}$ preheated for 10 min at 45°C **(a, c)** or 50°C **(b, d)**, or in the absence of nuclear extract **(e)**. The DNA products were separated by electrophoresis in 1.4% agarose gels and stained with ethidium bromide. The position of supercoiled (SC), linear (L), relaxed (R) and nicked (N) molecules is indicated. The length of the fragments is given in kb.

inactivation of other undesired enzymes in subsequent experiments (standard reaction conditions, see Materials and methods). Treatment of the nuclear extract for 10 min at 45°C (Figure 1C, lanes a and c) or 50°C (lanes b and d) completely inactivates the topoisomerase, assayed in the presence of EDTA (lanes c and d), whereas the activity of the nuclease, assayed in the presence of Mg^{2+} is only slightly affected by the higher temperature (lanes a and b). Since under our reaction conditions endonuclease G is the main Mg^{2+}-dependent DNase activity present in IE-$E_{0.3}$ (see below), the results suggest that the enzyme recognizes the target sequences in supercoiled as well as in relaxed DNA. This view is supported by the higher rate of nicking of supercoiled p554-3 (containing the homopolymer, see Figure 2) as compared to the vector pGEM4 (unpublished data). However, we cannot rule out the possibility that some of the DNA is relaxed by nicks introduced at minor sites in the plasmid before the attack at major sites occurs.

Specificity of endonuclease G

Figure 2D shows plasmids p554-5, p554-3 and pβG-5 used to determine the cleavage pattern of endonuclease G. p554-5 and

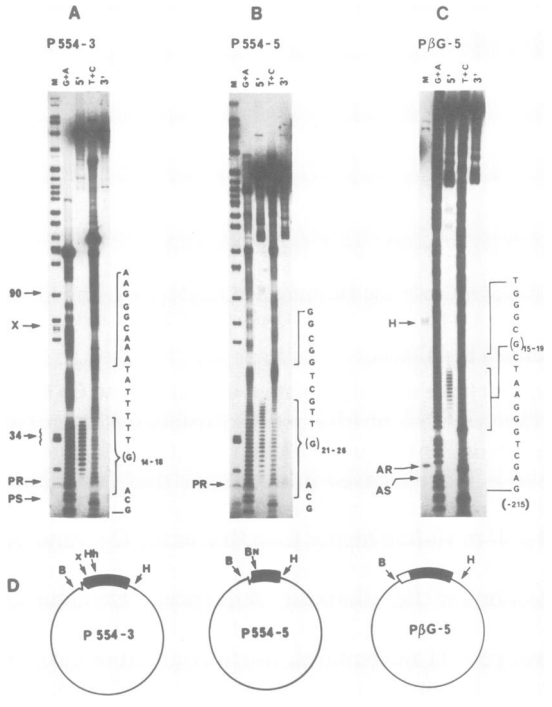

Fig. 2. Strand and sequence specificity of endonuclease G activity. **(D)** Plasmids used in the analysis. The *Pst*I 5′ and 3′ fragments of the H5 cDNA of p554 (see Figure 1B) and a *Pst*I fragment of the chicken β^A-globin gene (pβ1BR15) spanning from −215 to +757 (Dolan *et al.*, 1983) were inserted in the polylinker *Pst*I site of pGEM4 to generate p554-5, p554-3 and pβG-5, respectively. In all cases, the oligo(dG)·(dC) tracts (wavy lines) have the same orientation with respect to the *Bam*HI (B) site. The position of relevant restriction nuclease sites is indicated: *Asp*718 (A), *Bst*NI (BN), *Hin*dIII (H), *Xmn*I (X) and *Hha*I (Hh). p554-3 **(A)**, p554-5 **(B)** and pβG-5 **(C)** DNA were incubated with IE-$E_{0.3}$ in a standard reaction. *Hin*dIII-linearized full length molecules were digested with *Bam*HI and labeled at the 5′ or 3′ ends to comparable specific radioactivities. The autoradiographies show the electrophoretic separation in a denaturing gel of the DNA fragments. Equal amounts of radioactivity were loaded in all cases. Lanes G+A and T+C are sequencing ladders of the respective molecules labeled at the 5′ *Bam*HI ends. **Lanes M**: *Bam*HI 5′ end-labeled products of restriction nuclease digestion of control fragments mixed with 5′ end-labeled *Hpa*II fragments of pBR322 **(A, B;** the position of the 34- and 90-nucleotide long *Hpa*II fragments is indicated in A). X: *Xmn*I; H: *Hae*III. PR, PS and AR, AS indicate the position of *Pst*I (PR) and *Alu*I (AR) restriction nuclease fragments and the corresponding sites in the sequencing ladders (PS and AS) to illustrate the change in mobility of the respective fragments. The variable length of the oligo(dG)·(dC) stretches and the multiple bands of restriction fragments including them is due to length heterogeneity of the homopolymer runs introduced during propagation in *E. coli* C-600.

p554-3, respectively, carry the synthetic $(dG)_{24}\cdot(dC)_{24}$ and $(dG)_{17}\cdot(dC)_{17}$ of the parental p554, whereas pβG-5 carries the $(dG)_{18}\cdot(dC)_{18}$ found at the 5′ flanking region of the chicken β^A-globin gene (McGhee *et al.*, 1981; Dolan *et al.*, 1983). In the three cases the homopolymers have the same orientation with respect to the polylinker of the vector. In the experiments shown in Figure 2A–C, equal amounts of supercoiled plasmid DNA were incubated, in standard reactions, with IE–$E_{0.3}$ until most of the DNA was converted to nicked circular molecules. At the same time, 30–35% of the DNA remained supercoiled while 5–10% was linearized. Single-stranded cuts in the nicked molecules were mapped from the proximal *Bam*HI site (see Materials and methods) labeled either at the 5′ end (G-strand) or at the 3′ end (C-strand). Figure 2A–C shows that endonuclease G selectively attacked the three plasmid DNAs at vir-

