lop-11 *lig*$^+$-1 were then isolated as described (*2*). Since these phages contain the attachment site and *int* and *xis* genes, they can integrate into the host chromosome and can be maintained in the prophage state (*6*). Recombinants containing *lig* can be easily detected since a conditionally lethal ligase mutant, *lig*ts 7 (*7*), will grow at 43°C when λgt4-*lop*-11 *lig*$^+$ is carried as a prophage. A double lysogen of λ$^+$ and λgt4-*lop*-11 *lig*$^+$ was used to prevent the induction of phage growth at high temperature that occurs as a result of the temperature-sensitive mutation (*c*I857) in the λ repressor carried by the λgt vector. Lysogens containing λgt4 or λgt4-*lop*-11 *lig*$^+$ were induced by a shift to 43°C, and the cell extracts were assayed for DNA ligase activity (Table 1). This modified vector system produced an increase in ligase levels to approximately 100-fold over that in uninfected *lig*$^+$ cells, a significant improvement over the tenfold increase observed with λgt-*lop*-11 *lig*$^+$ (*2*).

λgt4-*lop*-11 *lig*$^+$ was further modified by introducing an amber mutation (S7) in the S gene (which prevents cell lysis) (*8*) by genetic recombination (*9*) between λgt4-*lop*-11 *lig*$^+$ and λ*c*I857 *nin*5 S7. Lysogens containing λgt4-*lop*-11 *lig*$^+$ S7 can be incubated for up to 2.5 hours after induction without significant cell lysis. DNA ligase levels in these cells increased continuously from the time of induction until lysis finally occurred (Fig. 2). Extracts of such cells harvested 2 hours after induction showed DNA ligase levels that exceeded those in uninfected *lig*$^+$ extracts by more than 500-fold. Since there is also overproduction of λ capsid proteins, the increase in specific enzyme activity is somewhat less (400-fold). We estimate that ligase represents approximately 5 percent of the protein in these extracts.

Early work with the lactose repressor carried on a defective λ phage demonstrated the correlation between the number of copies of lactose *i* gene and the amount of repressor present (*10*). Since then, a variety of phages have been used as vehicles for gene amplification. For example, a λ transducing phage that contains the DNA ligase gene has been isolated which, during vegetative growth, produced a 30-fold increase in ligase activity (*11*). The λ vector system that we have constructed in vitro has a number of very attractive features. Cells containing the vector as a prophage are stable, and λ carries no known translocons capable of the *rec*A-independent transfer of DNA segments known to occur with some plasmids (*12*). Since the hybrid DNA is present in only a single copy in the prophage, it should be possible to clone genes in λ which would be lethal to the host in multiple copies. Furthermore, the presence of the cloned gene in a single, unique sequence within the cell lessens the chance of sequence variants arising among a pool of gene copies. Amplification of the gene and, hence, the amount of enzyme coded for by the gene, is achieved, when desired, by induction of the prophage. The amber mutation in the S gene allows a long period of gene amplification and expression before extracts are prepared and, thus, even further enhances overproduction. It is obvious that the enzyme overproduction attained with this system significantly reduces the amount of purification necessary to obtain homogeneous protein and thus should make it substantially easier to prepare the large quantities of pure DNA ligase required for the detailed analysis of its structure and mechanism.

