MOLECULAR AND CELLULAR BIOLOGY, May 1985, p. 964–971
0270-7306/85/050964-08$02.00/0
Copyright © 1985, American Society for Microbiology
Vol. 5, No. 5

Site-Specific DNA Binding of Nuclear Factor I: Analyses of Cellular Binding Sites

RICHARD M. GRONOSTAJSKI,† SAMIT ADHYA, KYOSUKE NAGATA, RONALD A. GUGGENHEIMER, AND JERARD HURWITZ*

Sloan Kettering Cancer Institute, New York, New York 10021

Received 5 December 1984/Accepted 28 January 1985

Nuclear factor I is a cellular site-specific DNA-binding protein required for the efficient in vitro replication of adenovirus DNA. We have characterized human DNA sequences to which nuclear factor I binds. Three nuclear factor I binding sites (FIB sites), isolated from HeLa cell DNA, each contain the sequence TGG(N)$_{6-7}$GCCAA. Comparison with other known and putative FIB sites suggests that this sequence is important for the binding of nuclear factor I. Nuclear factor I protects a 25- to 30-base-pair region surrounding this sequence from digestion by DNase I. Methylation protection studies suggest that nuclear factor I interacts with guanine residues within the TGG(N)$_{6-7}$GCCAA consensus sequence. One binding site (FIB-2) contained a restriction endonuclease *Hae*III cleavage site (GGCC) at the 5′ end of the GCCAA motif. Digestion of FIB-2 with *Hae*III abolished the binding of nuclear factor I. Southern blot analyses indicate that the cellular FIB sites described here are present within single-copy DNA in the HeLa cell genome.

Considerable progress has been made in the purification and characterization of procaryotic DNA replication proteins. (12). However, the lack of both genetic analysis and in vitro replication systems has hampered similar studies on DNA synthesis in eucaryotes. Adenovirus (Ad) DNA replication is the only system in eucaryotes that is amenable to genetic manipulation (5, 22, 30) and for which the initiation and elongation of DNA chains have been reconstituted with purified proteins (20; for reviews see references 3, 9, and 28). We have fractionated and purified the viral and cellular proteins required for the in vitro replication of Ad DNA with the goal of isolating proteins involved in host DNA synthesis (7, 9, 16). A protein of particular interest is nuclear factor I, a cellular site-specific DNA-binding protein that is required for the efficient in vitro initiation of Ad DNA replication (19, 21, 23).

Nuclear factor I was purified from nuclear extracts of uninfected HeLa cells by its ability to support the replication of Ad DNA in vitro (19). The synthesis of full-length Ad DNA in vitro requires five purified proteins (4, 20). Three of these proteins are viral coded, the 80,000-dalton precursor to the 55,000-dalton terminal protein found at the 5′ end of the Ad genome (pTP), the Ad DNA polymerase, and the Ad DNA-binding protein. The two remaining proteins, nuclear factors I and II, are host coded and have been purified from nuclear extracts of uninfected HeLa cells. The initiation of Ad DNA synthesis occurs by the covalent attachment of dCMP, the 5′-terminal deoxynucleotide of Ad DNA, to the pTP (2, 15). This initiation reaction is catalyzed by the Ad DNA polymerase and, in the presence of the Ad DNA-binding protein, is completely dependent on nuclear factor I (19). Nuclear factor I has been shown to specifically bind, and protect from DNase I digestion, a 32-base-pair (bp) DNA sequence located in the replication origin present at the termini of the 36,000-bp Ad genome (21, 23). The ability of nuclear factor I to stimulate the initiation of Ad DNA replication has been directly correlated with its ability to bind this region of Ad DNA (10, 23). The pTP-dCMP initiation complex formed in the presence of nuclear factor I can be elongated by the Ad DNA polymerase to replication intermediates approximately one-fourth the size of viral DNA (20). Nuclear factor II, which posesses an intrinsic type I topoisomerase activity, has no effect on the formation of the pTP-dCMP initiation complex, but it is required for the elongation of replicating intermediates, formed in the presence of the three viral proteins and nuclear factor I, to full-length Ad DNA (20).

The host-coded proteins identified with the in vitro Ad DNA replication system are of interest with respect to their possible role in host nucleic acid metabolism. Because nuclear factor I binds specifically to DNA sequences present in the Ad genome, it appeared likely that this protein interacted with cellular DNA sequences. Recently, such cellular DNA sequences were isolated and cloned (8), and it was estimated that one nuclear factor I binding (FIB) site is present about every 100 kilobases (kb) in the HeLa cell genome. In this report, we have analyzed in detail three cellular DNA sequences to which nuclear factor I binds. DNA sequencing, DNase I protection, and dimethyl sulfate methylation protection studies have been performed to examine the binding of nuclear factor I to the cellular FIB sites. We also describe the use of a novel technique, termed "elongated primer selection", which may be useful in determining the minimal DNA sequence required for proteins that bind to specific sites on duplex DNA. The results of these studies provide a precise determination of those DNA sequences required for site-specific DNA binding by nuclear factor I.

MATERIALS AND METHODS

Preparation of nuclear factor I. Nuclear factor I was purified from extracts of uninfected HeLa cell nuclei as previously described (19). Denatured DNA-cellulose or glycerol gradient-purified nuclear factor I preparations were used unless otherwise noted.

Preparation of plasmid DNA. Plasmids were propagated in the DH-1 strain of *Escherichia coli* (11). DNA was prepared by alkaline lysis followed by centrifugation to equilibrium in

* Corresponding author.
† Present address: University of Toronto, Department of Medical Genetics, Toronto, Ontario, Canada, M5S 1A8.