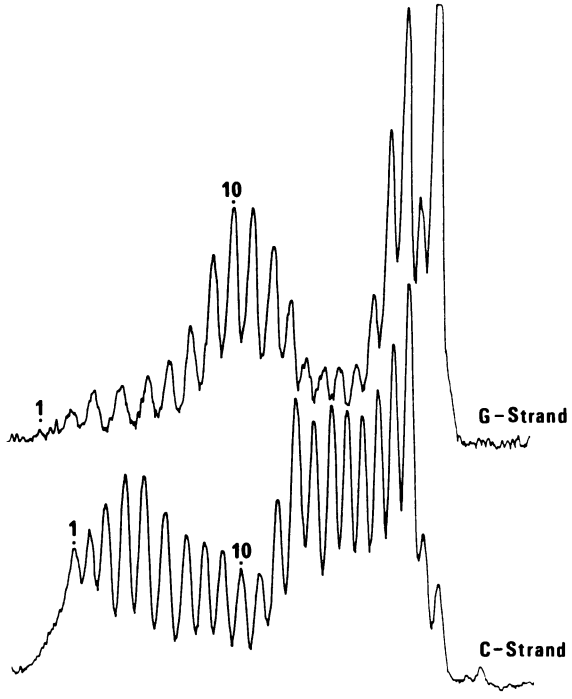

Fig. 3. Modal nicking pattern of endonuclease G. Densitometric tracings of the distribution of G- and C-strand cleavages in $(dG)_{24} \cdot (dC)_{24}$ of p554-5 grown in *E. coli* DH1. The numbers (5′ to 3′ for the G-strand and 3′ to 5′ for the C-strand) refer to the position of $(dG)_{24} \cdot (dC)_{24}$ from the *Pst*I junction. The experimental design was as in Figure 2 except that nicked circular DNA molecules were isolated after incubation with IE-$E_{0.3}$. Four times as much radioactivity was loaded for the 3′ end-labeled molecules (C-strand) than for the 5′ end-labeled molecules (G-strand).

tually every single nucleotide of the homopolymer sequences with a marked preference for the G-strand over the C-strand (see below an explanation for the smearing of the sequence ladders). Since the specific activities of the 3′ and 5′ labeled fragments were comparable, the relative extent of G- and C-strand cleavages was determined by scintillation counting of the respective gel fragments. The radioactivity recovered in the C-strand was 7% (p554-3), 24% (p554-5) and 25% (pβG-5) of that recovered in the respective G-strands, confirming a definite strand bias in the nicking pattern of the nuclease. The nucleotide composition of the homopolymer flanking sequences appear to influence the relative rate of strand cleavage. Thus, despite the equivalent length of the homopolymer tracts in p554-3 and pβG-5 the C-strand is cleaved 2.5 times faster in the former case.

Control experiments (not shown) indicated that the internal nicks were not detectably labeled during the 5′ or 3′ labeling reactions. Analysis of end-labeled mock-incubated control molecules also indicated that none of the nicks at the homopolymer sequences were introduced by the enzymes and procedure used in the mapping, and that the cleavage sites were not sensitive to RNases or alkali treatment (see Materials and methods).

The autoradiograms in Figure 2 also show that the rate of cleavage at each nucleotide along the $(dG)_n$ tracts is not identical but higher at the 3′ end than at the 5′ end. This is true even after correction for the faster mobility of the sequence ladders with respect to that of the nicked fragments (that is 1.5−2.0 nucleotides in this region of the gel, as determined by comparison with the mobility of restriction fragments). Figure 2B also reveals that the $(dG)_{24}$ in p554-5 is cleaved in a modal way. However, a faithful analysis of the distribution of nicks was impaired by a heterogeneity in the length of the oligo(dG)·(dC) runs that occurred during propagation in *Escherichia coli* C600. Thus, restriction fragments originating at the labeled *Bam*HI end and finishing before the homopolymers give single bands (e.g. Figure 2, PR and AR) whereas those ending past them give multiple bands (e.g. Figure 2, X and H). This is why the sequencing ladders are unreadable at the 3′ side of the homopolymer tracts as fragments of equal length have different sequences. Sequencing from the 3′ side of the oligomers confirmed that the amplification/deletion was confined to the $(dG)_n \cdot (dC)_n$ runs (unpublished data). As it will be described elsewhere, this expansion/shortening of the homopolymers is *recA*-dependent.

The densitometric tracings in Figure 3 show the nicking pattern at $(dG)_{24} \cdot (dC)_{24}$ of homogeneous length from p554-5 amplified in *E. coli* DH1 or HB101 (i.e. *recA*$^-$ hosts). The relative position of the nicks on each strand is given, 1 being the first nucleotide of the homopolymer sequence starting at the *Pst*I proximal side. This was determined by comparing the mobility of the fragments with that of sequence ladders from control molecules labeled at the respective ends. The patterns are modal and the peaks in one strand correspond to valleys in the opposite. Although the variation in cutting frequency is less regular at the 3′ side of the G-strand and corresponding 5′ side of the C-strand, it is possible to conclude that in each strand the distance between maxima is 9−11 nucleotides and between a maximum and a minimum is 5−6 nucleotides. Interestingly, the nucleotides at both ends of the homopolymer tract are cut at lower frequency.