SHARON M. PANASENKO
JOHN R. CAMERON
RONALD W. DAVIS, I. R. LEHMAN
Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

References and Notes

1. D. Vapnek, N. K. Alton, C. L. Bassett, S. R. Kushner, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3492 (1976); V. Hershfield, H. W. Boyer, C. Yanofsky, M. A. Lovett, D. R. Helinski, *ibid.* **71**, 3455 (1974); A. S. Hopkins, N. E. Murray, W. J. Brammar, *J. Mol. Biol.* **107**, 549 (1976); L. Clarke and J. Carbon, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4361 (1975).
2. J. R. Cameron, S. M. Panasenko, I. R. Lehman, R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3416 (1975).
3. M. Thomas, J. R. Cameron, R. W. Davis, *ibid.* **71**, 4579 (1974).
4. M. M. Gottesman, M. L. Hicks, M. Gellert, *J. Mol. Biol.* **77**, 531 (1973).
5. J. Shapiro, L. MacHattie, L. Eron, G. Ihler, K. Ippen, J. Beckwith, *Nature (London)* **224**, 768 (1969); A. Rambach and P. Tiollais, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3927 (1974).
6. M. E. Gottesman and M. B. Yarmolinsky, *Cold Spring Harbor Symp. Quant. Biol.* **33**, 735 (1968).
7. C. Pauling and L. Hamm, *Proc. Natl. Acad. Sci. U.S.A.* **60**, 1495 (1968).
8. A. R. Goldberg and M. Howe, *Virology* **38**, 200 (1969).
9. J. S. Parkinson, *Genetics* **59**, 311 (1968).
10. W. Gilbert and B. Müller-Hill, in *The Lactose Operon*, J. R. Beckwith and D. Zipser, Eds. (Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y., 1970), p. 93.
11. M. M. Gottesman, *Virology* **72**, 33 (1976).
12. S. Cohen, *Nature (London)* **263**, 731 (1976).
13. P. Modrich and I. R. Lehman, *J. Biol. Chem.* **245**, 3626 (1970).
14. O. H. Lowry, N. J. Rosebrough, A. L. Farr, R. J. Randall, *ibid.* **193**, 265 (1951).
15. Supported by National Cancer Institute contract NO1 CP61026 and NCI postdoctoral fellowship CA 09151 (to J.R.C.).

31 January 1977

Interspersion of Short Repetitive Sequences Studied in Cloned Sea Urchin DNA Fragments

Abstract. *The length and spacing of repetitive sequence elements were studied in several cloned sea urchin DNA fragments. Sequence organization in these fragments was found to be of the short interspersed form expected for sea urchin DNA, and for one clone the evidence shows that nearby repeats do not belong to the same repetitive sequence family.*

Interspersion of short repetitive DNA sequences with single copy sequences appears to be a widespread feature of higher animal genomes. DNA sequence organization in many animal genomes, representing a wide range of phylogenetic groups, has now been examined in some detail (*1–8*). In the typical case a majority of the genome consists of short repetitive sequences about 300 nucleotide pairs long interspersed with single copy sequences about 1000 to 3000 nucleotide pairs long. Exceptions to this general pattern of DNA sequence organization have so far been found only in the DNA's of several insects, including *Drosophila melanogaster* (*5, 6*), the honeybee *Apis mellifera* (*7*), and the midge *Chironomus tentans* (*8*). The genomes of these organisms are relatively small, and they do not contain detectable quantities of short repetitive DNA sequences interspersed with single copy DNA.

Current knowledge of DNA sequence organization derives largely from measurements of binding to hydroxyapatite of partially renatured DNA fragments of various lengths which contain repetitive duplex. The fraction of these fragments included in repetitive duplex regions and in the flanking single-strand nonrepetitive "tails" has been assayed optically, and by measuring the amount of renatured DNA that is resistant to digestion with single-strand specific nuclease. Additional evidence has been obtained by electron microscopy of the renatured DNA structures (*3–5*). These methods have yielded a quantitative insight into the average patterns of sequence organization in animal DNA's. We now describe recent progress in analyzing the specific features of interspersed sequence organization in several cloned fragments of sea urchin DNA. In such fragments, it is possible to determine the spacing and the exact length of the individual repetitive sequence elements, as well as their repetition frequency in the genome. An additional question of significant theoretical interest (*9*), which can be investigated in cloned DNA fragments, is whether neighboring

interspersed repetitive sequences are members of the same or of different repetitive sequence families.