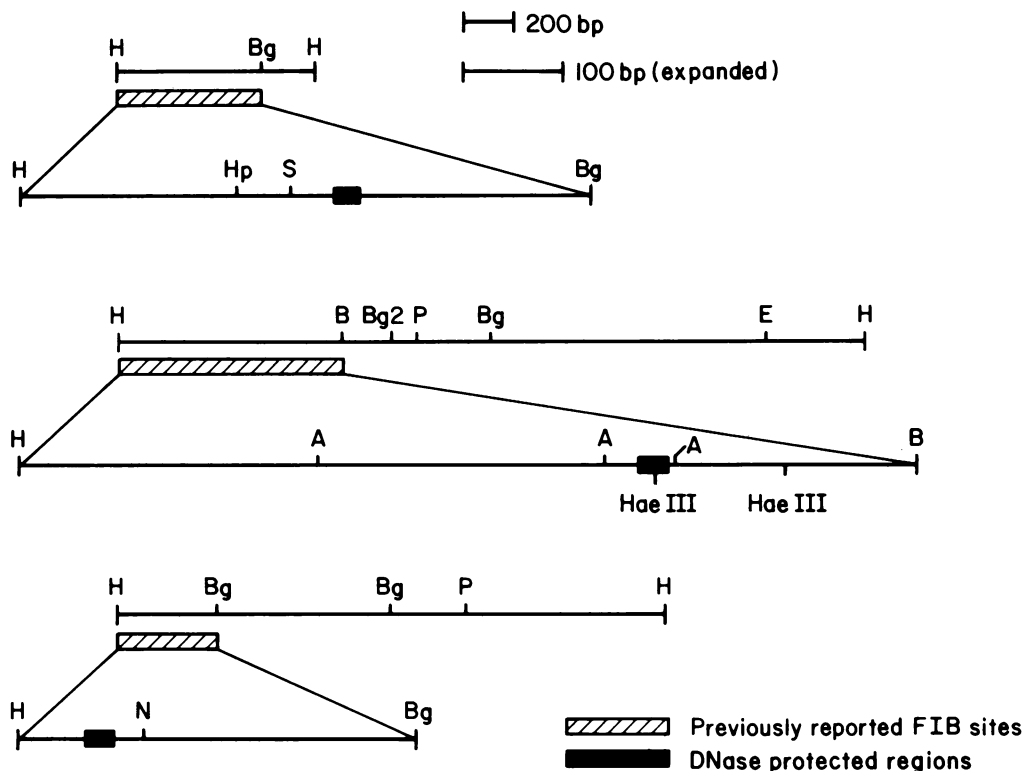

FIG. 1. Restriction maps of HeLa cell FIB sites. FIB-1, FIB-2, and FIB-3 are 0.8-, 3.0-, and 2.2-kb *Hin*dIII fragments, respectively, that contain nuclear FIB sites and were cloned from HeLa cell DNA (8). The nuclear FIB sites were previously localized to the regions shown by the hatched bars. The expansions of these areas show additional restriction enzyme cleavage sites and the region of DNA which is protected from DNase I digestion by nuclear factor I (black bars, see Fig. 2 and 3). The abbreviations for restriction enzyme sites shown are as follows: A, *Alu*I; B, *Bam*HI; Bg, *Bgl*I, Bg 2, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; *Hae*III; Hp, *Hpa*I; N, *Nco*I. The cleavage sites for several of these enzymes are shown only when they lie within the previously described nuclear FIB region (hatched bars). The coding region for the tetracycline resistance gene of the parent vector pBR322 would lie to the left of the figure.

cesium chloride-ethidium bromide gradients (17). Plasmids pFIB-1, pFIB-2, and pFIB-3 contain 0.8-, 3.0-, and 2.2-kb inserts, respectively, of HeLa cell DNA cloned into the *Hin*dIII site of pBR322. These plasmids contain nuclear FIB sites FIB-1, FIB-2, and FIB-3, respectively (8). pFIB-2α contains the 0.9-kb *Hin*dIII-*Bam*HI fragment of pFIB-2 cloned into the *Hin*dIII-*Bam*HI site of M13mp9. pFIB-2β and pFIB-2γ contain the 70-bp *Alu*I fragment (see Fig. 1) from the insert of pFIB-2 cloned into the *Sma*I site of M13mp9 in the same and opposite orientations as pFIB-2α, respectively. All M13 clones were grown in *E. coli* strain JM103. Single-strand and superhelical plasmid DNAs were prepared by standard procedures (17).

Nitrocellulose filter-binding assays and elongated primer selection. The retention of labeled DNA fragments on nitrocellulose filters in the presence of nuclear factor I was performed as previously described (8, 21). DNA eluted from the filters was analyzed by electrophoresis on agarose or acrylamide gels. To map the DNA sequences required for the binding of nuclear factor I to duplex pFIB-3 DNA, dideoxynucleotide sequencing reactions were performed by the method of Wallace et al. (32) with the synthetic pBR322 *Hin*dIII site sequencing primer (New England BioLabs, Beverly, Mass.) and heat-denatured pFIB-3 DNA. After completion of the sequencing reactions, the samples were extracted once with phenol-chloroform and once with water-saturated ether. A portion of each sample was incubated with nuclear factor I, filtered through nitrocellulose membranes (HAWP; Millipore Corp., Bedford, Mass.), and washed; the DNA retained on the filter was eluted as previously described (21). The eluted DNA was analyzed on 6% polyacrylamide–50% urea sequencing gels along with the unselected products of the sequencing reactions.

DNase protection and miscellaneous techniques. Plasmid DNA was digested with the indicated restriction endonuclease and the resulting fragments were labeled at their 5′ or 3′ termini with [γ-^{32}P]ATP or [α-^{32}P]dNTPs by standard techniques (17). The labeled fragments were digested with a second restriction enzyme to generate molecules with uniquely labeled ends, and these were purified by electrophoresis through agarose or acrylamide gels. DNase I protection experiments were performed with uniquely labeled fragments as previously described (21). Sequencing of the FIB sites was performed by the chemical cleavage (18) or chain termination (24) methods. Analysis of cellular DNA homologous to the cloned FIB sites was performed as previously described (27).