These distributions can be directly related to the probability of phosphodiester bond hydrolysis if only one nick was introduced into most of the substrate molecules. We believe that this is the case because of the very limited digestion. We can deduce that 65% of the nicked molecules have a single nick and 35% have two (Boseley *et al.*, 1980), not necessarily in the homopolymer tract. Further, the cleavage rate for both strands is highest towards the side of the homopolymer furthest from the label. Since these arguments apply for both strands, the pattern of the C-strand is unlikely to be much affected by that of the G-strand, even though the latter is cleaved at a 4-fold higher rate.

It is evident for p554-5 and pβG-5 (Figure 2B and C) that these molecules are nicked at sequences other than the major sites. As it will be shown elsewhere, the nicks represent minor cleavage sites of the nuclease at $(dG)_n \cdot (dC)_n$ sites where $3 \leq n \leq 7$. Under our standard conditions, the nuclease rate constant for the major sites was estimated to be between 20 and 100 times higher than that for the minor sites.

Requirements of endonuclease G activity

The enzyme used in the following experiments was about 50-fold purified by column chromatography (see Materials and methods) and it was heated at 50°C for 10 min prior to the reactions. The substrate was either unlabeled linear p554 DNA (Figure 4A and C) or fragment L3, containing the $(dC)_{17} \cdot (dG)_{17}$ tract, 5′ end-labeled at the *Bst*EII site (Figure 4B, see also Figure 1B). The qualitative effects of several components of the reaction buffer on the kinetics of digestion were examined by comparing the production of specific fragments. As shown in Figure 1C, no detectable activity is observed in the presence of EDTA. The optimal Mg^{2+} concentration is rather broad (0.5−8.0 mM, Figure 4B), and Ca^{2+} up to 2 mM has no effect (not shown). Mn^{2+} (1−4 mM) enhances the rate of cleavage at secondary or minor sites (Figure 4B), an effect similar to that observed by an increase in the temperature of the reaction from 25°C to 37°C (not shown). KCl from 7.5 to 75 mM is optimal, whereas higher concentrations inhibit the nuclease activity (Figure 4B). ATP (from 1−4 mM, Figure 4A), dGTP or dCTP (each at 2 mM), NAD^+

(2 mM), glycerol (up to 10%), iodacetamide (4 mM), and *E. coli* RNA (up to 4 μg/20 μl reaction) have no effect (not shown). The results in Figure 4C also demonstrate that the products of the reaction contain 5′-phosphoryl termini.

Is endonuclease G ubiquitous?

The widespread occurrence of homopolymer (dG)·(dC) sequences suggested that endonuclease G could be present in a variety of tissues and species. This prediction is born out by the results summarized in Table I. From the systems tested, only the transformed chicken pre-erythroblast cell-line has no detectable activity. However, since the reactions were performed with whole nuclear extracts, it is possible that in the negative case the activity is either very low or perhaps masked by endogenous inhibitors.

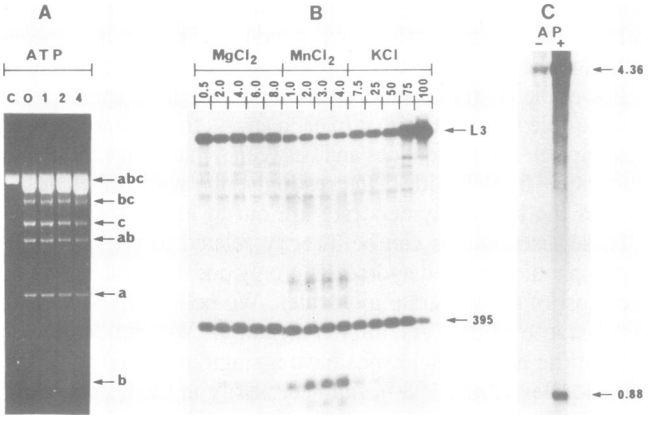

Fig. 4. Properties of partially purified endonuclease G activity from IE. **(A)** Effect of ATP. 2 μg of p554 DNA linearized at the *Pvu*II site of pBR322 were reacted with 0 (**c**) or 1 μg of the enzyme for 30 min in a standard reaction in the presence of 0, 1, 2 and 4 mM ATP (0−4, respectively. **a** and **c** correspond to molecules cleaved at the proximal and distal homopolymer tract from the *Pvu*II end, respectively. **b** corresponds to the insert released by cleavage at both homopolymer connectors (see Figure 1B for details). **(B)** Ionic strength and divalent ion requirements. A 1938-bp *Bst*EII−*Pvu*II fragment (L3) of p554 (see Figure 1B) labeled at the 5′ *Bst*EII end was mixed with 0.5 μg of pBR322 and reacted with 1 μg of the enzyme in a standard reaction except that the mM concentrations of $MgCl_2$, $MnCl_2$ and KCl were varied as indicated. **(C)** Endonuclease G generates 5′ phosphomonoesters. 6 μg of supercoiled p554 DNA were reacted with 5 μg of the enzyme in standard buffer containing 5 mM potassium phosphate. Purified DNA was either treated (+) or mock treated (−) with calf intestine phosphatase (AP) and subsequently labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase. The extent of labeling was determined by autoradiography after separation of equal amounts of DNA in a 1% agarose gel. The low levels of labeling of the DNA non-treated with the commercial phosphatase are presumably due to a contaminating phosphatase which is not completely inhibited by 5 mM phosphate.

