I. METHODS FOR RESTRICTION ENDONUCLEASE STUDIES OF DNA STRUCTURE

II. RESTRICTION ENDONUCLEOLYTIC CHARACTERIZATION OF ANIMAL MITOCHONDRIAL DNAs AND HUMAN GLOBIN GENES

Thesis by

Richard Carl Parker

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1979

(Submitted November 20, 1978)

ACKNOWLEDGEMENTS

I think I'll begin at the end. The existence of this thesis and the reason that it looks so good are because of Connie Katz. Without her typing and advice on compiling this dissertation, I am sure it would not be finished. I'd also like to thank Jane Chacon for her help in seeing that my papers went to press, that my correspondence was sent to the right people, and that my thesis was finished.

I'd like to acknowledge the Jean Weigle Memorial Fund for its help in defraying some of the expenses of preparing this thesis; the Earle C. Anthony Fellowship for supporting me during my first year at Caltech; and the NIH Training Grants that have supported me ever since. None of these ever seemed to cover all of the expenses but they certainly helped.

While I was at Caltech, I shared labs with three people, John Wells, David Goldberg, and Vann Parker, with whom I never had collaborations. Each, however, played a major role in my graduate career providing great times, lots of insight, stimulating discussions, and an atmosphere that made work very pleasant.

My first collaborator was Bob Watson. Together with Jerry Vinograd he was my introduction into the beauty and pleasure of research science. Bob was the first person with whom I shared the joy of sitting around at all hours of the day talking about experiments, both ours, other peoples, and those not yet claimed. It was not necessary to always be correct in those discussions, just intellectually alive. Working with Bob was truly a pleasure and highly productive.

I also collaborated with Brian Seed for a long time. It, too, was a terrific experience marked by endless scientific discussion. The time with Brian produced only one paper for this thesis but, in reality, those experiences were a major

contributor to everything that has followed and everything that will. Brian was the first person I met who demanded that those around him be thinking at all times and also lived up to his own standards.

During the past year I have worked with Dick Lawn and Ed Fritsch. Together, we were able to obtain a great deal of data very rapidly. Without their help I do not know when I would have finished this thesis.

Part of the reason for coming to a place like Caltech is the opportunity to be surrounded by exceptional people. Two such people, Sue Conrad and Mike Klymkowsky, added tremendously to my graduate career. I am very lucky that I got the chance to know them.

In addition to the people already mentioned, I would like to thank a few faculty members. Giuseppe Attardi, James Bonner, Jim Strauss, and for a short time Robert Sinsheimer were members of my thesis committee. Each offered me ideas and support that made my time here easier.

In effect I had three advisors while I was at Caltech. Initially, I worked with Jerry Vinograd. Our relationship was very special and it is something that I will always treasure. He taught me about science and about life. He imparted the love of exploring the unknown that makes science so interesting. It is impossible for me to summarize all that I gained from knowing him for two, short years.

After Jerry's death I began working on straightforward projects. The greater plan was always to obtain a Ph.D. rather than to understand the functioning of a biological system. I regret that my career followed that course; although, it was of tremendous value in developing my ability to function independently. I feel that I succeeded pretty well, however, there definitely was a void which two people tried to fill. One was John Baldeschweiler who assured me that I would

have the chemistry department's support while I finished my thesis. This assurance was very important to me.

The other was Norman Davidson. Although I never was a member of his lab, he gave me more time and input than students often receive from their advisors. I learned a great deal from him and am very grateful that I got the chance to know him so well.

For the past year I have been in Tom Maniatis' laboratory. Our relationship has had its stormy moments, but out of them has come a great deal of respect. I am very glad that I was able to spend the last year in Tom's laboratory and to get to know him as an advisor, a scientist, and a friend. I have learned a lot from him and hope to be able to share a lot with him in the years to come.

Finally, there are four people from outside the world of science whose contributions to this thesis were invaluable. They were available when I needed support, they assured me of the value of what I was doing, and they gave the whole experience meaning. They are, of course, my immediate family, Paul, Lenore and Judith Parker and the woman who has shared my highs and lows most directly, Beverly Schlimme. It is to them that I dedicate this thesis.

ABSTRACT

An initial approach to the structural organization of DNA is restriction endonuclease site mapping and gel electrophoretic analysis. This is true for genomes of the simplest or greatest complexities.

Two techniques are presented in this thesis that facilitate this approach. The first uses ethidium bromide to limit the action of a restriction endonuclease on a closed circular DNA in order to derive a set of circularly permuted linear molecules. These molecules, after appropriate treatment, can be used to orient the restriction endonuclease sites and to calibrate the relationship between electrophoretic mobility and DNA fragment size without the introduction of external standards.

The second technique utilizes a low melting temperature agarose. It provides a simple system for two-dimensional electrophoretic analysis of DNA molecules with restriction endonuclease digestion of the DNA occurring after the first electrophoretic separation and before the second.

These techniques and others were used to study mitochondrial DNA from mice and rats. Some of these data explore the evolutionary divergence of the mtDNA in these animals. This information can be compared to evolutionary studies with nuclear DNA.

Additionally, chromosomal DNA from patients with normal hemoglobin and hemoglobin Lepore was studied. Using this type of analysis we were able to demonstrate that a change in the amino acid sequence of some of the β-related globin chains in hemoglobin Lepore is associated with a change in DNA structure.

TABLE OF CONTENTS

CHAPTER ONE

Mapping of Closed Circular DNAs by Cleavage with Restriction Endonucleases and Calibration by Agarose Gel Electrophoresis

INTRODUCTION

Near the end of my first year of graduate school, I decided to prove that the ribonucleotides present in mouse and human mitochondrial DNA (mtDNA) were located at specific sites in the genome. Naturally, I recognized that my experiments might prove the contrary to be true, or even a mixture of these two possibilities.

Lacking in gel expertise, I turned to Bob Watson and proposed a collaboration. He readily agreed and we set out to develop suitable assays and gel systems. Before much time had passed, we were using ethidium bromide (EtdBr) to limit the action of restriction endonucleases in digestions of covalently closed circular DNA.

Initially, we intended to make singly nicked molecules that could be used for developing gel systems capable of resolving Watson and Crick linear DNA molecules and Watson and Crick circular molecules. As the following paper reports, we were able to use EtdBr and restriction endonucleases to form singly nicked circles. However, in the process of our studies, we learned that the singly cleaved molecules were also very interesting.

Technically, it was very difficult in the Spring of 1976 to isolate large quantities of Form III DNA from an agarose gel in order to digest it with a restriction endonuclease before re-electrophoresing the DNA. That was one of the largest drawbacks to using this technique for mapping small circular DNAs. That is no longer a problem because of the development of the low melting temperature agarose two-dimensional analysis technique (see Chapter 3, this thesis).

When we wrote the paper we felt that we had developed an interesting method for restriction endonuclease site mapping that also allowed for highly

accurate calibration of gel electrophoretic molecular weight-mobility relationships. We did not realize that we had developed a powerful technique for in vitro mutagenesis; although we did begin experiments that used this approach to alter SV40 DNA in vitro. These experiments were discontinued when Jerry died.

Since that time, other investigators have applied this technique to studies requiring in vitro mutagenesis of pBG-1 DNA (1) and SV40 DNA (2). In the future, it is quite likely that EtdBr limitation of restriction endonucleases will prove to be more useful for these experiments than for site mapping.

The information included in this paper will be published once again in Methods in Enzymology. Since this manuscript did not have the page limitations imposed by PNAS, it was possible to present the data in greater detail. This copy contains two figures that were not published in PNAS in addition to the computer program that is used for analysis of the cleaved permeated linears. A preprint of the Methods in Enzymology paper is presented in this chapter.

References

1. Weissman, C., Benzon Symposium, in press (1978).

2. Shortle, D. and Nathans, D., Proc. Nat. Acad. Sci. USA **75**, 2170-2174 (1978).

Reprinted from
Proc. Natl. Acad. Sci. USA
Vol. 74, No. 3, pp. 851–855, March 1977
Biochemistry

Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis

(ethidium bromide/DNA of simian virus 40/DNA of bacteriophage PM2)

RICHARD C. PARKER*, ROBERT M. WATSON†, AND JEROME VINOGRAD*†‡

* Division of Biology and † Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, Calif. 91125

Communicated by Norman Davidson, November 22, 1976

ABSTRACT **The cleavage of DNA by restriction endonucleases can be limited by the addition of ethidium bromide. When closed circular DNA is used as a substrate, DNA with one-site cleavages of one or both strands can be made by adding appropriate amounts of dye.**

The singly cleaved DNA is a complete set of full-length permuted linear molecules. Fractionation of the products of a digestion of the permuted linears with a single-hitting restriction endonuclease by gel electrophoresis yields a series of bands that can be used to determine relative molecular weights of the DNA fragments in the gel without the introduction of standards.

It is possible to determine the relative molecular weight of a fragment to within ±2.5%. These molecular weights immediately allow the determination of the *Hin*dIII and *Hpa* I maps of simian virus 40. The *Hin*dIII map of bacteriophage PM2 was determined by this method with one ambiguity that was resolved by using traditional techniques.

Restriction endonucleases are useful for the analysis of the structure of DNA. For a review of the techniques developed for ordering the DNA fragments that result from a DNA/restriction endonuclease digest, see the paper by Nathans and Smith (1). Ideally, a restriction endonuclease map can be determined by knowing the molecular weights of the complete digest products and of many of the partial digest products. At present, it is impossible to determine molecular weights with enough accuracy to reduce mapping to a simple numerical problem.

Slab gel electrophoresis is a powerful tool for the separation of DNA molecules of different molecular weights. Much work has been done in an attempt to describe the electrophoretic mobility of a linear DNA as a function of its molecular weight (2, 3).

Molecular weights have been determined by electron microscopy (4) and radioactivity (4, 5), and by use of synthetic molecules (2). The relationship has been described as electrophoretic mobility being a linear function of the logarithm of the molecular weight. We believe that this analysis is useful in some cases but does not adequately describe the relationship, which is more accurately approximated by a third-order exponential function.

We have developed sets of DNA molecules to study this relationship. They have known molecular weight relationships that are not dependent upon measurement by other techniques. Their inherent relationships also allow for rapid restriction endonuclease mapping of closed circuit DNAs.

Abbreviations: EtdBr, ethidium bromide; Form I DNA, covalently closed circular DNA; Form II DNA, noncovalently closed circular DNA; Form III DNA, linear DNA; *Eco*RI, *Escherichia coli* restriction endonuclease RI; *Hin*dIII *Haemophilus influenzae*, strain d, restriction endonuclease III; *Hpa* I and *Hpa* II, *Haemophilus parainfluenzae* restriction endonucleases I and II; SV40, simian virus 40.

‡ *Deceased.*

MATERIALS AND METHODS

Enzymes and DNA. *Eco*RI endonuclease was a gift from H. Boyer. Other restriction endonucleases were purchased from New England Biolabs. Bacteriophage PM2 DNA was prepared according to Espejo *et al.* (6). Bacteriophage λ DNA was a gift from B. Seed. Simian virus 40 (SV40) DNA was a gift from J. Jordan and H. Kasamatsu and was later prepared as described (7). A unit of enzyme will degrade 1 μg of λ DNA to completion in 60 min at 37° in a volume of 50 μl.

Electrophoresis. A modification of the Aquebogue vertical slab gel electrophoresis apparatus was used, allowing long gels (30 cm) or short gels (15 cm) to be run. A fan was added to cool the gel during the run; tapered combs improved the sample wells. This equipment is available from EPT, Pasadena, Calif.

Agarose (Sea Kem) gels were prepared according to Helling *et al.* (3). They were run at a constant voltage of 3.3 V/cm. The DNA was recovered from preparative gels by the freeze–thaw technique described by Pulleyblank *et al.* (8), who also described conditions for photographing the gels. Electrophoretic mobilities were measured from the photographs.

Restriction Endonuclease Digests. Reactions were carried out as previously reported: *Eco*RI (9), *Hin*dIII (4), *Hpa* I and *Hpa* II (5). Reaction temperature, amounts of enzyme and DNA, and concentration of ethidium bromide (EtdBr) varied according to the experiment.

Dialysis. In sequential enzyme reactions the first enzyme was inactivated by the addition of 0.1 volume of 10% sodium dodecyl sulfate. EtdBr was extracted with 1% Sarkosyl in water-saturated butanol. The DNA was dialyzed by a modification of the centrifugation procedure described by Neal and Florini (10). Volumes of 100–500 μl were dialyzed through an SW 50.1 tube containing P-2 gel beads, 100–200 mesh (Bio-Rad) equilibrated in the desired buffer.

RESULTS

In the absence of ethidium bromide, incubation of six units of restriction enzyme with 1 μg of closed-circular DNA overnight at 37° gives a complete digest of the DNA. The presence of ethidium bromide inhibits the reaction. By varying the amount of EtdBr in the reaction it is possible to make many partial digest products, or a mixture that is predominantly open circles (Form II) and full length linear molecules (Form III), or mainly Form II alone.

A cursory examination of temperature effects for *Hin*dIII indicated that 55° gave maximal activity and highest conversion to Form III. At 4° about half of the DNA was converted to Form II; most of the rest remained as Form I during a 16-hr incubation. It was possible to convert greater than 90% of the DNA to Form II at 20°.

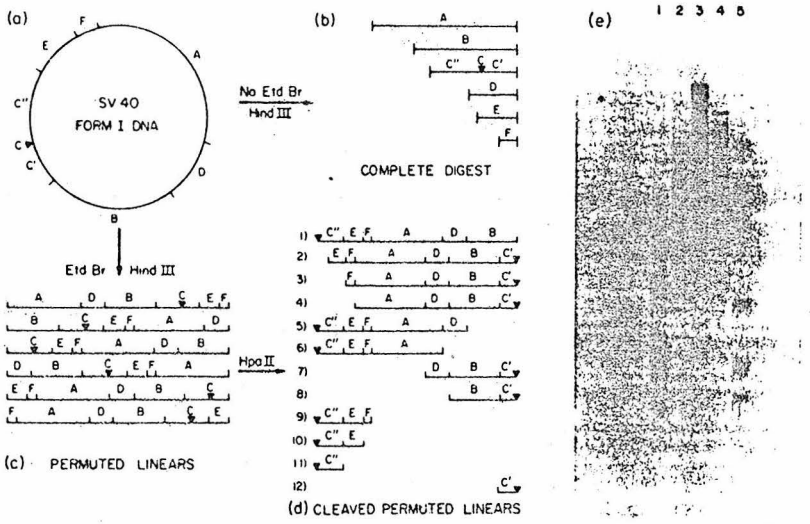


FIG. 1. (a) *Hin*dIII (lines) and *Hpa* II (triangle) restriction enzyme sites on SV40 DNA. (b) Complete digest products from a *Hin*dIII digestion of SV40 DNA. (c) Permuted linears from an EtdBr-limited *Hin*dIII digestion of SV40 DNA. (d) The permuted linears after digestion with *Hpa* II. The 12 fragments are ordered, as they would be resolved in a gel, from largest to smallest. Note that fragment 12 differs from fragment 8 by complete digest product B; similar reasoning yields the map that appears in (a). (e) Gel photograph; migration is from top to bottom. Slot 1: SV40/*Hin*dIII and SV40/*Hin*dIII/*Hpa* II complete digests. Slot 2: SV40/(EtdBr) *Hin*dIII after cleavage of permuted linears with *Hpa* II—the slowest migrating band is SV40 Form III. Slot 3: λ/*Eco*RI and λ/*Hin*dIII complete digests. Slot 4: PM2/(EtdBr) *Hin*dIII after cleavage of permuted linears with *Hpa* II—the slowest migrating band is PM2 Form III. Slot 5: PM2/*Hin*dIII and PM2/*Hin*dIII/*Hpa* II complete digests.

In studying other enzymes we found different temperature optimums. To maximize yields it will be necessary to study each enzyme's activity as a function of temperature. Maximum yields were not a necessity for our work and 37° incubations produced enough Form III DNA. It is possible in the systems we studied to find conditions that convert greater than 50% of the DNA to Form III; these systems utilized *Hin*dIII with PM2 or SV40 DNA and *Hpa* I with SV40 DNA.

Optimum ethidium bromide concentrations also varied in each case. At higher temperatures more EtdBr is needed to achieve the same effect. For PM2 and *Hin*dIII at 55°, 7.8 μg of EtdBr are needed per μg of DNA with 4.7 units of enzyme. For SV40 and *Hin*dIII the corresponding numbers for 1 μg of DNA are 55°, 18.4 μg of EtdBr, 5.5 units of *Hin*dIII. In the SV40–*Hpa* I system 1 μg of DNA was reacted with 20 μg of EtdBr and 4.55 units of *Hpa* I at 37°. All incubations were for at least 8 hr.

Increasing the level of EtdBr lowered the number of molecules receiving more than one cut, as did decreasing the amount of enzyme or DNA in the system. Pre-incubation of the enzyme with EtdBr had no effect on the results.

The Form II DNA was analyzed in an alkaline CsCl velocity experiment (11). It was found to be singly nicked (have one single-strand break). The Form III results from cleavage at any one of the possible restriction endonuclease sites. The sites are cleaved with differing frequencies. Each site is recognized often enough so that a complete set of full length permuted linear molecules is formed (Fig. 1).

The permuted linears upon cleavage with a single-hitting restriction endonuclease make it possible to determine the relationship between molecular weight and electrophoretic mobility, independent of other techniques, assuming that the migration of linear DNA in a constant concentration gel is a smooth function of the molecular weight of the DNA. They also simplify the problem of restriction enzyme mapping.

In an EtdBr-limited digest *Hin*dIII yields seven permuted linears with PM2 DNA and six permuted linears with SV40 DNA; *Hpa* I gives four permuted linears of SV40 DNA. Upon cleavage with a single-hitting enzyme in the absence of EtdBr these systems have 14, 12, and 8 bands. The sets of 14 and 12 bands can be resolved in a 1% agarose slab gel. Only six of the eight bands (12) formed in the SV40/*Hpa* I system can be resolved in 1% agarose because of the closeness of two of the *Hpa* I sites.

For any one digest, the resolved electrophoretic bands can be analyzed in pairs. The largest DNA fragment is the slowest migrating band; when paired with the smallest DNA fragment, which is the fastest migrating band in the slot, the result is a full-length molecule. The next to the fastest can be paired to make a full-length molecule with the next to the slowest, etc. We therefore have n equations,

$$MW_i + MW_{(2n+1-i)} = MW_{\text{Form III}} \quad (i = 1, 2, \ldots, n) \quad [1]$$

where $i = 1$ for the slowest band, etc., MW_i is the molecular weight of band i, and n is the number of restriction endonuclease sites.

We explored various functions in an attempt to relate the molecular weight to mobility. We finally adopted the general form:

$$MW_i = \exp(a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3) = f(x_i) \quad (i = 1, 2, \ldots, n) \quad [2]$$

in which x_i is the distance migrated by band i and the function

Table 1. Molecular weights of restriction endonuclease products from bacteriophage λ

	Thomas	Wellauer	Our values
*Hin*dIII B	—	5.84	5.97
*Eco*RI B	4.74	—	4.79
*Hin*dIII C	—	4.05	4.22
*Eco*RI C	3.73	—	3.73
*Eco*RI D	3.48	—	3.59
*Eco*RI E	3.02	—	3.07
*Hin*dIII D	—	2.67	2.73
*Eco*RI F	2.13	—	2.18
*Hin*dIII E	—	1.40	1.47

The measurements by Thomas and Davis (9) were done by electron microscopy using phage φX174 DNA as a standard. Wellauer *et al.* (13) also used electron microscopy to determine molecular weights with SV40 DNA as a standard. They used 3.28×10^6 as the molecular weight of SV40; we use 3.27×10^6 (3, 5, 13). Our values are averages from five PM2 calibration curves.

$f(x_i)$ is defined by the equation. In addition, there is the relation

$$MW_{\text{Form III}} = \exp(a_0 + a_1 x_{III} + a_2 x_{III}^2 + a_3 x_{III}^3) = f(x_{III}) \quad [3]$$

The above equations lead to the relations

$$f(x_i) + f(x_{(2n+1-i)}) = f(x_{III}) \; (i = 1, 2, \ldots, n) \quad [4]$$

This provides n equations for determining the four coefficients a_0, a_1, a_2, a_3. These coefficients will vary from experiment to experiment as agarose, buffer, voltage, run length, etc. change.

An additional equation is obtained by the following procedure. A complete digest by the enzyme for which there are multiple sites (sample A) is digested by the second single-hitting enzyme to give sample B. The two samples are run in adjacent slots of the gel used for resolving the cleaved permuted linears and measuring x_i above. One band (p) from sample A will not be present in sample B; instead, there will be two bands, p′ and p″, where p″ is the same as band 2n of Eqs. 1 and 2. Fragment p contains the site for the second enzyme. These observations lead to the equation,

$$f(x_p) = f(x_{p'}) + f(x_{p''}) \quad [5]$$

The $(n + 2)$ Eqs. 3, 4, and 5 (with the normalization that $MW_{III} = 1$) were used to determine a_0, a_1, a_2, and a_3. This over-determined nonlinear system of equations was solved with the aid of a computer program, by a least squares technique. The quality of a set of coefficients was determined by using the coefficients to solve the equation $MW_i = \exp(a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3)$ for each of the cleaved permuted linears. The appropriate pairs were then summed. The deviation of the sum of each pair from 1 was then squared and the squares of the deviations were summed. The program then changed the values for $a_0, \ldots, a_3$ until the sum of the deviations squared was minimized.

If it is assumed that the logarithm of the molecular weight is a linear function of the electrophoretic mobility, the best values for relative molecular weights of the recleaved permuted linears give pairs that sum to 1 ± 0.06. This is an error of 6%. By changing the equation, as described, it is possible to obtain molecular weights that upon summing are within ±1.5%. The gel data from the PM2/*Hin*dIII system were used to solve for the coefficients $a_0, \ldots, a_3$. From the general Eq. 2 it was then possible to determine the molecular weights of λ restriction endonuclease fragments run in a parallel slot in the gel (see Table 1). The calibration curve can also be used to map the cleaved permuted linears.

Table 2. Molecular weights of cleaved permuted linears and complete digest products

PM2/(EtdBr) *Hin*dIII/*Hpa* II	PM2/ *Hin*dIII	SV40/ (EtdBr) *Hpa* I/*Eco*RI	SV40/ *Hpa* I	SV40/ *Hpa* I*
III 6.27	A 3.53	III 3.22	A 1.33	1.353
1 4.95	A′ 1.98	1 2.77	B 1.24	1.261
2 4.65	A″ 1.48	2 2.53	A′ 0.75	0.785
3 4.35	B 1.42	3 2.08	C 0.63	0.656
4 3.99	C 0.61	4 1.20	A″ 0.53	0.572
5 3.80	D 0.265	5 0.75		
6 3.74	E 0.245	6 0.53		
7 3.45	F 0.18			
8 2.86	G --			
9 2.59				
10 2.52				
11 2.34				
12 1.98				
13 1.72				
14 1.48				

All data have been converted to molecular weights $\times 10^{-6}$. We determined from our curves that SV40 = 51.6% of PM2. The Form III values in the table are less than 100% because of the mathematical function used. All fragments smaller than 1.42×10^6 were measured from SV40 calibration curves. The SV40 curves slightly overestimated molecular weights of fragments greater than 50% of SV40 and slightly underestimated smaller fragments. The PM2 data are accurate to within ±2.5%. It is not possible to determine the molecular weights of the smallest *Hin*dIII complete digest products because they are smaller than any of the cleaved permuted linears. Columns 1 and 3 are cleaved permuted linears; 2, 4, and 5 are complete digests.
* These values for the complete digest products are from ref. 5.

Each of the 2n cleaved permuted linears has one end of the molecule in common—the site for the second, single-hitting restriction endonuclease (Fig. 1). The other end of these molecules is a restriction site for the first, EtdBr-limited enzyme. All possible permutations are represented. Each molecule differs in size from one of the other molecules by the size of a complete digest product of the DNA with the first enzyme.

By calculating the differences in the relative molecular weights of the cleaved permuted linears and by knowing approximate molecular weights of the complete digest products (these can often be determined from Eq. 2) it is possible to orient the complete digest fragments with respect to the single-hitting enzyme. There may be some ambiguity when some complete digest products are very similar in size. It may be necessary to orient those final few pieces by more traditional methods (14). It was necessary to resolve one ambiguity in the *Hin*dIII map of PM2 by cutting out a PM2/*Hin*dIII partial digest fragment and redigesting it with *Hin*dIII. The map that had already been developed from the cleaved permuted linears clearly indicated which partial digest fragment had to be isolated from a gel and cleaved.

The relative sizes of the cleaved permuted linears and the maps this information leads to are shown in Table 2 and Fig. 2. The data presented in Table 2 are averages from many gels. Each gel contained the cleaved permuted linears, the double enzyme digest, and Form III DNA. For brevity, the SV40/*Hin*dIII data are not presented in Table 2. The map and the fragment sizes that the data imply are shown in Fig. 2.