Eco RI fragments of sea urchin DNA which contain both unique and repetitive sequences were cloned in the plasmid pSC101 and in an ampicillin resistant derivative of Col E1, RSF2124 (*10*). The methods used and the general characteristics of the clones obtained have been described. Eight cloned fragments have been examined to various extents in the present work. These range in size from 3300 to 8300 nucleotide pairs. The repetitive sequence content of each fragment was determined by reacting labeled cloned DNA fragments with excess 2000 nucleotide long sea urchin DNA to C_0t 20 (C_0 is the concentration of nucleotides in moles per liter, and t is the time in seconds), and then measuring the fraction of the cloned fragment resistant to digestion with the single-strand specific S1 nuclease. The resistant fraction measures the amount of duplex formed between the repetitive sequences of the cloned DNA fragments and the excess of sea urchin DNA. The result of these analyses are summarized in Table 1. Here it can be seen that seven out of the eight cloned fragments contain repetitive sequences, and that these constitute 6 to 20 percent of the total sea urchin DNA in the cloned inserts. One fragment contains no detectable repetitive sequence. These results are approximately consistent with expectation for random samples of the sea urchin genome. Thus, at least 75 percent of the total DNA of this organism consists of repetitive sequences averaging 300 nucleotides in length interspersed with longer single copy sequences (*2*). Regions displaying this form of sequence organization (*2*) include on the average about 14 percent of the genome as repetitive sequence, and ≥ 61 percent as single copy sequence, or about 300 nucleotides of repetitive sequence per ≥ 1300 nucleotides of single copy DNA. In some cases the fraction of repetitive sequence will be significantly smaller since many single copy sequences are well over 2000 nucleotides long (*2*). Similarly, some interspersed regions where the single copy sequences are only 1000 nucleotides or less (*2*) could have as much as 25 percent repetitive sequences. An interesting question that can be answered by studying cloned sequences is whether relatively short and relatively long single copy sequence elements tend to be contiguous to each other, or whether they tend to be scrambled.

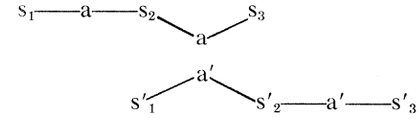
Table 1. Repetitive sequence content of cloned sea urchin DNA fragments. Abbreviation: NTP, nucleotide pairs (*18*).

Clone number*	Fragment size† (NTP)	S1 resistance (%)‡	Σ repetitive sequences (NTP)
CS0870	3300	7	230
CS0895	3700	0	0
CS0849§	5200	20	1040
	6400	6	390
CS0859	5700	15	855
PSC34	7000	13	900
CS0025	8000	12	960
CS0818	8300	18	1440

*The designation CS indicates a sea urchin fragment cloned in plasmid RSF2124. PSC34 indicates a sea urchin fragment cloned in plasmid pSC101. †Estimated from electrophoretic mobility in agarose gels. ‡The S1 resistant duplex content at C_0t 20 was measured by binding to hydroxyapatite. §Clone CS0849 contains two sea urchin fragments, one 5200 nucleotide pairs long, the other 6400 nucleotide pairs long.

To determine the length of the repetitive sequence elements in the cloned fragments, we measured the size of the S1 nuclease resistant DNA duplexes on polyacrylamide gels. Experimental procedures are described briefly in the legends of Figs. 1 and 2. Self-reaction of the cloned sea urchin fragments was avoided by the use of strand-separated cloned fragments. Results for two of the cloned fragments are shown in Fig. 1, A and B. It can be seen that the individual repetitive sequences in these clones, PSC34 and CS0859, are short. Their length is about 300 to 500 nucleotide pairs. Together with the measurements shown in Table 1, these data indicate that there are two or three such sequences in each of the two cloned fragments. The cloned fragments thus appear fairly typical of the interspersed repetitive sequences of the sea urchin genome.

If they are interspersed with single copy sequence, the individual repetitive sequence elements on each fragment could be homologous; for example, s_1-a-s_2-a-s_3, where "s" is single copy sequence and "a" is a given repetitive sequence, or they could be heterologous, for example, s_1-a-s_2-b-s_3, where "a" and "b" belong to different repetitive sequence families. If they are homologous, several different registrations are possible when the DNA is renatured, and a branched or looped renaturation product will sometimes be formed; for example,



If the repeats are heterologous, however, only one registration is possible:

s_1 -a -s_2 -b -s_3

s'_1-a′-s'_2-b′-s'_3

To distinguish between these alternatives the cloned fragments were renatured and the structures formed were examined in the electron microscope. Out of hundreds of examples of structures of the length of the sea urchin DNA inserts inspected, only long, perfectly paired, nonbranched duplexes were observed. A sample is shown in Fig. 2A. This experiment in it-