Discussion

Relationship of endonuclease G to other DNA endonucleases from higher eukaryotes

The results in Table I suggest that endonuclease G is a common DNase. It is also the most important Mg^{2+}-dependent DNA endonuclease activity in all systems examined (e.g. Figure 5). A few Mg^{2+}-dependent endonucleases that appear to introduce a limited number of cuts in DNA have been previously reported (e.g. Wang and Furth, 1977; McKenna *et al.*, 1981; Nakayama *et al.*, 1981; Hope *et al.*, 1986). However, to our knowledge none has been shown to cleave at $(dG)_n \cdot (dC)_n$ and possess the high selectivity of the nuclease described here. In fact, endonuclease G does not cut to any significant extent other poly(dPu)·poly(dPy) sequences like $(dA)_n \cdot (dT)_n$ (e.g. $n = 5$, Figure 2A, or longer, unpublished data) or $(dGdA)_{34} \cdot (dTdC)_{34}$ [Figure 5, cleavage at the copolymer should produce a fragment with a mobility close to that of the *Pst*I (P) marker, lane a].

A DNase found in mouse L cells and in an Abelson transformed pre-B cell line (Desiderio and Baltimore, 1984) or in chicken bursa and adult mouse liver (Kataoka *et al.*, 1984) has been reported to cleave close to the 5′ ends of the immunoglobulin Jκ segments (i.e. $J\kappa_1 - J\kappa_5$). Precise mapping indicated that several of the sites in fact corresponded to $(dG)_{5-6} \cdot (dC)_{5-6}$ clusters located outside the conserved heptammer and nonamer sequences thought to be involved in immunoglobulin gene rearrangement. We have examined the digestion pattern of pVJ-1 DNA (Kataoka *et al.*, 1984) by $E_{0.3}$ nuclear extracts from IE, bursa of Fabricius, and purified adult rat hepatocytes. Figure 5 shows the results obtained with the 5.5-kb *Bst*EII−*Eco*RI fragment of pVJ-1, containing segments $J\kappa_1 - J\kappa_5$, and as a comparison those obtained with fragment L3 under identical conditions. If endonuclease G were to cut close to the Jκ segments, major bands at 242, 580, 955, 1208 and 1565 bp are then to be expected (lane b shows the position of a 568-bp *Bst*EII−*Ava*I fragment used as a marker for the position of the $J\kappa_4$ segment). Instead, weak bands were observed in the overexposed autoradiography, none of which corresponds to the predicted fragments. This is also true for the 240-bp region of the gel removed for illustration purposes. Given the large difference between the rate of cleavage at $(dG)_{17} \cdot (dC)_{17}$ of L3 and at the sequences of pVJ-1 (Figure 5), we conclude that endonuclease G is different from the activity that cleaves at the Jκ segments described by the Baltimore and Honjo groups. It is probable that those cleavages mapping at $(dG)_{5-6} \cdot (dC)_{5-6}$ sequences correspond to minor sites introduced by contaminating endonuclease G activity present in their enzyme preparation (see also Hope *et al.*, 1986).

Features recognized by endonuclease G

X-ray fiber diffraction studies suggest that linear poly(dG)·poly(dC) has preference for the A−DNA conformation even at high humidities (Arnott and Selsing, 1974). Recent X-ray work on

Table I. Relative endonuclease G activity in nuclear extracts from several tissues and species

Chicken ME	1	H-32 (transformed fibroblasts)	580
Chicken IE	23	Calf thymus	8
Chicken erythroleukemia (41/2)	ND	Rat hepatocytes	485
Chicken bursa of Fabricius	48	Rat hepatome (H4IIEC3)	93
Chicken T-lymphoblastoma (MSB-1)	263	Fetal calf serum	ND
Chicken brain	85		

0.3 M KCl nuclear extracts were prepared as described in Materials and methods. 5′ end-labeled L3 fragment from p554 was incubated, in standard buffer containing 5 mM potassium phosphate, with different amounts of the nuclear extracts. The extent of cleavage at $(dG)_{17} \cdot (dC)_{17}$ was determined by densitometry of the autoradiograms. The values are expressed as activity/μg DNA relative to that found in mature chicken erythrocytes (ME). ND, non-detectable.

heavily hydrated crystals has also demonstrated that the oligomer d(GGGGCCCC) adopts a right-handed structure similar, but not identical, to classical A−DNA (McCall *et al.*, 1985). Other evidence indicates that these structural characteristics persist in solution (Bram, 1971; Peck and Wang, 1981; Drew and Travers, 1984, 1985; Fox and Waring, 1984). Hence, the nuclease could recognize the distinct structural conformation of the A-type DNA, presumably the wide minor groove because of its shallowness and higher atomic order (McCall *et al.*, 1985).