Methylation protection. The 82-bp *Eco*RI-*Bam*HI DNA fragment (2 fmol) from pFIB-2β, uniquely labeled with ^{32}P at the 5′ terminus of the *Bam*HI site as described above, was incubated in reaction mixtures (50 μl) containing 50 mM sodium cacodylate buffer (pH 8.1), 5 mM $MgCl_2$, 10 μg of bovine serum albumin, 150 mM NaCl, and various amounts of nuclear factor I (denatured DNA-cellulose fraction) for 20 min at 0°C. The reaction mixtures were then transferred to 21°C for 5 min, dimethyl sulfate was added to a final concentration of 0.5%, and the incubation was continued for an additional 5 min. Reactions were halted by the addition of

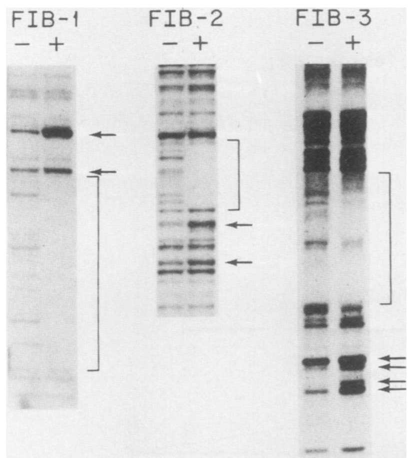

FIG. 2. DNase I protection of HeLa cell FIB sites. DNA fragments (2 to 10 fmol) from within the hatched areas of Fig. 1 were labeled at a unique end and digested with DNase I in the absence (−) and presence (+) of nuclear factor I (5 ng) as described previously (21). The resulting fragments were analyzed on 6% polyacrylamide sequencing gels in the presence of DNA restriction fragment markers or markers produced by subjecting the fragments to limited chemical cleavage sequencing reactions. The DNA fragments used are designated as follows: FIB-1, the *Sau*3A1-*Bgl*I fragment of pFIB-1 3′ labeled at the *Sau*3A1 site; FIB-2, the *Hin*dIII-*Bam*HI fragment of pFIB-2 3′ labeled at the *Bam*HI site; FIB-3, the *Hin*dIII-*Bgl*I fragment of pFIB-3 3′ labeled at the *Hin*dIII site.

20 µl of a solution containing 1.5 M sodium acetate and 1.0 M 2-mercaptoethanol. *E. coli* tRNA (10 µg) was added, and the nucleic acid was precipitated twice with ethanol, dried in vacuo, and subjected to reactions which favored

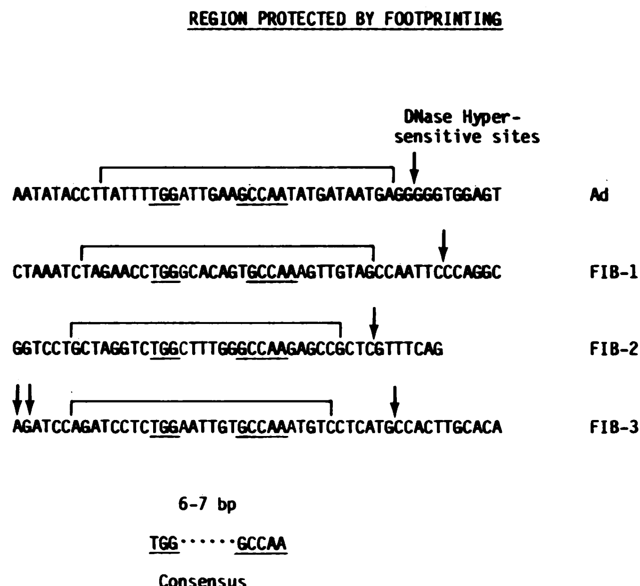

FIG. 3. DNA sequence of Ad5 and HeLa FIB sites. The DNA sequences of the Ad5 and HeLa FIB sites are aligned at regions protected from DNase I digestion by nuclear factor I. The brackets above each sequence denote the protected region, and the arrows indicate sites of enhanced DNase I cleavage in the presence of nuclear factor I. The consensus sequence $TGG(N)_{6\text{-}7}GCCAA$ is underlined in each FIB site.

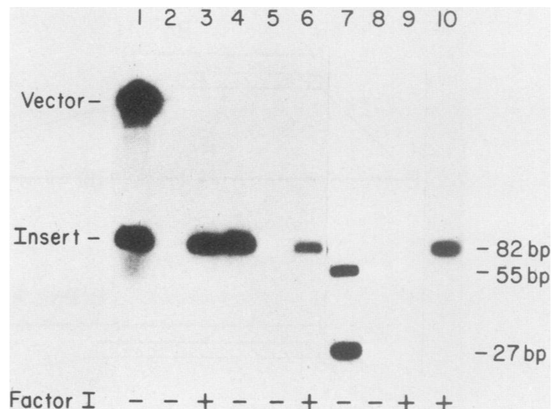

FIG. 4. Binding of nuclear factor I to clone FIB-2β DNA fragments. Nitrocellulose filter binding assays of 3′-^{32}P-labeled restriction fragments were performed as described in the text. DNA fragments employed were as follows: lanes 1 to 3, clone pFIB-2β DNA (125 ng) digested with *Eco*RI and *Bam*HI; lanes 4 to 6, gel-purified 82-bp insert (1.4 ng); lanes 7 to 9, a *Hae*III digest of an 82-bp fragment (1.4 ng); and lane 10, a mixture of the 82-bp fragment and its *Hae*III digest (1.4 ng each). Lanes 1, 4, and 7 show the input DNA; the remaining lanes show DNA bound to nitrocellulose filters in the absence (lanes 2, 5, and 8) or presence (lanes 3, 6, 9, and 10) of 44 ng of nuclear factor I. Filter-bound DNA was eluted and electrophoresed on a 10% polyacrylamide gel. Vector and insert fragments are indicated at the left, and restriction fragment lengths (in base pairs) are shown at the right of the figure.

cleavage at guanine residues (18). DNA fragments were lyophilized three times and analyzed by electrophoresis on 8% polyacrylamide–50% urea sequencing gels.

