Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2432–2436, April 1982
Biochemistry

Sequence-specific endonuclease *Bam*HI: Relaxation of sequence recognition

(solvent effect/sequence relaxation/type II endonuclease)

JAY GEORGE AND JACK G. CHIRIKJIAN

Division of Molecular Genetics, Vincent T. Lombardi Cancer Center, and Department of Biochemistry, Georgetown University Medical Center, Washington, D.C. 20007

Communicated by Max Tishler, November 9, 1981

ABSTRACT **The effect of glycerol on the specificity of DNA cleavage by the restriction endonuclease *Bam*HI has been examined. In addition to the canonical G ↓ G-A-T-C-C site, *Bam*HI cuts DNA at several sites that we have named noncanonical *Bam*HI.1 sites. The number of *Bam*HI.1 sites in simian virus 40 and pBR322 was determined to be 13 for each DNA. Cutting sites determined by DNA sequence analysis include G ↓ G-A-A-C-C, G ↓ G-C-T-C-C, G ↓ G-G-T-C-C, and G-A-A-T-C-C with the complementary strand sequence assignments of G-G-T-T-C-C, G-G-A-G-C-C, G-G-A-C-C-C, and G-G-A-T-T-C. The relaxation in specificity was related to hydrogen bond acceptor and donor sites in the recognition sequence, in an attempt to generate a model of *Bam*HI recognition of cognate sites in DNA.**

Type II restriction enzymes are sequence-specific endonucleases that require only Mg^{2+} as cofactor (1). Much of the current progress in molecular biology has been made possible by the specificity with which DNA fragments can be generated by these enzymes. The enzymes also provide a system for studying protein–nucleic acid interactions at the molecular level. *Bam*HI and its cognate methylase are reported to be the restriction-modification system in *Bacillus amyloliquefaciens* (2, 3). The site of cleavage on DNA is the sequence G ↓ G-A-T-C-C, which is a hexanucleotide that has 2-fold rotational symmetry. DNA is cleaved at this site by hydrolysis of phosphodiester bonds in a staggered fashion that generates 5′-phosphoryl termini (4).

We have previously reported that the cleavage specificity of homogeneous *Bam*HI preparations is altered in the presence of hydrophobic solvents such as glycerol and dimethyl sulfoxide (5). When applied to various DNAs, the activity of the enzyme in which the specificity is altered, designated *Bam*HI.1, generates fragments that are different from those generated by ordinary *Bam*HI. We present here data that allow assignment of specific recognition sites for the various relaxed recognition sequences and propose a model that accounts for the site preferences of *Bam*HI.1.

MATERIALS AND METHODS

All materials were obtained from Bethesda Research Laboratories: ultrapure reagents; pBR322; λ, φX174, and simian virus 40 (SV40) DNAs; restriction endonucleases; polynucleotide kinase; and the M13 sequence analysis kit.

B. amyloliquefaciens (RUB500) cells were grown to midlogarithmic phase, harvested, and stored as described (6).

The standard *Bam*HI reaction mixture was 20 mM Tris·HCl, pH 8.5/10 mM $MgCl_2$/2 mM 2-mercaptoethanol containing 1 μg of DNA and 1 unit of enzyme in 50 μl. To alter the specificity of *Bam*HI, we used a reaction mixture that was identical except that various amounts of hydrophobic reagent were added and the amount of enzyme was 50 units (5). All incubations were carried out at 37°C for 1 hr.

Locations of the altered recognition sites on SV40 and pBR322 DNAs were determined through the procedure of Smith and Birnstiel (7), a mapping technique that has some