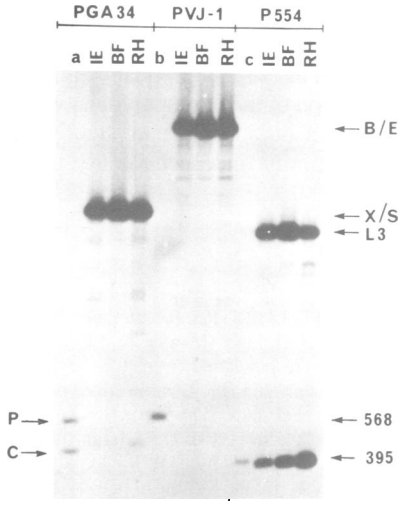

Fig. 5. Endonuclease G does not cleave preferentially the immunoglobulin Jχ segments, or (dGdA)·(dTdC) tracts. Autoradiography of the digestion products of pGA_{34}, pVJ-1 and p554 DNA incubated with nuclear extracts from several sources. A 2-kb *Xmn*I−*Sca*I (X/S) fragment from pGA34 [containing $(dGdA)_{34} \cdot (dTdC)_{34}$], a 5.5 kbp *Bst*EII−*Eco*RI (B/E) fragment from pVJ-1 (containing $J\chi_1 - J\chi_5$) and fragment L3 from p554 [containing $(dG)_{17} \cdot (dC)_{17}$] were labeled to comparable specific radioactivities at the 5′ *Xmn*I and *Bst*EII ends, respectively. Equimolar amounts of each DNA were mixed with 0.5 μg of salmon sperm DNA and reacted with nuclear extracts from IE (3 μl), bursa of Fabricius (BF, 1 μl) and purified rat hepatocytes (RH, 0.1 μl) for 30 min in a standard reaction in the presence of 5 mM potassium phosphate. **(a)** *Xmn*I−*Cla*I (C) and *Xmn*I−*Pst*I (P) fragments flanking the homocopolymer sequence of pGA34; **(b)** 568-bp *Bst*EII−*Ava*I fragment from pVJ-1; **(c)** 395-bp *Bst*EII−*Pst*I fragment of L3 [cleaved at the $(dG)_{17} \cdot (dC)_{17}$ boundary].

The high specificity of endonuclease G for $(dG)_n \cdot (dC)_n$ and the distinct structural features of the homopolymer make it difficult to distinguish which of the two components of helix recognition (i.e. sequence or structure) plays the more dominant role. However, there are indications that the structure of the substrate is important in the recognition/catalysis mechanism. Endonuclease G shows progressively lower cutting frequencies towards the ends of the homopolymer sequence (Figures 2 and 3), an effect that can be interpreted in terms of structural transition and change in minor groove width occurring at the junction sequences (Drew and Travers, 1985). Further, the nuclease requires at least nine consecutive (dG)·(dC) to display a high cleavage rate constant (unpublished data), suggesting that a minimum length of the homopolymer is needed to adopt the conformation preferred by the enzyme.

The higher preference for the G-strand implies asymmetry either in the double helical structure and/or the nuclease. The asymmetry of the recognition sequence might explain some of the observations. The model for poly(dG)·poly(dC) proposed by McCall *et al.* (1985) has symmetric strands but the authors note that to obtain the helical repeat of 10.7 observed in solution (Peck and Wang, 1981) the helix has to be slightly twisted by 1.5° about an axis centered on, and perpendicular to, the guanine bases. This would extend the cytosine strand relative to the guanine resulting in a certain amount of asymmetry. There are indications that endonuclease G recognizes a 'single-strand'-like character of the homopolymer since its activity is strongly inhibited by heat-denatured DNA (i.e. binds to it or acts as a competitive inhibitor, unpublished data). Further, its activity is stimulated by Mn^{2+} and higher temperatures and inhibited by 0.1−0.2 M KCl. It is therefore possible that the nuclease binds selectively to one of the two strands and cuts the other across. The ~5 bp staggering between the G- and the C-strand is compatible with this view.

The other distinctive characteristic of the mechanism of endonuclease G action is its nicking periodicity, approximately every 10 nucleotides. This nicking pattern indicates that the cuts are introduced preferentially in one side of the homopolymer, presumably reflecting the helical repeat of the DNA.

Table II. Genetic instability of $(dG)_n \cdot (dC)_n$ tracts and nearby sequences

Sequence (5′−3′)	Locus	Observed variability (sequence and/or length at or near homopolymer)	$(dG)_n \cdot (dC)_n$ length heterogeneity	Ref.[a]
G_{13}	Upstream mouse c-*myc*	Fusion at G_{13} (translocation 15/12)	No	1
G_{14}	1st intron chicken c-*myc*	Proviral ALV integration	No	2
C_n	ETS[b] *Xenopus* rDNA	Insertions/deletions, substitutions	$n = 11-14$	3
C_nACG_n	ITS[b]-1 *Xenopus* rDNA		$n = 10-11$	3
C_3G_4	ITS-1 *Xenopus* rDNA	Insertion/deletion C-rich sequence	Yes[c]	3
C_n	ITS-2 *Xenopus* rDNA		$n = 9-11$	3
C_n	ITS-2 *Xenopus* rDNA		$n = 10-15$	3
C_n	D-loop mammalian mit DNA	Insertion/deletion/substitutions	Yes[d]	4−6
C_{12}	NTS[b] mouse rDNA	Next to hypervariable repeated DNA	No	7
C_n	Crab satellite DNA	Insertions/deletions/substitutions	$n = 22-23$	8
C_{13}	Crab satellite DNA	Insertions/deletions/substitutions	No	8
C_n	Sea urchin repeated DNA	Sequence variation	$n = 0,16$	9
C_n	Polyoma	Frameshift mutation	$n = 8-9$	10
G_n	Upstream chicken β^A globin		$n = 16,18$	11,12

[a]1, Gerondakis *et al.* (1984); 2, Shih *et al.* (1984); 3, Stewart *et al.* (1983); 4, Anderson *et al.* (1981); 5, Anderson *et al.* (1982); 6, Walberg and Clayton (1981); 7, Kuehn and Arnheim (1983); 8, Fowler and Skinner (1986); 9, Posakony *et al.* (1981); 10, Wilson *et al.* (1986); 11, McGhee *et al.* (1981); 12, Day *et al.* (1981).
[b]ETS, external transcribed spacer; ITS, internal transcribed spacer; NTS, non-transcribed spacer.
[c]Variations include: $C_3GCTC_3GC_5$, $C_3GC_5GC_6$, $C_3GCTC_3GC_7$, $C_3GC_5GC_6$.
[d]Variations include: C_{12} (bovine), C_6AC_5 (mouse), C_7AC_5 (rat), C_9TC_6 (human KB cell), C_7TC_5 (human placenta).