RESULTS

DNase I cleavage protection and enhancement. It was shown previously that nuclear factor I binds specifically to the HeLa cell DNA fragments present in pFIB-1, pFIB-2, and pFIB-3 (8). The binding sites noted previously by restriction endonuclease analysis are delineated by the hatched bars in Fig. 1. DNase I protection experiments were performed to futher localize the FIB sites on each of these DNA fragments. DNA fragments were uniquely labeled at one terminus and were incubated with DNase I in the presence or absence of nuclear factor I. In the presence of nuclear factor I, a 25- to 30-bp region of each clone was protected from DNase I digestion (Fig. 2). In addition, the presence of nuclear factor I caused enhanced cleavage by DNase I at sites adjacent to the protected regions of each clone (Fig. 2, arrows). For each DNA fragment, the size of the region protected from DNase I digestion by nuclear factor I is similar to that seen previously for the FIB site present on Ad serotype 5 DNA (21, 23). Analysis of the DNA sequence encompassing the protected regions revealed that the sequence $TGG(N)_{6\text{-}7}GCCAA$ appeared in all four FIB sites (Fig. 3.) A defined sequence of this size and complexity would appear randomly in a genome approximately once every 16,000 bp ($1/4^7$; uncorrected for base composition). Because this sequence is present in each FIB site, it is likely to be important for the binding of nuclear factor I. The nuclear FIB sites are designated by the solid black bars on the expanded restriction maps of pFIB-1, pFIB-2, and pFIB-3 (Fig. 1).

Determination of the minimal sequence required for binding of nuclear factor I. Guggenheimer et al. (10) demonstrated

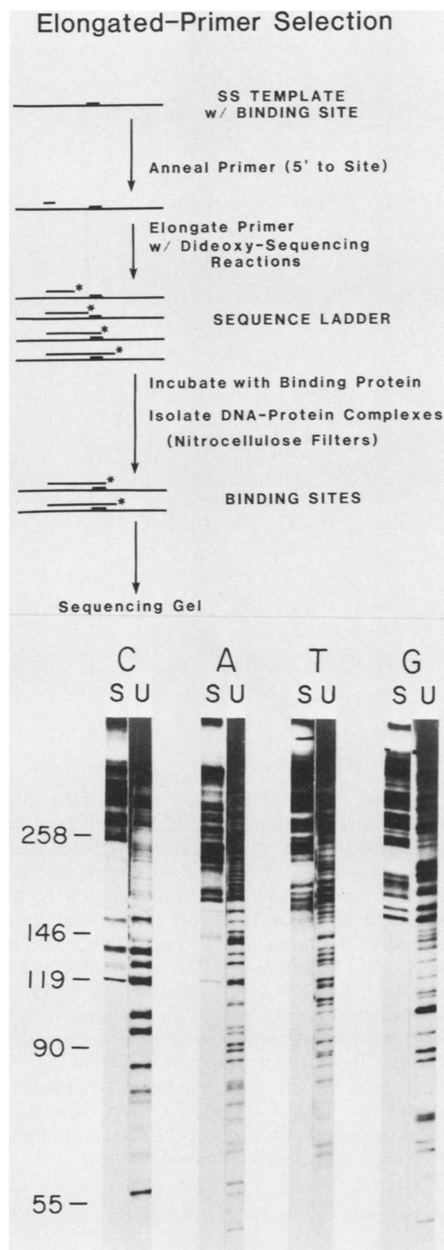

FIG. 5. Analysis of FIB-3 by elongated primer selection. pFIB-3 DNA was linearized with *Eco*RI and heat denatured, and the pBR322 *Hin*dIII sequencing primer was annealed by the method of Wallace et al. (32). The annealed primer was elongated with dideoxy-sequencing reactions, and the DNA was isolated as described in the text. Portions of the DNA (45 ng) were incubated with 10 ng of nuclear factor I and filtered through alkali-washed nitrocellulose membranes as described previously (8, 21). DNA retained on the filters in the presence of nuclear factor I (lanes marked S) was eluted and analyzed on a 6% polyacrylamide sequencing gel along with DNA not subjected to nuclear factor I selection (lanes marked U). Lanes C, A, T, and G are sequencing reactions performed in the presence of ddCTP, ddATP, ddTTP, and ddGTP, respectively.

that a 48-bp fragment of the Ad5 FIB site was sufficient for the binding of nuclear factor I. We therefore attempted to isolate a small fragment of a cellular FIB site that would bind nuclear factor I. An *Alu*I digest of the *Hin*dIII-*Bam*HI fragment of pFIB-2 yielded a 70-bp DNA fragment (Fig. 1) to which nuclear factor I bound specifically (data not shown). This fragment was cloned into the *Sma*I site of M13mp9 to generate pFIB-2β. Digestion of pFIB-2β with *Eco*RI and *Bam*HI liberated an 82-bp fragment that contained FIB-2. This 82-bp fragment contains a *Hae*III site (GGCC) at the 5′ end of the second motif (GCCAA) of the consensus sequence (Fig. 1 and Fig. 3). We therefore tested whether cleavage of FIB-2 at this *Hae*III site would prevent binding by nuclear factor I. The 82-bp *Eco*RI-*Bam*HI fragment containing FIB-2 was labeled at its 3′ termini with [α-^{32}P]dNTPs and the large fragment of DNA polymerase I (Klenow fragment). This labeled fragment was retained by nuclear factor I in a nitrocellulose filter binding assay (Fig. 4, lanes 3 and 6). Cleavage of the fragment with *Hae*III generated subfragments of 55 and 27 bp (Fig. 4, lane 7). Neither of these subfragments were retained on nitrocellulose filters by nuclear factor I (Fig. 4, lane 9). Mixing of the 55- and 27-bp fragments with the intact 82-bp fragment did not inhibit the binding of nuclear factor I to the intact fragment (Fig. 4, lane 10). Thus, cleavage at the 5′ end of the second motif (GCCAA) destroyed the ability of FIB-2 to be bound by nuclear factor I.