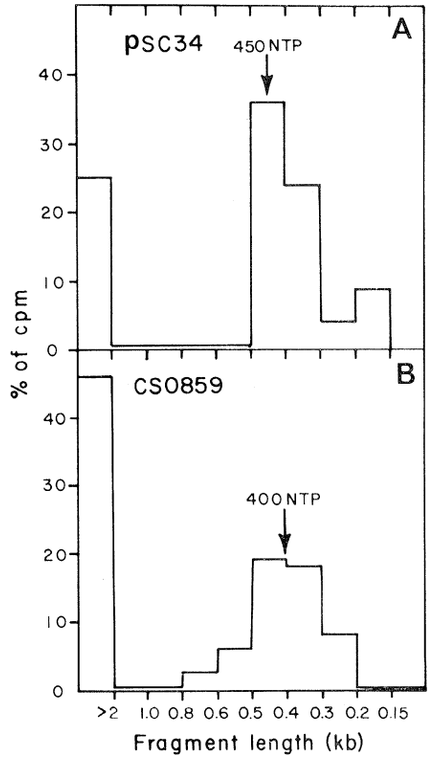


Fig. 1. Size determinations for repetitive sequence elements in clones PSC34 and CS0859. Plasmid DNA's were labeled in vivo with [^{3}H]-thymine. (A) PSC34 supercoils were treated with Eco RI, and the 7000-nucleotide-pair sea urchin fragment was isolated by gel electrophoresis. The (+) and (−) strands of the 7000-nucleotide-pair fragment were prepared by alkaline denaturation of the isolated fragment, followed by electrophoresis in a 1 percent agarose gel (*14*). The separated strands were renatured to C_0t 20 with excess sea urchin DNA 2000 nucleotides in length, and the reaction mixtures were treated with S1 nuclease (*15*). The S1 resistant DNA duplexes were collected on hydroxyapatite columns and sized on a 4 percent polyacrylamide gel with ϕX174 RF restriction fragments as size markers (*16*). DNA at the top of the gel is renatured PSC34 derived from a small percent of cross contamination with the other strand. (B) CS0859 supercoils were treated with Bam HI, which cuts the RSF2124 genome once, but does not cleave this particular sea urchin DNA sequence. The linear molecules of unit length containing both the plasmid and the sea urchin DNA fragment sequence were then strand separated in a poly-ribo(uridylate-guanylate) CsCl gradient (*17*). The separated strands were hybridized to sea urchin DNA, treated with S1 nuclease, and sized on an acrylamide gel as described in (A); *NTP*, nucleotide pairs.

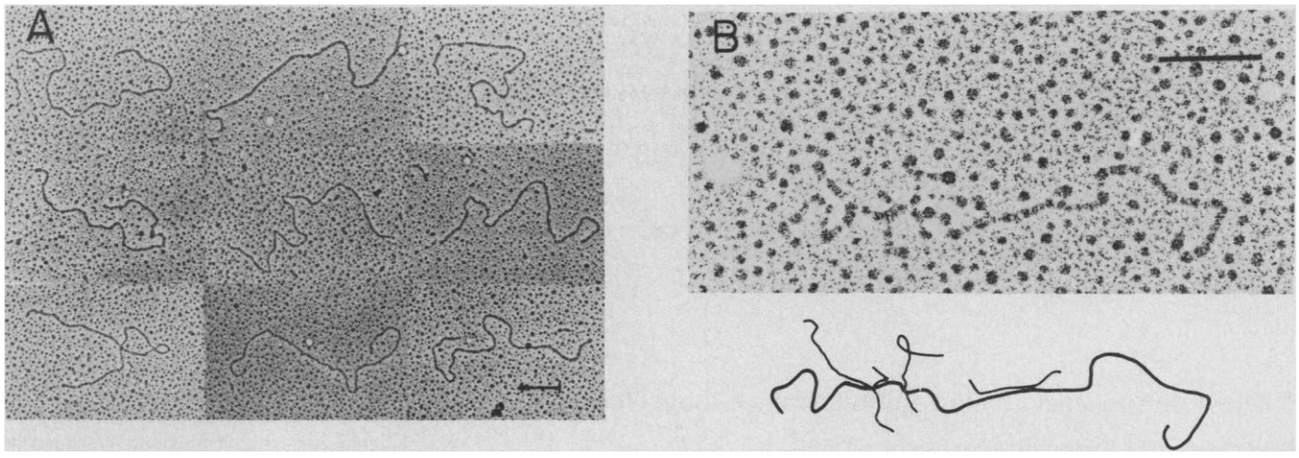


Fig. 2. Electron micrographs of renatured cloned sea urchin DNA fragments. (A) Unit-length sea urchin DNA fragments from plasmid PSC34 were denatured with alkali and renatured. They were then dialyzed into a solution of 57 percent formamide, 0.1*M* tris, 0.01*M* EDTA, *p*H 8.5 (*12*). They were spread for electron microscopy from this medium by a modification (*12*) of the Kleinschmidt procedure. Several renatured PSC34 fragments are shown. The bar represents 1 kilobase pair. (B) A PSC34 molecule is shown after renaturation with excess sea urchin DNA to C_0t 20. A terminal poly(A) tail had been added to the PSC34 fragment as described in text, and most of the excess sea urchin DNA was removed by passing the reaction mixture over an oligo(dT) column. The bar represents 1 kilobase. An interpretation of the complex structure formed is shown below the photograph.