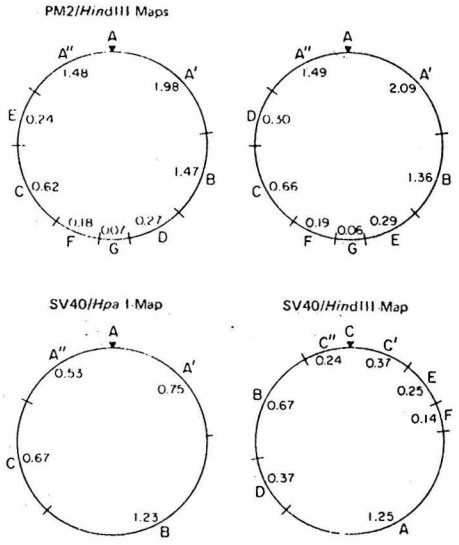


FIG. 2. *Upper left.* PM2/*Hin*dIII map from the information in the sizes of the smaller half of the cleaved permuted linears. *Upper right.* PM2/*Hin*dIII map from the information in the sizes of the larger half of the cleaved permuted linears—note that fragments D and E are reversed compared to the first map. *Lower right.* SV40/*Hin*dIII map from the information in the sizes of the smaller half of the cleaved permuted linears—the information in the larger half gives the same map. *Lower left.* SV40/*Hpa* I map from the information in the sizes of the smaller half of the cleaved permuted linears—the information in the larger half gives the same map.

In order to map the DNA, the molecular weight data describing the cleaved permuted linears can be analyzed by determining the differences in the molecular weights of either the smaller half of the molecules or the larger half. These differences should be sizes of complete digest products and, in principle, these approaches should give the same map.

Digestion of PM2 DNA with *Hin*dIII gives two pieces (fragments D and E) (Table 2) that are very similar in size. When the smaller half of the cleaved permuted linears is analyzed the resulting map places F, the smaller of D and E, beside the largest fragment in the complete digest, A, and fragment D beside fragment B. The larger half of the cleaved permuted linears indicates that fragment D is connected to fragment A and E is connected to B (Fig. 2).

To clarify the situation we partially digested PM2 with *Hin*dIII and isolated the fragment containing A and its neighbor and the fragment containing B and its neighbor. Upon digestion of the partials with *Hin*dIII the map derived from the larger half of the cleaved permuted linears, which placed D beside A and E beside B, was proven to be correct.

While this manuscript was in preparation Brack *et al.* published a map of the *Hin*dIII and *Hpa* II sites of PM2 DNA (15). Our map is in agreement with theirs. The SV40/*Hin*dIII and SV40/*Hpa* I maps confirm data published elsewhere (16, 17).

It is possible to map the sizes of DNA fragments from the calibration curve because the electrophoretic mobility of a linear DNA does not change from one slot of the gel to the next when certain precautions are taken. Mapping and gel calibration are dependent upon being able to compare samples in one gel slot with those in other slots. We minimized slot-to-slot variation in mobility as a function of molecular weight by modifying the equipment, which reduced the uneven heating in the gel that leads to DNA in the center slots running faster than in the side slots. Sample concentration is also important. If the mass of an individual sample of DNA is too great, the band that it forms in a gel will run faster than a smaller amount of the same material. For a sample run halfway into our 4 mm thick, 1% agarose short gels (running distance ≈7.5 cm) the band preferably should contain less than 50 ng of DNA. We found that band shape after the run was a function of the volume of the sample applied. Our sample wells had a minimal cross-sectional area of 10 mm^2, so we limited our sample volumes to 25 μl.

DISCUSSION

The function we have used to relate electrophoretic mobility and molecular weight was arbitrarily chosen. A linear logarithmic function was used with less satisfactory results. A cubic polynomial was almost as successful as the cubic exponential function.

Regardless of which function is chosen, the methodology will only yield reliable data if mobility is a smooth function of molecular weight. This precludes the possibility that different DNAs within the sample may have greatly varying G+C composition if, as has been reported (9, 18), G+C bias alters mobility in gel electrophoresis. Assuming that mobility is a smooth function of molecular weight, the experiments presented here offer a simple method for gel calibration without the introduction of standards and provide a rapid mapping technique for circular DNAs.

It is possible to derive molecular weights of fragments only within the range determined by the mobility of Form III and the mobility of the smallest cleaved permuted linear. An additional point, as noted in the legend to Table 2, is that our SV40 curves slightly underestimate the size of fragments smaller than 50% of the DNA and overestimate the larger fragments. The complementary nature of the problem is due to the constraint imposed by Eq. 4, demanding that pairs sum to 100%. In our experience so far, the error is never greater than 2% of the total size of the DNA.

A partial remedy for this problem is provided by using the added constraint $MW_{p'} + MW_{p''} = MW_p$ (Eq. 5). The usefulness of this constraint is dependent upon the size of the "p" fragment. In the SV40 *Hin*dIII system "p" equals the C fragment, which is approximately 20% of the genome. The calibration curve in this system deviates from expected values by as much as 2% of the full length of SV40 DNA. In the cases where the "p" fragment was either 40% or 55% of the genome, as in the SV40/*Hpa* I or PM2/*Hin*dIII systems, respectively, the resulting calibration curves were very good. They give SV40/*Hpa* I complete digest product sizes that are less than 1% smaller than accepted molecular weights. Additional constraints may further improve the mathematics. One is readily available: the sum of the molecular weights of complete digest products equals 1; another, the sum of two products equals a partial, requires knowing the map of the calibrating DNA. SV40 DNA is so well mapped this need not be a problem.

The calibration curve from the PM2/*Hin*dIII system is also very good. All of the λ restriction endonuclease fragment sizes presented in Table 1 were determined from PM2 calibration

curves. Our values, which span the range of the calibration curve, are greater than or equal to previously published values; but do not deviate greatly from those values. If the curves were failing in the overestimating/underestimating manner, some of our values would be higher than literature values while others would be lower.

There is further evidence for the accuracy of the PM2 calibration curve. The six largest cleaved permuted linears from the SV40/*Hin*dIII system and the three largest cleaved permuted linears from the SV40/*Hpa* I system fall within the range defined by the PM2/*Hin*dIII system. Therefore, it is possible to determine the molecular weights of those nine SV40 fragments from SV40 calibration curves and from PM2 calibration curves. This is done by using Eq. 2.

These nine fragments are all greater than 50% of SV40. Their sizes are slightly overestimated by the SV40 calibration curves. The PM2 calibration curves give lower values for all of the fragments. All of these lower values, when compared to expected values, are within the 1.5% error margin that is inherent in the best fit for Eq. 2. This error margin is much smaller than the normal standard deviation in sizes of large fragments measured by other methods. Therefore, we recommend the use of cleaved permuted linears for the mapping of closed circular DNAs and molecular weight determinations of linear DNA molecules. Our 1% agarose gel only resolved six of the 8 SV40/*Hpa* I-cleaved permuted linears. Those numbered 2 and 5 in Table 2 are doublets that did not resolve. We treated them accordingly and therefore our *Hpa* I map of SV40 contains only three fragments. The fourth fragment is located between the A and C fragments.

Even with all inathematical problems solved, DNA with many restriction enzyme sites will not be easy to analyze with this technique until gels with better resolution are developed. A gradient agarose gel system may expand the permissible size range for cleaved permuted linears.

While this manuscript was in preparation Nosikov *et al.* reported that distamycin A and actinomycin D can inhibit restriction endonuclease activity (19). We have also used propidium diiodide and actinomycin D. Site preference, but not exclusion, is observed in the EtdBr-limited systems at the dye levels we have explored. We have not determined whether this differs from the site specificity that often occurs in the absence of EtdBr.

This research was done under the leadership of Prof. Jerome Vinograd. Without his contributions, direction, and inspiration this paper would not exist. While this manuscript was in its early stages of preparation his unfortunate death occurred.

We thank Dr. J. Franklin for his advice about the mathematical aspects of this paper and M. Klymkowsky for his technical assistance. We thank Dr. N. Davidson for all of the help he gave us. This work was supported in part by National Institutes of Health Grants CA08014 and GM15327. This is Contribution no. 5417 from the Division of Chemistry and Chemical Engineering.

1. Nathans, D. & Smith, H. O. (1975) *Annu. Rev. Biochem.* **44**, 273–293.
2. Maniatis, T., Jeffrey, A. & van de Sande, H. (1975) *Biochemistry* **17**, 3787–3794.
3. Helling, R. B., Goodman, H. M. & Boyer, H. W. (1974) *J. Virol.* **14**, 1235–1244.
4. Danna, K. & Nathans, D. (1971) *Proc. Natl. Acad. Sci. USA* 68, 2913–2917.
5. Subramanian, K. N., Pan, J., Zain, S. & Weissman, S. M. (1974) *Nucleic Acids Res.* **1**, 727–752.
6. Espejo, R. T. & Canelo, E. S. (1968) *Virology* **34**, 738–747
7. Rush, M. G., Eason, R. & Vinograd, J. (1971) *Biochim. Biophys. Acta* **228**, 585–594.
8. Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. & Vosberg, H.-P. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4280–4284.
9. Thomas, M. & Davis, R. W. (1975) *J. Mol. Biol.* **91**, 315–328.
10. Neal, M. W. & Florini, J. R. (1973) *Anal. Biochem.* 55, 328–330.
11. Vinograd, J., Lebowitz, J., Radloff, R., Watson, R. & Lapis, P. (1965) *Proc. Natl. Acad. Sci. USA* 53, 1104–1111.
12. Botchan, M., Topp, W. & Sambrook, J. (1976) *Cell* 9, 269–287.
13. Wellauer, P. K., Reeder, R. H., Carroll, D., Brown, D. D., Deutch, A., Higashinakagawa, T. & Dawid, I. B. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2823–2827.
14. Lee, A. S. & Sinsheimer, R. L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2882–2886.
15. Brack, C., Eberle, H., Bickle, T. A. & Yuan, R. (1976) *J. Mol. Biol.* **104**, 305–309.
16. Lai, C.-J. & Nathans, D. (1974) *J. Mol. Biol.* 89, 179–193.
17. Sharp, P. A., Sugden, W. & Sambrook, J. (1973) *Biochemistry* 12, 3055–3063.
18. Zeiger, R. S., Salomon, R., Dingman, C. W. & Peacock, A. C. (1972) *Nature New Biol.* 238, 65–69.
19. Nosikov, V. V., Braga, E. A., Karlisher, A. V., Zhuze, A. L. & Polyanovsky, O. L. (1976) *Nucleic Acids Res.* 3, 2293–2301

VII. Determination of DNA Fragment Sizes

3. Conversion of Circular DNA to Linear Strands for Mapping

by Richard Parker

Division of Biology

California Institute of Technology

Pasadena, California 91125

Running head: DNA Restriction Fragment Sizing and Mapping

Circular DNA molecules are a suitable substrate for restriction endonucleases. When properly treated, covalently closed circular DNAs (Form I DNA) can be converted by a multi-hitting restriction endonuclease to full-length linear molecules (Form III DNA). After a subsequent reaction, the resulting molecules contain the information necessary to form a restriction endonuclease site map of the DNA and to ascertain the relative molecular weights of DNA fragments resolved by gel electrophoresis without the introduction of external standards.[1]

A restriction endonuclease digestion of DNA can be inhibited by Ethidium Bromide (EtdBr). Altering the concentration of EtdBr in a reaction permits covalently closed circular DNA to be singly nicked or singly cleaved by an enzyme that in the absence of EtdBr would cut the DNA many times.

The singly cleaved DNA molecules form a complete set of full-length permuted linear molecules. This set of molecules, each of which contains the entire genome, contains members with ends formed by cleavage at any site recognized by the restriction endonuclease. If there are six sites in the circular DNA, there will be six types of full-length linears, each with a different 5' end, formed by cleavage in the presence of EtdBr. This is the case with SV40 DNA and Hind III.

Digestion of the permuted linears with a restriction endonuclease that has only one recognition site in the DNA (a single-hitting enzyme) will yield twice as many fragments as there are types of linear molecules in the population. Therefore, digestion of a circular DNA in the presence of EtdBr with an enzyme that can cleave the DNA n times yields n types of permuted linears. Cleavage of the n types of permuted linears with a single-hitting enzyme will give 2n fragments.

After electrophoretic separation of the 2n fragments, the inherent size relationships of the fragments makes it possible to determine a restriction endonuclease map of the DNA. It is also possible to independently determine the relative size of the fragments to within ± 2.5%.

Conversion of Form I DNA to Form III DNA

The optimal concentration of ethidium bromide required to limit the cleavage of covalently closed circular DNA by a multi-hitting restriction endonuclease can only be determined by titration. In order to obtain optimal yields, it is necessary to determine the temperature at which the restriction endonuclease is most active. A simple examination showed that different enzymes have enhanced activities at widely varying temperatures: Pst I (23^{o}), Eco RI (37^{o}), Hind III (55^{o}).

After determining the optimal temperature, the yield of Form III can be maximized by titrating the EtdBr while maintaining all other variables (temperature, buffer, enzyme and DNA concentrations) constant. Increasing the level of EtdBr minimizes the number of molecules receiving more than one cut. The same effect can be achieved by decreasing the amount of enzyme or DNA in the reaction. If too much EtdBr is present, very little double-stranded cleavage will occur and most of the molecules formed in the reaction will be singly nicked non-covalently closed circles (Form II DNA). It is possible to obtain over 90% of the population as Form II DNA. These molecules can be shown to be singly nicked by the presence of approximately equal amounts of single-stranded linear molecules and single-stranded circles in alkaline CsCl velocity experiments.

If the temperature is changed to optimize enzymatic activity, the level of EtdBr will also have to be changed. At temperatures with greater enzymatic activity, more EtdBr is required to produce similar effects to those achieved at

other temperatures. In no case does preincubation of the enzyme with EtdBr affect the results.

Sizing and Mapping of Cleaved Permuted Linears

Cleavage of the permuted linears with a single-hitting enzyme, followed by electrophoretic separation of the resulting fragments, makes it possible to determine the relative molecular weights of the fragments in the gel as a function of electrophoretic mobility. These sizes are determined independent of other techniques, assuming that the mobility of a linear DNA in a constant concentration gel is a smooth function of its molecular weight.

The resolved bands can be analyzed in pairs. The smallest DNA fragment migrates faster than the others. Its complement, the largest DNA fragment, migrates slower than the others. When these two fragments are paired, the result is a full-length molecule. The next to the fastest can be paired with the next to the slowest to form a full-length molecule, etc. Therefore, the following n equations can be formed:

$$(1) \qquad MW_i + MW_{(2n+1-i)} = MW_{\text{Form III}} \qquad (i = 1,2,.....,n)$$

where i = 1 for the slowest band, MW_i = the molecular weight of band i, and n is the number of sites recognized by the restriction endonuclease in the EtdBr limited digest.

After exploring many functions that related molecular weight to mobility, we adopted the general form:

$$(2) \quad MW_i = \exp (a_o + a_1 x_i + a_2 x_i^2 + a_3 x_i^3) = f(x_i) \qquad (i = 1,2,.....,n)$$

where x_i is the distance migrated by band i. The mobility of full-length linear

molecules (Form III DNA) can be applied to Eq. (2) yielding:

$$(3) \qquad MW_{\text{Form III}} = \exp (a_o + a_1 x_{III} + a_2 x_{III}^2 + a_3 x_{III}^3) = f(x_{III})$$

The preceding equations and the information from the electrophoretic separation of the n cleaved permuted linears lead to the following n equations:

$$(4) \qquad f(x_i) + f(x_{(2n + 1 - i)}) = f(X_{III}) \qquad (i = 1,2,.....,n)$$

These n equations can be used to determine the four coefficients a_o, a_1, a_2 and a_3. In different experiments, these coefficients will change as agarose, buffer, run length, voltage, etc., vary.

To obtain the best curve relating molecular weight to mobility, it is necessary to introduce one additional equation. This equation is derived from fractionating the products of a complete digest of the DNA with the multi-hitting enzyme that is used in the EtdBr limited reaction (sample A) and the products of a complete digest from a reaction containing both the multi-hitting enzyme and the single-hitting enzyme that was used to cleave the permuted linears (sample B). The two samples are run in two slots of the gel that was used for fractionating and determining the mobilities (x_i above) of the cleaved permuted linears.

One of the bands (p) present in sample A will not be present in sample B. In sample B, band p will have been cleaved by the single-hitting enzyme giving rise to two bands not present in sample A, p' and p". The smaller of these bands, p', will be present in the cleaved permuted linears as band 2n. These relationships lead to the equation:

$$(5) \qquad f(x_p) = f(x_{p'}) + f(x_{p''})$$

Equations 3, 4 and 5 provide n + 2 relationships that were used (with the normalization that $MW_{III} = 1$) to determine the coefficients in Eq. (2), a_0, a_1, a_2, and a_3. That is an over-determined nonlinear set of equations. It was solved by a least squares technique with the aid of a computer program (see Appendix).

In order to solve for the coefficients a_0, a_1, a_2, and a_3, the program first solves the linear fit:

$$f(x) = \exp (a_0 + a_1 x_i) \tag{6}$$

After determining the best coefficients for this equation, the program adds another term and using the predetermined a_0 and a_1, solves the equation:

$$f(x) = \exp (a_0 + a_1 x_i + a_2 x_i^2) \tag{7}$$

In the process of solving this quadratic exponential, the values of a_0 and a_1 are not maintained as they were when solving Eq. (6). However, their values for Eq. (6) are the starting point, along with $a_2 = 0$, for solving Eq. (7).

Finally, the values of a_0, a_1, and a_2 that best fit Eq. (7) are used as the starting point, along with $a_3 = 0$, for solving the cubic exponential that is Eq. (2). The quality of the fit improves greatly between Eqs. (6) and (7) but changes very little from the quadratic Eq. (2) to the cubic Eq. (3).

The computer output consists of the coefficients that solve Eqs. (2), (6) and (7), the relative sizes of each of the cleaved permuted linears, Form III DNA and band p of Eq. (6), an error term which equals the difference between 1 and the sum of the relative sizes of the pair of bands that form a full-length molecule, and the sum of the error squared for the entire population. A sample of the output is presented with the program in the Appendix.

Mobilities of the relevant DNA fragments were determined from negative photographs of EtdBr stained gels (Fig. 1). The negatives were optically scanned and mobilities were measured as the distance from the top of the gel to the peak of the band on the trace (Fig. 2).

Each set of coefficients was tested by using them to solve the equation $MW_i = (a_0 + a_1x_i + a_2x_i^2 + a_3x_i^3)$ for each of the cleaved permuted linears. The molecular weights for the appropriate pairs were then summed. The deviation of this sum from 1 (the desired total) was then squared and the squares of the deviations were summed. The program then changed the values for $a_0,\ldots,a_3$ until the sum of the deviations squared was minimized.

Using this approach, it is possible to obtain molecular weights that sum to within $\pm$ 1.5% of the expected value. If it is assumed that the logarithm of the molecular weight is a linear function of the electrophoretic mobility (as opposed to the cubic function in Equation 2), it is only possible to sum within 6% in the systems tested.

Hind III cleaves SV40 DNA six times while Hpa II cleaves this DNA only once. Digestion of 1 µg of SV40 Form I DNA with 5.5 units of Hind III at 55°C in the presence of 18.4 µg of EtdBr in a 50 µl reaction maximizes the yield of permuted linears. These molecules when cut with Hpa II produce 12 cleaved permuted linears that can be electrophoretically resolved on a 1% agarose gel (Fig. 1). The sizes of the fragments formed, along with similar data from experiments with other systems, are presented in Table I.

The fragment sizes, determined by the third order exponential function, can be used to construct a restriction endonuclease map. Each of the 2n molecules formed in the second digest has one end in common, the site of the single-hitting

enzyme. The other end of the molecule is a restriction endonuclease site for the first, multi-hitting enzyme (Fig. 1). Each of the possible permutations is represented. Additionally, any given molecule differs in size from one of the other molecules by the size of a complete digest product of the DNA with the first enzyme. In Fig. 1d, cleaved permuted linear #12 differs in size from cleaved permuted linear #8 by the size of complete digest product "B." Similarly, fragments #10 and #11 differ only by complete digest product "E."

By calculating the differences in the relative molecular weights of the cleaved permuted linears and by knowing approximate molecular weights of the complete digest products (these can be determined from Eq. 2), it is possible to construct a restriction endonuclease map of the DNA. There may be some ambiguity in the final map if complete digest products are similar in size. If such ambiguities arise, they can be resolved by other methods, such as the isolation of a specific partial digest product.

Equation (2) can also be used to size additional DNA fragments run on the same gel as the cleaved permuted linears. The molecular weights of the Eco RI and the Hind III complete digest products of λ DNA were determined using the data from the analysis of cleaved permuted linears of PM2 DNA. These fragments were resolved on a 1% agarose gel (Fig. 1), and the fragment molecular weights were determined according to Eqs. (1)-(5) (Tables I and II). The form of Eq. (2) as determined by these data is shown in Fig. 3.

The data presented in Table II allow for the determination of three restriction endonuclease maps. The Hind III map of SV40 DNA presented in Fig. 1 can be ascertained from these data. The size of smallest fragment, 0.24×10^6 daltons (which equals .36 kilobases), implies that there is a Hind III site that

distance from the Hpa II site. This fragment is also found in the double enzyme complete digest and is labeled C'. Fragment #11 of the cleaved permuted linears is fragment C" of the double enzyme digest and implies that there is a Hind III site approximately .54 kb from the Hpa II site. These two Hind III sites define the "C" fragment.

Cleaved permuted linear #10 has a molecular weight of 0.62×10^6 daltons (.93 kb). That is the distance between a Hind III site and the sole Hpa II site in SV40 DNA. Between those two sites, however, there is another Hind III site which is either .54 kb away from the Hpa II site (fragment #11) or .36 kb away (fragment #12). Therefore, placement of a Hind III site .93 kb away from the Hpa II site will form a Hind III to Hind III distance of either .39 kb (.93 - .54) or .57 kb (.93 - .36). Table II indicates that there is a Hind III complete digest product of 0.25×10^6 daltons (approximately .39 kb) and there is not a Hind III complete digest product of 0.38×10^6 daltons. Therefore, the next Hind III site has been mapped; fragment E (0.25×10^6 daltons) is beside fragment C" (0.37×10^6 daltons) and their molecular weights sum to 0.62×10^6 daltons, the size of cleaved permuted linear #10. The next largest cleaved permuted linear has a molecular weight of 0.76 daltons. When complete digest fragment F is placed adjacent to cleaved permuted linear 10, the result is a fragment that is approximately 0.76×10^6 daltons. The size of fragment F is not shown in Table II. It is smaller than 0.24×10^6 daltons and is outside of the range of accurate fragment sizing. It is apparent, however, that it should be placed adjacent to cleaved permuted linear #10 in order to form cleaved permuted linear #9 because the alternative would require forming fragment #9 by placing a complete digest fragment beside fragment #12. This would be impossible for the complete digest fragment would have to be 0.52×10^6 daltons

($0.76 \times 10^6 - 0.24 \times 10^6$). This assignment can be confirmed by other approaches or by more accurate sizing of complete digest F. The next piece to be added must form a cleaved permuted linear of 0.91×10^6 daltons. This can be accomplished by placing fragment B (0.70×10^6 daltons) beside fragment C' (0.24×10^6 daltons) which is cleaved permuted linear #12. At this point, the following order has been established:

$$- F - E - C'' - C' - B -$$

where C' and C'' are formed by cleavage of Hind III fragment C with Hpa II.

By continuing this analysis of the data in Table II and extending it to the other information provided, it is possible to determine the restriction maps for SV40 DNA with Hind III, Hpa I, and Hpa II; additionally, the restriction map of PM2 DNA with Hind III can be determined (Fig. 4). The data for the latter map leave one assignment ambiguous, that is, the positioning of fragments D and E which are very similar in size. As published before, the final location of these fragments places D beside A and E beside B.

DISCUSSION

The third order exponential function that is used in this paper to study the relationship between electrophoretic mobility and molecular weight was arbitrarily chosen. Other, non-exponential functions could have been employed; a cubic polynomial met with only slightly less success.

Regardless of which mathematical relationship is used, the methodology contains the inherent assumption that mobility is a smooth function of molecular weight. If this assumption is valid (to date there is not sufficient evidence to assess the quality of the assumption), the method presented here offers a simple

method for gel calibration and restriction endonuclease site mapping of close circular DNAs.

A limitation of the technique is that it only allows for gel calibration over a defined range of molecular weight. Fragments that migrate faster than the smallest cleaved permuted linear, or slower than Form III DNA, cannot be accurately sized using Eq. (2). It is possible, however, to size DNAs of widely differing origins when they are electrophoresed simultaneously in different slots of a gel.

If DNAs are to be sized in this manner, it is necessary to minimize slot-to-slot variation in fragment mobility. Many factors involved in the mobility of a DNA fragment in a gel must be controlled before sample mobilities can be used to infer molecular weights.

Uneven heating in gels frequently leads to samples in center slots migrating faster than identical samples in side slots. Heating can be reduced by using thin glass plates, by having water or air circulation, and by running at low voltages. High sample concentrations can lead to overloaded bands which migrate anomalously rapidly. To prevent rapid migration due to overloading, we tried to load less than 50 nanograms of DNA with a sample run halfway (running distance $\sim$7.5 cm) into our 4 mm thick, 1% agarose gels.

It is also possible to alter DNA mobility by placing too great a sample volume on the gel. Band shape after the run is a function of the sample volume applied. With sample wells having a cross-sectional area of 10 mm^2, it is desirable to load no more than 25 μl. Finally, band shape is also altered by salt concentration. When sample mobility is going to be compared, the different DNAs should be layered on in approximately the same salt and it should be less than 100 mM.

The technique presented allows for gel calibration and restriction endonuclease site mapping of circular DNAs without the introduction of external standards. This can most easily be accomplished by:

1) An EtdBr titration designed to find a reasonable level of conversion to Form III DNA. "Reasonable" is partially determined by the availability of the DNA. Greater than 50% conversion to Form III was obtained in the systems described in this paper.