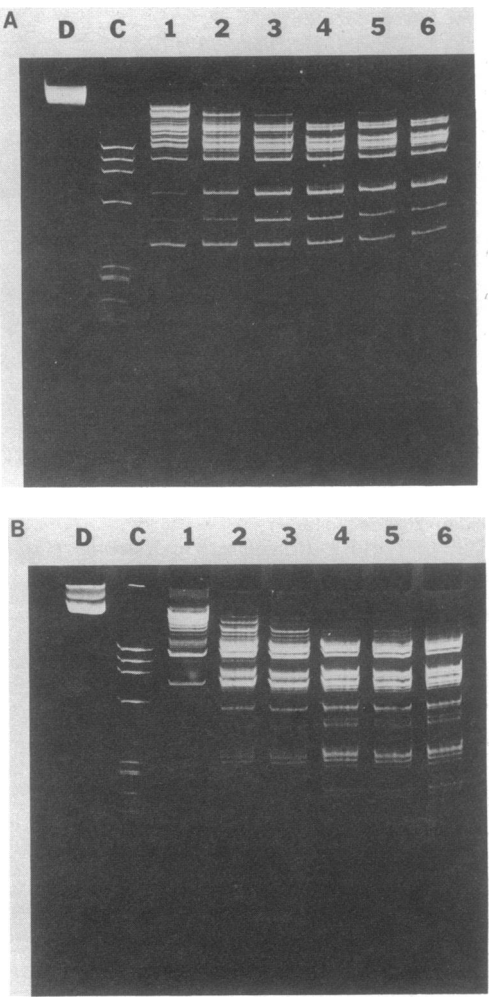

FIG. 1. Digestion of pBR322 and SV40 DNAs in the presence of glycerol: Conditions that generate preferred and nonpreferred fragments. SV40 DNA (*A*) and pBR322 DNA (*B*) were digested with 40 units of *Bam*HI in the presence of 35% (vol/vol) glycerol. Lanes: 1–6, reaction products from incubations of 10, 20, 30, 40, 50, and 60 min, respectively; C, *Hae* III digestive products of φX174; D, intact SV40 DNA.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: SV40, simian virus 40.

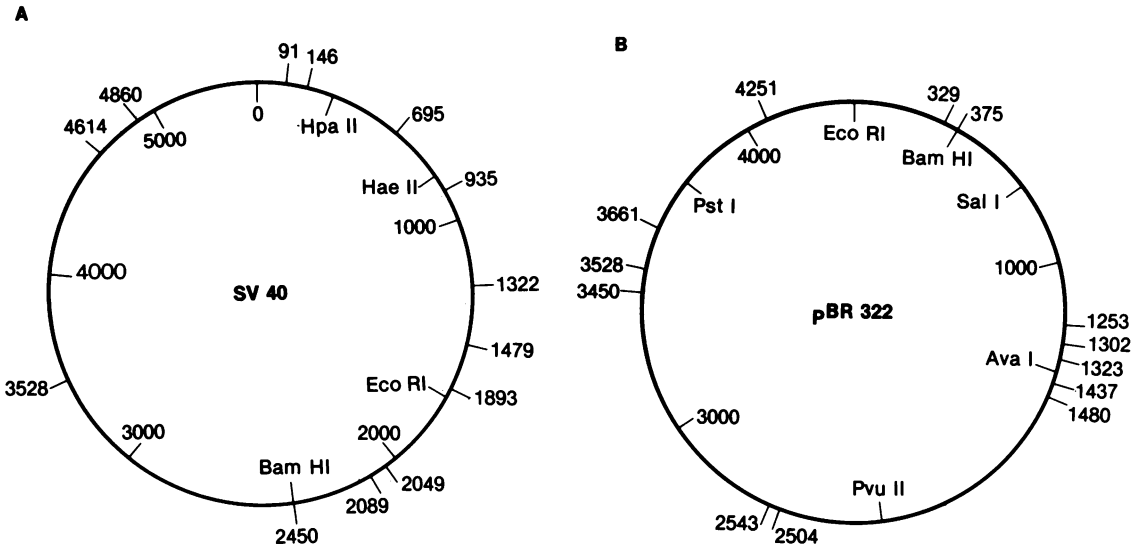

FIG. 2. Restriction maps identifying *Bam*HI and *Bam*HI.1 recognition sites as determined by the method of Smith and Birnstiel (7). (*A*) SV40 DNA. (*B*) pBR322 DNA.

elements in common with the Maxam and Gilbert DNA sequence analysis technique (8). DNA fragments ≈1000 base pairs long were end labeled with polynucleotide kinase and [γ-^{32}P]ATP. One terminus per DNA fragment was then removed by endonuclease cleavage. The fragments that were labeled at a single 5′ end were digested with *Bam*HI.1 under conditions such that the entire spectrum of intermediate products was obtained; these products were later separated by electrophoresis on polyacrylamide gels. Only the intermediate digestion products that contain 5′-end label appear on the autoradiogram and the distance travelled correlates to the distances of the *Bam*HI.1 cleavage sites from the labeled end of the DNA fragment.