Possible biological functions

The absence of tissue specificity suggests a general role for endonuclease G. However, the specific activity of the nuclease in different cells provides no obvious clue as to its function *in vivo*, but the high selectivity displayed *in vitro* makes it ineffective as a degradative enzyme.

Whatever its involvement in general processes, the role of the nuclease is likely to be tightly related to that of its target sequences. We have found in the current DNA sequence data banks 21 cases in which $(dG)_{12-18}\cdot(dC)_{12-18}$ occurs. Despite the bias of data banks for gene sequences, the homopolymers have been found exclusively in repeated DNA, gene spacers, gene flanking regions and introns. More importantly, in a remarkably high number of cases they appear to be closely associated to genetically unstable and highly recombinogenic regions. The examples given in Table II are significant because cumulative information from repeated analysis of the loci permits a more meaningful and less prejudicial correlation to be made.

There are two types of event that appear to be related to the presence of the homopolymeric sequences. One is the hypervariability in length and composition of the homopolymer itself. Although it can be argued that this could be due to cloning artifacts, variability of the *Xenopus* ribosomal DNA was also observed with uncloned chromosomal DNA (Stewart *et al.*, 1983). Further, in the case of polyoma the spontaneous frameshift mutation $C_9 \rightarrow C_8$ is responsible for the revertant phenotype of the transformed cells (Wilson *et al.*, 1986). Significantly, no homopolymer tract of length comparable to those found in the *Xenopus* ribosomal ETS and ITS is found in the structural regions of the highly GC-rich gene.

The second type of event is the hypervariability of the sequences next or close to the homopolymer. In some cases this is a consequence of translocation [e.g. in a murine plasmacytoma, the fusion J558 took place next to the homopolymer stretch in the 5′ flanking region of c-*myc* (Gerondakis *et al.*, 1984)], or to proviral insertion [e.g. in avian leukosis virus-induced lymphomas more than 70% of proviral integration sites are clustered in a 250-bp region of the first c-*myc* intron that contains a $(dG)_{14}\cdot(dC)_{14}$ (Shih *et al.*, 1984)].

Other hypervariability appears to be a consequence of deletions, insertions and substitutions. The gene organization of all mammalian mitochondria examined is virtually identical, and the gene sequences are relatively homologous. On the other hand, the extremities of the displacement-loops (D-loops) are variable in sequence and in size. One of the most significant homologies in the most variable region is a C-rich sequence (L-strand, 5′ to 3′) which is itself hypervariable (Anderson *et al.*, 1981, 1982; Walberg and Clayton, 1981).

The high proportion of cases in which the homopolymer appears next to regions in which either rearrangement, translocation, deletion and hypervariability of the DNA occurs, strongly suggests that short homopolymeric $(dG)_n\cdot(dC)_n$ sequences might be the target of the mechanisms leading to these changes. In addition, the homopolymer itself is a recombinational 'hot spot'. We have found that $(dG)_n\cdot(dC)_n$ tracts of $10 < n \leq 24$ become heterogeneous in length during propagation in *E. coli* C-600 (e.g. Figure 2). The mutation results in the insertion or deletion of one or a few G/C base pairs so that individual *E. coli* clones harbor plasmids that differ in the length of the homopolymer tract. Our present evidence indicates that this phenomenon is dependent on the *recA* function and does not originate during plasmid DNA replication. The hypervariability in length could be a consequence of slippage and originate during the synthesis that transforms two recombining molecules into hybrids.

Although the above evidence is only circumstantial, it strongly suggests that endonuclease G activity and the homopolymer sequences could easily play a role in recombination. In view of the properties of the nuclease, it is possible to envisage its involvement at one or more stages of recombination processes. It could initiate strand exchange by the introduction of single or double strand cuts at homopolymer sequences, be responsible for the resolution of Holliday structures or it could be involved in translocation and illegitimate recombination processes. Clearly, the type of recombinational event (homologous or non-homologous) as well as the final frequency in a given locus could depend on the location of the homopolymers with respect to other recombinogenic elements such as reiterated DNA sequences, and on the spatial proximity and accessibility of the $(dG)_n\cdot(dC)_n$ tract.

Materials and methods

Materials

Enzymes were purchased from BRL Inc. (Gaithesburg, MD), New England BioLabs (Beverly, MA) and Boehringer (Mannheim, FRG). Radioisotopes (sp. act. $3-4 \times 10^3$ Ci/mmol) were purchased from ICN (Irvine, CA).

The construction of plasmid p554 has been described previously (Ruiz-Vazquez and Ruiz-Carrillo, 1982). Plasmids p554-5, p554-3 and pβG-5 (Figure 2) were constructed by ligation of the respective DNA fragments to the polylinker *Pst*I site of pGEM4 (Promega Biotec, Madison, WI). Plasmid pGA_{34} contains a stretch of $(dGdA)_{34}\cdot(dTdC)_{34}$ inserted at the *Sal*I site of pDP by *Xho*I linkers (Pulleyblank *et al.*, 1985). Plasmid pVJ-1 contains the mouse immunoglobulin Jκ and Vκ segments (Kataoka *et al.*, 1984). Plasmids were propagated in *E. coli* C-600, HB101 or DH-1 as indicated in the text.