2) Fractionation of the permuted linears formed by EtdBr limitation of the restriction endonuclease digest on a low melting temperature agarose gel.

3) Cleavage of the permuted linears in the presence of the low melting temperature agarose (see Parker and Seed, this volume).

4) Fractionation of the cleaved permuted linears and complete digest products by gel electrophoresis.

5) Analysis of mobilities and determination of fragment sizes by a non-linear, least squares analysis with the aid of a computer.

6) Determination of restriction endonuclease sites from the relative fragment sizes.

ACKNOWLEDGEMENTS

I would like to thank Robert Watson and Jerome Vinograd with whom I originally collaborated in doing much of this work (see reference 1). I would also like to thank Brian Seed for his invaluable help and suggestions.

REFERENCE

1. R. C. Parker, R. M. Watson, and J. Vinograd, PNAS (USA) **74**, 851 (1977).

Table I

Molecular Weights of Bacteriophage Lambda Restriction Endonuclease Products

		Thomas[a]	Wellauer[b]	Our Values
Hind III	B	--	5.84	5.97
Eco RI	B	4.74	--	4.79
Hind III	C	--	4.05	4.22
Eco RI	C	3.73	--	3.73
Eco RI	D	3.48	--	3.59
Eco RI	E	3.02	--	3.07
Hind III	D	--	2.67	2.73
Eco RI	F	2.13	--	2.18
Hind III	E	--	1.40	1.47

The measurements by Thomas and Davis were done by electron microscopy using ØX174 DNA as a standard. Wellauer et al. also used electron microscopy to determine molecular weights with SV40 DNA as a standard.

They used 3.28×10^6 as the molecular weight of SV40; we use 3.27×10^6. Our values are averages from five PM2 calibration curves.

[a]M. Thomas and R. W. Davis, J. Mol. Biol. **91**, 315 (1975).

[b]P. K. Wellauer, R. H. Reeder, D. Carroll, D. D. Brown, A. Deutch, T. Higashinakagawa, and I. B. Dawid, PNAS (USA) **71**, 2823 (1974).

Table II

Molecular Weights of Cleaved Permuted Linears and Complete Digest Products

PM2/(EtdBr) Hind III/Hpa II		PM2/ Hind III		SV40/(EtdBr) Hind III/Hpa II		SV40/ Hind III		SV40/(EtdBr) Hpa I/Eco RI		SV40/ Hpa I		SV40/ Hpa I[a]
III	6.27	A	3.53	III	3.26	A	1.08	III	3.22	A	1.33	1.353
1	4.95	A"	1.98	1	3.07	B	0.70	1	2.77	B	1.24	1.261
2	4.65	A'	1.48	2	2.89	C	0.66	2	2.53	A"	0.75	0.785
3	4.35	B	1.42	3	2.63	C"	0.37	3	2.08	C	0.63	0.656
4	3.99	C	0.61	4	2.50	D	0.29	4	1.20	A'	0.53	0.572
5	3.80	D	0.265	5	2.35	E	0.25	5	0.75			
6	3.74	E	0.245	6	2.01	C'	0.24	6	0.53			
7	3.45	F	0.18	7	1.28	F	--					
8	2.86	G	--	8	0.91							
9	2.59			9	0.76							
10	2.52			10	0.62							
11	2.34			11	0.37							
12	1.98			12	0.24							
13	1.72											
14	1.48											

All data have been converted to daltons x 10^{-6}. We determined from our curves that SV40 = 51.6% of PM2. The Form III values in the table are less than 100% because of the mathematical function used. All fragments smaller than 1.42 x 10^{6}

Table II (continued)

were measured from SV40 calibration curves. The SV40 curves slightly overestimated molecular weights of fragments greater than 50% of SV40 and slightly underestimated smaller fragments. The PM2 data are accurate to within ± 2.5%. It is not possible to determine the molecular weights of the smallest Hind III complete digest products because they are smaller than any of the cleaved permuted linears. Columns 1, 3 and 5 are cleaved permuted linears; 2, 4, 6, and 7 are complete digests.

[a]K. N. Subramanian, J. Pan, S. Zain, and S. M. Weissman, Nucleic Acids Research **1**, 727 (1974).

Figure 1

a) Hind III (slashes) and Hpa II (triangle) restriction enzyme sites on SV40 DNA.

b) Complete digestion products from a Hind III digestion of SV40 DNA.

c) Permuted linears from an EtdBr limited Hind III digestion of SV40 DNA.

d) The permuted linears after digestion with Hpa II. The 12 fragments are ordered, as they would be resolved in a gel, from largest to smallest. Note that fragment 12 differs from fragment 8 by complete digest product B; similar reasoning yields the map that appears in (a).

e) Gel photograph

Slot 1: SV40/Hind III and SV40/Hind III/Hpa II complete digests.

Slot 2: SV40/(EtdBr) Hind III after cleavage of permuted linears with Hpa II - the slowest migrating band is SV40 Form III.

Slot 3: λ/Eco RI and λ/Hind III complete digests.

Slot 4: PM2/(EtdBr) Hind III after cleavage of permuted linears with Hpa II - the slowest migrating band is PM2 Form III.

Slot 5: PM2/Hind III and PM2/Hind III/Hpa II complete digests.

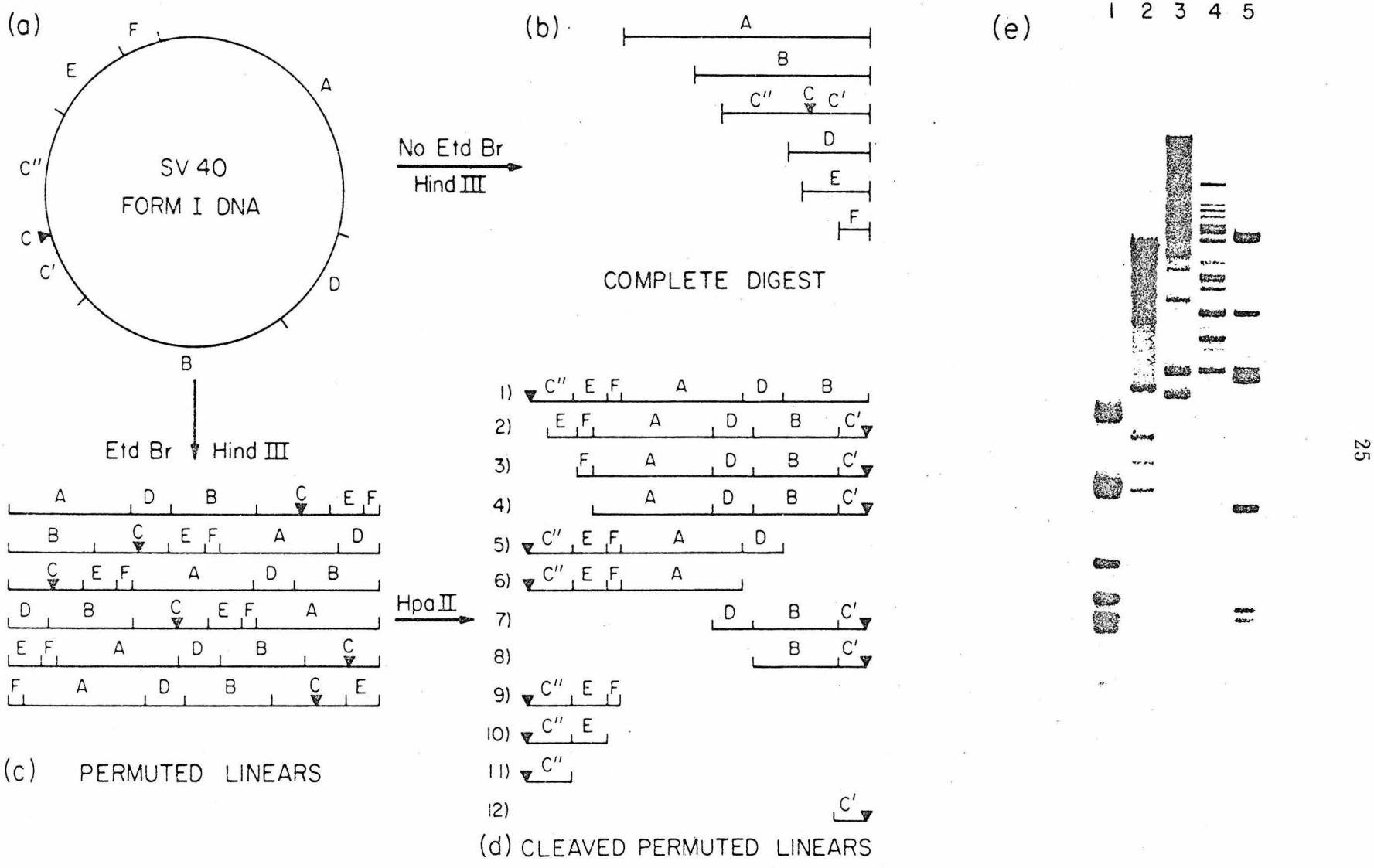

(a)
SV 40
FORM I DNA
No Etd Br
Hind III
(b)
COMPLETE DIGEST
Etd Br
Hind III
(c) PERMUTED LINEARS
Hpa II
(d) CLEAVED PERMUTED LINEARS
(e)
1 2 3 4 5

Figure 2. Slot 2 of the negative for the photograph in Figure 1e was optically scanned yielding this trace of the SV40/Hind III permuted linears after cleavage with Hpa II. Note that the peaks corresponding to fragments 4 and 9 are disproportionately small. This indicates that the rate of the EtdBr limited cleavage of SV40 DNA with Hind III is not the same at each of the enzyme's recognition sites.

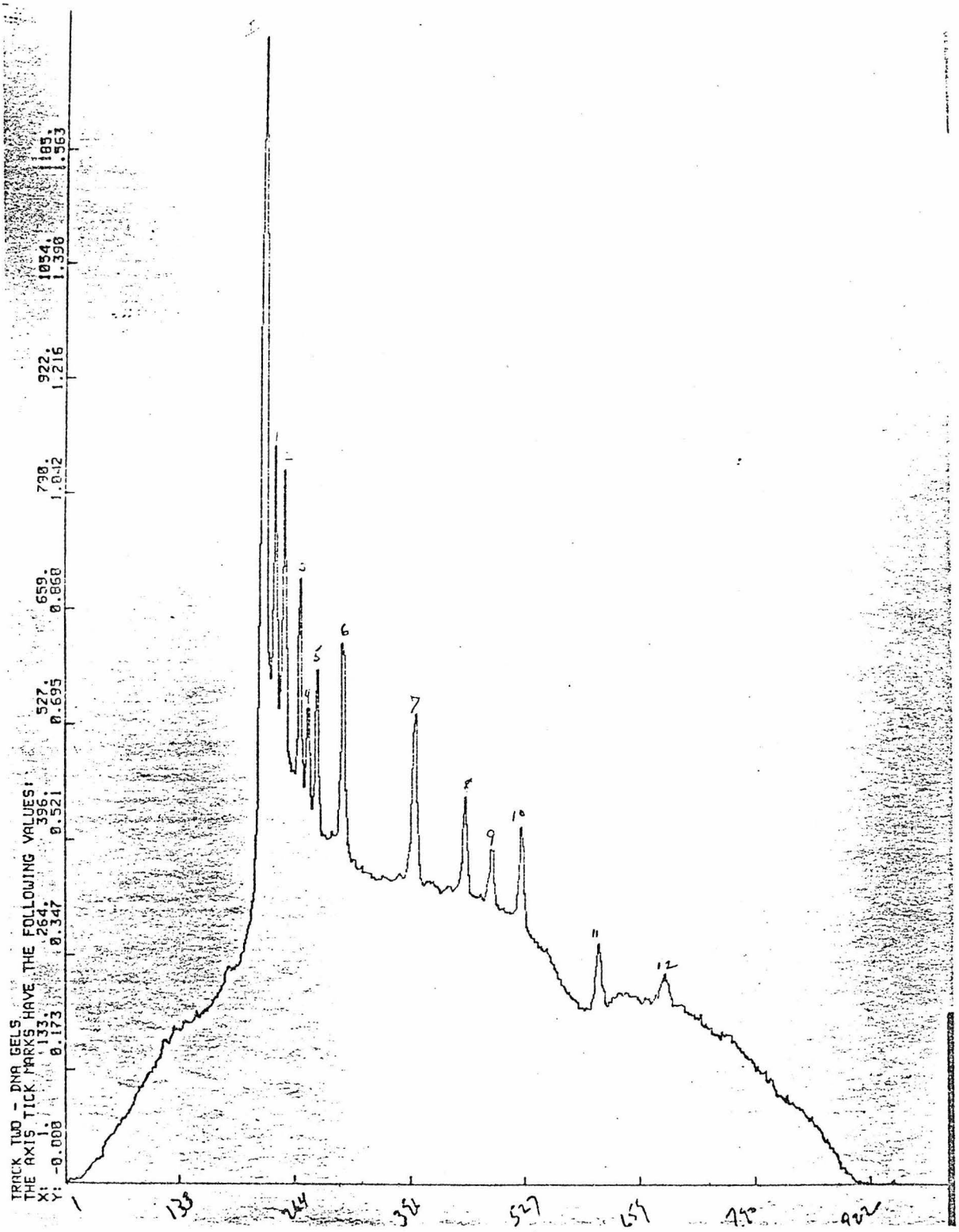

TRACK TWO - DNA GELS
THE AXIS TICK MARKS HAVE THE FOLLOWING VALUES:
X: 1. 133. 264. 396. 527. 659. 790. 922. 1054. 185.
Y: -0.000 0.173 0.347 0.521 0.695 0.868 1.042 1.216 1.390 1.563
1
133
264
396
527
659
790
922
1054
1185
1
2
3
4
5
6
7
8
9
10
11
12

Figure 3. This is a graph of the third order exponential function describing the electrophoretic mobility molecular weight relationship. The data used to derive this curve come from the PM2 fragments in slots 4 and 5 of Figure 1e.

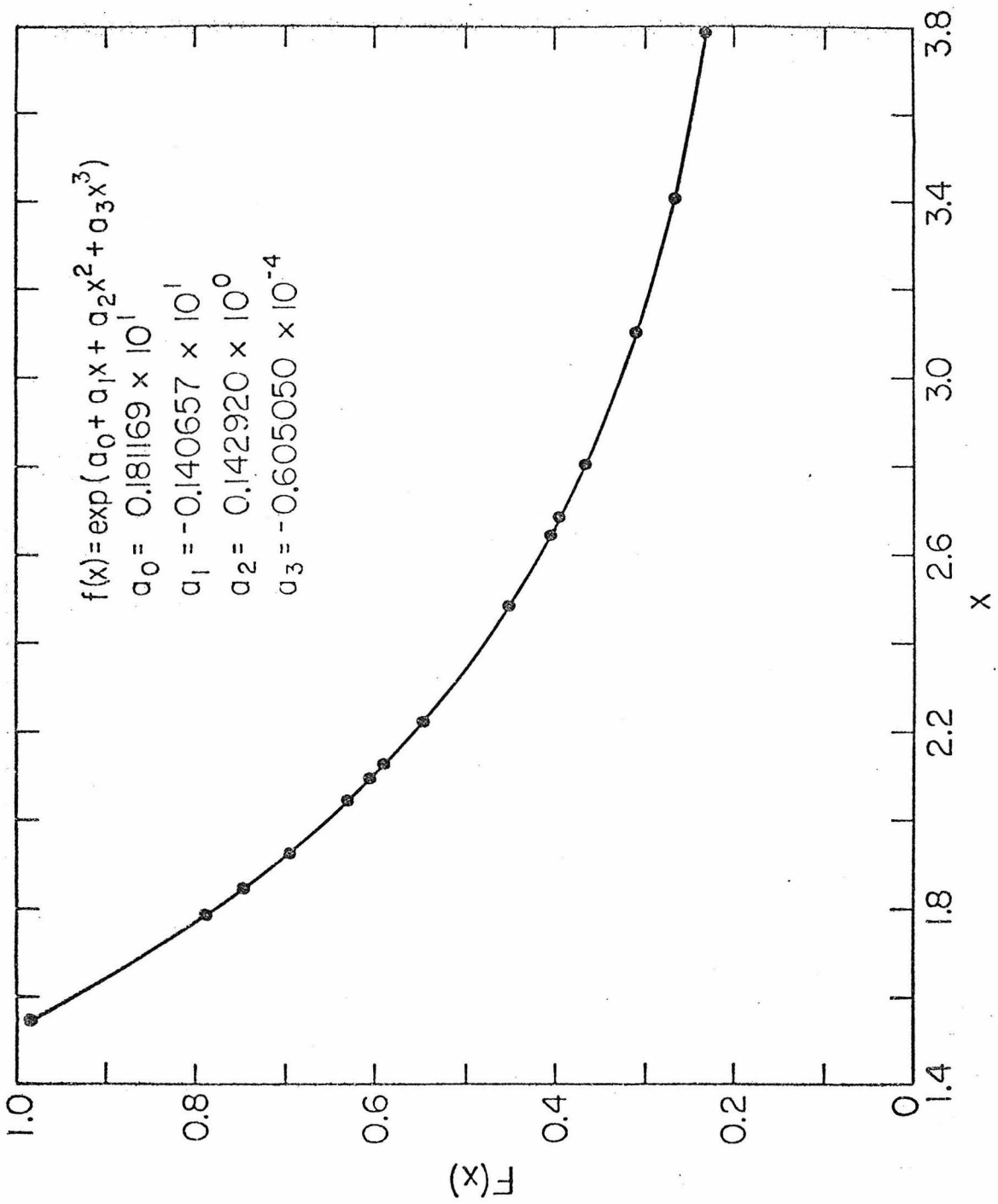

f(x)= exp (a0+ a1x + a2x2 + a3x3)
a0 = 0.181169 x 10^1
a1 = -0.140657 x 10^1
a2 = 0.142920 x 10^0
a3 = -0.605050 x 10^-4
F(x)
x
1.0
0.8
0.6
0.4
0.2
0
1.4
1.8
2.2
2.6
3.0
3.4
3.8

Figure 4

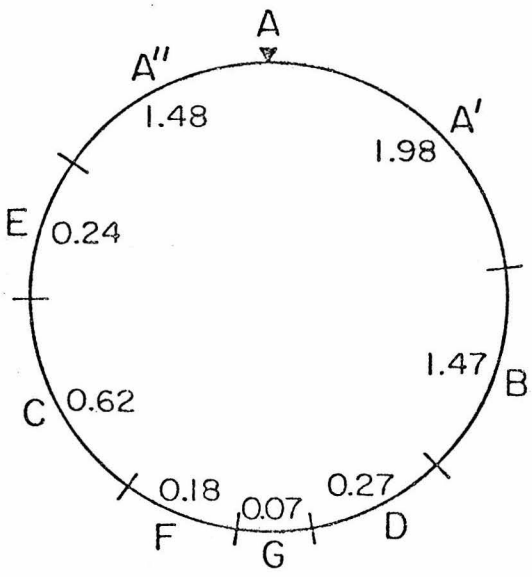
a) PM2/Hind III map from the information in the sizes of the smaller half of the cleaved permuted linears.

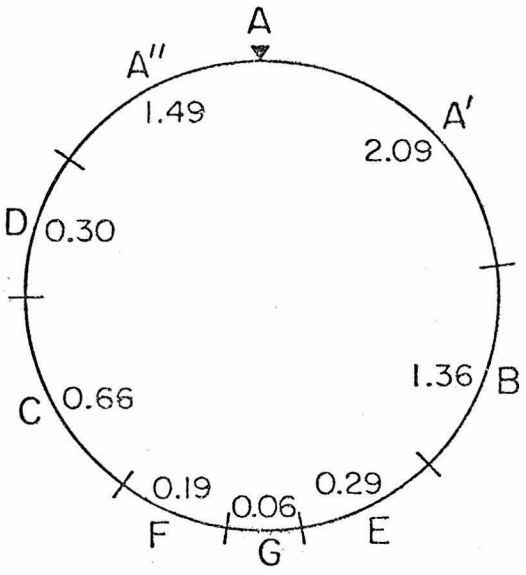
b) PM2/Hind III map from the information in the sizes of the larger half of the cleaved permuted linears - note that fragments D and E are reversed compared to Figure 4a.

c) SV40/Hpa I map from the information in the sizes of the smaller half of the cleaved permuted linears - the information in the larger half gives the same map.

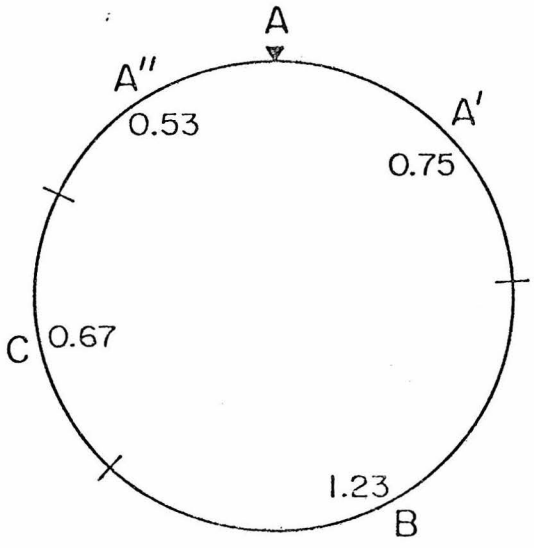
d) SV40/Hind III map from the information in the sizes of the smaller half of the cleaved permuted linears - the information in the larger half gives the same map.