To determine the minimal size of duplex DNA required for binding by nuclear factor I, a novel technique was employed that took advantage of the fact that nuclear factor I binds tightly to duplex DNA but poorly to single-stranded DNA (21). The FIB site on pFIB-3 lies within 150 bp of the end of the segment of cloned HeLa cell DNA (Fig. 1). To determine the minimal sequence of duplex DNA required for the binding of nuclear factor I, a synthetic DNA primer was

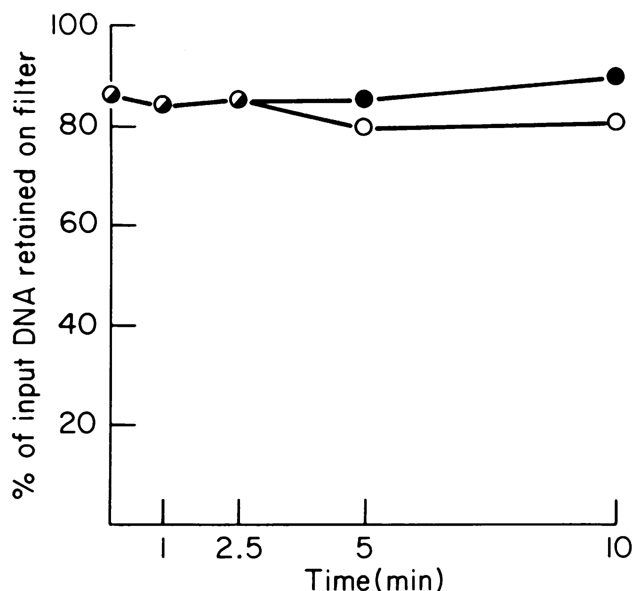

FIG. 6. Effect of dimethyl sulfate on the binding of nuclear factor I. Parallel experiments were performed in the presence or absence of dimethyl sulfate to assay the binding of nuclear factor I to an 82-bp *Eco*RI-*Bam*HI DNA fragment containing FIB-2. Reaction mixtures (50 μl) contained 50 mM sodium cacodylate buffer (pH 8.1), 5 mM $MgCl_2$, 10 μg of bovine serum albumin, 50 mM NaCl, ^{32}P-labeled DNA (2 fmol; 2,000 cpm/fmol), and nuclear factor I (0.1 μg; denatured DNA-cellulose fraction). ○, Reaction shifted from 0 to 21°C without the addition of dimethyl sulfate; ●, reaction shifted from 0 to 21°C, followed by the addition of dimethyl sulfate to a final concentration of 0.5%. The binding of nuclear factor I was assayed as previously described (21).

hybridized to *Eco*RI-digested, heat-denatured pFIB-3 DNA (32) and was elongated in the presence of a mixture of deoxy- and dideoxynucleoside triphosphates. The minimum length of the primer-template that was retained on nitrocellulose filters by nuclear factor I was determined by incubating the ddNMP chain-terminated primer-template with nuclear factor I and performing a standard nitrocellulose filter binding assay. The DNA selectively retained on the nitrocellulose filter in the presence of nuclear factor I (marked S) was compared with the unselected DNA from the sequencing reactions (marked U) in Fig. 5. No DNA was retained on the nitrocellulose filters in the absence of nuclear factor I (data not shown). In the presence of nuclear factor I, only those DNA primer-templates elongated to within or past the consensus sequence shown in Fig. 3 were efficiently retained on nitrocellulose filters (Fig. 5). The dark band at ca. 119 bp that is selected by nuclear factor I (Fig. 5, lane C_s) represents a DNA chain elongated to the second cytosine of the second motif of the consensus sequence (GC*C*AA). Thus, elongation of the DNA primer only 3 nucleotides into the second part of the sequence TGG$(N)_{6-7}$GCCAA was sufficient to generate a nuclear FIB site from the single-strand DNA template. This finding suggests that, although nuclear factor I protects a 25- to 30-bp sequence from DNase I digestion, only those regions 5′ to and including most of the TGG$(N)_{6-7}$GCCAA consensus sequence are required in a duplex form for the binding of nuclear factor I. These results confirm DNA sequencing and restriction endonuclease analysis data which suggested an essential role for the conserved TGG$(N)_{6-7}$GCCAA DNA sequence in the binding of nuclear factor I.

Methylation protection of FIB sites by nuclear factor I. Dimethyl sulfate (M_r = 126), an alkylating agent that methylates the N-7 position of guanine and N-1 and N-3 positions of adenine residues (18), was used to determine regions of close contact between nuclear factor I and guanine residues of the 82-bp duplex DNA fragment containing FIB-2. Under conditions in which dimethyl sulfate had no effect on the extent of the binding of nuclear factor I (Fig. 6), nuclear factor I protected three guanine residues on one strand of the 82-bp fragment from methylation (Fig. 7). The DNase I-protected region on this strand of FIB-2 is 26 bases long and contains seven guanine residues. Because of the labeling protocol, the sequence shown in Fig. 7 is complementary to the sequence of FIB-2 given in Fig. 3. The three protected guanine residues are located entirely within the complement of the TGG$(N)_{6-7}$GCCAA consensus sequence of FIB-2, suggesting that these nucleotides are critical for the site-specific binding of nuclear factor I and may represent contact points between nuclear factor I and DNA. Methylation protection studies performed with DNA labeled uniquely on the opposite strand of the 82-bp fragment demonstrated that only those guanine residues present in the consensus sequence on this strand of FIB-2 were protected from methylation by nuclear factor I (data not shown). Enhanced methylation of guanine residues in the presence of nuclear factor I was not observed at any region on either strand of the duplex DNA fragment containing FIB-2 (data not shown).

Analysis in FIB sites in cellular DNA. To determine the nature of the cellular DNA surrounding the FIB sites, total HeLa cell DNA was digested with restriction endonuclease *Hin*dIII and analyzed by Southern hybridization. Because the pFIB plasmids used as probes contained inserts of 0.8, 3.0, and 2.2 kb (pFIB-1, pFIB-2, and pFIB-3, respectively), hybridization between conserved regions of limited length, such as the TGG$(N)_{6-7}$GCCAA consensus sequence, would not be detected. When *Hin*dIII digests of HeLa genomic