Mature (ME) and immature erythrocytes (IE) were obtained and purified by filtration through SP-Sephadex as described by Perucho *et al.* (1979). Any remaining buffy coat was discarded during the subsequent washings. Bursa of Fabricius and brain were obtained from 1-week-old chicks. Both tissues were extensively rinsed in saline before isolation of nuclei. Adult rat hepatocytes were purified from liver (Deschenes *et al.*, 1980). H-32 (LSCC-H32) was established from primary quail embryo fibroblasts that were spontaneously transformed (Kaaden *et al.*, 1982). Rat hepatoma cell line H4-II-E-C3, derived from a primary tumor induced by *N*-2-acetylaminofluorine, was purchased from American Type Culture Collection (Rockville, MD). Clone 41/2 is a chicken pre-erythroblast cell line transformed by ts34 avian erythroblastosis virus (Beug *et al.*, 1979). MSB-1 is a chicken T-lymphoblastoma cell line transformed by Marek's disease virus. Culture conditions have been described (Renaud and Ruiz-Carrillo, 1986).

Nuclear extracts and enzyme purification

Cells were lysed in buffer A [0.25 M sucrose, 0.4 mM phenylmethylsulfonyl fluoride (PMSF), 3 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 0.2 mM 1,2-di(2-aminoethoxy)ethane-*N,N,N′,N′*-tetra-acetic acid, sodium salt (EGTA), 20 mM *N*-2-hydroxyethyl-piperazine-*N′*-2-ethanesulfonic acid, sodium salt (Hepes), pH 7.3] containing 0.2% Nonidet P 40. Nuclei were washed several times in buffer A, once in buffer B (20 mM Hepes, pH 7.9, 3 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 0.4 mM PMSF, 0.2 mM EGTA), and finally extracted with one volume of buffer B containing 10% glycerol and 0.2–0.8 M KCl. The nuclear debris was removed by centrifugation (12 000 *g*, 20 min) and the supernatant was frozen at −80°C.

Nuclear extracts from IE were fractionated by ammonium sulfate precipitation (350 mg/ml plus 7 μl of 20 mM NaOH/ml). The precipitate was dissolved in buffer C (5% glycerol, 5 mM $MgCl_2$, 50 mM KCl, 1 mM 2-mercaptoethanol, 0.1 mM PMSF, 0.2 mM EGTA, 20 mM Hepes, pH 7.9), dialyzed for 18 h against buffer C and loaded on a DEAE-52 (Whatman) column equilibrated in the same buffer. Endonuclease G activity, eluted in the flow through fraction, was loaded on a phosphocellulose P 11 (Whatman) column. After washing with 0.1 M KCl in buffer C, the activity was eluted with 0.3 M KCl-containing buffer C, dialyzed against buffer C and stored at −80°C. This procedure resulted in a 30- to 50-fold increase in the specific activity of the nuclease.

Enzymatic assays

The standard reactions were carried out in buffer C at 25°C for 30–60 min in a volume of 20 μl. Unless otherwise specified, fractionated endonuclease G or whole extracts ($E_{0.3}$) were pre-heated at 50°C for 10 min immediately before the start of the incubations. Reactions were quenched with EDTA, SDS and Pro-

teinase K (final concentrations 10 mM, 0.2% and 0.5 mg/ml, respectively) and the incubations were continued at 50°C for 2 h. At the end of the treatment, samples were directly loaded in agarose gels.

Mapping of endonuclease G cleavage sites

To determine the location of double-stranded DNA cleavages, supercoiled and 5′ end-labeled or unlabeled linear plasmid DNA was incubated with nuclear extracts and the fragments produced during the reaction were separated by electrophoresis in 0.7–1.4% agarose or 6% denaturing polyacrylamide gels. For autoradiography, agarose gels were treated with cold 5% trichloroacetic acid for 60 min and dried.

To map the major endonuclease G nicking sites, supercoiled plasmid DNA was incubated with nuclear extracts in standard reactions. The digested DNA was cleaved with *Hin*dIII (Figure 2) and the full length molecules were isolated by electrophoresis in low melting point agarose (BRL, Gaithesburgh, MD); alternatively, nicked circular molecules were isolated (Figure 3). In either case, DNA was digested with *Bam*HI and labeled at the 5′ or 3′ ends with T4 polynucleotide kinase and [γ-^{32}P]ATP or with the Klenow fragment and [α-^{32}P]dATP, respectively. Molecules only labeled at the *Bam*HI site proximal to the insert were purified by PAGE after cleavage with the appropriate restriction nucleases. Chemical DNA sequencing reactions were essentially carried out as described by Maxam and Gilbert (1980). Kodak XAR-5 films were used for autoradiography. Films were scanned with a Chromoscan-3 microdensitometer (Joyce-Loebl).

Control experiments

Several control experiments were routinely carried out. The location of nicks present in untreated plasmids, and those possibly introduced by the enzymes used for the mapping, was determined in 3′ or 5′ end-labeled control molecules by analysis in denaturing gels. The lack of labeling at internal nicks during the end-labeling reactions was monitored by analysis of *Bam*HI labeled molecules digested with *Pst*I, that cleaves between the homopolymer tract and the labeled end. Even though chloramphenicol was not used for plasmid amplification, the level of ribosubstitution in DNA was examined. Molecules end-labeled at the *Bam*HI site were alkali-denatured and after neutralization they were digested overnight either with RNase A, RNase T1, or 0.15 M KOH at 37°C. The patterns of treated and mock-treated molecules were the same as judged from analysis in denaturing gels, whereas chicken erythrocyte total RNA was digested in control reactions. In no case was there any evidence that the bands produced by endonuclease G corresponded to ribosubstitution sites. Finally, mock incubations and subsequent treatments of supercoiled or linear DNA were always carried out in parallel to the reactions. No evidence for chemical degradation of DNA by the buffer components was ever obtained.