PM2/Hind III Maps

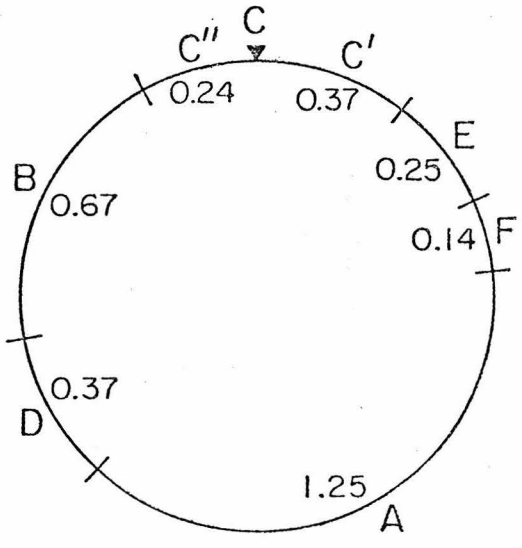




SV40/Hpa I Map



SV40/Hind III Map



APPENDIX

```
00010    C THIS PROGRAM FINDS THE BEST FIT TO AN EQUATION OF THE FORM
00020    C Y=EXP(A0+A1*X+A2*X*X+...) BETWEEN THE MOLECULAR WEIGHTS, Y, OF THE
00030    C PERMUTED CLEAVAGE PRODUCTS AND THE DISTANCES MIGRATED, X. TO USE THE
00040    C PROGRAM IT IS NECESSARY TO ENTER THE FOLLOWING VARIABLES:
00050    C       NX=THE NO. OF PERMUTED CLEAVAGE PRODUCTS + 1
00060    C       NN=THE HIGHEST POWER OF X DESIRED + 1
00070    C       NC1=THE SUBSCRIPT IN THE ARRAY X(I) OF ONE OF THE TWO PERMUTED
00080    C       CLEAVAGE PRODUCTS WHICH HAVE THE SAME LENGTH AS TWO OF THE
00090    C       PRODUCTS FOUND IN A LIMIT DOUBLE DIGEST.
00100    C       NC2=THE ANALOGOUS SUBSCRIPT OF THE OTHER DOUBLE DIGEST PRODUCT
00110    C       XC=THE MOBILITY OF THE SINGLE DIGEST PRODUCT CORRESPONDING TO THE
00120    C       SUM OF PIECES NC1+NC2
00130    C       X(I)=AN ARRAY OF THE MOBILITIES OF THE PERMUTED CLEAVAGE PRODUCTS
00140    C       IN ORDER OF DECREASING MOLECULAR WEIGHT. X(1) IS THE MOBILITY
00150    C       OF THE FULL LENGTH MOLECULE (FORM III). THERE MUST BE NX
00160    C       ENTRIES IN THIS ARRAY.
00170    C
00180          DIMENSION X(100),A(20),Y(100),XP(101),YP(101),P(20,24),DA(22),
00190         ,DOC(3),ERR(100)
00200          EXTERNAL F
00210          COMMON NA,NH,NX,N1,NC1,NC2,XC,YC,ERC,X,ERR
00220          DATA DOC,EPS/3*0.0,1.0E-6/
00230          CALL ERRSET(207,256,-1,1)
00240          CALL ERRSET(208,256,-1,1)
00250       10 CONTINUE
00260            WRITE(5,3)
00270        3 FORMAT(' ENTER NX,NN,NC1,NC2,XC, FORMAT(4I5,F10.0)')
00280          READ (5,1,END=100)NX,NN,NC1,NC2,XC
00290        1 FORMAT(4I5,F10.0)
00300          IF(NN .LT. 2 .OR. NN .GT. 20) GO TO 100
00310          IF(NX .LE. 0 .OR. NX .GT. 99) GO TO 100
00320          N1=NX+1
00330          NH=(NX-1)/2
00340            WRITE (5,4)
00350        4 FORMAT(' ENTER X(I),FORMAT(F10.0)')
00360          READ (5,2)(X(I),I=1,NX)
```

```
00370        2 FORMAT(F10.0)
00380          DO 95 I=1,NX
00390          IF(X(I) .LT. 10.0)GO TO 95
00400          WRITE(5,96)
00410       96 FORMAT('0  X IS LARGER THAN 10.0, POSSIBLE DATA ERROR')
00420          STOP
00430       95 CONTINUE
00440          DO 205 NA=2 ,NN
00450          NC=NA+2
00460          IF(NA .GT. 2) GO TO 210
00470          DO 25 I=1,NA
00480          P(I,1)=0.5**I
00490       25 CONTINUE
00500          GO TO 220
00510      210 DO 215 I=1,NA
00520      215 P(I,1)=A(I)
00530          P(NA,1)=0.0
00540      220 DO 225 I=1,NA
00550      225 DA(I)=0.2
00560          CALL AMOEBA(P,DA,NA,EPS,F)
00570          DO 45 I=1,NA
00580          A(I)=P(I,NC)
00590       45 CONTINUE
00600          CALL F(A,RES)
00610          DO 5 I=1,NX
00620          Y(I)=FEXP(X(I),NA,A)
00630        5 CONTINUE
00640          WRITE(5,6)(A(I),I=1,NA)
00650        6 FORMAT('1  THE COEFFICIENTS A0, A1, A2, ... ARE'/(5X,5E15.6))
00660          WRITE(5,8) RES
00670        8 FORMAT('0  SUM OF ERROR SQUARE = ',E14.6)
00680          WRITE (5,7)(X(I),Y(I),ERR(I),I=1,NX),XC,YC,ERC
00690        7 FORMAT(1H0,12X,'X',12X,'Y',12X,'ERROR'/(3X,3F15.6))
00700          DX=(X(NX)-X(1))/100
00710          DO 15 I=1,101
00720          XP(I)=X(1)+(I-1)*DX
```

```
00730          YP(I)=FEXP(XP(I),NA,A)
00740       15 CONTINUE
00750      205 CONTINUE
00760            GO TO 99
00770    C       SYSTEM PLOTTING SUBROUTINES
00780          CALL SCALE(X(NX),X(1),XMX,XMN,15,IE)
00790          CALL LABEL(0.,0.,XMN,XMX,15.,15,'X',1,0)
00800          CALL LABEL(0.,0.,0.,1.,10.,10,'F(X)',4,1)
00810          CALL XYPLT(NX,X,Y,XMN,XMX, 0.0,1.0,DOC,0,4)
00820          CALL XYPLOT(101,XP,YP,XMN,XMX, 0.0,1.0,DOC,1)
00830       99 GO TO 10
00840      100 STOP
00850          END
00860          FUNCTION FEXP(X,NA,A)
00870          DIMENSION A(1)
00880          ARG=A(1)
00890          DO 5 I=2,NA
00900          ARG=ARG+A(I)*X**(I-1)
00910        5 CONTINUE
00920          FEXP=EXP(ARG)
00930          RETURN
00940          END
00950          SUBROUTINE F(A,RES)
00960          DIMENSION A(1),Y(100),X(100),ERR(100)
00970          COMMON NA,NH,NX,N1,NC1,NC2,XC,YC,ERC,X,ERR
00980          DO 5 I=1,NX
00990          Y(I)=FEXP(X(I),NA,A)
01000        5 CONTINUE
01010          ERR(1)=Y(1)-1.0
01020          RES=ERR(1)**2
01030          DO 15 I=1,NH
01040          ERR(I+1)=Y(I+1)+Y(N1-I)-1.0
01050          ERR(N1-I)=ERR(I+1)
01060          RES=RES+ERR(I+1)**2
01070       15 CONTINUE
01080          IF(NC1*NC2 .LE. 0) RETURN
```

```
01090          YC=FEXP(XC,NA,A)
01100          ERC=YC-Y(NC1)-Y(NC2)
01110          RES=RES+ERC**2
01120          RETURN
01130          END
01140          SUBROUTINE AMOEBA(P,Y,N,E,F)
01150    C
01160    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
01170    C
01180    C  UPON ENTRY THE FOLLOWING PARAMETERS MUST BE PASSED:
01190    C
01200    C      N -- THE NUMBER OF VARIABLES FOR THE MINIMIZATION.
01210    C      E -- ABSOLUTE DELTA FUNCTION VALUE FOR DETERMINING CONVERGENCE.
01220    C      F -- THE NAME OF A SUBROUTINE WHICH WHEN CALLED BY CALL F(V,X)
01230    C           WHERE V IS AN ARRAY OF N VALUES WILL RETURN WITH THE COR-
01240    C           RESPONDING FUNCTION VALUE IN X.
01250    C      P -- AN ARRAY OF DIMENSION (N,N+4) WITH INITIAL VALUES FOR THE
01260    C           VARIABLES X(I) IN P(I,1) FOR I=1(1)N.
01270    C      Y -- A VECTOR OF DIMENSION (N+2) CONTAINING N DISPLACEMENTS
01280    C           DX(I) IN Y(I) FOR I=1(1)N.  THESE VALUES DX(I) WILL BE USED
01290    C           TO CONSTRUCT THE INITIAL SIMPLEX IN THE ARRAY P:
01300    C                P(I,J)=X(I)+DELTA(J-1,I)*DX(I) FOR I=1(1)N,J=1(1)N+1,
01310    C           WHERE DELTA(I,J) IS THE KRONECKER DELTA FUNCTION.
01320    C
01330    CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
01340    C
01350          INTEGER N,I,J,NS,NH,NL,NC,NR,NT,NEW
01360          REAL P,Y,E,F,YH,YL,YC,YR,YT,X,FLN,ERR
01370          REAL RFACT1,RFACT2,CFACT1,CFACT2,EFACT1,EFACT2
01380          DIMENSION P(20,1),Y(1)
01390          DATA RFACT1,RFACT2/-1.0,2.0/
01400          DATA CFACT1,CFACT2/0.5,0.5/
01410          DATA EFACT1,EFACT2/2.0,-1.0/
01420    C
01430    CCC  INITIALIZE SUBROUTINE PARAMETERS.
01440    C
```

```
01450          FLN=FLOAT(N)
01460          NS=N+1
01470          NC=N+2
01480          NR=N+3
01490          NT=N+4
01500    C
01510    CCC  CONSTRUCT INITIAL SIMPLEX.
01520    C
01530      100 DO 101 J=1,N
01540          DO 102 I=1,N
01550          P(I,J+1)=P(I,1)
01560          IF(I.EQ.J) P(I,J+1)=P(I,J+1)+Y(I)
01570      102 CONTINUE
01580      101 CALL F(P(1,J),Y(J))
01590          CALL F(P(1,NS),Y(NS))
01600    C
01610    CCC  FIND CURRENT MAX AND MIN.
01620    C
01630      200 NH=NS
01640          NL=NS
01650          YH=Y(NS)
01660          YL=Y(NS)
01670          DO 201 I=1,N
01680          IF(Y(I).LE.YH) GO TO 202
01690          YH=Y(I)
01700          NH=I
01710          GO TO 201
01720      202 IF(Y(I).GE.YL) GO TO 201
01730          YL=Y(I)
01740          NL=I
01750      201 CONTINUE
01760    C
01770    CCC  COMPUTE CENTROID.
01780    C
01790      300 DO 301 I=1,N
01800          X=0.0
```

```
01810          DO 302 J=1,NS
01820          IF(J.EQ.NH) GO TO 302
01830          X=X+P(I,J)
01840        302 CONTINUE
01850        301 P(I,NC)=X/FLN
01860          CALL F(P(1,NC),YC)
01870      C
01880      CCC  REFLECT CURRENT MAX THROUGH CENTROID.
01890      C
01900        400 DO 401 I=1,N
01910        401 P(I,NR)=RFACT1*P(I,NH)+RFACT2*P(I,NC)
01920          CALL F(P(1,NR),YR)
01930          IF(YR.LT.YL) GO TO 500
01940          X=YL
01950          DO 402 I=1,NS
01960          IF(I.EQ.NH) GO TO 402
01970          IF(Y(I).GT.X) X=Y(I)
01980        402 CONTINUE
01990          IF(YR.GT.X) GO TO 600
02000          NEW=NR
02010          GO TO 800
02020      C
02030      CCC  EXPAND.
02040      C
02050        500 DO 501 I=1,N
02060        501 P(I,NT)=EFACT1*P(I,NR)+EFACT2*P(I,NC)
02070          CALL F(P(1,NT),YT)
02080          NEW=NT
02090          IF(YT.GE.YL) NEW=NR
02100          GO TO 800
02110      C
02120      CCC  CONTRACT.
02130      C
02140        600 NEW=NH
02150          IF(YR.GT.YH) GO TO 601
02160          NEW=NR
```

```
02170          YH=YR
02180        601 DO 602 I=1,N
02190        602 P(I,NT)=CFACT1*P(I,NEW)+CFACT2*P(I,NC)
02200          CALL F(P(1,NT),YT)
02210          IF(YT.GT.YH) GO TO 700
02220          NEW=NT
02230          GO TO 800
02240      C
02250      CCC  SHRINK SIMPLEX TOWARD CURRENT MIN.
02260      C
02270        700 DO 701 J=1,NC
02280          IF(J.EQ.NL) GO TO 701
02290          DO 702 I=1,N
02300        702 P(I,J)=0.5*(P(I,J)+P(I,NL))
02310          CALL F(P(1,J),Y(J))
02320        701 CONTINUE
02330          YC=Y(NC)
02340          GO TO 802
02350      C
02360      CCC  TEST FOR CONVERGENCE.
02370      C
02380        800 DO 801 I=1,N
02390        801 P(I,NH)=P(I,NEW)
02400          YH=YR
02410          IF(NEW.EQ.NT) YH=YT
02420          Y(NH)=YH
02430        802 ERR=0.0
02440          DO 803 I=1,NS
02450          ERRI=(Y(I)-YC)**2
02460          IF(ERRI .GT. ERR)ERR=ERRI
02470        803 CONTINUE
02480          ERR=SQRT(ERR)
02490          IF(ERR.LT.E) GO TO 1000
02500      C
02510      CCC  UPDATE CYCLE DATA.
02520      C
```

```
02530      900 GO TO 200
02540    C
02550    CCC  END OF MINIMIZATION.
02560    C
02570     1000 RETURN
02580          END
02590
```

```
ENTER NX,NN,NC1,NC2,XC, FORMAT(4I5,F10.0)
   15    4   13   15  3.1

ENTER X(I),FORMAT(F10.0)
 2.44
 2.72
 2.79
 2.86
 2.96
 3.04
 3.06
 3.14
 3.39
 3.53
 3.56
 3.66
 3.92
 4.15
 4.46
```

THE COEFFICIENTS A0, A1, A2, ... ARE
0.183334E+01 -0.768442E+00

SUM OF ERROR SQUARE = 0.785824E-02

X	Y	ERROR
2.440000	0.959198	-0.040802
2.720000	0.773506	-0.023364
2.790000	0.732998	-0.009234
2.860000	0.694611	0.002212
2.960000	0.643233	0.018861
3.040000	0.604881	0.010511
3.060000	0.595656	0.010746
3.140000	0.560141	0.022377
3.390000	0.462267	0.022377
3.530000	0.415090	0.010746
3.560000	0.405650	0.010511
3.660000	0.375627	0.018861
3.920000	0.307601	0.002212
4.150000	0.257768	-0.009234
4.460000	0.203130	-0.023364
3.100000	0.577625	0.066895

THE COEFFICIENTS A0, A1, A2, ... ARE
0.292751E+01 -0.147393E+01 0.111100E+00

SUM OF ERROR SQUARE = 0.384683E-03

X	Y	ERROR
2.440000	0.992605	-0.007395
2.720000	0.771356	0.009176
2.790000	0.726201	0.005391
2.860000	0.684434	0.003272
2.960000	0.630088	0.005847
3.040000	0.590679	-0.007451
3.060000	0.581348	-0.008424
3.140000	0.545957	-0.001252
3.390000	0.452791	-0.001252
3.530000	0.410228	-0.008424
3.560000	0.401869	-0.007451
3.660000	0.375759	0.005847
3.920000	0.318838	0.003272
4.150000	0.279190	0.005391
4.460000	0.237820	0.009176
3.100000	0.563275	0.006616

THE COEFFICIENTS A0, A1, A2, ... ARE
0.292544E+01 -0.147298E+01 0.111210E+00 -0.585645E-04

SUM OF ERROR SQUARE = 0.384422E-03

X	Y	ERROR
2.440000	0.992637	-0.007363
2.720000	0.771454	0.009065
2.790000	0.726304	0.005368
2.860000	0.684539	0.003308
2.960000	0.630191	0.005942
3.040000	0.590776	-0.007341
3.060000	0.581443	-0.008309
3.140000	0.546044	-0.001118
3.390000	0.452838	-0.001118
3.530000	0.410248	-0.008309
3.560000	0.401883	-0.007341
3.660000	0.375751	0.005942
3.920000	0.318769	0.003308
4.150000	0.279063	0.005368
4.460000	0.237611	0.009065
3.100000	0.563366	0.006986

CHAPTER TWO

Restriction Endonucleolytic Studies with Animal (Human, Mouse and Rat) Mitochondrial DNA

INTRODUCTION

After Jerry died, Bob Watson and I shifted our attention to an evolutionary study of mtDNA. We decided to study the DNA from Old World and New World rodents by restriction endonuclease analysis.

We began by studying the standard lab mouse (Mus musculus), a mouse cell line (LA9), the lab rat (Rattus norvegicus), and two rat cell lines that were a kind gift from Dr. Murray Gardner of USC (Rattus rattus and Rattus norvegicus).

Early in the studies, we found that our cell line, LA9, was contaminated with HeLa cells. That led to a brief communication in Nucleic Acids Research about the sensitivity of mtDNA restriction endonuclease patterns as an assay for cell contamination.

After solving that problem and re-establishing pure lines of LA9 cells, we determined in a rapid screening with restriction endonucleases that mtDNA from LA9 cells produced the same patterns as mtDNA from the lab mouse. Although the cell line (LA9) was derived from the lab mouse (1), it seemed possible that mtDNA from these two sources might not be identical.

mtDNA was isolated from three sources of rat cells. One was the lab animal (R. norvegicus) and two were cell lines (R. norvegicus and R. rattus). The DNAs from these three sources produced three distinct cleavage patterns after digestion with Hae III. These results were reported in the following paper. They were put aside while the detailed restriction endonuclease maps presented in that paper were constructed. We never returned to the problem of evolutionary divergence, however, those maps do provide a foundation for studies investigating this problem.

While constructing the restriction endonuclease maps, we were able to learn more about mtDNA than simply where enzyme recognition sites exist. This paper demonstrated mtDNA heterogeneity within an inbred line of rats. The heterogeneity, later confirmed by Francisco and Simpson (2), could have occurred within one individual or simply within the population. Our experiments were done with DNA isolated from a pool of 20 rat livers; therefore, we could not determine which of these models was correct.

The second important point that should be stressed is that by restriction endonuclease analysis, approximately 24% base divergence occurs between these mtDNAs. This number is essentially the same as that reported by Kohne et al. (3) for nuclear, single-copy DNA. Therefore, mtDNA seems to be diverging at a rate similar to that of nuclear DNA although there are thousands of mtDNA molecules per cell (4).

References

1. Earle, W. R., Schilling, E. L., Stark, T. H., Straus, N. P., Brown, M. E. and Shelton, E., J. Nat. Cancer Inst. **4**, 165-212 (1943).
2. Francisco, J. F. and Simpson, M. V., FEBS Letters **79**, 291-294 (1977).
3. Kohne, D. E., Quart. Rev. Biophys. **3**, 327-375 (1970).
4. Bogenhagen, D. and Clayton, D. A., J. Biol. Chem. **249**, 7991-7995 (1974).

Volume 4 Number 5 1977 Nucleic Acids Research

Restriction endonuclease cleavage maps of rat and mouse mitochondrial DNAs

Richard C.Parker[1] and Robert M.Watson[2]

[1]Division of Biology and [2]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Received 28 February 1977

ABSTRACT

Mitochondrial DNA from an Old World mouse, Mus musculus, and from an Old World rat, Rattus norvegicus, contain 19 and 22 distinct sites, respectively, for the 8 restriction endonucleases, BamHI, EcoRI, HaeII, HhaI, HincII, HindIII, HpaI and PstI. The relative positions of the sites have been mapped by the study of partial and double enzyme digests. Some sites may have been conserved between the mouse and rat mitochondrial genomes.

INTRODUCTION

Animal mtDNA is a closed circular molecule of approximately 15000 base pairs. It contains genes that code for tRNAs (7), poly (A) containing RNA (8, 9), and two rRNAs. The development of restriction endonuclease maps should facilitate the process of locating these genes.

This technique is a valuable tool for studying, among other things, genome organization and expression (1, 2) and evolution (3). Various techniques have been employed for ordering the fragments produced in a restriction endonuclease digest (4, 5, 6). Among the simplest of these techniques is a gel electrophoretic analysis of pieces formed in partial enzyme and double enzyme digests. We have relied predominantly upon such analysis for the determination of 8 sets of restriction endonuclease sites in mtDNA isolated from mice and rats.

The genes coding for the two rRNAs have been shown to be adjacent to one another in mtDNA from X. laevis (10), Hela cells (11), and D. melanogaster (12). It has been reported that the 16S RNA and 12S RNA molecules are almost 180° apart on the rat mtDNA genome (13). The restriction endonucleases maps for Rattus norvegicus presented here differ from those in the previous study. Our restriction endonuclease maps, combined with the hybridization data presented by Saccone et al. (13), permit the

inference that the genes for 16S RNA and 12S RNA are also adjacent in rat mtDNA.

MATERIALS AND METHODS

Enzymes and DNA

EcoRI endonuclease was a gift from Dr. H. Boyer. All other restriction endonucleases were purchased from New England Biolabs. PM2 DNA was prepared according to Espejo et al. (14). λ DNA was a gift from B. Seed. SV40 DNA was prepared as described (15). Mitochondrial DNA was prepared from LA9 cells and livers from white mice (Mus musculus, Swiss Webster) and white rats (Rattus norvegicus, Sprague Dawley) by the procedures of Smith et al. (16); the sucrose gradient purification of mitochondria was eliminated.

Electrophoresis

A modification of the Aquebogue vertical slab gel electrophoresis apparatus was used. In the modified apparatus the upper reservoir is supported by two side panels. The panels are removable and can be interchanged with ones of different lengths. Changing the side panels alters the distance between the upper and lower reservoirs allowing long gels (30 cm) or short gels (15 cm) to be run. A fan was placed beneath the upper reservoir to cool the gel during the run. Tapered combs were made to improve the sample well. This equipment is now available from EPT, Pasadena, California.

Agarose (SeaKem, Marine Colloids) gels with concentrations from 0.4 to 2.5% were prepared in 40 mM Tris, 5 mM sodium acetate, 1 mM EDTA with the pH adjusted to 7.4 by addition of glacial acetic acid (E buffer).

Acrylamide gradient gels (all supplies from Biorad) were made by mixing equal volumes of 4% and 20% acrylamide solutions in a linear gradient maker; a 0-5% sucrose gradient was included to provide density stabilization. The 4% solution contained: 2.5 ml 10 x E buffer, 2.5 ml 40% acrylamide (acrylamide to bis-acrylamide ratio of 20:1), 25 μl 100% TEMED, 75 μl 10% ammonium persulfate (APS), and 20 ml of water. The 20% solution contained: 2.5 ml 10 x E buffer, 10 ml 50% acrylamide (acrylamide to bis-acrylamide ratio of 50:1), 7.5 μl 100% TEMED, 20 μl 10% APS, 1.25 gm sucrose, and 12.5 ml of water. Samples (10-50 μl for analytical gels and 500-1000 μl for preparative gels) were layered into the sample wells in a solution containing approximately 10% Ficoll 70 (Pharmacia) and 2 mM EDTA at pH 8. The gels were run at a constant voltage of 3.3 v/cm.

After electrophoresis the gels were stained for 10 min. in EtdBr

(2 μg/ml) and then de-stained for 10 min. in water. They were photographed with Kodak Plus-X film under short wave ultraviolet light excitation.

DNA was recovered from preparative gels after staining by cutting out the band of interest. The agarose was then minced and frozen at -20°C for at least 8 hours. After thawing, the expelled supernatant was collected and the DNA was ethanol precipitated by addition of two volumes of cold (-20°C) ethanol. NaCl was added to a final concentration of 0.15 M. After a 20 min. incubation at -20°C the DNA was pelleted in a Beckman SW 50.1 rotor at 40,000 RPM for 30 min. The pellet was then suspended in 10 mM Tris (pH 7.5), 1 mM EDTA.

Restriction Endonuclease Digests

Reactions were carried out in 0.1 M Tris (pH 7.4), 50 mM NaCl, 7 mM $MgCl_2$ (EcoRI); in 7 mM Tris (pH 7.5), 7 mM $MgCl_2$, 60 mM NaCl, 5.7 mM β-mercaptoethanol (BamHI, HincII, and HindIII); in 10 mM Tris (pH 7.5), 10 mM $MgCl_2$, 6 mM KCl, 1 mM DTT (HpaI); in 6 mM Tris (pH 7.5), 6 mM NaCl, 6 mM $MgCl_2$, 6 mM β-mercaptoethanol (HaeII, HaeIII and HhaI).

Electron Microscopy

The location of the D-loop was determined by measurement of micrographs of BamHI and HaeII restricted, glyoxal-fixed rat mtDNA as described in Brown and Vinograd (3).

RESULTS

Complete double enzyme digests and partial single enzyme digests were used to construct the restriction endonuclease maps presented in Figures 1 and 2. The site locations presented in Tables I and II were determined from log molecular weight versus electrophoretic mobility in either agarose or gradient acrylamide gels. The relationship in agarose gels is best approximated by a 3rd order exponential function as is described in detail (6). Standards of known molecular weights--PM2 digested with HindIII (6), λ digested with either HindIII or EcoRI (6), and SV40 digested with HaeIII (17)--were used for calibration.

When mtDNA replicates a D-loop is formed. It expands unidirectionally from a fixed point (18) which is defined as 0/100 map units. One map unit equals 1% of the full length DNA molecule. Map units increase from 0 to 100 in the direction of D-loop expansion.

Maps were compiled by analysis of the sizes of partial enzyme digest products or double enzyme digest products. All fragments produced in a

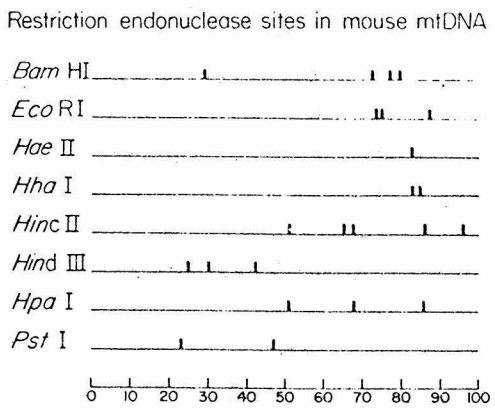


TABLE I. Restriction Endonuclease Sites in Mouse mtDNA

BamHI	29, 72, 77, 79
EcoRI	74, 75, 87
HaeII	82
HhaI	82, 83.5
HincII	51, 65, 67, 86.5, 95
HindIII	25, 30, 42
HpaI	51, 67, 86.5
PstI	23, 47

Figure 1 and Table I. Site locations are presented in map units and have an error of ± 1 map unit. Underlined sites are considered to be conserved through evolution.

partial enzyme digest must be combinations of complete digest products formed by the same enzyme.

Partial Enzyme Digest Analysis

One of the techniques used to determine restriction endonuclease maps was the analysis of the sizes of partial digest products. As an example of this technique the results of a HincII digest of mouse mtDNA will be described in detail.

HincII cleaves mouse mtDNA five times. The sizes of the complete digest products are: A = 55%, B = 20%, C = 14%, D = 9%, and E = 2%. A partial digest of HincII yields 6 partial products smaller than the 55% piece. Since only the 4 smallest fragments can be used to make these partial products (any partial product containing the 5th fragment would be larger than that fragment) all possible partial products are present. Therefore,

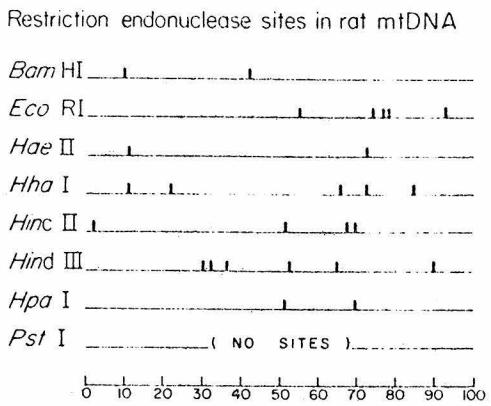


TABLE II. Restriction Endonuclease Sites in Rat mtDNA

BamHI	10, 42
EcoRI	55, 74, 76.5, 77.5, 93
HaeII	11, 72.5
HhaI	11, 22, 66.5, 72.5, 84
HincII	1, 52, 67.5, 69
HindIII	30, 32, 36.5, 52.5, 66, 91
HpaI	52, 69
PstI	No Sites

Figure 2 and Table II. Site locations are presented in map units and have an error of ± 1 map unit. Underlined sites are considered to be conserved through evolution.

the largest of the partials must contain the 4 smallest fragments. The approximate sizes of the HincII partial products of LA9 mtDNA are: 16, 22 29, 31, 36, and 45%. One can immediately deduce that the smallest partial contains the "C" and "E" complete digest fragments. Analysis of the possible permutations of the fragments leads to the conclusion that A joins D which abuts B which is adjacent to E which is connected to C which is joined to A (-A-D-B-E-C-) because mitochondrial DNA is circular.