The high-resolution gels used for DNA sequence analysis were 8% polyacrylamide/8 M urea/Tris borate buffer (8). They measured 33 cm × 40 cm × 4 mm and were subjected to electrophoresis at 45 mA and 1700 V. The sequence of the *Bam*HI.1 cleavage sites were determined as described (9).

*Bam*HI was purified to homogeneity, as shown by denaturing $NaDodSO_4$/polyacrylamide gel electrophoresis. In some preparations, a protein with a molecular weight of 29,000 copurified with the *Bam*HI; it was later removed by chromatography on Sephadex G-100 (4 × 70 cm) in 20 mM K phosphate, pH 7.0/7 mM 2-mercaptoethanol/7 mM EDTA.

RESULTS

A number of restriction endonucleases are known to undergo a relaxation in their assigned recognition sites. Sequence recognition has been changed by varying the pH and ionic strength, by substituting metal ions, and by introducing organic solvents (10, 11). Attempts to identify the probable recognition sites for *Bam*HI.1 by analysis of sized DNA fragments from SV40 and pBR322 were inconclusive because final reaction products could not be obtained (5). Bands that lightly stained with ethidium bromide were obtained in nonequimolar quantities when constant amounts of *Bam*HI.1 and either SV40 DNA or pBR322 DNA were incubated for 30–150 min (Fig. 1 *A* and *B*, lanes 1–6). These light-staining bands were of higher molecular weight as compared with some of the more intensely staining bands and are indicative of incomplete DNA digestion products. Based on visual examination, the relative concentrations of these products changed very slowly between 90 and 180 min of digestion (lanes 4–6). The relatively slow rate of production of these DNA digestion products suggests that *Bam*HI.1 activity cleaves certain sites more slowly than others. We chose to concentrate our effort on determination of the faster-cleaved or preferred sites by constructing restriction site cleavage maps as described by Smith and Birnstiel (7). This mapping procedure requires that all sites be cleaved by the restriction enzyme at approximately the same rate. Because *Bam*HI.1 has site preference, weaker sites may not be detected by this technique (Fig. 1).

Mapping was facilitated by the availability of the entire sequences of pBR322 and SV40 DNAs (12, 13). To ensure that the mapping would be accurate, we chose restriction fragments up to 1100 base pairs from SV40 and pBR322 DNAs. By analyzing the sizes of the intermediate digestion products of known fragments, we constructed restriction site cleavage maps of SV40 DNA and pBR322 that indicated the *Bam*HI.1 sites (Fig. 2). pBR322 and SV40 DNA each contain a single G ↓ G-A-T-C-C *Bam*HI site. Under our experimental conditions, 13 strong *Bam*HI.1 recognition sites, including G-G-A-T-C-C, were detected in both SV40 DNA and pBR322 DNA.

Sequence analysis of DNA fragments obtained from pBR322 and ϕX174 replicative form DNAs indicated that G-G-A-N-C-

Table 1. Diversity of *Bam* HI.1 cleavage sites

	Noncanonical			
Canonical	G-G-A-N-C-C	G-G-N-T-C-C	G-N-A-T-C-C	N-G-A-T-C-C
G ↓ G-A-T-C-C	G - G-A-G-C-C	G ↓ G-C-T-C-C	G ↓ A-A-T-C-C	?
	G - G-A-C-C-C	G ↓ G-G-T-C-C	?	?
	G ↓ G-A-A-C-C	G - G-T-T-C-C	?	?

↓, Specific cleavage sites as determined by DNA sequence analysis (also listed are sequences on the complementary strand); ?, potential *Bam*HI.1 sites based on preliminary observations.