Acknowledgements

We are grateful to C.d'Auteuil, H.LaRue and M.Nöel for providing us with some of the cells and tissues used in this work, and to M.Lambert for plasmid DNA preparation. We are thankful to Dr Kaaden for a gift of the cell line H-32. We are indebted to Drs D.Engel (pβ1BR15), D.E.Pulleyblank (pGA$_{34}$), and T.Honjo (pVJ-1) for gifts of the indicated plasmids. We also thank N.Lemieux and G.Langlois for graphic work. Parts of the manuscript benefited from a revision by Dr T.Moss. This work was supported in part by a grant from the NCI and MRC of Canada. Parts of this work were presented at the 29th Annual CFBS meeting held 16–20 June 1986 at the University of Guelph, Ontario, Canada.

References

Anderson,S., Bankier,A.T., Barrell,B.G., de Bruijn,M.H.L., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schreier,P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1981) *Nature*, **290**, 457–465.

Anderson,S., de Bruijn,M.H.L., Coulson,A.R., Eperon,I.C., Sanger,F. and Young,I.G. (1982) *J. Mol. Biol.*, **156**, 683–717.

Arnott,S. and Selsing,E. (1974) *J. Mol. Biol.*, **88**, 551–552.

Beug,H., von Kirchbach,A., Döderlein,G., Conscience,J.-F. and Graf,T. (1979) *Cell*, **18**, 375–390.

Boseley,P.G., Moss,T. and Birnstiel,M.L. (1980) *Methods Enzymol.*, **65**, 478–494.

Bram,S. (1971) *Nature, New Biol.*, **233**, 161–164.

Day,L.E., Hirt,A.J., Lai,E.C., Mace,M. and Wod,S.L.C. (1981) *Biochemistry*, **20**, 2091–2098.

Deschenes,J., Valet,J.P. and Marceau,N. (1980) *In Vitro*, **16**, 722–730.

Desiderio,S. and Baltimore,D. (1984) *Nature*, **308**, 860–862.

Dolan,M., Dodgson,J.B. and Engel,J.D. (1983) *J. Biol. Chem.*, **258**, 3983–3990.

Drew,H.R. and Travers,A.A. (1984) *Cell*, **37**, 491–502.

Drew,H.R. and Travers,A.A. (1985) *Nucleic Acids Res.*, **13**, 4445–4467.

Fowler,R.F. and Skinner,D.M. (1986) *J. Biol. Chem.*, **261**, 8994–9001.

Fox,K. and Waring,M. (1984) *Nucleic Acids Res.*, **12**, 9271–9285.

Gerondakis,S., Cory,S. and Adams,J.M. (1984) *Cell*, **36**, 973–982.

Hope,T.J., Aguilera,R.J., Minie,M.E. and Sakano,H. (1986) *Science*, **231**, 1141–1145.

Kaaden,O.-R., Lange,S. and Stiburek,B. (1982) *In Vitro*, **18**, 827–834.

Kataoka,T., Kondo,S., Nishi,M., Kodaira,M. and Honjo,T. (1984) *Nucleic Acids Res.*, **12**, 5995–6010.

Kuehn,M. and Arnheim,N. (1983) *Nucleic Acids Res.*, **11**, 211–224.

Lobban,P.E. and Kaiser,A.D. (1973) *J. Mol. Biol.*, **78**, 453–471.

McCall,M., Brown,T. and Kennard,O. (1985) *J. Mol. Biol.*, **183**, 385–396.

McGhee,J.D., Wood,W.I., Dolan,M., Engel,J.D. and Felsenfeld,G. (1981) *Cell*, **27**, 45–55.

McKenna,W.G., Maio,J.J. and Brown,F.L. (1981) *J. Biol. Chem.*, **256**, 6435–6444.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–580.

Nakayama,J., Fujiyoshi,T., Nakamura,M. and Anai,M. (1981) *J. Biol. Chem.*, **256**, 1636–1642.

Nickol,J.M. and Felsenfeld,G. (1983) *Cell*, **35**, 467–477.

Peck,L.J. and Wang,J.C. (1981) *Nature*, **292**, 375–378.

Perucho,M., Molgaard,H.V., Shevac,A., Pataryas,T. and Ruiz-Carrillo,A. (1979) *Anal. Biochem.*, **98**, 464–471.

Posakony,J.W., Scheller,R.H., Anderson,D.M., Britten,R.J. and Davidson,E.H. (1981) *J. Mol. Biol.*, **149**, 41–67.

Pulleyblank,D.E., Haniford,D.B. and Morgan,A.R. (1985) *Cell*, **42**, 271–280.

Renaud,J. and Ruiz-Carrillo,A. (1986) *J. Mol. Biol.*, **189**, 217–226.

Ruiz-Vazquez,R. and Ruiz-Carrillo,A. (1982) *Nucleic Acids Res.*, **10**, 2093–2108.

Schon,E., Evans,T., Welsh,J. and Efstratiadis,A. (1983) *Cell*, **35**, 837–848.

Shih,C.-K., Linial,M., Goodenow,M.M. and Hayward,W.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4697–4701.

Simpson,R.T. and Künzler,P. (1979) *Nucleic Acids Res.*, **6**, 1387–1415.

Stewart,M.A., Hall,L.M.C. and Maden,B.E.H. (1983) *Nucleic Acids Res.*, **11**, 629–646.

Walberg,M.W. and Clayton,D.A. (1981) *Nucleic Acids Res.*, **9**, 5411–5421.

Wang,E.-C. and Furth,J.J. (1977) *J. Biol. Chem.*, **252**, 116–124.

Wilson,J.B., Hayday,A., Courtneidge,S. and Fried,M. (1986) *Cell*, **44**, 477–487.

Received on 27 October 1986