Double Enzyme Digest Analysis

Analysis of the sizes of double enzyme digest products was extensively used to determine restriction endonuclease sites. An illustration of this technique is the development of the HpaI map of rat mtDNA. HpaI cleaves rat mtDNA twice, forming pieces approximately 83% and 17% in length. In order to determine the HpaI map of rat mtDNA, by use of double enzyme

digests, it is necessary to know that BamHI cleaves rat mtDNA at 10 and 42 map units and that HaeII cleaves rat mtDNA at 11 and 73 map units.

A HpaI/HaeII double digest produces 4 fragments, 41%, 38%, 17%, and 4% in length. This implies that there are no HpaI sites in the 38% fragment produced by a HaeII digest of rat mtDNA which begins at 73 map units and ends at 11 map units.

A double digest of rat mtDNA with BamHI and HpaI also produces 4 fragments. They are approximately 41%, 32%, 17%, and 10% in length. This implies that there are no HpaI sites in the 32% fragment produced by a BamHI digest of rat mtDNA which begins at 10 map units and ends at 42 map units. Therefore, both of the HpaI sites must be between 42 and 73 map units. The presence of two sites within this region implies that a double enzyme digest of HpaI with either HaeII or BamHI must form a complete digest product from this region and a double digest product having one HpaI site and either the HaeII site at 73 map units or the BamHI site at 42 map units as its ends. There is only one possible placement of HpaI sites within this region that can produce a BamHI/HpaI double digest product of 10%, a HaeII/HpaI double digest product of 4%, and a HpaI complete digest product of 17%. The HpaI sites must be at 69 map units and 52 map units.

Arguments similar to those presented in the sections on partial enzyme digests and double enzyme digests were used to deduce all of the maps presented in Figures 1 and 2. The information that was used is presented in the appendices. They contain the sizes of the fragments produced by partial enzyme digests and double enzyme digests.

DISCUSSION

Tissue culture cells from Mus musculus (LA9) and Rattus norvegicus (Amsterdam rat, Schmidt-Ruppin sarcoma) were also used as sources of mtDNA. The EcoRI, HinII and III, and HaeIII restriction patterns of mtDNA from these cells were compared with those patterns obtained from live animal mtDNA. No differences were observed in any of the mouse digests nor in the HinII and III digests of the rat DNA. Slight changes were noted in the rat/HaeIII system.

The EcoRI digest of the rat tissue culture line produced 6 bands (any small bands would not have been detected in the system employed). These are the same 6 fragments whose locations were mapped by Kroon et al. (19). We disagree with their map.

The animal DNA contains a fragment of approximately 150 base pairs

that was not cited by Kroon et al. Their maps were determined by size analysis of partial digest products. The presence of this small fragment might resolve the differences between our map and theirs. While the EcoRI digest of the animal DNA produced all of the bands that the tissue culture DNA produced, 4 of these bands appeared as minor species. Two new bands appeared: one seems to be the sum of the two largest minor species and the other seems to be the sum of the two smaller species.

The minor species that appear in the animal DNA may be due to heterogeneous DNA; we pooled DNA isolated from 20 rats. If a small percentage of the DNA contained one or two additional EcoRI sites, sites that are contained in all of the tissue culture DNA, the four extra bands would appear. The content of the minor species is so low (<5%) that if their presence is due to heterogeneous DNA there must either be heterogeneity within an individual or the unusual individual(s) must have yielded greatly lowered amounts of mtDNA.

The alternatives, that these minor species were caused by a contaminating enzymatic activity or a difficult to hit EcoRI site, were examined by an incubation with a 20 fold excess of enzyme for a 70 fold increase in time. It did not change the results. Other DNAs (SV40, ϕX174-RF) were digested with the rat mtDNA and gave traditional EcoRI patterns.

The HindIII map of rat mtDNA presented in this paper also differs from that in Kroon et al. Our map was determined by analysis of the data presented in the appendices and confirmed by the isolation of the fragments produced by a HaeII digest of rat mtDNA followed by digestion of those fragments with HindIII.

The two maps presented here allow for an estimation of sequence divergence between rat and mouse mtDNA. Previous work (21) has indicated between 21% and 37% mismatch by analysis of thermal denaturation of hybrids between the heavy and light strands of rat and mouse mtDNA.

By assuming restriction endonucleases recognize sites that have neither an unusual selective advantage nor an unusual selective disadvantage the data presented here can be used with a binomial analysis to determine the percentage of base divergence between these two animals.

Of the 8 restriction endonucleases studied one, HhaI, recognizes a tetranucleotide sequence; two, HincII and HaeII, have relaxed specificities within their recognition sequences and at some sites they do not differentiate between the two purines; the other five enzymes, BamHI, EcoRI, HindIII, HpaI, and PstI, recognized unique hexanucleotides.

The set of five enzymes cleave both mouse and rat mtDNA 15 times. Three of the 15 sites are approximately the same distance from the origin of replication in both systems.

If the probability of a base change equals "p" the probability that a base will not change is (1 - p). Of the 15 unique hexanucleotide sites studied 3 are not changed, therefore, "p" which is an estimate of sequence divergence is approximately 24%. Further work can now be done to determine whether or not limited regions of the genome are highly conserved.

ACKNOWLEDGEMENTS

We wish to express our appreciation to K. Moore, P. Johnson, S. Chandler, and L. Grossman for informing us of their preliminary mouse mtDNA mapping results. We thank S. Conrad, M. Klymkowsky, and A. Sodja for their invaluable assistance, L. Matsumoto for his electron microscopy results that confirmed the HaeII site on mouse mtDNA, and Professors J. Strauss and N. Davidson for helpful discussions. This work was supported in part by USPHS NIH Grants GM 15327 and CA 08014. This is contribution no. 5525 from the Division of Chemistry and Chemical Engineering.

This work was done in Jerry Vinograd's laboratory after his unfortunate death. His leadership and the standards he set made it possible. We dedicate this paper to him.

APPENDIX 1: Approximate fragment sizes (in % of genome) of restriction endonuclease products of mouse mtDNA.

EcoRI 87, 12, 1; EcoRI/HaeII 87, 8, 4, 1; HindIII 83, 12, 5; HindIII/HaeII 42, 41, 12, 5; HhaI (a subset of HaeII) 98.5, 1.5; HhaI/EcoRI 87, 8, 2.5, 1.5, 1; HpaI 64, 20, 16; HpaI/EcoRI 64, 16, 11.5, 7, 1, 0.5; HpaI/HaeII 64, 16, 15, 5; HindII 55, 20, 14, 9, 2; HindII (partial) ..., 45, 36, 29, 22, 16; HinII and III 30, 20, 14, 12, 9, 8, 5, 2; BamI 49, 45, 4, 2; BamI (partial) ..., 6; BamI/HaeII 46, 45, 4, 3, 2; BamI/EcoRI 44, 42, 8, 2, 2, 1, 1; BamI/HindIII cleaves 5% HindIII piece into 4% and 1%; PstI 76, 24; PstI/HaeII 40, 36, 24; PstI/HhaI 40, 34, 24, 1.5.

APPENDIX 2: Approximate fragment sizes (in % of genome) of restriction endonuclease products of rat mtDNA.

HaeII 62, 38; BamHI 68, 32; BamHI/HaeII 37, 31, 31, 1; HpaI 83, 17; HpaI/HaeII 41, 38, 17, 4; HpaI/BamHI 41, 32, 17, 10; HindIII 39, 25, 16, 13, 5, 2; HindIII (partial) ..., 6; HindIII/BamHI 24, 20.5, 19.5, 13, 10.5, 5.5, 5.1; HindIII/HaeII 21.5, 18.5, 18, 16, 13, 7.5, 5, 1; HincII 51, 32,

15.5, 1.5; HincII/HhaI 30, 17, 14.5, 12, 11, 9, 4, 1.5, 1; HincII/BamHI 32, 32, 15.5, 10, 8, 1.5; EcoRI 62.5, 19.5, 15.5, 2.5, 1.0; EcoRI (partial) ..., 22, ..., 4; EcoRI/HaeII 43.5, 19, 18.5, 15.5, 2.5, 1.0, 1.0; EcoRI/BamI 32, 19.5, 17, 15.5, 12.5, 2.5, 1.0; PstI does not cleave.

REFERENCES

1. Carroll, D. and Brown, D. D. (1976) Cell 7, 467-476.
2. Acheson, N. H. (1976) Cell 8, 1-12.
3. Brown, W. M. and Vinograd, J. (1974) Proc. Nat. Acad. Sci. USA 71, 4617-4621.
4. Nathans, D. and Smith, H.O. (1975) Ann. Rev. Biochem. 44, 273-293.
5. Smith, H. O. and Birnstiel, M. L. (1976) Nucleic Acids Res. 3, 2387-2398.
6. Parker, R. C., Watson, R. M. and Vinograd, J. (1977) Proc. Nat. Acad. Sci. USA, in press.
7. Angerer, L., Davidson, N., Murphy, W., Lynch, D. and Attardi, G. (1976) Cell 9, 81-90.
8. Hirsch, M. and Penman, S. (1974) Cell 3, 335-339.
9. Ojala, D. and Attardi, G. (1974) J. Mol. Biol. 88, 205-219.
10. Wellauer, P. K. and Dawid, I. B. (1973) Carnegie Institute Yearbook 72, 45-46.
11. Wu, M., Davidson, N., Attardi, G. and Aloni, Y. (1972) J. Mol. Biol. 71, 81-93.
12. Klukas, C. K. and Dawid, I. B. (1976) Cell 9, 615-625.
13. Saccone, C., Pepe, G., Cantatore, P., Terstra, P. and Kroon, A. M. (1976) in The Genetic Function of Mitochondrial DNA, pp. 27-36.
14. Espejo, R. T. and Canelo, E. S. (1968) Virology 34, 738-747.
15. Rush, M. G., Eason, R. and Vinograd, J. (1971) Biochim. Biophys. Acta 228, 585-594.
16. Smith, C. A., Jordan, J. M. and Vinograd, J. (1971) J. Mol. Biol. 59, 255-272.
17. Subramanian, K. N., Pan, J., Zain, S. and Weissman, S. M. (1974) Nucleic Acids Res. 1, 727-752.
18. Kasamatsu, H. and Vinograd, J. (1973) Nature New Biol. 241, 103-105.
19. Kroon, A. M., Bakker, H., Holtrap, M. and Terpstra, P. (1977) Biochim. Biophys. Acta, 474, 61-68.
20. Potter, S. S., Newbold, J. E., Hutchison, C. A. and Edgell, M. H. (1975) Proc. Nat. Acad. Sci. USA 72, 4496-4500.
21. Jakovcic, S., Casey, J. and Rabinowitz, M. (1975) Biochemistry 14, 2043-2050.

Volume 4 Number 5 1977 Nucleic Acids Research

Detection of a contaminant cell culture line by restriction endonuclease cleavage patterns of mitochondrial DNA*

Lawrence I.Grossman[1], Richard C.Parker[2], Robert M.Watson[3], Sarah E.W.Chandler[1], Marlyn Teplitz[3]

[1]Department of Biochemistry, Wayne State University School of Medicine, Detroit, MI 48201, [2]Division of Biology, [3]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Received 10 February 1977

ABSTRACT

A putative HeLa cell culture line was discovered to be contaminated with mouse cells by examination of agarose gel profiles of restriction endonuclease digests of mitochondrial DNA. The contamination was confirmed by karyotypic analysis, and by observation of the mouse satellite band in an analytical buoyant density centrifugation of total cellular DNA. Restriction endonuclease analysis of mitochondrial DNA is suggested as a useful method for monitoring the species of cells in culture.

INTRODUCTION

Numerous reports have documented the accidental contamination of long-term cell cultures with unrelated cell lines (1, 2). Various methods have been applied for monitoring cell lines for clonal purity. Among methods in common use are biochemical procedures, karyotypic analysis and surface antigen identification (3). These are reviewed by Stulberg (4).

We suggest here the use of restriction endonuclease cleavage patterns of mitochondrial DNA (mtDNA) for species identification of cells in culture. A contamination of a HeLa cell line by mouse cells was detected in our laboratories using this method. This contamination was subsequently confirmed by two other procedures.

MATERIALS AND METHODS

Growth of Cells. Cells were grown in suspension culture with Dulbecco's modification of Eagle's Phosphate Medium (Grand Island Biological Co.) and 5% calf serum.

Isolation of Mitochondrial DNA, *Restriction Endonuclease Digestions* and *Gel Electrophoresis* were carried out as described (5, 6).

Analytical Buoyant Density Centrifugation. A 1 ml cell pellet was mixed with 8 ml of buffer (10 mM Tris-HCl [pH 7.5], 50 mM NaCl, 1 mM EDTA) and lysed by the addition of 1 ml 10% sodium dodecyl sulphate. After vigorous mixing the viscosity was reduced by several passages through a 25 gauge needle. An aliquot was mixed with a CsCl solution and centrifuged at 44,770 rev./min, 25° C for 24 hr in a Beckman model E ultracentrifuge equipped with a photoelectric scanner.

Karyotype Analysis. Preparations for karyotype analysis were carried out by a modification of a standard method (7).

RESULTS

A cell line putatively identified as HeLa was in use in one of our laboratories for the construction of restriction endonuclease cleavage maps. This cell line was sent to the second laboratory for similar studies. Both laboratories isolated mtDNA from these cells, digested it with several restriction enzymes, and analyzed the digestion products by agarose gel electrophoresis.

The results of the analysis in the first laboratory are shown in Fig. 1. Slots 3 and 6 show the cleavage products generated by the enzymes *Hin*III and *Hae*II, respectively. The molecular weights of the fragments in each of these slots total approximately twice those of mouse and human mtDNAs. Dimer mtDNA is known to consist of a head-to-tail arrangement of monomer molecules (8).

The *Hae*II digest of mtDNA from the cell line in question in slot 6 is bordered by a *Hae*II digest of LA9 mtDNA (slot 5) and HeLa mtDNA (slot 7). The largest species in slot 6 migrates indistinguishably from the products of *Hae*II on LA9 mtDNA. The latter DNA contains a single *Hae*II site (5). The remaining species in slot 6 co-migrate with the products of *Hae*II on HeLa mtDNA (slot 7).

The *Hin*III digest of the cell line in question in slot 3 is similarly compared to digests by this enzyme of mouse LA9 (slot 2) and HeLa (slot 4) mtDNAs. It is clear that the bands in slot 3 are composed of the *Hin*III products of each DNA alone.

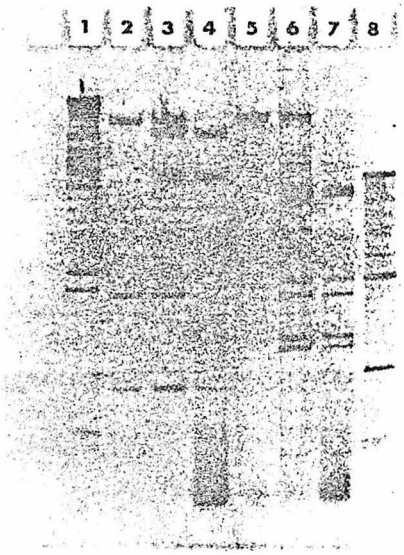


Fig. 1. Gel electrophoresis of restriction endonuclease digests of mtDNA. Electrophoresis was carried out at 50 volts, 6 hr in 1% agarose.
(1) λ x *Eco*RI + λ x *Hin*III.
(2) LA9 mtDNA x *Hin*III.
(3) "HeLa" mtDNA x *Hin*III.
(4) HeLa mtDNA x *Hin*III.
(5) LA9 mtDNA x *Hae*II.
(6) "HeLa" mtDNA x *Hae*II.
(7) HeLa mtDNA x *Hae*II.
(8) PM2 DNA x *Hin*III.

The HeLa mtDNA used for comparison with the mtDNA from the contaminated cell line was not extensively purified. The streak near the bottom of slots 4 and 7 is nuclear DNA. Slots 1 and 8 contain molecular weight standards. The *Eco*RI and *Hin*III products of bacteriophage λ DNA are shown in slot 1 (9, 10); the *Hin*III products of bacteriophage PM2 DNA are shown in slot 8 (10).

Approximately equal masses of mouse and human mtDNA are present in the isolate from the contaminated cell line, as judged by the relative intensities of appropriate DNA species in slots 3 and 6. The mtDNA content of a HeLa cell has been shown to be at least eight times that of a mouse cell (11). Assuming equal efficiencies of extraction of mtDNA from the mixed cell population, we estimate that the "HeLa" cell line contained 90% mouse cells at the time of DNA isolation.

The cleavage patterns in the recipient laboratory were generated with *Eco*RI and *Hin*III. The digests consisted exclusively of species generated by these enzymes on mouse mtDNA. We attribute the loss of the human cells to inadvertent selection during initial growth in culture dishes.

Two additional procedures were used to confirm the restriction endonuclease analysis. First, total cell DNA was extracted from the contaminated cell line and analyzed by

buoyant centrifugation in a CsCl solution, as described in Materials and Methods. The satellite DNA found in mouse cells was observed.

In addition, metaphase chromosome preparations were examined. Most spreads showed the same characteristics as those of the abnormal mouse cells (LA9) being carried in our laboratories. These cells are distinguishable from HeLa cells on the basis of several karyotypic features, including fewer total chromosomes and, most strikingly, a much higher proportion of acrocentric forms --approximately 60% *versus* 20% in HeLa cells (unpublished results).

DISCUSSION

The necessity for periodic monitoring of long-term cell cultures for contamination by other cell lines is now well established. The choice of an available method is often determined by the expertise in a given laboratory. Increasing use of restriction endonucleases and their commercial availability, as well as the widespread use of gel electrophoresis, makes the application of mtDNA cleavage patterns practical as a routine procedure, particularly in laboratories growing cells for studies with nucleic acids. The possibility of using several restriction endonucleases to more closely define a cleavage pattern enhances the appeal of this method, which has also been suggested by others (12, 13). The growing availability of restriction endonuclease cleavage maps of various mtDNAs may allow positive identification of contaminant cell species.

Variations in the cleavage patterns of mtDNAs within the same species have been documented (12). Thus, the production of reference cleavage patterns with several enzymes for cell lines in use in a given laboratory is recommended. It may be possible in this way to monitor contamination between cells of the same species.

ACKNOWLEDGEMENTS

We are grateful to Mrs. Paula Berry, City of Hope National Medical Center, for help with the karyotype preparations. This work was supported by grants GM 21704 from NIH at Wayne State University and GM 15327 and CA 08014 from NIH at Caltech. This is contribution No. 5501 from the Division of Chemistry and Chemical Engineering.

REFERENCES

*Dedicated to the memory of Professor Jerome Vinograd.

1. Gartler, S.M. (1967) Nat. Cancer Inst. Monogr. 26, 167-178.
2. Nelson-Rees, W.A., Flandermeyer, R.R. and Hawthorne, P.K. (1974) Science 184, 1093-1096.
3. Kruse, P.F., Jr. and Patterson, M.K., Jr., eds. (1973) Tissue Culture Methods and Applications. pp. 744-821. Academic Press, New York.
4. Stulberg, C.S. (1973) in Contamination in Tissue Culture, ed. Fogh, J. Academic Press, New York.
5. Parker, R.C. and Watson, R.M. (1977) Nucleic Acids Res. 4, 1291 - 1299.
6. Moore, K.H., Johnson, P.H., Chandler, S.E.W. and Grossman, L.I. (1977) Nucleic Acids Res. 4,1273 - 1289.
7. Moorehead, P.S. (1973) in Tissue Culture Methods and Applications, eds. Kruse, P.F., Jr. and Patterson, M.K., Jr. Academic Press, New York.
8. Clayton, D.A., Davis, R.W. and Vinograd, J. (1970) J. Mol. Biol. 47, 137-153.
9. Thomas, M. and Davis, R.W. (1975) J. Mol. Biol. 91, 315-328.
10. Parker, R.C., Watson, R.M. and Vinograd, J. (1977) Proc. Nat. Acad. Sci. U.S.A., in press.
11. Bogenhagen, D. and Clayton, D.A. (1974) J. Biol. Chem. 249, 7991-7995.
12. Potter, S.S., Newbold, J.E., Hutchison, C.A. and Edgell, M.H. (1975) Proc. Nat. Acad. Sci. U.S.A. 72, 4496-4500.
13. Brown, W.M. (1976) Ph.D. Thesis, California Institute of Technology.

CHAPTER THREE

Two-Dimensional Gel Electrophoretic Analysis of Restriction Endonuclease Cleaved DNA

INTRODUCTION

The following paper is a preprint of an article that will appear in Methods in Enzymology. It explains how one can use a low melting temperature (LMT) agarose to facilitate two-dimensional electrophoretic analysis of restriction endonuclease cleaved DNA. This is possible because DNA electrophoretically isolated and maintained in low melting temperature agarose can be digested to completion with restriction endonucleases and electrophoresed on a second gel.

The power of this technique is very great and still untapped. It has been used extensively in our laboratory to facilitate restriction endonuclease site mapping of simple DNAs. However, this is just one of many potential uses.

Various experiments can and should be tried using low melting temperature agarose. Among them are experiments dealing with:

1) Hybridization - a driver or a probe could be added to an LMT agarose solution containing an isolated nucleic acid.

2) Other enzymatic reactions - S1, polynucleotide kinase, DNA ligase.

3) Characterization of eukaryotic chromosomes.

a) Isolation of DNA fragments in LMT agarose known (e.g., by blotting experiments) to contain a gene followed by restriction endonuclease digestion, gel electrophoresis, and blotting. This simplifies the interpretation of double digest results.

b) Two-dimensional analysis of wild-type versus mutant eukaryotic chromosomes. Restriction digests of each DNA can be run in adjacent lanes of an LMT agarose gel. Slices of the gel can be removed from the adjacent slots, redigested with a different restriction endonuclease and re-run on a second gel.

Adjacent lanes of the second gel can then be compared to find differences in the wild type and mutant patterns. Polymorphisms, of course, will complicate interpretation.

Some of these experiments can be done immediately using the following protocol. Others demand that experiments be done to test the feasibility of the ideas.

VI. Endonuclease Cleavage Mapping Techniques

12. Two-Dimensional Agarose Gel Electrophoresis

"SeaPlaque" Agarose Dimension

by Richard C. Parker and Brian Seed

Division of Biology

California Institute of Technology

Pasadena, California 91125

Running head: Restriction Digests in the Presence of Agarose

There are many techniques currently available for mapping the restriction endonuclease sites of various DNAs. Most of these involve gel electrophoretic fractionation of the products of enzymatic digestion. Though the simplicity and resolution of gel electrophoresis are unrivalled, it is often difficult to manipulate or isolate DNA electrophoretically embedded in gel media. When large numbers of restriction fragments are to be analyzed, enzymatic redigestion of DNA electrophoresed through gels becomes an unprofitable approach to restriction mapping. Methods for the in situ digestion of fragments trapped in gels have been presented elsewhere. In general, these techniques have not become popular because they require large amounts of enzyme for complete digestion, and either have not been shown to be applicable, or have been shown to be inapplicable to a broad range of enzymes with different specificities.

In the following we present a simple method for redigestion of restriction fragments electrophoresed through a low melting temperature agarose. The technique is both easy to use and economical of the second-dimension enzyme. It is applicable to all the enzymes we have tested and does not require extensive modification of conventional electrophoretic apparatuses. The resolving characteristics of the low melting temperature (LMT) agarose used are essentially identical to those of ordinary agaroses; however, the low melting temperature agarose is both more expensive and more fragile than commonly used agaroses.

In the following procedure, restriction endonuclease products are fractionated in a gel made from hydroxyethyl agarose (SeaPlaque Agarose, Marine Colloids, Inc.). SeaPlaque Agarose melts at 65^{o} and remains in solution at 37^{o}, allowing the gel to be dissolved without denaturing DNA. It is possible to find conditions which permit complete digestion of DNA with restriction endonucleases in the

presence of SeaPlaque Agarose at 37°. Enzyme efficiencies are not greatly reduced under these conditions. Following fractionation, the bands of interest can be visualized either by ethidium bromide fluorescence or autoradiography. Discrete regions of the gel are excised and placed in tubes, the agarose melted, buffers and restriction endonucleases added, and the DNA cleaved by the added enzymes.

Pouring the Gel

Gels for the first dimension of two-dimensional DNA electrophoresis are made by dissolving hydroxyethyl agarose in E buffer (40 mM Tris, 5 mM sodium acetate, 1 mM EDTA, pH adjusted to 7.4 with glacial acetic acid) by heating to at least 65°C. Before the gel is poured, the agarose should be cooled to 37°C to minimize shrinkage while the gel solidifies. If this is not done, the gel may crack or detach from the apparatus.

Pouring an LMT agarose gel in a horizontal apparatus presents no special problems. When pouring vertical gels in forms with insertable combs, however, care must be taken to remove the comb without tearing the gel. Two steps may be taken to accomplish this. The first is to clamp the comb a few millimeters above the top of the gel. The second is to chill the gel in a refrigerator before removing the comb. After the gel has solidified, buffer can be pipetted around the exposed teeth and the comb removed without damage to the gel.