self does not necessarily prove that the adjacent repetitive sequences are heterologous, since it is difficult to exclude the possibility of branch migration. The interpretation given here has been confirmed in an independent manner for clone PSC34 as follows. The sea urchin DNA insert was digested with restriction endonuclease Hha I, yielding nine discrete fragments. When reacted with sea urchin DNA, six of these proved to contain repetitive sequence. None of the Hha I fragments were able to react with each other. Therefore, there is no homology among the repetitive sequences on this fragment. At least for this case, the proximal repeat elements must belong to distinct repetitive sequence families.

The total length of the six fragments that contain repetitive sequence is 3200 nucleotide pairs. However, as Table 1 shows, PSC34 contains only about 900 nucleotide pairs of repetitive sequences. Therefore, 2300 nucleotide pairs of single copy DNA must be included in these fragments. Since the largest two Hha I fragments containing repetitive sequences are 640 and 700 nucleotide pairs, the single copy DNA must be included in more than two fragments. It follows that more than one repetitive sequence element exists on PSC34. This conclusion is consistent with the data of Fig. 1A, where it is shown that the individual repetitive sequence elements are only 300 to 500 nucleotides in length.

To investigate the spacing of the repetitive sequences of the cloned fragments, we developed a technique for visualizing the interspersion pattern in the electron microscope. The 7000-nucleotide sea urchin fragment was purified from the plasmid sequence after Eco RI treatment by gel electrophoresis, and polyadenylate [poly(A)] tracts of a few hundred nucleotides in length were enzymatically attached to the 3′ termini of the fragment by the terminal transferase method (*11*).

The cloned sea urchin DNA fragment was reacted to C_0t 20 with excess sea urchin DNA of average fragment length of 1400 nucleotides, and the partially renatured molecules were then passed over an oligo(dT) (dT, deoxythymidylate) cellulose column. The bound fraction was eluted with a solution of 10 m*M* tris, *p*H 7.4, 5 m*M* NaCl, and 1 m*M* EDTA at room temperature; under these conditions, the repetitive DNA duplexes remained intact. The bound fraction was recycled over another oligo(dT) cellulose column to further reduce the driver DNA background. It was then concentrated and prepared for electron microscopy (*12*). Molecules of the length of the PSC34 fragment were observed in the spreads, and these always displayed branched structures formed by reaction with sea urchin DNA. Shorter DNA driver molecules displaying typical interspersed branched structures (*3*) were also observed. An example of a complex of PSC34 fragment and sea urchin DNA is reproduced in Fig. 2B. There appear to be two duplex regions separated by about 1000 nucleotides. More than 30 such molecules were analyzed, and the structures found in all were consistent with one particular sequence organization. One end of the PSC34 fragment appears as a long (3000 nucleotides) "tail" consisting mainly of single copy DNA, while the other contains interspersed repetitive sequences. In some cases, a short repetitive sequence was observed in the "tail" region of the molecule. Since the driver sea urchin DNA fragments form branched single-stranded structures with the PSC34 fragment, the two families of repetitive sequence represented in clone PSC34 generally appear in the sea urchin genome in an environment of other sequences, almost certainly single copy sequences.

The PSC34 fragment hybridizes with sea urchin oocyte RNA, and thus may contain a transcribed sequence that may be a structural gene responsible for the synthesis of a maternal RNA (T. Thomas and R. Angerer, unpublished results). Although still in progress, our studies demonstrate short period repetitive sequence interspersion in cloned DNA fragments. This, to our knowledge, has not previously been observed since most studies so far published have been carried out on *Drosophila* DNA, which lacks short period interspersion (*13*). The sequence organization observed in these fragments is approximately in accord with average predictions made from physical measurements on the total DNA.