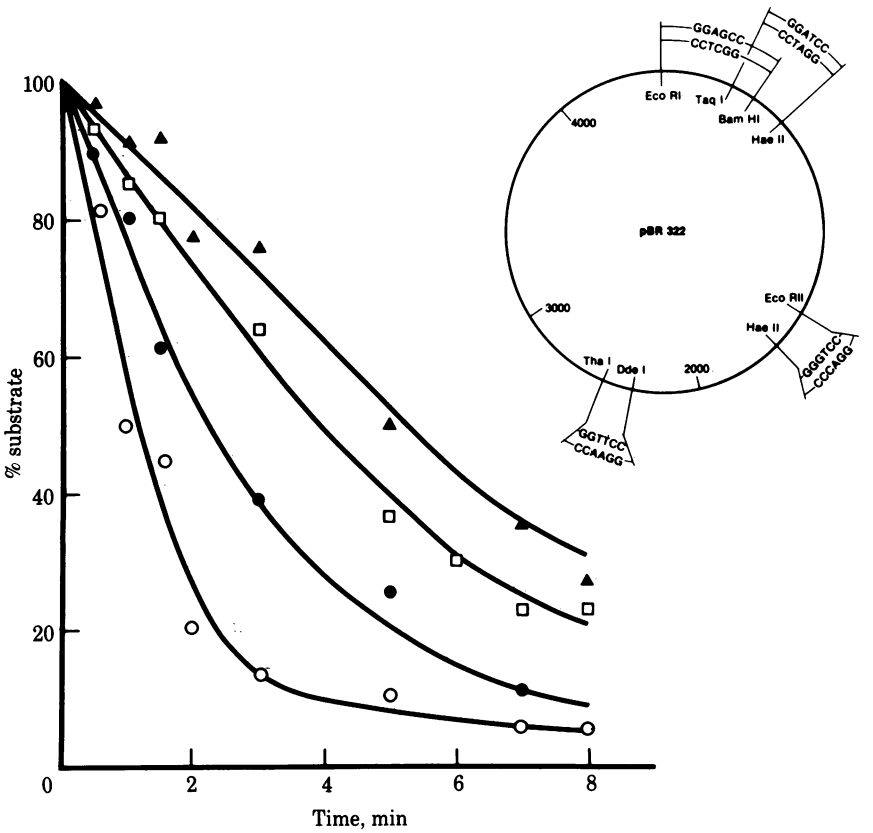

FIG. 3. Relative rates of catalysis of the strong *Bam*HI.1 sites by *Bam*HI. Restriction fragments from pBR322 containing single *Bam*HI.1 recognition sites were isolated as follows: G-G-C-T-C-C (□) at position 329 was isolated by digestion with *Eco*RI/*Bam*HI, G-G-G-T-C-C (●) at position 1480 was obtained by digestion with *Eco*RII/*Hae* II, the canonical *Bam*HI site G-G-A-T-C-C (○) at position 375 was isolated by digestion with *Taq* I/*Hae* II, G-G-T-T-C-C (▲) at position 2504 was obtained by digestion with *Dde* I/*Tha* I. DNA fragments were end labeled with polynucleotide kinase and [γ-^{32}P]ATP. Reactions were carried out in 100-μl of standard reaction mixture/36% glycerol and the appropriate restriction fragments containing the specific *Bam*HI.1 sites. Aliquots were removed at various times and quantitated after separation by polyacrylamide gel analysis.

C, G-G-N-T-C-C, and G-A-A-T-C-C are the sites responsible for *Bam*HI.1 activity. As representative sites, we determined the sequences of G ↓ G-A-A-C-C (G-G-A-N-C-C), G ↓ G-C-T-C-C and G ↓ G-G-T-C-C (G-G-N-T-C-C), and G ↓ A-A-T-C-C (G-N-A-T-C-C). All fragments whose sequences were determined were obtained from pBR322 except the fragment that contained G ↓ G-A-A-C-C, which was from ϕX174 replicative form DNA. Additional preferred sites were also assigned in SV40 DNA on the basis of computer analysis in conjunction with the mapping data. The sites cleaved in the presence of glycerol are given in Table 1.