LMT agarose gels are more slippery than ordinary gels, and tend to slide from vertical holders more easily. Depending on the design of the apparatus, a mechanical support at the bottom of the gel, or a conventional agarose plug, may be necessary.

Running the Gel

The gel should be run for at least 10 minutes before samples are applied. During this time, the voltage should be gradually increased. LMT agarose gels frequently crack when high voltage gradients (5 V/cm) are applied to the gel without this gradual increase. If the voltage is slowly increased, the gels are stable to at least 6.6 V/cm. In our apparatus, this gradient spans a 15 cm long, 4 mm thick gel and requires approximately 60 milliamps of current. Horizontal apparatuses with thin (e.g., paper) wicks may exhibit sufficient ohmic heating at the ends of the gel to melt LMT agarose. Running at lower voltages, or in the cold, can avoid melting.

Samples are layered on LMT agarose gels in the same manner as on other agarose or acrylamide gels. Mobilities of both circular and linear DNA molecules in this agarose are similar to their mobilities in gels made from more commonly used agaroses.

Band Detection and Isolation

Depending upon the quantity of DNA available, it may be more convenient to determine the location of the DNA in the gel by autoradiography of ^{32}P-labeled DNA or by ethidium bromide (EtdBr) fluorescence. If bands are to be detected by autoradiography, the gel should be soaked for 10 minutes in dH_2O in buffer. The bands can be cut out of the gel by measuring their mobility on the autoradiogram and removing that region of the gel. If bands are to be detected by EtdBr fluorescence, the gel should be stained for 10 minutes in 0.2 μg/ml of EtdBr and then destained for 10 minutes without EtdBr. The solutions for staining and destaining can be made in distilled water or in a buffer appropriate for the second restriction endonuclease.

The bands can then be visualized by either long wave or short wave ultraviolet light depending upon the quantity of DNA in the bands. It is preferable to use long wave light in order to minimize nicking of the DNA. The bands should be cut from the gel in slices that are as thin as possible so that the entire volume of the slice can be loaded on the next gel.

Restriction Endonuclease Digestion in SeaPlaque

After slices of the gel have been placed into individual test tubes, their volume should be estimated and 1/9 volume of a stock of ten times concentrated enzyme buffer added. This step is unnecessary if the gel was already soaked in buffer.

The samples are then placed at 65°C until the agarose melts. When liquid, the sample is placed at 37°C and allowed to thermally equilibrate. After it has equilibrated, a sterile solution of nuclease-free bovine serum albumin (Pentex BSA, Miles Laboratories) is added to a final concentration of 0.1%. The addition of BSA is essential for complete digestion.

The presence of agarose does not seem to greatly inhibit restriction endonucleases under these conditions. To insure complete digestion, two to three times more enzyme should be added than would be used to digest the DNA under standard conditions. If the number of restriction sites in the DNA is known, we generally normalize the amount of enzyme per microgram of DNA so that the ratio of enzyme to concentration of sites is equivalent to that found in the standard assay with lambda DNA. When the number of sites in lambda is known, we calculate the amount of enzyme required by the following formula:

$$(1)\quad \frac{\lambda_{mw}}{X_{mw}} \times \frac{X_{sites}}{\lambda_{sites}} \times \frac{\text{units of enzyme}}{1\ \mu g\ \text{of}\ \lambda\ \text{DNA}} = \frac{\text{units of enzyme}}{1\ \mu g\ \text{of X DNA}}$$

The Second Dimension

If the second gel is a vertical gel, it is important that the LMT agarose remain as a liquid until the DNA has migrated into the gel. If the agarose is allowed to set up, the bands in the second dimension will be very broad. If the second gel is a horizontal gel, the agarose may be allowed to set up. In the event that a vertical gel is used, there are two ways to insure that the agarose stays in solution. The first, and easiest, is to heat both the samples and the running ("E") buffer for the upper reservoir to 65°C. Alternatively, the samples may be loaded in 33% formamide. The latter has the disadvantage of increasing the sample volume.

RESULTS

Four of the enzymes that work well under our standard second dimension conditions were used to obtain the patterns shown in Figure 1. This gel contains the electrophoretically resolved products of digestions of λcI_{857} DNA in the presence or absence of SeaPlaque agarose. The restriction endonucleases used were: Ava I, Bam HI, Eco RI, and Hind III. In all cases, the complete digest products obtained in the presence of LMT agarose were identical to those in the parallel digestion.

At the time of writing, we have tested the following seventeen enzymes: Alu I, Ava I, Bam HI, Eco RI, Hae II, Hae III, Hha I, Hinc II, Hind III, Hpa I, Hpa II, Kpn I, Pst I, Pvu II, Sal I, Sma I, and Xho I. All were capable of producing complete digests in the presence of LMT agarose.

The potential uses of LMT agarose are clearly broader than those presented here. We anticipate that as the need arises, various other applications will be devised. In our experience, the single most important step for obtaining

good enzymatic activity, in the presence of LMT agarose, has been the addition of bovine serum albumin to 1 mg/ml. This knowledge may be helpful in the design of future applications.

ACKNOWLEDGEMENTS

We would like to thank Norman Davidson and Tom Maniatis for their suggestions and encouragement.

Figure 1. A 0.7% agarose gel was used to electrophoretically resolve the products of restriction endonuclease digests of $\lambda c I_{857}$ DNA in the presence (slots 1,3,5,7) and absence (slots 2,4,6,8) of SeaPlaque agarose. The enzymes used were: Ava I (slots 1,2), Bam HI (slots 3,4), Eco RI (slots 5,6), and Hind III (slots 7,8).

CHAPTER FOUR

Cloning and Characterization of Normal Adult Human β-Related Globin Genes

INTRODUCTION

About a year ago, I went to Tom Maniatis and asked him if he would be interested in my working on the organization of human globin genes. As expected, he had given a great deal of thought to the project and shortly after Thanksgiving I began working on this problem.

We decided that our immediate need was an understanding of the location of restriction endonuclease sites in and around the genes. Our collaborator, Bernard Forget of Yale University, provided us with human chromosomal DNA and sequenced cDNA plasmids corresponding to the α-, β-, and γ-globin mRNAs.

When the project began, it was not possible to clone recombinant DNA molecules containing human DNA. Therefore, our first experiments involved chromosomal blots of DNA isolated from a patient with normal hemoglobin.

Shortly after I began working on this project, two post-docs, Dick Lawn and Ed Fritsch, joined our lab and began working with me. Together with Tom, we decided what experiments needed to be done and by whom.

Dick and Ed were invaluable collaborators. We were also helped by a freshman, Geoff Blake, who spent about eight months in our lab learning about biochemistry and molecular biology.

The globin genes in human beings present an interesting system to the basic researchers because of the wide variety of mutants that exist. Until very recently, it has been impossible to study the DNA of an eukaryote at a very detailed level.

It has been possible to study certain RNAs and proteins. Experiments with human hemoglobin have demonstrated that a wide variety of mutants exist

and that the mutants probably contain many varying DNA sequences.

The proposed changes in DNA structure that we hoped to study were manifested by many changes in globin mRNA and in globin proteins. Diseases had been characterized where:

1) Neither globin mRNA nor globin protein was found (1);

2) The RNA was present as was a very low amount of protein (2);

3) No DNA was detected by hybridization experiments (3);

4) The globin protein contained one altered amino acid (4);

5) The globin protein contained many extra amino acids at the C-terminal probably due to either a point or a frameshift mutation (5a,b); and

6) The globin protein contained the N-terminal amino acids of one globin protein and the C-terminal amino acids of another globin protein (6).

This wide variety of mutations makes human hemoglobin a very attractive system as do the applied aspects which are also of great interest. The first abnormal DNA that we chose to study came from a patient with hemoglobin Lepore.

This patient did not produce any of the normal adult β-related globin chains (the δ-protein and the β-protein). Instead, the only adult β-related protein found in this person's blood was the Lepore protein. This polypeptide contains the N-terminal amino acids of the δ-protein and the C-terminal amino acids of the β-protein.

We decided to explore whether this protein was coded for by a fused gene containing 5'-end δ-gene sequences and 3'-end β-gene sequences or by some alternative mechanism. The results of that work are not presented in this thesis.

There is one other preprint in this chapter. It is a copy of a manuscript that will appear in two symposia volumes. All of the data on human globin gene information presented in this paper were also presented in the previous article found in this chapter. I have decided to include this paper because it provides a valuable insight into the basic research of globin genes and the rest of the globin work that was done while I was in the Maniatis laboratory.

References

1. Ottolenghi, S., Lanyon, W. G., Williamson, R., Weatherall, D. J., Clegg, J. B., and Pitcher, C. S., Proc. Nat. Acad. Sci. USA **72**, 2294-2299 (1975).
2. Ramirez, F., Natta, C., O'Donnell, J. V., Carala, V., Bailey, G., Sanguensermsri, T., Maniatis, G. M., Marks, P. A. and Bank, A., Proc. Nat. Acad. Sci. USA **72**, 1550-1554 (1975).
3. Forget, B. G., Hillman, D. G., Lazarus, H., Barell, E. F., Benz, E. J., Caskey, C. T., Huisman, T. H. J., Schroeder, W. A. and Housman, D., Cell **7**, 323-329 (1976).
4. Ingram, V. M., Biochim. Biophys. Acta **36**, 402-411 (1959).

5a. Bunn, H. F., Schmidt, G. J., Haney, D. N. and Dluhy, R. G., Proc. Nat. Acad. Sci. USA **72**, 3609-3613 (1975).

5b. Clegg, J. B., Weatherall, D. J. and Milner, P. F., Nature **234**, 337-340 (1971).

6. Baglioni, C., Proc. Nat. Acad. Sci. USA **48**, 1880-1886 (1962).

The Isolation and Characterization of Linked δ- and β-Globin Genes from a Cloned Library of Human DNA

Richard M. Lawn, Edward F. Fritsch, Richard C. Parker,
Geoffrey Blake and Tom Maniatis

Division of Biology, California Institute of Technology
Pasadena, California 91125

Running Title: Linked Human Globin Genes

Summary

A cloned library of large, random embryonic human DNA fragments was constructed and screened for β-globin sequences using the cloned human β-globin cDNA plasmid pJW102 (Wilson et al., 1978) as a hybridization probe. Two independent clones were obtained and then characterized by restriction endonuclease cleavage analysis, hybridization experiments and partial DNA sequencing. Each of the clones carry both the adult δ- and β-globin genes. The two genes are separated by approximately 5.4 kilobases (kb) of DNA and their orientaton with respect to the direction of transcription is 5' - δ - β - 3'. Both the δ- and β-globin genes contain a large non-coding intervening sequence (950 and 900 bp, respectively) located between the codons for amino acids 104 (arginine) and 105 (leucine). Although the location of the large intervening sequence within the coding regions of the two genes is identical, the two non-coding sequences bear little sequence homology. A second, smaller intervening sequence similar to that found in other mammalian β-globin genes was detected near the 5' end of the human β-globin gene. The two independently isolated β-globin clones differ from each other by the presence of a Pst I restriction enzyme cleavage site within the large intervening sequence of the δ-globin gene of one of the clones. This suggests that the human DNA carried in the two clones was derived from two homologous chromosomes which were heterozygous for the Pst I restriction enzyme recognition sequence.

Introduction

We have recently established a procedure for isolating eukaryotic genes which involves the construction and direct screening of permanent, cloned libraries of large (15-20 kb), random fragments of genomic DNA (Maniatis et al., 1978). The feasibility of isolating single copy mammalian genes by this approach was demonstrated by purifying β-globin genes from a library of rabbit DNA. With this procedure, different members of a gene family can be isolated by screening a library with a mixed hybridization probe. In addition, if enough independent recombinants are screened, it is possible to obtain a set of overlapping chromosomal segments which contain the gene of interest plus adjacent sequences extending many kilobases from the gene in the 5' and 3' directions. Thus, it is possible that two or more closely linked genes will be carried in a single recombinant clone. For example, a clone containing both an adult β-globin gene and a β-related globin gene was obtained from a library of rabbit DNA by hybridization to an adult β-globin probe (Maniatis et al., 1978; Lacy et al., 1978).

The human globin gene family, comprised of 3 α-related and 5 non-α genes, is an example of a small group of functionally and evolutionarily related genes that are differentially expressed during development (see Bunn, Forget and Ranney, 1977; and Nienhuis and Benz, 1977 for reviews). Most individuals carry two adult α-globin genes (Hollan et al., 1972; DeJong et al., 1975) which are located on chromosome 16 (Deisseroth, 1977). The chromosomal locations of ζ and ε, the α-related and non-α embryonic globin genes, respectively, are not known. The remaining non-α genes ($^{G}\gamma$, $^{A}\gamma$, δ and β) are thought to be closely linked (Weatherall and Clegg, 1972; Bunn et al., 1977) on chromosome 11 (Deisseroth et al., 1978). The close linkage of the δ- and β-globin genes (the minor and

major β-related adult globin genes, respectively) was deduced from studies of the mutant Hb Lepore which is a fusion protein containing N-terminal δ and C-terminal β amino acids (Baglioni, 1962). Similarly, the linkage of the $^{A}\gamma$- (one of the two non-allelic fetal genes) and β-globin genes was deduced from structural and genetic analysis of the mutant Hb Kenya, a fused protein whose N-terminal and C-terminal amino acids are derived from the $^{A}\gamma$- and β-globin genes, respectively (Huisman et al., 1972; Kendall et al., 1973). Assuming that these fused proteins result from unequal crossing over between homologous genes at meiosis, the 5' - $^{G}\gamma$ - $^{A}\gamma$ - δ - β - 3' or 5' - $^{A}\gamma$ - δ - β - $^{G}\gamma$ - 3' configurations are both consistent with the structural data.

In this paper, we provide definitive evidence for the close physical linkage of the human δ- and β-globin genes in the order: 5' - δ - β - 3'. This is accomplished by isolating and characterizing a cloned human DNA fragment which carries both the δ- and β-globin genes.

Results

Construction of a Human Genome Library

A library of approximately 1 x 10^{6} independently derived bacteriophage λ clones consisting of large (15-20 kb) fragments of human DNA covalently joined to a bacteriophage λ vector was prepared as previously described (Maniatis et al., 1978). In brief, human fetal liver DNA was subjected to a non-limit digestion with the restriction endonucleases Hae III and Alu I and the products size fractionated by sucrose gradient centrifugation. Large fragments (15-20 kb) were isolated, and treated with Eco RI methylase to render Eco RI sites within the human DNA resistant to cleavage with Eco RI . Synthetic dodecameric DNA molecules bearing

an Eco RI cleavage site (Eco RI linkers) were ligated to the methylated DNA, and digested with Eco RI to generate Eco RI cohesive ends. Following an additional size selection (15-20 kb), the human DNA was suitable for insertion into the bacteriophage λ cloning vector, Charon 4A (λCh4A) (Blattner et al., 1977).

Foreign DNA can be inserted into the λ Ch4A vector after removal of two internal Eco RI fragments which contain genes nonessential for phage growth (Blattner et al., 1977). The two "arms" of the bacteriophage DNA are annealed through their 12 base pair cohesive ends and joined to the human DNA by ligation of the Eco RI cohesive ends. The ligation reaction is performed at a high DNA concentration to promote the formation of long concatemeric DNA molecules which are the substrate for in vitro packaging (Sternberg et al., 1977). Approximately 1×10^6 in vitro packaged phage were amplified 10^6-fold by low density growth on agar plates to establish a permanent library of cloned human DNA fragments. Table 1 presents a brief summary of the essential features of the human library.

Isolation and Characterization of Genomic DNA Clones Bearing β-globin Sequences

The human library was screened for clones containing globin sequences using the in situ plaque hybridization procedure of Benton and Davis (1977) (see also Maniatis et al., 1978). ^{32}P-labeled nick-translated cDNA plasmids containing the human α-, β- or γ-globin genes (Wilson et al., 1978) were used as hybridization probes. On the first screening of 300,000 recombinant phage, two independently derived clones bearing β-globin sequences were identified and plaque purified. We have designated these clones HβG1 and HβG2.

The location of various restriction endonuclease sites in the recombinant DNA was determined as a means of characterizing the DNA sequence organization

of the human β-globin gene. Figure 1A shows the agarose gel electrophoresis pattern of HβG1 and HβG2 DNA digested with six different enzymes. Although the two clones are similar, they clearly are not identical. For example, digestion of HβG1 DNA with Xba I generates 4 fragments, while only 3 Xba I fragments are found in HβG2 DNA. Two of the Xba I fragments are common to both clones. Similarly, Eco RI digestion of the two DNAs produces 7 fragments in each case but only 5 are common to both clones. The size of the Eco RI fragments present in the two clones indicates that HβG1 and HβG2 contain 15.9 and 14.5 kb of human DNA, respectively.

Localization of β-globin Sequences in HβG1 and HβG2

The position of β-globin sequences within the cloned human DNA was determined by digesting the DNA with various restriction enzymes, fractionating the products by agarose gel electrophoresis, blotting the DNA directly from the gel onto nitrocellulose filter paper (Southern, 1975) and hybridizing the filter with the human β-globin cDNA plasmid pJW102 (Wilson et al., 1978) labeled with ^{32}P by nick translation (Maniatis et al., 1975a). HβG1 and HβG2 DNA were first digested with four restriction endonucleases which do not cleave within the human β-globin coding sequence (Xba I, Hha I, Hind III, Bgl I) (Marrota et al., 1977). As seen in the autoradiogram of Figure 1B, the β-globin probe hybridizes to only one fragment in the Bgl I, Hha I and Xba I digests. The 10.75 kb Xba I fragment is generated by two Xba I sites within the inserted human DNA since this enzyme does not cleave λ Ch4A DNA (data not shown). The Hha I fragment which hybridizes the β-globin probe is larger than the human DNA inserts of HβG1 and HβG2. Unexpectedly, two Hind III fragments hybridize to the β-globin probe even though no Hind III sites are present in the β-globin mRNA sequence.

When the experiment was performed using restriction enzymes known to cleave only once within the β-globin mRNA sequence, more than two hybridizing fragments were observed. For example, Eco RI recognizes a single site in the β-globin mRNA sequence corresponding to the codons for amino acids 121 and 122. Thus, when the cloned DNAs are digested with Eco RI, two fragments which hybridize to the β-globin probe are expected. However, as shown in Figure 1B, Eco RI digestion of HβG1 and HβG2 DNA produces four fragments which hybridize the β-globin probe. In HβG2 the sizes are: C (5.2 kb), D (3.2 kb), E (2.25 kb), and G (1.75 kb). Similarly, Bam HI recognizes a single site in the β-globin mRNA sequence corresponding to the codons for amino acids 98 to 100. Digestion of both DNAs with Bam HI also produces four fragments which hybridize to the β-globin probe. In HβG2 the sizes are: A (19 kb), B (8.3 kb), D (4.4 kb), and F (1.8 kb). There are at least two possible explanations for these unexpected results. First, the cloned β-globin gene may contain at least two non-coding intervening sequences, two of which are cleaved at least once by Eco RI and Bam HI, and one of which is cleaved by Hind III. Alternatively, the human DNA insert in both clones may contain two closely linked β-globin related gene sequences.

To discriminate between these two alternatives, we constructed a detailed map of the restriction sites within and surrounding the sequences which hybridize to the β-globin probe.

Mapping Restriction Endonuclease Cleavage Sites in HβG2 DNA

Restriction mapping of HβG2 DNA was accomplished by digesting the cloned DNA with various restriction enzymes singly or in combination and determining the sizes of the digestion products by agarose gel electrophoresis. In addition, we

made use of a two-dimensional agarose gel electrophoresis procedure which involves redigestion of restriction fragments electrophoresed through low melting temperature "SeaPlaque" agarose gels (Parker and Seed, 1979). In this procedure, DNA digested with one restriction enzyme is fractionated by electrophoresis in SeaPlaque agarose. The bands are identified by ethidium bromide staining and excised from the gel. The gel fragment is melted by heating at 65°C. When the temperature is lowered to 37°C, the agarose remains in solution and the DNA can be digested to completion with a second enzyme and the entire mixture layered onto a second gel for electrophoresis. In the following, we summarize the data which were used to derive the maps of various restriction endonuclease cleavage sites shown in Figure 4.

Hind III: Digestion of HβG2 DNA with Hind III produces three fragments of 27, 13, and 5.6 kb which we designate Hind A, B, and C, respectively (Figure 1A). The left arm of λ Ch4A (20 kb) is not cleaved by Hind III, while the 10.5 kb right arm is cleaved into two fragments of 5.6 kb and 4.9 kb (data not shown). Thus, the Hind III A fragment of HβG2 (27 kb) includes the entire left arm of the cloning vector and a portion of the human DNA insert while the Hind C fragment derives solely from the right arm. The Hind B fragment therefore lies between the A and C fragments (Figure 4).

Xba I: Digestion of HβG2 DNA with endonuclease Xba I yields three fragments designated Xba A (21.5 kb), B (13.5 kb), and C (10.75 kb) (Figures 1 and 2). There are no Xba I sites in either arm of the vector (data not shown). Therefore, the Xba A fragment (21.5 kb) must include the 19 kb left arm of λ Ch4A and a portion of the human DNA insert. To orient fragments B and C, the products of an Xba I digest were electrophoresed in a 0.5% SeaPlaque agarose gel and the B and C fragments were separately excised from the gel and digested

with Eco RI. The Xba B fragment yielded the right arm of λ Ch4A DNA (10.5 kb) plus two smaller fragments (Figure 2). The Xba C fragment yielded several smaller Eco RI fragments from within the human DNA insert (data not shown). Thus, the order of the Xba I fragments is B-C-A (Figure 4).

Eco RI: Digestion of HβG2 DNA with Eco RI generates seven fragments designated RI A (20 kb), B (10.5 kb), C (5.2 kb), D (3.2 kb), E (2.25 kb), F (2.05 kb), and G (1.75 kb) (Figure 1). The 20 and 10.5 kb fragments are the left and right arms, respectively, of λ Ch4A DNA. The remaining five Eco RI fragments contain only human DNA. The five smaller fragments were positioned by employing double enzyme digests, done either simultaneously, or sequentially by the SeaPlaque agarose technique.

Double digestion of HβG2 DNA with Eco RI and Xba I produces 9 fragments (Figure 2). Five of these fragments are the Eco RI complete digest products RI A, B, C, F, and G. The other four fragments were formed by Xba I cleavage of both RI D and RI E, producing fragments of 1.9 kb, 1.85 kb, 1.4 kb, and 0.4 kb. These sizes indicate that the 0.4 kb and either the 1.9 kb or 1.85 kb fragments were derived from RI E.

The Xba B fragment, already shown to contain the right arm of λ Ch4A, was isolated from a 0.6% SeaPlaque gel and redigested with Eco RI. Three fragments were resolved by agarose gel electrophoresis, the 10.5 kb right arm of λ Ch4A, the 2.0 kb Eco RI F fragment, and the 0.4 kb portion of RI E formed by cleaving RI E with Xba I (Figure 2). Digestion of the isolated Xba B fragment with Eco RI, therefore, demonstrated that RI B (the right arm of λ Ch4A) is positioned next to RI F, which, in turn, is positioned next to RI E.

The remainder of the Eco RI fragments were oriented by digesting HβG2

DNA with Eco RI and Hind III. Double digestion of HβG2 DNA with Eco RI and Hind III produces 9 fragments (Figure 2). As described above, there are two Hind III sites in HβG2 DNA, one of which is in the right arm of λ Ch4A producing fragments of 4.9 kb and 5.6 kb. A comparison of the digests in lanes 4 and 6 indicates that the other Hind III site is in the RI C fragment, producing fragments of 3.7 and 1.5 kb.

After digestion of HβG2 DNA with Hind III, the products were resolved on a 0.6% SeaPlaque gel and the 13 kb Hind B band was isolated. This Hind B band was digested with Eco RI and the products were fractionated on a 1.15% agarose gel (Figure 2). Five fragments of 4.9 kb, 2.25 kb, 2.0 kb, 1.75 kb, and 1.5 kb were formed. The middle three of these fragments are the Eco RI complete digest products - RI E, RI F, and RI G. The largest fragment comes from the right arm of λ Ch4A and the smallest is part of RI C. This result indicates that RI G is positioned next to RI E and that RI C is next to RI G. By elimination, RI D must be located between RI C and RI A. The Eco RI fragments in HβG2 DNA are thus arranged in the order B - F - E - G - C - D - A (see Figure 4).

Bam HI: Digestion of HβG2 DNA by Bam HI produces 9 fragments designated Bam A (19 kb), B (8.3 kb), C (5.2 kb), D (4.4 kb), E (3.8 kb), F (1.8 kb), G (1.5 kb), H (0.7 kb), and I (0.5 kb) (Figure 1A). Digestion of the left arm of λ Ch4A DNA with Bam HI generates fragments of 5.2 and 14.8 kb (data not shown). The 5.2 kb fragment corresponds to fragment C of Bam HI digested HβG2 DNA. The 14.8 kb fragment is larger than all of the Bam HI fragments of HβG2 DNA except the A fragment (19 kb). Thus, the A fragment must consist of 14.8 kb of λ DNA and 4.2 kb of human DNA. The Bam HI fragment A must then be positioned next to Bam C in the left arm and must span the junction between the λ Ch4A left arm

and the inserted human DNA. Digestion of the right arm of λ Ch4A generates fragments of 4.7, 3.8, 1.5, and 0.5 kb (data not shown). The latter three fragments correspond to the HβG2 Bam HI fragments E, G and I. The 4.7 kb fragment is larger than all of the Bam HI fragments of HβG2 not yet assigned a map position except for fragment B (8.3 kb). Thus, Bam B must be the junction fragment between the right arm of the vector and the inserted human DNA, consisting of 4.7 kb of λ DNA and 3.6 kb of human DNA.