AMY S. LEE, ROY J. BRITTEN
ERIC H. DAVIDSON
Division of Biology, California Institute of Technology, Pasadena 91125

References and Notes

1. E. H. Davidson, B. R. Hough, C. S. Amenson, R. J. Britten, *J. Mol. Biol.* **77**, 1 (1973); E. H. Davidson, G. A. Galau, R. C. Angerer, R. J. Britten, *Chromosoma* **51**, 253 (1975); R. C. Angerer, E. H. Davidson, R. J. Britten, *Cell* **6**, 29 (1975).
2. D. E. Graham *et al.*, *Cell* **1**, 127 (1974).
3. M. E. Chamberlin, R. J. Britten, E. H. Davidson, *J. Mol. Biol.* **96**, 317 (1975).
4. P. L. Deininger and C. W. Schmid, *ibid.* **106**, 773 (1976).
5. J. E. Manning, C. W. Schmid, N. Davidson, *Cell* **4**, 141 (1975).
6. W. R. Crain, F. C. Eden, W. R. Pearson, E. H. Davidson, R. J. Britten, *Chromosoma* **56**, 309 (1976).
7. W. R. Crain, E. H. Davidson, R. J. Britten, *ibid.* **59**, 1 (1976).

8. R. Wells, H. D. Royer, C. P. Hollenberg, *Mol. Gen. Genet.* **147**, 45 (1976).
9. E. H. Davidson and R. J. Britten, *Q. Rev. Biol.* **48**, 565 (1973); E. H. Davidson, W. H. Klein, R. J. Britten, *Dev. Biol.* **55**, 69 (1977).
10. M. So, R. Gill, S. Falkow, *Mol. Gen. Genet.* **142**, 239 (1975). Our work was done in compliance with the NIH guidelines for recombinant DNA research.
11. R. Roychoudhurg, E. Jay, R. Wu, *Nucleic Acids Res.* **3**, 101 (1976).
12. R. W. Davis, M. Simon, N. Davidson, *Methods Enzymol.* **21D**, 413 (1971).
13. D. M. Glover, R. L. White, D. J. Finnegan, D. S. Hogness, *Cell* **5**, 149 (1975).
14. G. S. Hayward, *Virology* **49**, 342 (1972).
15. R. J. Britten, D. E. Graham, F. C. Eden, D. M. Painchaud, E. H. Davidson, *J. Mol. Evol.* **9**, 1 (1976).
16. A. S. Lee and R. L. Sinsheimer, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2882 (1974).
17. W. Barnes, unpublished observations.
18. Five of these clones were previously studied by W. R. Crain in this laboratory. The results of both studies are in close agreement.
19. Supported by NIH grants HD 05753 and GM 20927, and by NSF grant BMS 75-07359. A.S.L. has been supported by a fellowship from the American Cancer Society, California Division, J-289.

7 February 1977

Nucleotide Sequences from a Rabbit Alpha Globin Gene Inserted in a Chimeric Plasmid

Abstract. *Rabbit alpha globin gene copies have been made, using reverse transcriptase and DNA polymerase I, and cloned in bacterial plasmids. Plasmid pHb72 has been shown to contain the alpha gene sequence by restriction enzyme analysis and nucleotide sequencing studies, and therefore has been approved for propagation under P2 plus EK1 conditions by the National Recombinant DNA Committee.*

Reverse transcriptase has been used to synthesize complementary DNA (cDNA) copies of a variety of polyadenylated messenger RNA's (mRNA's). Several years ago, we discovered that reverse transcriptase leaves a short double-stranded hairpin structure when it terminates the synthesis of such cDNA's. This structure provides an efficient primer for second-strand synthesis by DNA polymerase I (*1*) so that it is possible to synthesize a double-stranded DNA copy of the mRNA sequence. This, and similar observations (*2*), prompted us as well as others to use the self-priming characteristics of cDNA's made with reverse transcriptase to develop general methods for cloning sequences copied from mammalian mRNA's in bacterial plasmids (*3–5*). Other methods have been developed for cloning mRNA sequences that did not utilize the self-priming nature of cDNA's synthesized by reverse transcriptase (*6*). In some of these cases, however, the use of complicated steps to force priming artificially at the desired location resulted in cloning efficiencies one to two orders of magnitude lower than that which we reported.