The *Bam*HI.1 sites were characterized with respect to preferential rate of cleavage. pBR322 was searched for sequences that contain *Bam*HI.1 sites. By using a combination of restriction endonuclease digests, we obtained fragments that contained a single strong site. The sequence G-G-A-G-C-C was obtained by *Eco*RI/*Bam*HI digestion, and G-G-A-T-C-C was obtained by *Taq* I/*Dde* I digestion (Fig. 3 *Inset*). The fragments containing unique *Bam*HI.1 sites were digested with 50 units of enzyme in the presence of 36% glycerol at 37°C for 8 min. At various times, aliquots were removed and the reaction products were separated by gel electrophoresis. The resulting bands were excised and quantitated. To determine the relative rates of digestion, the decrease in substrate was plotted versus time (Fig. 3). The order of the rates of digestion of the *Bam*HI.1 sites tested was G-G-A-T-C-C > G-G-G-T-C-C > G-G-C-T-C-C > G-G-T-T-C-C.

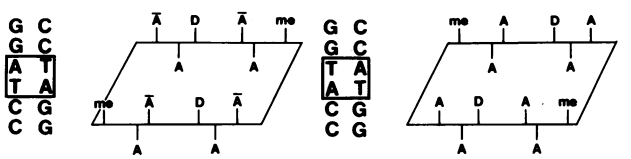

FIG. 4. Schematic of A·T and T·A base pairs, illustrating methyl groups and positions available for hydrogen bond interactions between the DNA and the enzyme. A and D, hydrogen bond acceptor and donor, respectively; me, 5′-thymine methyl group. Overbars indicate proposed sites for *Bam*HI.1 interactions. Hydrogen bond acceptor positions are *N*-7 of adenine and guanine, *O*-4 of thymine, and *O*-6 of guanine. Hydrogen bond donor positions are *N*-6 of adenine and *N*-4 of cytosine. The horizontal line represents the hydrogen bond between the *N*-1 of the pyrimidine and the *N*-3 of the purine of the base pair. Hydrogen bond acceptors and donors above the horizontal line face into the major groove of the DNA and those below are in the minor groove.

DISCUSSION

Several type II restriction endonucleases undergo a relaxation in their sequence specificity under specific conditions (10, 11). Interpretation of published results is difficult because most of the studies have involved nonhomogeneous enzyme preparations (11, 14). Changes in the reaction environment could result in activation of trace amounts of contaminating nucleases that, in turn, could alter the DNA fragmentation pattern. Among the enzymes that have been purified to homogeneity, *Eco*RI (15), *Bam*HI (5), *Hha* II (16), and *Bst* I (17) reportedly recognize sites other than their assigned canonical palindromes; *Bsu* I, *Eco*RI, and *Bam*HI have been studied in detail.

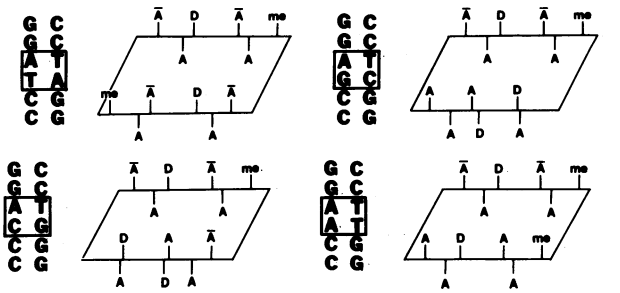

FIG. 5. Schematic representation of proposed hydrogen bond interactions between *Bam*HI and the internal position of *Bam*HI.1. Symbols used are as in Fig. 4.

Polisky *et al.* (15) reported an activity of *Eco*RI—designated *Eco*RI*—that recognized and cleaved the sequence A-A-T-T, the internal tetranucleotide of the *Eco*RI site G-A-A-T-T-C. Woodbury *et al.* (18) subsequently deduced with DNA mapping the *Eco*RI* activity to cleave a hierarchy of both weak and strong cleavage sites.

Under standard reaction conditions, ϕX174 replicative form DNA is not cleaved by *Eco*RI. However, at low ionic strength and high pH, it was readily cleaved at five sites (18, 19). On the basis of information from mapping and end-group analysis, the strong sites were identified; several weaker sites were detected but not studied further. There were comparable effects when Mn^{2+} was added to *Eco*RI reaction mixtures (20).