By elimination, Bam F, H, and D are located entirely within the human DNA insert. To establish the order of these fragments, the Hind III fragments A and B were isolated by SeaPlaque agarose gel electrophoresis and redigested with Bam HI (Figure 3). Digestion of Hind A yielded Bam A, C, F, and H, plus an additional 0.2 kb fragment. The relative order of the Bam HI fragments F and H was determined by partial Bam HI digestion of the isolated Hind III fragment A. A partial digestion product of 0.9 kb in length was observed (data not shown). The presence of the 0.9 kb partial indicates that the 0.7 kb Bam fragment H is positioned next to the 0.2 kb fragment observed in a complete Bam digestion of Hind A. The 0.2 kb fragment must be positioned within the human DNA at the end of Hind A because it is not found in the Bam HI digest of HβG2 DNA. Therefore, the order of the Bam HI fragments in Hind A is Bam H-F-A-C.

Digestion of the Hind III B fragment yielded the Bam B and an additional 4.2 kb fragment. The Bam HI D fragment was not found in the digestion products of either of the Hind III fragments, suggesting that Bam D is cleaved by Hind III. Double digestion of HβG2 DNA by Hind III and Bam HI confirmed this deduction (data not shown) and further demonstrated that the 4.2 kb Hind III/Bam HI fragment was derived from Bam D. Therefore, Bam D is located within the human DNA

insert, adjacent to Bam B. The Hind III site in Bam D is located 4.2 kb from the end of the Bam B. The complete map of Bam HI sites in HβG2 DNA is presented in Figure 4.

Orientation of β-globin Related Sequences with Respect to the Direction of Transcription

Comparison of the hybridization results of Figure 1 and the maps of Figure 4 reveals that the two Hind III fragments (A and B) and the 4 Eco RI fragments (D, C, G, and F) which hybridize to the β-globin probe are contiguous. The 4 Bam HI fragments (A, F, D, and B) which hybridize are similarly arranged except that a small Bam HI fragment (H) lies between the Bam F and D fragments. These data alone do not allow us to distinguish between the presence of two linked β-related genes and the presence of more than one intervening sequence containing Hind III, Eco RI and Bam HI cleavage sites. To distinguish between these possibilities and to orient the gene(s) in the cloned DNA, we prepared hybridization probes specific to the 5' or 3' ends of the β-globin sequence. (This corresponds to the 5' or 3' end of the encoded mRNA sequence). The β-globin cDNA plasmid pJW102 was digested with endonucleases Bam HI and Hha I to produce fragments containing the β-globin gene portion on either side of the single Bam HI site located near the middle of the β-globin message sequence (see Figure 9B for a map of these sites in β-globin cDNA). A fragment containing that part of the gene 3' to the Eco RI site was also prepared by Eco RI and Hha I digestion. These fragments were fractionated by polyacrylamide gel electrophoresis, recovered, and labeled by nick translation for use as hybridization probes.

When HβG2 DNA is digested with Bam HI, fractionated by agarose gel electrophoresis, and transferred to nitrocellulose, the A (19 kb) and D (4.4 kb)

fragments hybridize to the 3' probe and the B (8.3 kb) and F (1.8 kb) fragments hybridize to the 5' probe (Figure 5). Due to cross-contamination of the probes, some hybridization to all 4 fragments is observed. Figure 5 also includes digests of HβG2 DNA with Eco RI. Here, probe representing the β-globin message sequence 3' to the Eco RI site hybridizes to the D (3.2 kb) and G (1.75 kb) Eco RI fragments, while 5' probe hybridizes to the C (5.2 kb) and E (2.25 kb) fragments. Thus, the 4 Eco RI and Bam HI fragments of Figure 5 which hybridize β-globin probe are arranged in order 5'-3'-5'-3' from left to right in Figure 4. This result definitively shows that HβG2 contains two closely linked β-globin related genes which are transcribed from the same DNA strand. The Eco RI site in one gene is separated from the corresponding Eco RI site in the second gene by approximately 7 kb of DNA. Approximately the same distance separates the corresponding Bam HI sites. Considering the number of base pairs between the Eco RI or Bam HI site in the mRNA sequence and the 5' and 3' ends of β-globin mRNA, we calculate that there is about 5.4 kb of non-mRNA coding sequence between the two genes.

Identification of the Linked β-related Globin Genes

The two β-related globin genes in HβG2 can be tentatively identified on the basis of the mapping and hybridization data described above. As mentioned previously, the arrangement 5' - δ - β - 3' was deduced from structural analysis of the Hb Lepore fusion protein (Baglioni, 1962). The amino sequence of the δ- and β-globin proteins differ in 10 out of the 146 amino acid residues (Dayhoff, 1972). Thus, the δ-globin gene should hybridize the cloned β-globin cDNA probe less efficiently than does the β-globin gene. The data of Figures 1 and 5 show that, as expected, the Hind III fragment which encompasses the gene to the left

(5' - δ) in the map of Figure 4 displays a weaker hybridization signal than the Hind III fragment encompassing the gene to right (3' - β). To confirm the tentative 5' - δ - β - 3' identification of the β-related globin genes in HβG2, the nucleotide sequence of a region within both genes corresponding to codons 105 to 122 was determined. This region of the sequence was chosen because amino acid residues 116 and 117 differ in the δ- and β-globin proteins. These amino acids are encoded within a region of the β-globin mRNA sequence between a Bam HI recognition site in codons 98-100 and an Eco RI site in codons 121-122 (see Figure 9B for a map of these sites in β-globin cDNA). The restriction maps of Figure 4 indicate that single Bam HI and Eco RI cleavage sites are similarly located within the two genes of HβG2. We therefore presumed that these sites delineate the region between codons 98-122 in the β-globin gene and the corresponding region of the δ-globin gene.

A comparison of the Bam HI and Eco RI maps presented in Figure 4 indicated that the distance between the Eco RI and Bam HI sites within each gene was approximately 1000 bp. Digestion of cloned DNA with both Bam HI and Eco RI produces fragments of 1000 and 950 bp, among others. The 1000 and 950 bp fragments were identified as the internal Bam HI/Eco RI fragment in the 5' and 3' β-related gene, respectively, by restriction endonuclease mapping experiments (data not shown) and by hybridization of the isolated 1000 or 950 bp fragments to the Pst I fragments of HβG1 DNA, which contain the 5' or 3' genes (see Figures 11 and 13).

To sequence the Eco RI ends of the internal Bam HI/Eco RI fragments from the two genes, HβG2 DNA was digested with Eco RI, and the ends labeled with ^{32}P using $\gamma^{32}P$ labeled ATP and T4 polynucleotide kinase (Berkner and Folk,

1977). The labeled DNA was then digested with Bam HI and subjected to electrophoresis in a 3-1/2% polyacrylamide gel. Following autoradiography, the 1000 base pair (5') and the 950 base pair (3') fragments were excised, eluted, and sequenced using the procedure of Maxam and Gilbert (1977). Based on the orientation of the Bam HI/ Eco RI fragments in the maps of Figure 4, the coding strand should be end-labeled. As shown in Figures 6 and 7, the first 53 nucleotides of the 950 base pair 3' fragment corresponds exactly to the nucleotide sequence of human β-globin mRNA, beginning within the Eco RI site in codon 122 and extending towards the 5' end of the mRNA through codon 105. The first 53 nucleotides of the 1000 base pair (5') fragment are identical to the β-globin sequence in all but 7 positions. Three changes occurring in one region, from $\overset{3'}{\underset{5'}{\text{GTAGTG}\atop\text{CATCAC}}}$ in the 950 bp fragment to $\overset{3'}{\underset{5'}{\text{GCGTTG}\atop\text{CGCAAC}}}$ in the 1000 bp fragment, correspond to the known replacement of amino acids 116 (histidine) and 117 (histidine) in β-globin with arginine and asparagine in these positions in δ-globin. These are the only amino acid differences in the two proteins between positions 105 and 122. The four other observed nucleotide changes represent "silent" nucleotide substitutions in redundant codons for leucine (position 106), asparagine (108), valine (111), and lysine (120). Thus, in agreement with previous assumptions, the 5' and 3' β-related globin genes are definitively identified as the δ- and β-globin genes, respectively. Inspection of Figures 6 and 7 also shows that the two sequences diverge beyond the 53rd nucleotide. This point of divergence represents the beginning of the non-coding intervening sequences in each gene (see below).

Non-coding Intervening Sequences in the δ- and β-Globin Genes

Studies of the rabbit (Flavell et al., 1978) and mouse (Tilghman et al., 1978a) β-globin genes have revealed the presence of non-coding DNA sequences (intervening

sequences) within the coding regions of the two genes. These intervening sequences do not appear in mature cytoplasmic β-globin mRNA but are present in nuclear RNA (Tilghman et al., 1978b). The presence of intervening sequences within the human β- and δ-globin genes was examined by comparing the locations of restriction endonuclease cleavage sites in cloned genomic DNA to the corresponding sites in the β-globin mRNA sequence.

The Eco RI and Bam HI sites in the β-globin mRNA sequence are separated by 67 base pairs of coding sequences (see Figure 9B for a map of these sites in β-globin cDNA) (Marrota et al., 1977). However, as indicated in the Eco RI and Bam HI maps of Figure 4 and as discussed above, these sites are separated by 950 bp in the coding region of the cloned β-globin DNA. The 950 bp separation of the Bam HI and Eco RI sites within the coding region of the β-globin gene indicates that the β-globin gene contains a large (∿900 bp) intervening sequence between these sites. Large (600-700 bp) intervening sequences are also found in the rabbit and mouse β-globin genes in approximately the same positions.

The precise location of the junction between the coding and intervening sequences near the Eco RI site in the human β-globin gene can be determined from the nucleotide sequence data presented in Figures 6 and 7. The first 53 nucleotides (corresponding to codons 122 through 105) are exactly complementary to the β-globin mRNA sequence. However, the nucleotide sequence at positions 54 through 81 is not complementary to the β-globin mRNA sequence indicating that the junction between the coding and large intervening sequences is located between codons 104 and 105 which specify the amino acids arginine and leucine.

The nucleotide sequence of δ-globin mRNA is not known. Therefore, the presence of an insert(s) in the coding sequence cannot be directly demonstrated

by examining the location of restriction endonuclease cleavage sites. However, the β- and δ-globin protein (Dayhoff, 1972) and nucleic acid (Comi et al., 1977) sequences are closely related and the δ-globin chromosomal gene also contains Bam HI and Eco RI recognition sites (Figure 4). If the locations of the Bam HI and Eco RI sites in the δ-globin message are analogous to those in β-globin mRNA (which is also consistent with the known amino acid sequence), then the separation of the Bam HI and Eco RI sites in the cloned δ-globin gene by approximately 1000 base pairs suggests that this region of the δ-globin coding sequence is also interrupted by approximately 950 base pairs of intervening sequence.

The presence of an intervening sequence near the Eco RI site in the δ-globin gene was formally demonstrated by a comparison of the known amino acid sequence of the δ-globin protein and the nucleotide sequence data presented in Figures 6 and 7. The first 53 nucleotides from the Eco RI site in the 5' direction encode the amino acid sequence of the δ-globin chain from amino acids 122 to 105. The next 9 codons do not correspond to the known amino acid sequence. Thus, an intervening sequence is present in the δ-globin gene, also located between codons 104 and 105. It is interesting to note that these codon positions also correspond to the amino acids arginine and leucine.

An additional intervening sequence was detected in the 5' region of the β-globin gene by comparing the sizes of Hae III restriction fragments in cloned genomic DNA and in the β-globin cDNA plasmid (Figure 8). The 1.8 kb Bam HI fragment F of HβG2 DNA, which contains the portion of the β-globin gene 5' to the internal Bam HI site, was isolated, digested with Hae III, and electrophoresed on 5% polyacrylamide gel. As a control, the Bam HI/Hha I fragment of JW102, containing the portion of the β-globin cDNA plasmid 5' to the Bam HI site in the

message sequence (see Figure 5), was also digested with Hae III and coelectrophoresed.

The Hae III digest of the Bam HI/Hha I fragment of JW102 produced, among others, a 141 base pair fragment (B), spanning the Hae III sites at codon positions 75 and 26 (see Figure 9B for a map of the Hae III sites in β-globin cDNA). Digestion of the 1.8 kb Bam HI fragment F from HβG2 DNA did not produce a fragment identical in mobility to this 141 bp fragment. The absence of a 141 bp fragment from the 1.8 kb HβG2 DNA indicates there is an intervening sequence in the cloned genomic β-globin DNA between the Hae III sites at codons 75 and 26. Similarly, a Hinf fragment, 120 bp long, spanning the region between codons 4 and 44 (Wilson et al., 1978) is expected but not observed in the cloned DNA (unpublished results). This result in combination with the Hae III data locates an intervening sequence between codons 26 and 44. (The possibility that both the Hae III and Hinf sites found in the cloned mRNA sequence are missing in the cloned genomic β-globin DNA due to genetic polymorphism is unlikely but is not ruled out by this experiment.) The size of the intervening sequence could not be determined because the possible presence of additional Hae III and Hinf cleavage sites within the intervening sequence could not be ascertained. Nucleotide sequencing of the 5' region of the cloned human β-globin gene can be used to establish the size and exact location of the intervening sequence. A small intervening sequence has also been detected in a similar location in both the mouse and rabbit β-globin genes (see Discussion).

In similar experiments, no further intervening sequences were detected in the human β-globin gene (unpublished results). Similar experiments were not performed on the cloned δ-globin gene, due to the unavailability of the δ-globin

mRNA sequence. A detailed map of the globin gene containing region in HβG2 DNA, including the location of non-coding intervening sequences, is presented in Figure 9A.

Heterozygosity in the δ-Globin Genes in HβG1 and HβG2

An interesting difference in the Pst I cleavage pattern of HβG1 and HβG2 DNA was observed. As shown in Figure 10, digestion of HβG1 DNA with Pst I yields two fragments which hybridize the β-globin probe (Pst B and Pst G). However, when HβG2 DNA is similarly digested, three hybridizing fragments are produced (Pst B, J and M) only one of which (Pst B) comigrates with a Pst I fragment from HβG1 DNA. The combined sizes of the smaller hybridizing fragments from HβG2 DNA [Pst J (1.35 kb) and Pst M (0.95 kb)] are equal to the size of the HβG1-Pst G fragment (2.3 kb). This result indicates that either the δ- or β-globin gene in HβG2 DNA is cleaved by Pst I.

A Pst I restriction map of HβG2 DNA was derived to distinguish which of the two genes was cleaved. As shown in Figure 11, the Pst B fragment contains the entire β-globin gene while the Pst J and M fragments each contain part of the δ-globin gene. The Pst I map of HβG1 DNA is identical to that of HβG2 except that the Pst I site between Pst J and Pst M is absent (indicated by an arrow in Figure 11).

To determine if the Pst I site in the δ-globin gene of HβG2 DNA is within the coding or non-coding sequences of the gene, the Bam HI/Eco RI 1000 bp fragment containing the large non-coding intervening sequence of the δ-globin gene was isolated from both clones and digested with Pst I. As shown in Figure 12, the Bam HI/Eco RI fragment from HβG2 DNA is cut by Pst I while the HβG1 derived fragment is not. The sizes of the two fragments produced (530 bp and 470 bp)

indicates that the Pst I site is located near the center of the Bam HI/Eco RI fragment, presumably within the intervening sequence. Thus, it appears that the human DNA fragments in HβG1 and HβG2 were derived from homologous chromosomes which are heterozygous with respect to the Pst I site in the intervening sequence of the δ-globin gene. This heterozygosity is confirmed by the fact that all four Pst I fragments which hybridize to the β-globin probe in HβG1 and HβG2 are also observed in genomic blots of the fetal liver DNA which was used in the construction of the library (manuscript in preparation).

Comparison of the Large Intervening Sequences in the β- and δ-Globin Genes

Cross hybridization experiments were performed to determine whether the large intervening sequences in the β- and δ-globin genes are homologous. The Bam HI/Eco RI fragments containing the β- or δ-globin intervening sequences and 67 bp of coding sequence (Figure 9) were purified by polyacrylamide gel electrophoresis, and labeled by nick-translation for use as hybridization probes. These probes were separately hybridized to the filter-bound DNA of a Pst I digestion of HβG1 DNA. The Pst B and G fragments from HβG1 contain, respectively, the entire β- and δ-globin gene (Figure 11). To minimize cross hybridization between the 67 nucleotides of homologous coding sequence carried in the Bam HI/Eco RI fragments from each gene, the filters were washed under stringent conditions after the hybridization.

The Eco RI/Bam HI fragment containing the β-globin intervening sequence hybridized efficiently to the Pst B fragment, which contains the β-globin gene, and did not hybridize to the Pst G fragment, which contains the δ-globin gene (Figure 13). Conversely, the Eco RI/Bam HI fragment containing the δ-globin

intervening sequence hybridized well to the Pst G fragment and only weakly to the Pst B fragment. The lack of significant hybridization between the Eco RI/Bam HI fragment containing the β- or δ-globin intervening sequence and the Pst I fragment containing the opposite gene is consistent with the nucleotide sequence data of Figures 6 and 7. Thus, it appears that the two intervening sequences have not been conserved in evolution.

Discussion

The linkage arrangement of human δ- and β-globin genes predicted by genetic analysis and structural studies of mutant globins (Weatherall and Clegg, 1972; Baglioni, 1962; Huisman et al., 1972) has now been verified at the molecular level by detailed characterization of a cloned human DNA fragment carrying both genes. Unambiguous identification of the δ- and β-globin genes in the linked gene complex was accomplished by partial nucleotide sequence analysis of restriction fragments which map within each of the genes.

The structure and organization of these genes have also been studied using the procedure of genomic blotting (Southern, 1975; Botchan et al., 1976; Jeffreys and Flavell, 1977a,b) to map restriction sites within and surrounding the two genes in normal and Hb Lepore DNA (Mears et al., 1978; Flavell et al., 1978; unpublished results from this lab). In one of these studies (Flavell et al., 1978), the physical linkage, approximate intergene distance and relative orientation of the δ- and β-globin genes described here were independently determined. The sizes of various globin-containing restriction fragments in HβG2 DNA are in agreement with those derived from our genomic blotting experiments (manuscript in preparation). Slightly different sizes were derived from other genomic blotting

experiments (Mears et al., 1978; Flavell et al., 1978). However, the relative positions of the fragments in the proposed restriction maps are identical. We can therefore conclude that the organization of the δ- and β-globin genes in HβG1 and HβG2 has not been altered by cloning and amplification in the bacteria.

The significance of the tandem arrangement of the two genes with regards to their coordinate and/or sequential activation during development is not known. Analysis of the types of hemoglobin tetramers present at different times during development indicates that small amounts of the major adult hemoglobin [HbA ($\alpha_2 \beta_2$)] can be detected as early as 6-8 weeks of gestation, but that substantial amounts are not detected until birth (Hollenberg et al., 1971; Pataryas and Stamatoyannopoulos, 1972; Wood and Weatherall, 1973). The minor adult hemoglobin [HbA_2 ($\alpha_2 \delta_2$)] is detected in small amounts at 6 months of gestation and increases to the normal level of 2.5% of the total hemoglobin at 6 months to one year after birth (Bunn et al., 1977). The δ-globin chain is synthesized in immature red cells during adult erythropoiesis, but, in contrast to the β-globin chain, is not produced in mature reticulocytes (Roberts, Weatherall and Clegg, 1972). The level at which the control of adult δ- and β-globin synthesis is mediated is not known since only the end products of the genetic activation could be studied. Using well-defined segments of cloned δ- and β-globin genes as hybridization probes, it will now be possible to determine whether the activation of the two genes during embryonic and adult red cell development occurs at the level of transcription, post-transcriptional processing or translation.

A large, non-coding intervening sequence of approximately 600-700 bp was found in the mouse (Tilghman et al., 1978a) and rabbit (Jeffreys and Flavell, 1977b) β- globin genes. We and others (Flavell et al., 1978; Mears et al., 1978)

have located a larger (900 bp) intervening sequence in the human β-globin gene. By analogy to the β-globin gene, a similar insert was tentatively located in the δ-globin gene. The nucleotide sequence of Figures 6 and 7 verifies this prediction. Nucleotide sequence analysis of the large intervening sequence in mouse (Tilghman et al., 1978), rabbit (A. Efstratiadis, E. Lacy and T. Maniatis, unpublished results), and human δ- and β-globin genes indicate that the non-coding insert is, in every case, located between the codons for amino acids 104 (Arg) and 105 (Leu). Examination of the amino acid sequences of a large number of mammalian β-globin genes (Dayhoff, 1976) reveals that this region of the amino acid sequence is highly conserved in evolution. For example, the amino acids Arg/Leu/Leu are found in position 104 through 106 in 14 out of 17 of the mammalian species examined. The three β-related globins which do not fit this pattern (human γ, bovine β, and horse β) contain Lys/Leu/Leu in these three positions, a sequence which could result from a single base change from the arginine codon (AG A/G) to the lysine codon (AA A/G). In contrast to the location of the insert, the majority of nucleotides within the intervening sequences of the β-related mammalian globin genes are not highly conserved (Tiemeier et al., 1978; Figures 6 and 7). There appears to be only slight sequence homology between the intervening sequences in the mouse β_{major}- and β_{minor}-globin genes and we detect little, if any, cross hybridization between the large intervening sequences in the human δ- and β-globin genes. This is consistent with the fact that many of the nucleotides at the 3' end of the large intervening sequences are different (Figures 6 and 7). In addition, some divergence has occurred in the large intervening sequence in the δ-globin genes on homologous chromosomes (Figure 12). Similar polymorphism in the intervening sequence of the ovalbumin gene has been observed by Weinstock et al. (1978) and

Garapin et al. (1978). It appears that the only sequences which are conserved are those which lie near the junction between the coding and non-coding sequences within the gene (Figures 6 and 7; Tiemeier et al., 1978).

A second, smaller intervening sequence was detected in the mouse β_{major}-globin gene, located in the 5' direction from the large intervening sequence (Tilghman et al., 1978a). A similar interruption has been detected in the rabbit β-globin gene (A. Efstratiadis, E. Lacy and T. Maniatis, unpublished results). The non-coding intervening sequence in the rabbit gene is 126 bp long and located between the codons for amino acids 30 (Arg) and 31 (Leu). Restriction endonuclease cleavage analysis of the human β-globin gene described here tentatively identifies an intervening sequence located between the codons for amino acids 26 and 44.

The significance of this conservation of position of the two intervening sequences in β-globin genes between mammalian species may be related in some way to nuclear RNA splicing and processing. Tilghman et al. (1978b) have shown that the entire coding and non-coding sequences in both of the mouse β-globin genes are present in a 15S nuclear RNA globin precursor, indicating that both intervening sequences are transcribed and, presumably, cut and spliced to produce mature mRNA.

Hemoglobin Lepore (Boston-Washington) is characterized by a fusion of the δ- and β-globin peptides somewhere between codons 87 and 116. The large intervening sequence in Hb Lepore (Boston-Washington) DNA (Mears et al., 1978) could therefore be derived either from the δ- or β-globin gene or from a combination of both. The lack of homology between the intervening sequences of the two genes observed here argues against, but does not eliminate, the possibility that an unequal crossing over event within the intervening sequence produced the fused gene.

The existence of well-defined functional deficiencies in human hemoglobin expression (thalassemias) provides the opportunity for studying relationships between globin gene organization and function (Clegg and Weatherall, 1976). The availability of cloned genomic globin DNA with its associated non-coding sequences will extend the utility of this approach by making it possible to detect structural changes in sequences not accessible to cDNA probes. For example, the cloned δ- and β-globin intervening sequences can be used to identify the origin of the intervening sequence in Hb Lepore DNA. Once structural differences in DNA from any thalassemic individual have been detected by genomic blotting, the genetic defect can be studied at the level of nucleotide sequence by cloning the affected gene. The gene isolation procedure described should allow the rapid isolation of mutant globin genes. Relatively small amounts of DNA are required and the procedure is not effected by changes in restriction endonuclease cleavage sites which flank the affected genes since randomly fragmented DNA, rather than a specific restriction fragment, is cloned. Thus, thalassemias resulting from large deletions or sequence rearrangements as well as those resulting from single base changes will be accessible to structural analysis. Studies of this kind should identify the molecular basis of thalassemias and possibly provide insight into human genetic disorders in general.

Experimental Procedures

Construction of a Human Genomic DNA Library and Isolation of β-globin Clones

Detailed procedures used in the construction of the human library and the isolation of recombinant clones containing globin genes are identical to those previously described (Maniatis et al., 1978). A brief outline of the procedure has been presented in the Results section and selected data are presented in Table 1. Human fetal

liver DNA prepared according to the procedure of Blin and Stafford (1976) was kindly provided by B. Forget. Synthetic Eco RI linkers were a gift of C. K. Itakura, and Eco RI methylase was provided by J. Rosenberg.

The DNA was fragmented by performing a non-limit digestion with the enzymes Hae III + Alu I and fragments in the size range 18-25 kb were prepared by sucrose gradient centrifugation.

Approxiamtely 300,000 phage recombinants were screened as previously described (Maniatis et al., 1978). Two recombinants containing β-globin sequences were identified and plaque purified. Recombinant phage were grown and the DNA purified as previously described (Maniatis et al., 1978).