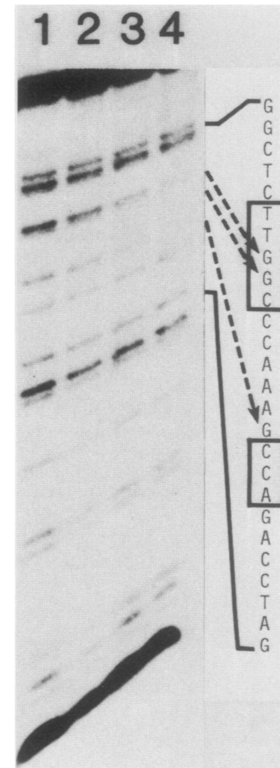

FIG. 7. Nuclear factor I-mediated methylation protection of FIB-2. Methylation protection of FIB-2 by nuclear factor I was assayed as described in the text with the ^{32}P-labeled *Bam*HI-*Eco*RI DNA fragment containing FIB-2 and no nuclear factor I (lane 1), 0.02 μg of nuclear factor I (lane 2), 0.1 μg of nuclear factor I (lane 3), and 0.4 μg of nuclear factor I (lane 4). These amounts of nuclear factor I resulted in the binding of 0, 16, 36, and 55% of the input DNA, respectively. Methylated DNA was subjected to guanine cleavage reactions. The DNA sequence of FIB-2 that is protected from DNase digestion by nuclear factor I is shown to the right of lane 4. Due to the labeling procedure, the sequence determined is the complement of the FIB-2 binding site shown in Fig. 3. The 5′ end of the sequence is at the top. The sequences complementary to the left and right motifs of the TGG$(N)_{6-7}$GCCAA consensus sequence are boxed, and the guanine residues protected from methylation by nuclear factor I are indicated by dashed arrows.

DNA and pFIB-1, pFIB-2, and pFIB-3 DNAs were analyzed on 0.8% agarose gels and probed with nick-translated pFIB-2, only the 3.0-kb FIB-2 insert, the corresponding 3.0-kb genomic fragment, and vector sequences were detected (Fig. 8B). Similar Southern analyses with pFIB-1 and pFIB-3 as probes confirmed that the inserts of these cloned DNAs do not cross-hybridize and show that each probe detects a single band in genomic digests corresponding in size to the cloned DNA fragments (Fig. 8A and C). The higher-molecular-weight bands seen in the genomic digest of HeLa DNA with pFIB-1 as the probe (Fig. 8A) are most likely due to partial digestion products and a shadow produced by intense hybridization of pBR322 vector sequences in adjacent lanes (Fig. 8D). It is possible, however, that sequences homologous to the pFIB-1 probe exist in a few copies in the HeLa cell genome. Quantitation of the amount of hybridization of probe DNA to the bands in the genomic digests indicates that the DNA fragments surrounding the sites FIB-1, FIB-2, and FIB-3 are present at one (or a few) copies per cell.

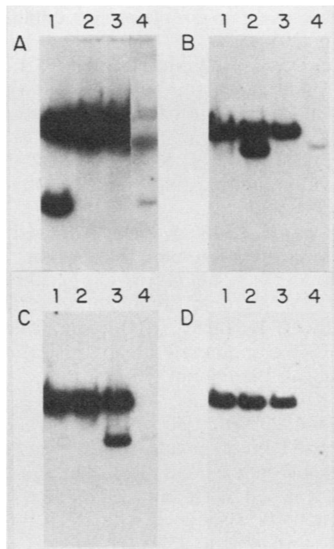

FIG. 8. Southern analysis of FIB sites. *Hin*dIII digests of pFIB-1, pFIB-2, pFIB-3, and HeLa cell DNA (lanes 1 to 4, respectively) were analyzed on a 0.8% agarose gel and subjected to Southern blot hybridization. The amount of DNA loaded in each lane was 400 pg of pFIB-1, pFIB-2, and pFIB-3 and 10 μg of HeLa cell DNA. The nick-translated ^{32}P-labeled DNAs used as probes were pFIB-1 (A), pFIB-2 (B), pFIB-3 (C), and pBR322 (D).

DISCUSSION

Nuclear factor I was initially characterized as a protein required for the initiation of Ad DNA replication in vitro (19). It was subsequently demonstrated that this host protein binds specifically to the replication origin of Ad DNA (21, 23). The results reported here and in other studies (10) have defined the DNA sequences required for the site-specific binding of nuclear factor I to DNA. The three cellular FIB sites described here bind to nuclear factor I with approximately the same efficiency as does the site located in the origin of Ad replication (8). Thus, the protein must recognize some feature or features common to each site. The presence of the sequence TGG$(N)_{6-7}$GCCAA in all of the FIB sites (Fig. 3) suggests an important role for this sequence in the binding of nuclear factor I.

The consensus sequence presented here differs from those proposed in previous studies that examined the binding of DNA fragments by proteins in crude nuclear extracts that presumably contained nuclear factor I (1, 26). The larger number of FIB sites examined here suggests that the previously determined sequences are related to and comprise subsets of the consensus sequence shown in Fig. 3 and 9. Since the cloned HeLa FIB sites were selected by tight binding to nuclear factor I, it is likely that the TGG$(N)_{6-7}$GCCAA consensus sequence is closer to the "parental" FIB site sequence than those determined by screening processes.

Results of dimethyl sulfate-mediated methylation protection studies demonstrated that, of the guanine residues present in the DNase I-protected region of FIB-2, only those residues within the TGG$(N)_{6-7}$GCCAA consensus sequence are sites of direct apposition between nuclear factor I and DNA (Fig. 7). However, it is unclear whether this consensus sequence is the only important component of a FIB site. Siebenlist et al. (26) and Borgmeyer et al. (1) have examined the specific binding of DNA fragments by a protein, presumed to be nuclear factor I, present in crude extracts of nuclei of various tissues. The DNA sequences described by these workers are compared with the Ad and HeLa cell FIB sites in Fig. 9. The BS1a, BS1b, and BS2 FIB sites are found in the 5′ flanking region of the chicken lysozyme gene (1) and contain either perfect or near-perfect copies of the TGG$(N)_{6-7}$GCCAA sequence. The c-*myc*II_1 and c-*myc*II_2 FIB sites, found upstream of the mRNA start sites of the human c-*myc* gene (26), contain a perfect and imperfect consensus sequence, respectively. The putative c-*myc*II_2 FIB site has two changes from the consensus sequence and was shown previously to bind only very poorly to proteins in nuclear extracts that presumably contain nuclear factor I (26). Surprisingly, the c-*myc*II_1 FIB site, which contains a perfect consensus sequence, binds to nuclear factor I with ca. 1/10 the affinity of the Ad or HeLa FIB sites (26; unpublished observations). It is therefore likely that features of the DNA other than the consensus sequence can affect the binding of nuclear factor I.