Similarly, secondary sites are cleaved by *Bam*HI when reaction conditions are changed. An additional activity that copurified with *Bam*HI and was present in homogeneous preparations has been reported (6), and it was concluded that this activity, designated HI.1, was intrinsic to *Bam*HI and expressed in the presence of glycerol.

The results presented here indicate that the *Bam*HI.1 recognition sites include G ↓ G-A-A-C-C, G ↓ G-C-T-C-C, G ↓ G-G-T-C-C, and G ↓ A-A-T-C-C. Because cleavage of DNA by restriction enzymes requires double-stranded DNA cleavage, the complementary sequences G-G-T-T-C-C, G-G-A-G-C-C, G-G-A-C-C-C, and G-G-A-T-T-C must also be recognized by *Bam*HI.1 (Table 1). The *Bam*HI.1 site we originally reported as C-C-A-G-A-A-G-G ↓ C-G-G-T-T-C (5) was reassigned as C-C-A-G-A-A-G-G-C-G ↓ G-T-T-C-C.

Seeman *et al.* (21) proposed that the information required for specific binding of a protein to a nucleic acid molecule resides in the intricate pattern of hydrogen bond donors and acceptors possessed by the individual base pairs. In their proposed model, a protein must make at least two hydrogen bond contacts per base if there is to be unambiguous discrimination between base pairs.

A schematic illustration of such base-pair interactions was suggested byWoodbury *et al.* (18). The relative positions available for hydrogen-bonding interactions of the A·T base pair in the *Bam*HI palindrome G-G-A-T-C-C and a T·A base pair are identical in the minor groove but are inverted in the major groove as illustrated in Fig. 4. *Bam*HI.1 does not cleave the sequence G-G-T-A-C-C, which suggests that *Bam*HI discriminates between A·T and T·A base pairs by probing nonidentical interaction points in the major groove.

Berkner and Folk (22) have shown that uracil-substituted DNA has little or no effect on the kinetic properties of *Bam*HI. This suggests that the methyl group of thymidine is not essential for recognition of G-G-A-T-C-C by *Bam*HI. Likewise, methylation at the *N*-6 position of the adenine in the palindrome has little effect on *Bam*HI activity (unpublished result). Given the requirement that two hydrogen bonds must be recognized, the only remaining available positions in the major groove are the *N*-7 of adenine and the *O*-4 of thymine.

It is attractive to interpret rate variations among the different *Bam*HI.1 sites in terms of hydrogen bond interactions at the A·T region in the major groove of the palindrome. The hydrogen bond capacity of the internal dinucleotide sequences that are recognized by *Bam*HI.1 are illustrated in Fig. 5. Under normal conditions, *Bam*HI could make four hydrogen bond contacts in the major groove at the A·T region during catalysis. When the specificity is relaxed, resulting in additional cleavage (as in the presence of glycerol), the enzyme may require fewer hydrogen bonds for activity. When purine-to-purine and pyrimidine-to-pyrimidine substitution occurs, as in the difference between the *Bam*HI $\frac{\text{G-G-A-T-C-C}}{\text{C-C-T-A-G-G}}$ site and one of the *Bam*HI.1 $\frac{\text{G-G-G-T-C-C}}{\text{C-C-C-A-G-G}}$ sites, the *N*-7 purine contact point remains unchanged. Therefore the $\frac{\text{G-G-G-T-C-C}}{\text{C-C-C-A-G-G}}$ *Bam*HI.1 sequence maintains three of the four original hydrogen bond contacts, *N*-7 of adenine, *O*-4 of thymidine, and *N*-7 of substituted purine, and has the second fastest cleavage rate compared with that of the *Bam*HI G-G-A-T-C-C site. In the *Bam*HI.1 sequence $\frac{\text{G-G-G-T-C-C}}{\text{C-C-C-A-G-G}}$, a single hydrogen bond contact point is lost when cytosine is substituted for thymidine. In contrast, in the case of $\frac{\text{G-G-A-G-C-C}}{\text{C-C-T-C-G-G}}$ and $\frac{\text{G-G-A-A-C-C}}{\text{C-C-T-T-G-G}}$, two of the four original hydrogen bond contact points are absent, *N*-7 of adenine and *O*-4 of thymidine. The latter two sequences have the lowest rate of cleavage among the sites examined. The observable cleavage-rate difference between the two sequences (Fig. 3) may be due to substitution of the *O*-4 thymidine position by the *O*-6 guanine position, which may occupy a stereochemically similar position in G-G-A-G-C-C and therefore may be recognized by the enzyme (23).