We have described the characterization of rabbit beta globin clones that we constructed using the cDNA technique (*3*, *7*). Our analysis included the determination of a 169-nucleotide sequence coding for the beta globin amino acid sequences from cysteine 93 through the carboxyl terminus of the structural gene (*7*, *8*). In the course of screening a series of 36 plasmids (pHb51 through pHb86) for the presence and the orientation of beta globin gene sequences, we identified nine plasmids that appeared to contain alpha globin gene sequences. Here we report the preliminary characterization of these plasmids.

Digestion with the two restriction endonucleases Eco RI and Bam HI provides a convenient method for screening beta globin gene sequences. A fragment 70 nucleotides in length is liberated from plasmids which carry that portion of the beta globin gene which codes for amino acid residues 98 through 122. In addition, it is possible to determine rapidly the orientation of beta globin gene sequences within these plasmids by comparing the fragments liberated by digestion with Bam HI with those liberated with both Bam HI and Eco RI. Since there is a Bam HI site within the beta globin gene and another within the plasmid gene for tetracycline resistance, Bam HI digestion liberates a fragment in the size range 600 to 1100 base pairs. If the orientation of the beta gene insertion is such that the Eco RI site is between the two Bam HI sites, then the corresponding fragment from a Bam HI plus Eco RI digestion will be 70 base pairs shorter. With the reverse orientation the Eco RI digestion does not affect this fragment.

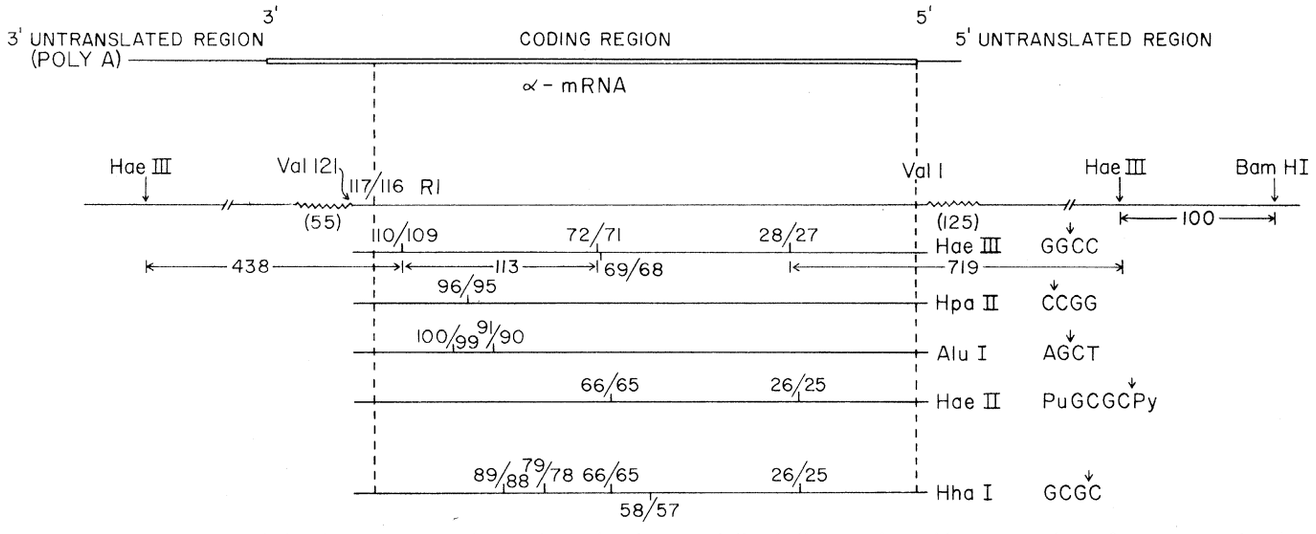


Fig. 1. Plasmid pHb72 restriction fragment map. The number pairs above each line indicate amino acid correspondence for sequences bearing confirmed sites for the restriction endonucleases cited. The number pairs under each line indicate amino acid correspondence for sequences bearing possible sites for the enzymes cited. Single numbers between sites indicate the number of nucleotides between sites. The inserted sequences in pHb72 extend from the codon for valine (Val) 121 to approximately 12 nucleotides beyond the initiation codon. A wavy line indicates poly(dA · dT) homopolymer. The lengths shown (numbers in parentheses) for these homopolymer regions are approximate and were estimated from ladder sequencing gels (data not shown). The arrangement of the fragments in the map was determined by a combination of conventional techniques and correlation of sequencing results with the known amino acid sequence (*10*).