The expected frequency of appearance of a consensus sequence of this complexity at random in a genome is about once every 16 kb ($1/4^7$; uncorrected for base composition). This frequency is considerably higher than the previously determined estimate for the occurrence of FIB sites in the HeLa cell genome (one every 100 to 200 kb [8]). The latter estimate was determined from the frequency of recovery of plasmids containing FIB sites from a library of HeLa cell DNA fragments. The library had been enriched in FIB sites by selecting DNA fragments retained on nitrocellulose filters in the presence of nuclear factor I (8). The lack of agreement between these two estimates (expected versus measured) is consistent with our suggestion that features other than the TGG$(N)_{6-7}$GCCAA sequence may be important for the binding of nuclear factor I.

The strongest evidence that the TGG$(N)_{6-7}$GCCAA sequence is not the only requirement for the binding of nuclear

	SEQUENCE	NAME
1)	tattt**TGG**att-gaa**GCCAA**tatga	Ad
2)	gaacc**TGG**gcacagt**GCCAA**agttg	FIB-1
3)	aggtc**TGG**ctt-tgg**GCCAA**gagcc	FIB-2
4)	tcctc**TGG**att-tgt**GCCAA**atgtc	FIB-3
5)	atgcc**TGG**aag-gca**GCCAA**atttt	c-MYCII_1
6)	ctggg**TGG**aag-gta**TCCAA**tccag	c-MYCII_2
7)	actgc**TGG**cac-tat**GCCAC**ggcct	BS1b
8)	gcagc**TGG**gcagatt**GCCAA**ggtta	BS2
9)	agaac**TGG**caa-gct**GTCAA**aaaca	BS1a
A	4231----453-223---9843113	
T	222319---23-32511---22442	Frequency
G	22331-992-214318----33212	of Base
C	13127---3211-2--89-1-1232	

TGG--------GCCAA CONSENSUS

FIG. 9. Comparison of FIB sites. Confirmed and putative FIB sites are aligned at the TGG$(N)_{6-7}$GCCAA consensus sequence. Nucleotides present in the highly conserved motifs are shown in upper case. The Ad FIB site is from Nagata et al. (21). FIB-1, FIB-2, and FIB-3 are present in the HeLa genome and are from this study. Sites c-*myc*II_1 and c-*myc*II_2 are located 5′ to the mRNA start site of the human c-*myc* gene (26). BS1b, BS2, and BS1a are located 5′ to the mRNA start site of the chicken lysozyme gene (1). The table below the sequences shows the frequency of occurrence of each base (from nine sequences) at every nucleotide within and around the consensus sequence.

factor I is the presence of this sequence between nucleotides 294 and 307 of pBR322 DNA (29). Previous studies (8, 23) demonstrated that pBR322 DNA neither binds to nuclear factor I nor inhibits the binding of FIB site DNA to the protein. The consensus sequence present in pBR322 is TGGGCGGCGGCCAA. One unusual feature of this sequence that may disrupt its ability to bind nuclear factor I is the exclusive presence of G · C base pairs within the central nonspecific region of the consensus sequence. Mutational analysis of this region in pBR322 may yield important information on the structural requirements for a FIB site.

Nagata et al. (21) showed that nuclear factor I bound poorly to FIB sites that were in a single-stranded form. We utilized this property in our elongated primer selection technique to determine the minimal duplex region required for the binding of nuclear factor I. These studies suggest that nuclear factor I can bind to a FIB site even when the last two A residues of the consensus sequence and the remainder of the DNase I-protected region are in a single-stranded form. Thus, the important contacts between nuclear factor I and the primer strand probably occur 5′ to the final two A residues of the consensus sequence. One objection to this technique is that the presence of a single-strand region adjacent to a binding site may disrupt possible secondary structures required for binding. The apparently weak binding of nuclear factor I to some primer-template DNAs elongated just beyond the consensus sequence (above the 119-bp marker in Fig. 5) may reflect such an inhibition. We are currently investigating the cause of this apparent weak binding. Although the applicability of this technique to the study of other site-specific DNA binding proteins is untested, its usefulness in determining duplex DNA sequences required for the binding of nuclear factor I is clear. This technique might be extended with limited nuclease digestion (6, 25) to produce single-stranded regions in the complementary DNA strand and 5′ side of the consensus sequence. These studies could yield additional information on the requirements for duplex DNA structure in the binding of nuclear factor I and possibly other site-specific DNA binding proteins.

Models for the function of FIB sites in the HeLa genome must take into account both the abundance of sites (1/100 to 1/200 kb [8]) and their location within nonrepetitive DNA (Fig. 8). Such models should also consider the only known role of nuclear factor I, as a protein required for the efficient initiation of Ad DNA synthesis. The frequency of FIB sites in the HeLa cell genome is in reasonable agreement with estimates for the number of origins of DNA replication (13) or nuclear matrix attachment points (31). However, the presence of putative low-affinity FIB sites 5′ to the mRNA start site of the human c-*myc* gene (26) may suggest a role for nuclear factor I in RNA transcription. The use of well-characterized cloned FIB sites should allow the direct testing of various models for the function of nuclear factor I and FIB sites in DNA or RNA metabolism.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant NIH-GM-13344-18 and Training Grant T32CA09060-09 from the National Institutes of Health.