The data presented here indicate that the *Bam*HI relaxation occurs at the internal nucleotide sequence of the G-G-A-T-C-C palindrome. The process appears to be different in *Eco*RI (18); the relaxation specificity occurs at the outer nucleotide positions of the recognition site. Under our experimental conditions DNA sequences in which the first guanine of the *Bam*HI recognition site can be replaced by another base have not been observed to serve as preferred substrates for *Bam*HI.1 activity (Table 1). We concentrated on determination of the preferred cleavage sites under our experimental conditions. Therefore, sequences that constitute nonpreferred cleavage sites may not have been detected. These secondary sites appear to vary in their susceptibility to *Bam*HI.1 cleavage and result in generation of nonequimolar amounts of DNA fragments.

We thank Drs. Robert Blakesley and Richard Woodbury for helpful discussions. This work was supported by U.S. Public Health Service Grant GM 27701.

1. Smith, H. O. (1979) *Science* **205**, 455–462.
2. Wilson, G. A. & Young, F. E. (1977) *J. Mol. Biol.* **97**, 123–125.
3. Shibata, T. & Ando, T. (1976) *Biochim. Biophys. Acta* **442**, 184–196.
4. Roberts, R. J., Wilson, G. A. & Young, F. E. (1977) *Nature (London)* **265**, 82–84.
5. George, J., Blakesley, R. W. & Chirikjian, J. G. (1981) *J. Biol. Chem.* **255**, 6521–6524.
6. Smith, L. A. & Chirikjian, J. G. (1979) *J. Biol. Chem.* **254**, 1003–1006.
7. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3**, 2387–2398.
8. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
9. McConnell, D. J., Searcy, D. G. & Sutcliffe, J. G. (1978) *Nucleic Acids Res.* **5**, 1729–1739.
10. Chirikjian, J. G. & George, J. (1981) in *Gene Analysis and Amplification*, ed. Chirikjian, J. G. (Elsevier/North-Holland, New York), Vol. 1, pp. 74–98.
11. Malyguine, E., Vannier, P. & Yot, P. (1980) *Gene (Amsterdam)* **8**, 163–177.
12. Sutcliffe, J. G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
13. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, S., Pna, J., Ghosh, P. K., Celma, M. L. & Weissman, S. M. (1978) *Science* **200**, 494–502.
14. Tikchonenko, T. I., Karamov, E. V., Zarizion, B. A. & Naroditsky, B. S. (1978) *Gene (Amsterdam)* **4**, 195–212.
15. Polisky, B., Greene, P., Garfin, D. E., McCarthy, B. J., Goodman, H. M. & Boyer, H. W. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3310–3314.
16. Mann, M. B. (1981) in *Gene Analysis and Amplification*, ed. Chirikjian, J. G. (Elsevier/North-Holland, New York), Vol. 1, pp. 230–237.
17. Clarke, C. M. & Hartley, B. S. (1979) *Biochem. J.* **177**, 49–62.
18. Woodbury, C. P., Jr., Hagenbuchle, O. & Von Hippel, P. H. (1980) *J. Biol. Chem.* **255**, 11534–11546.

19. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchinson, C. A., III, & Slocombe, P. M. (1978) *J. Mol. Biol.* **125**, 225–246.
20. Hsu, M. & Berg, P. (1978) *Biochemistry* **17**, 131–138.
21. Seeman, N. C., Rosenberg, J. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 804–808.
22. Berkner, K. L. & Folk, W. R. (1979) *J. Biol. Chem.* **254**, 2551–2560.
23. Rosenberg, M. J., Boyer, H. W. & Greene, P. (1981) in *Gene Analysis and Amplification*, ed. Chirikjian, J. G. (Elsevier/North-Holland, New York), Vol. 1, pp. 132–164.