LITERATURE CITED

1. **Borgmeyer, U., J. Nowock, and A. E. Sippel.** 1984. The TGGCA-binding protein: a eukaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. Nucleic Acids Res. **10**:4295–4311.
2. **Challberg, M. D., S. V. Desiderio, and T. J. Kelly, Jr.** 1980. Adenovirus DNA replication *in vitro*: characterization of a protein covalently linked to nascent DNA strands. Proc. Natl. Acad. Sci. U.S.A. **77**:5105–5109.
3. **Challberg, M. D., and T. J. Kelly, Jr.** 1982. Eukaryotic DNA replication: viral and plasmid model systems. Annu. Rev. Biochem. **51**:901–934.
4. **Friefeld, B. R., J. H. Lichy, J. Field, R. M. Gronostajski, R. A. Guggenheimer, M. D. Krevolin, K. Nagata, J. Hurwitz, and M. S. Horwitz.** 1984. The *in vitro* replication of adenovirus DNA. Curr. Top. Microbiol. Immunol. **110**:221–255.
5. **Friefeld, B. R., J. H. Lichy, J. Hurwitz, and M. S. Horwitz.** 1983. Evidence for an altered adenovirus DNA polymerase in cells infected with the mutant H5*ts*149. Proc. Natl. Acad. Sci. U.S.A. **80**:1589–1593.
6. **Guo, L.-H., and R. Wu.** 1982. New rapid methods for DNA sequencing based on exonuclease III digestion followed by repair synthesis. Nucleic Acids Res. **10**:2065–2084.
7. **Gronostajski, R. M., J. Field, and J. Hurwitz.** 1984. Purification of a primase activity associated with DNA polymerase α from HeLa cells. J. Biol. Chem. **259**:9479–9486.
8. **Gronostajski, R. M., K. Nagata, and J. Hurwitz.** 1984. Isolation of human DNA sequences that bind to nuclear factor I, a host protein involved in adenovirus DNA replication. Proc. Natl. Acad. Sci. U.S.A. **81**:4013–4017.
9. **Guggenheimer, R. A., K. Nagata, J. Field, J. Lindenbaum, R. M. Gronostajski, M. S. Horwitz, and J. Hurwitz.** 1983. *In vitro* synthesis of full length adenoviral DNA, p. 395–421. *In* N. R. Cozzarelli (ed.), Mechanisms of DNA replication and recombination. Alan R. Liss, Inc., New York.
10. **Guggenheimer, R. A., B. W. Stillman, K. Nagata, F. Tamanoi, and J. Hurwitz.** 1984. DNA sequences required for the *in vitro* replication of adenovirus DNA. Proc. Natl. Acad. Sci. U.S.A. **81**:3069–3073.
11. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
12. **Kornberg, A. (ed.).** 1980. DNA replication. Freeman, San Francisco.
13. **Lewin, B.** 1980. DNA replication, p. 570–582. *In* B. Lewin (ed.), Gene Expression, vol. 2. John Wiley and Sons, New York.
14. **Lichy, J. H., J. Field, M. S. Horwitz, and J. Hurwitz.** 1982. Separation of the adenovirus terminal protein precursor from its associated DNA polymerase: role of both proteins in the initiation of adenovirus DNA replication. Proc. Natl. Acad. Sci. U.S.A. **79**:5225–5229.
15. **Lichy, J. H., M. S. Horwitz, and J. Hurwitz.** 1981. Formation of a covalent complex between the 80,000-dalton adenovirus terminal protein and 5′-dCMP *in vitro*. Proc. Natl. Acad. Sci. U.S.A. **78**:2678–2682.
16. **Lichy, J. H., K. Nagata, B. R. Friefeld, T. Enomoto, J. Field, R. A. Guggenheimer, J.-E. Ikeda, M. S. Horwitz, and J. Hurwitz.** 1983. Isolation of proteins involved in the replication of adenoviral DNA *in vitro*. Cold Spring Harbor Symp. Quant. Biol. **47**:731–740.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
19. **Nagata, K., R. A. Guggenheimer, T. Enomoto, J. H. Lichy, and J. Hurwitz.** 1982. Adenovirus DNA replication *in vitro*: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc. Natl. Acad. Sci. U.S.A. **79**:6438–6442.
20. **Nagata, K., R. A. Guggenheimer, and J. Hurwitz.** 1983. Adenovirus DNA replication *in vitro*: synthesis of full-length DNA with purified proteins. Proc. Natl. Acad. Sci. U.S.A. **80**:4266–4270.
21. **Nagata, K., R. A. Guggenheimer, and J. Hurwitz.** 1983. Specific binding of a cellular DNA replication protein to the origin of

replication of adenovirus DNA. Proc. Natl. Acad. Sci. U.S.A. **80**:6177–6181.

22. **Ostrove, J. M., P. Rosenfeld, J. Williams, and T. J. Kelly, Jr.** 1983. *In vitro* complementation as an assay for purification of adenovirus DNA replication proteins. Proc. Natl. Acad. Sci. U.S.A. **80**:935–939.
23. **Rawlins, D. R., P. J. Rosenfeld, R. J. Wides, M. D. Challberg, and T. J. Kelly, Jr.** 1984. Structure and function of the adenovirus origin of replication. Cell **37**:309–319.
24. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
25. **Shortle, D., and D. Nathans.** 1978. Local mutagenesis: a method for generating viral mutants with base substitutions in preselected regions of the viral genome. Proc. Natl. Acad. Sci. U.S.A. **75**:2170–2174.
26. **Siebenlist, U., L. Henninghausen, J. Battey, and P. Leder.** 1984. Chromatin structure and protein binding in the putative regulatory region of the *c-myc* gene in Burkitt lymphoma. Cell **37**:381–391.
27. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
28. **Stillman, B. W.** 1983. The replication of adenovirus DNA with purified proteins. Cell **35**:7–9.
29. **Sutcliff, J. G.** 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43**:77–90.
30. **van Bergen, B. G. M., and P. C. van der Vliet.** 1983. Temperature-sensitive initiation and elongation of adenovirus DNA replication in vitro with nuclear extracts from H5ts36-, H5ts149-, and H5ts125-infected HeLa cells. J. Virol. **46**:642–648.
31. **Vogelstein, B., D. M. Pardoll, and D. S. Coffey.** 1980. Supercoiled loops and eucaryotic DNA replication. Cell **22**:79–85.
31. **Wallace, R. B., M. J. Johnson, S. V. Suggs, K. Miyoshi, R. Bhatt, and K. Itakura.** 1981. A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing in the plasmid vector pBR322. Gene **16**:21–26.