Preparation of in vitro Labeled DNA Hybridization Probes

The human β-globin cDNA plasmid pJW102 (Wilson et al., 1978) was provided by J. Wilson, L. Wilson and B. G. Forget. Plasmid DNA was prepared as described by Godson and Vapnek (1973). Hha I digested pJW102 DNA was labeled in vitro with ^{32}P by nick-translation (Maniatis, Jeffrey and Kleid, 1975) using high specific activity α-^{32}P deoxynucleotide triphosphates (~300 ci/mmole) purchased from New England Nuclear or Amersham. Specific activities of 2-20 x 10^7 cpm/μg were obtained. E. coli DNA polymerase I was purchased from Boehringer Mannheim and DNAse I from Worthington Biochemicals.

DNA Sequence Analysis

50 μg of HβG2 DNA was digested with Eco RI, and end labeled using the T4 polynucleotide kinase exchange reaction conditions of Berkner and Folk (1977). T4 polynucleotide kinase was purchased from PL Biochemicals. γ-^{32}P ATP was prepared by the method of Glynn and Chapell (1964) as described by Maxam and

Gilbert (1977) using ^{32}Pi purchased from ICN Radiochemical Corporation.

The end labeled DNA was extracted with phenol, ether extracted and ethanol precipitated. The precipitate was dissolved in Bam HI restriction enzyme buffer and the DNA digested with Bam HI. The digestion products were fractionated in a 20 cm/40 cm/.2 cm polyacrylamide slab gel, the appropriate DNA fragments identified by autoradiography and eluted from the gel (Maniatis, Jeffrey and van de Sande, 1975). Following ethanol precipitation, the labeled DNA was subjected to non-limit, base specific chemical degradation as described by Maxam and Gilbert (1977) and the samples divided and electrophoresed in 15% and 20% polyacrylamide gels containing 7 M Urea.

Restriction Endonuclease Digestion and Agarose Gel Electrophoresis

DNA was digested with restriction endonucleases Hha I, Hind III, and Pst I in 60 mM NaCl, 10 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 1 mM DTT, and 100 µg/ml BSA. Bam HI and Kpn I digestions were performed in the same buffer except that the NaCl concentration was 150 mM. Digestions with Bgl I were done in 10 mM Tris (pH 7.4), 10 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, and 100 µg/ml BSA. Eco RI digestions were carried out in 100 mM Tris (pH 7.4), 50 mM NaCl, 10 mM $MgCl_2$, 6 mM β-mercaptoethanol, and 100 µg/ml BSA. Restriction endonuclease digestions were incubated at 37°C for one hour except the temperature was changed with Pst I (23°C) and Hind III (55°C). Electrophoresis in agarose gels was carried out as described by Sharp et al. (1973).

Two-dimensional Electrophoresis with SeaPlaque Agarose

Isolation of restriction endonuclease fragments of DNA was carried out according to Parker and Seed (1979). Low melting temperature agarose gels were made

from SeaPlaque agarose (Marine Colloids, Inc.). Horizontal gels of 0.6% or 0.8% (w/w) were made by dissolving the agarose in E buffer (40 mM Tris, 5 mM NaOAc, 1 mM EDTA adjusted to pH 7.4 with glacial acetic acid) by heating to at least 70°C. The agarose was cooled to 37°C before the gel was cast. Gels were formed and run at 4°C. Depending upon the experiment, some gels contained 0.5 μg/ml of ethidium bromide (EtdBr). Gels without EtdBr were stained in a solution of EtdBr (2 μg/ml) for 10 min and destained in distilled H_2O for at least 10 min. Bands were visualized using either short- or long-wave ultraviolet light. DNA bands were excised from the gel with a razor blade and stored at -20°C.

Low melting temperature agarose melts at 65°C and remains in solution at 37°C. DNA fragments excised from SeaPlaque gels were incubated at 70°C to melt the agarose without denaturing the DNA. One-quarter volume of a five times concentrated restriction endonuclease buffer solution was added and the sample was allowed to equilibrate to 37°C. After equilibration, nuclease-free BSA (Pentex) was added to a final concentration of 1 mg/ml. Restriction endonuclease was added (approximately 2.5 units of enzyme per μg of DNA; 1 unit of enzyme digests 1 μg of λ DNA to completion in 60 min) and the samples were incubated for 2 hr. Reactions were terminated by cooling the samples to -20°C.

The products of the second restriction endonuclease digestion were fractionated by agarose gel electrophoresis. Both horizontal and vertical gels were used in the second dimension. DNA samples, still in a solution of SeaPlaque agarose, were layered directly into the wells of the second gel. When horizontal gels were used, the SeaPlaque agarose was allowed to solidify in the wells before electrophoresis began. When vertical gels were used, the SeaPlaque agarose was maintained in solution by pre-heating the upper reservoir buffer to 55°C. Electrophoresis began immediately after the samples were layered into the wells.

Blotting and Hybridization Experiments

DNA in agarose and acrylamide gels was transferred to nitrocellulose filter paper (0.22 μm and 0.45 μm, Millipore) and hybridized to radioactive probes by a modification of the procedures developed by Southern (1977) and Jeffreys and Flavell (1977b). The transfer buffer used was 10 x SSC (1 x SSC equals 0.15 M NaCl, 0.015 M Na_3 Citrate, pH 7.0). After the transfer, the filter paper was dried at 68°C for greater than 2 hours and baked, under vacuum, at 80°C for at least 8 hours. Filters were soaked in hybridization buffer (10 x Denhardt's solution, 0.1% SDS, 5 x SSC, 10 mM NaPPi, .125 M Na_2PO_4, pH 8) at 68°C for 4 hr before the denatured radioactive probe was added. Hybridizations were carried out at 68°C, with constant mixing for at least 10 times the $t_{1/2}$ of hybridization in solution. After hybridization, the filters were washed in 50 mM NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA at 68°C (except where noted in figure legends). Filters were used to expose Kodak X-Omat X-ray film. When necessary, Du Pont lightning-plus intensifying screens were used to enhance the signals.

Recombinant DNA Safety Procedures

The construction, screening, and propagation of recombinant bacteriophage was conducted in accordance with the NIH Guidelines for recombinant DNA research. The experiments were performed in a P3 facility using the EK2 vector host system λ Charon 4A/DP50SupF.

Acknowledgements

We are grateful to Bernard Forget for advice and discussion, and for providing fetal liver DNA. We thank J. Wilson, L. Wilson, B. Forget, and S. Weissman for

providing human globin cDNA clones and C. O'Connell for help in constructing and screening the human DNA library. We also thank J. Posakony for help with the DNA sequencing, J. Maniatis for assistance in preparing graphics, and A. Cortenbach for preparing media and materials.

The work was supported by a grant from the National Institutes of Health and by funds from a Biomedical Research Support grant to the California Institute of Technology. R.M.L. and E.F.F. were supported by postdoctoral fellowships from the American Cancer Society and the Damon Runyon-Walter Winchell Cancer Fund, respectively. T.M. is a recipient of a Rita Allen Foundation Career Development Award.

References

Baglioni, C. (1962). The fusion of two polypeptide chains in hemoglobin Lepore and its interpretation as a genetic deletion. Proc. Nat. Acad. Sci. USA 48, 1880-1885.

Berkner, K. L. and Folk, W. R. (1977). Polynucleotide kinase exchange reaction. J. Biol. Chem. 252, 3176-3184.

Benton, W. D. and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196, 180-182.

Blattner, F. R., Williams, B. G., Blechl, A. E., Thompson, K. D., Faber, H. E., Furlong, L. A., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., and Smithies, O. (1977). Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. Science 196, 161-169.

Blin, N. and Stafford, D. W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res. 3, 2303-2308.

Botchan, M., Topp, W. and Sambrook, J. (1976). The arrangement of SV40 sequences in the DNA of transformed cells. Cell 9, 269-287.

Bunn, F. H., Forget, B. G., and Ranney, H. M. (1977). Human Hemoglobins. W. B. Saunders Company, Philadelphia.

Clarke, L., and Carbon, J. (1976). A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire E. coli genome. Cell 9, 91-99.

Clegg, J. B. and Weatherall, D. J. (1976). Molecular basis of thalassaemia. Br. Med. Bull. 32, 262-269.

Comi, P., Giglioni, B., Barbarano, L., Ottolenghi, S., Williamson, R., Novakova, M., and Masera, G. (1977). The transcriptional and post-transcriptional defects in β^{o}-thalassemia. Europ. J. Biochem. 79, 617-622.

Dayhoff, M. O. (1972). Atlas of Protein Sequence and Structure. Washington, D. C., National Biomedical Research Foundation.

Dayhoff, M. O. (1976). Atlas of Protein Sequence and Structure. Vol. 5 Supplement. Washington, D. C., National Biomedical Research Foundation.

DeJong, W. W. W., Khan, P. M. and Bernine, L. F. (1975). Hemoglobin Koya Dora: High frequency of a chain termination mutant. Am. J. Hum. Genet. 27, 81-90.

Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W. F., Ruddle, F., Lawrence, J., Creagen, R., and Kucherlapate, R. (1977). Localization of the human alpha globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. Cell 12, 205-218.

Deisseroth, A., Nienhuis, A., Lawrence, J., Giles, R., Turner, P. and Ruddle, F. (1978). Chromosomal location of human β globin gene on human chromosome 11 in somatic cell hybrids. Proc. Nat. Acad. Sci. USA 75, 1456-1460.

Flavell, R. A., Kooter, J. M., De Boer, E., Little, P. F. R. and Williamson, R. (1978). Analysis of the human β-δ globin gene loci in normal and Hb Lepore DNA: Direct determination of gene linkage and intergene distance. Cell (in press).

Garapin, A. C., Cami, B., Rosham, W., Kourilsky, P., LePennec, J. P., Perrin, F., Gerlinger, P., Cochect, M., and Chambon, P. (1978). Electron microscopy and restriction enzyme mapping reveal additional intervening sequences in the chicken ovalbumin split gene. Cell 14, 629-639.

Glynn, I. M. and Chapell, J. B. (1964). A simple method for the preparation of ^{32}P-labeled adenosine triphosphate of high specific activity. Biochem. J. 90, 147-149.

Godson, G. N. and Vapnek, D. (1973). A simple method for preparing large amounts of ØX174 RF I supercoiled DNA. Biochem. Biophys. Acta 299, 516-520.

Hollan, S. R., Szelenyi, J. G., Brimball, B., Duerst, M., Jones, R. T., Koler, R. D. and Stockleniz, A. (1972). Multiple alpha chain loci for human haemoglobins: Hb J-Buda and Hb G-Pest. Nature 235, 47-50.

Hollenberg, M. D., Kalback, M. M. and Kazazian, H. H. (1971). Adult hemoglobin synthesis by reticulocytes from the human fetus at midtrimester. Science 174, 698-702.

Huisman, T. H. J., Wrightstone, R. N., Wilson, J. B., Shroeder, W. A. and Kendall, A. G. (1972). Hemoglobin Kenya, the product of fusion of δ and β polypeptide chains. Arch. Biochem. Biophys. 153, 850-853.

Jeffreys, A. J. and Flavell, R. A. (1977a). A physical map of the DNA regions flanking the rabbit β globin gene. Cell 12, 429-439.

Jeffreys, A. J. and Flavell, R. A. (1977b). The rabbit β-globin gene contains a large insert in the coding sequence. Cell 12, 1097-1108.

Kendall, A. G., Ojwang, P. J., Schroeder, W. A. and Huisman, T. H. J. (1973). Hemoglobin Kenya, the product of a δ-β fusion gene: Studies of the family. Am. J. Human Gen. 25, 548-563.

Lacy, E., Lawn, R. M., Fritsch, E., Hardison, R. C., Parker, R. C. and Maniatis, T. (1978). Isolation and characterization of mammalian globin genes. In Cellular and Molecular Regulation of Hemoglobin Switching, G. Stamatoyannopoulos and A. Nienhuis, eds. (New York: Grune and Stratton). In press.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975a). Nucleotide sequence of the rightward operator of phage λ. Proc. Nat. Acad. Sci. USA 72, 1184-1188.

Maniatis, T., Jeffrey, A. and Van de Sande, H. (1975b). Chain length determination of small double and single stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry 14, 3787-3794.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eukaryotic DNA. Cell, in press.

Marrota, C. A., Wilson, J. T., Forget, B. G. and Weissman, S. M. (1977). Human β-globin messenger RNA. III. Nucleotide sequences derived from complementary DNA. J. Biol. Chem. 252, 5040-5053.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA 74, 560-564.

Mears, J. G., Ramirez, F., Leibowitz, D. and Bank, A. (1978). Organization of human δ and β globin genes in cellular DNA and the presence of intragenic inserts. Cell, in press.

Nienhuis, A. W. and Benz, E. J. (1977). Regulation of hemoglobin synthesis during development of the red cell. New Eng. J. Med. 297, 1318-1328, 1371-1381, 1430-1436.

Parker, R. and Seed, B. (1979). Endonuclease cleavage mapping techniques: Two dimensional agarose gel electrophoresis "SeaPlaque" agarose dimension. In Methods in Enzymology, L. Grossman and K. Moldave, eds. (New York: Academic Press). In press.

Parker, R. C., Watson, R. M. and Vinograd, J. (1977). Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. Proc. Nat. Acad. Sci. USA 74, 851-855.

Pataryas, H. A. and Stamatoyannopoulos, G. (1972). Hemoglobin in human fetuses; evidence for adult hemoglobin production after the 11th gestational week. Blood 39, 688-696.

Roberts, A. V., Weatherall, D. J. and Clegg, J. B. (1972). The synthesis of human hemoglobin A2 during erythroid maturation. Biochim. Biophys. Res. Comm. 47, 81-87.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M. and Smith, M. (1977). Nucleotide sequence of bacteriophage ØX174 DNA. Nature 265, 687-695.

Sharp, P. A., Sugden, B. and Sambrook, J. (1973). Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose ethidium bromide electrophoresis. Biochemistry 12, 3055-3063.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Sternberg, N., Tiemeier, D. and Enquist, L. (1977). In vitro packaging of a λ Dam vector containing EcoRI DNA fragments of E. coli and phage P1. Gene 1, 255-280.

Thomas, M. and Davis, R. (1975). Studies on the cleavage of bacteriophage lambda DNA with Eco RI restriction endonuclease. J. Mol. Biol. 91, 315-328.

Tiemeier, D. C., Tilghman, S. M., Polsky, F. I., Seidman, J. G., Leder, A., Edgell, M. H. and Leder, P. (1978). A comparison of two cloned mouse β globin genes and their surrounding and intervening sequences. Cell 14, 237-245.

Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. and Leder, P. (1978a). Intervening sequence of DNA identified in the structural portion of a mouse β-globin gene. Proc. Nat. Acad. Sci. USA 75, 725-729.

Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P., and Weissman, C. (1978b). The intervening sequence of a mouse β globin gene is transcribed within the 15S β globin mRNA precursor. Proc. Nat. Acad. Sci. USA 75, 1309-1313.

Weatherall, D. J. and Clegg, J. B. (1972). The Thalassaemia Syndromes. 2nd Edition. Oxford, Blackwell Scientific Publications.

Weinstock, R., Sweet, R., Weiss, M., Cedar, H. and Axel, R. Intragenic DNA spacers interrupt the ovalbumin gene. (1978). Proc. Nat. Acad. Sci. USA 75, 1299-1303.

Wellauer, P. K., Reeder, R. H., Carroll, D., Brown, D. D., Deutch, A., Higashinakagawa, T. and Dawid, I. B. (1974). Amplified ribosomal DNA from Xenopus laevis has heterogeneous spacer lengths. Proc. Nat. Acad. Sci. USA 71, 2823-2827.

Wilson, J. T., Wilson, L. B., deRiel, J. K., Villa Komaroff, L., Efstratiadis, A., Forget, B. G. and Weissman, S. M. (1978). Insertion of synthetic copies of human globin genes into bacterial plasmids. Nucleic Acids Res. 5, 563-581.

Wood, W. G. and Weatherall, D. J. (1973). Haemoglobin synthesis during human fetal development. Nature 244, 162-165.

Table 1.

Molar ratio of Charon 4A arms to human DNA in ligation reaction[a]	0.6/1
Packaging efficiency of extracts for intact Charon 4A DNA[b]	5×10^6 pfu/μg
Packaging efficiency for human DNA[c]	2.5×10^4 pfu/μg human DNA
Background of nonrecombinant phage DNA packaged[d]	∿1%
Total number of independent recombinant phage recovered[e]	1×10^6
Number of recombinant phage required for a "complete library"[f]	8×10^5

[a]Conditions for the ligation reaction have been described in detail (Maniatis et al., 1978). The ligation reaction contained 40 μg of λ Ch4A arms and 40 μg of human DNA prepared for cloning as briefly described in the text (see also Maniatis et al., 1978) in a volume of 340 μl. The molar ratio of λ Ch4A DNA to human DNA was calculated assuming sizes of 31 kb for the λ Ch4A arms and 20 kb for the human DNA.

[b]Extracts for in vitro packaging of DNA were prepared as described by Sternberg et al. (1977) and modified for use in EK2 level recombinant DNA experiments as described by Maniatis et al. (1978). A portion of each extract was used to

test the packaging efficiency of intact λ Ch4A DNA. 0.35 μg of λ Ch4A DNA was packaged and 1.7×10^6 plaque-forming units of λ Ch4A were recovered.

[c]Fourteen separate packaging reactions were used to package the products of ligation reaction described above (footnote a). Following the packaging reaction, the in vitro packaged phage were pooled and purified on a CsCl step gradient. 1.08×10^6 plaque-forming units of packaged phage were recovered from the gradient.

[d]The percentage of non-recombinant phage in the in vitro packaged phage was determined by testing the phage for the lac 5 function as described by Blattner et al. (1977). The percentage of blue plaques did not change significantly following amplification of the packaged phage.

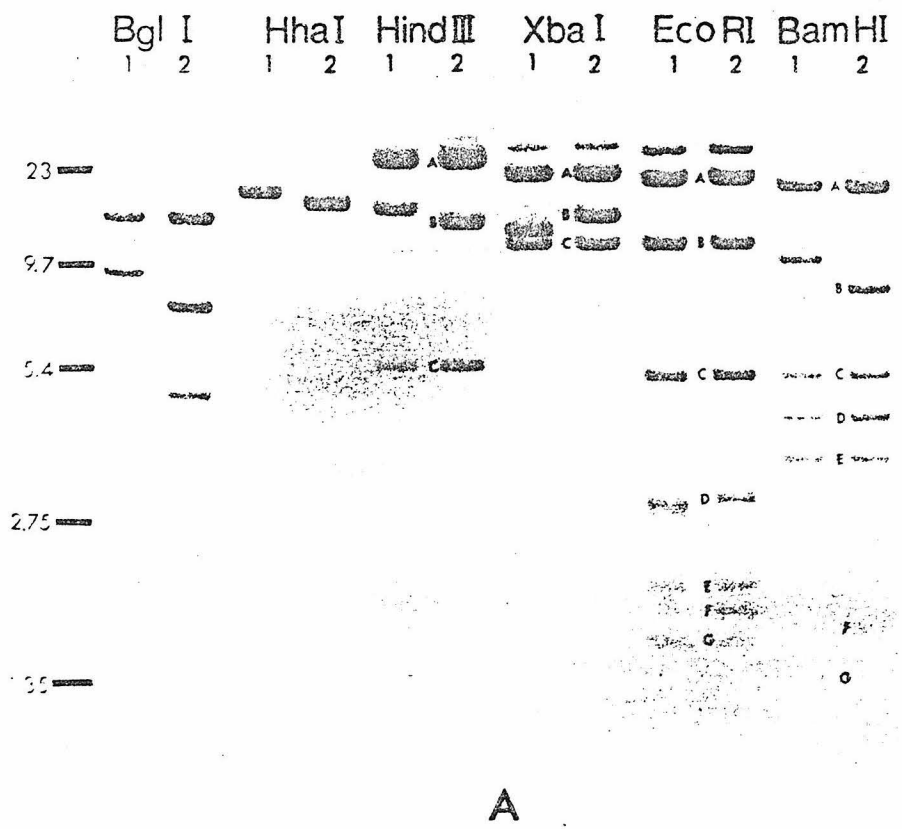
[e]The total number of independent recombinant phage was determined from the total number of in vitro packaged phage recovered and the percentage of non-recombinant phage.

[f]Calculated as described by Clarke and Carbon (1976) using 20 kb as the average length for the inserted human DNA and 3×10^9 bp for the human genome size. The average length of human DNA inserts was not determined experimentally but represents the average size of the human DNA before ligation with λ Ch4A DNA. A "complete library" is defined as a library having a $\geq$ 99% probability of containing any sequence present in the genome.

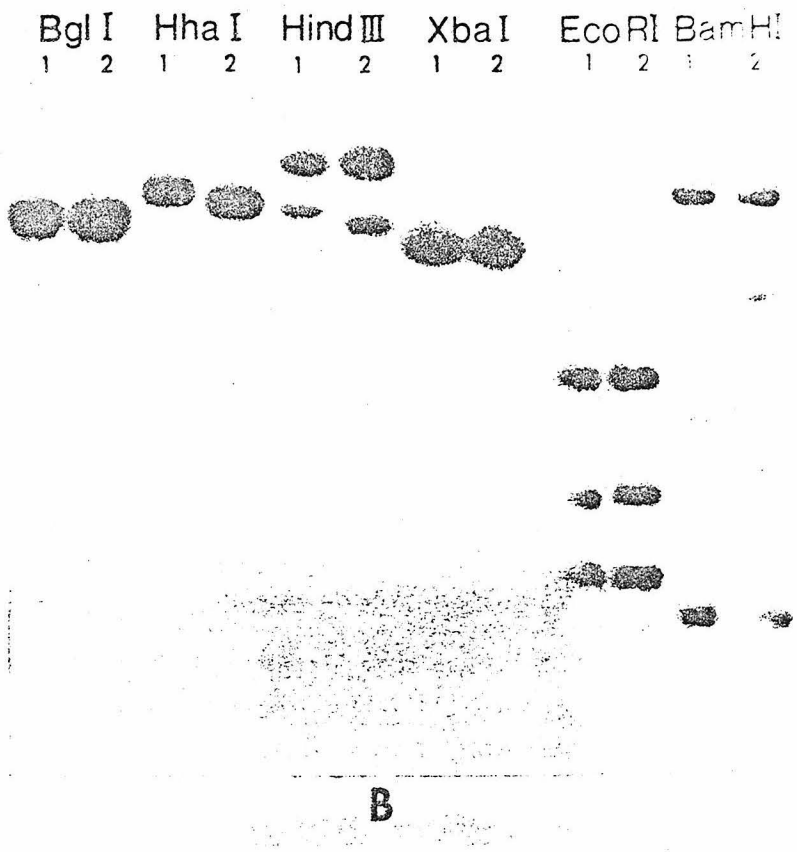
Figure 1. Restriction Endonuclease Cleavage Analysis of Cloned Human DNA and Identification of Fragments Containing β-Globin Related Sequences

(A) HβG1 (1) and HβG2 (2) DNA were digested with various restriction endonucleases and the products were electrophoresed in an 0.7% agarose gel and visualized by staining with ethidium bromide (EtdBr). The fainter bands in the Eco RI, Bam HI, and Xba I digests result from limited annealing of the cohesive ends of bacteriophage λ DNA. Two small Bam HI fragments (0.7 and 0.5 kb) were run off the gel. Restriction endonuclease digestion products of ØX174 DNA (Sanger et al., 1977) and λ DNA (Parker et al., 1977; Wellauer et al., 1974; and Thomas and Davis, 1975) were coelectrophoresed as size standards. The position and size (in kb) of some of the markers are indicated. Letters refer to fragments of HβG2 DNA discussed in the text.

(B) The DNA from the gel in Figure 1A was transferred to nitrocellulose by the method of Southern (1975) and hybridized with ^{32}P-labeled pJW102 DNA (specific activity 5 x 10^7 cpm/μg) as described in Experimental Procedures. Incomplete transfer reduced the width of bands in the Eco RI digest of HβG1. The faintly hybridizing bands observed in the Bgl I digests were assumed to be due to contaminating endonuclease activity and were not examined further.


Bgl I
Hha I
Hind III
Xba I
Eco RI
Bam HI
1 2
23
9.7
5.4
2.75
A
A
B
C
D
E
F
G

A


Bgl I
Hha I
Hind III
Xba I
Eco RI
Bam HI
1 2

B

Figure 2. Mapping Eco RI Sites in HβG2 DNA

The Xba B and Hind B fragments of HβG2 DNA were isolated from a 0.6% SeaPlaque gel and redigested with Eco RI. The Eco RI digestion products were electrophoresed on a 1.15% agarose gel and the DNA was visualized by staining with EtdBr. Sizes of the lettered Eco RI fragments are indicated in kb.

Lane 1: HβG2 DNA digested with Eco RI.

Lane 2: The Xba B fragment of HβG2 DNA digested with Eco RI.

Lane 3: HβG2 DNA digested with Eco RI and Xba I.

Lane 4: HβG2 DNA digested with Eco RI.

Lane 5: The Hind B fragment of HβG2 DNA digested with Eco RI.

Lane 6: HβG2 DNA digested with Eco RI and Hind III.

A 0.6 kb fragment produced by Eco RI digestion of the Xba B fragment is too faint to be seen in Lane 2. Lanes 1-3 and 4-6 are from different experiments performed under similar conditions.

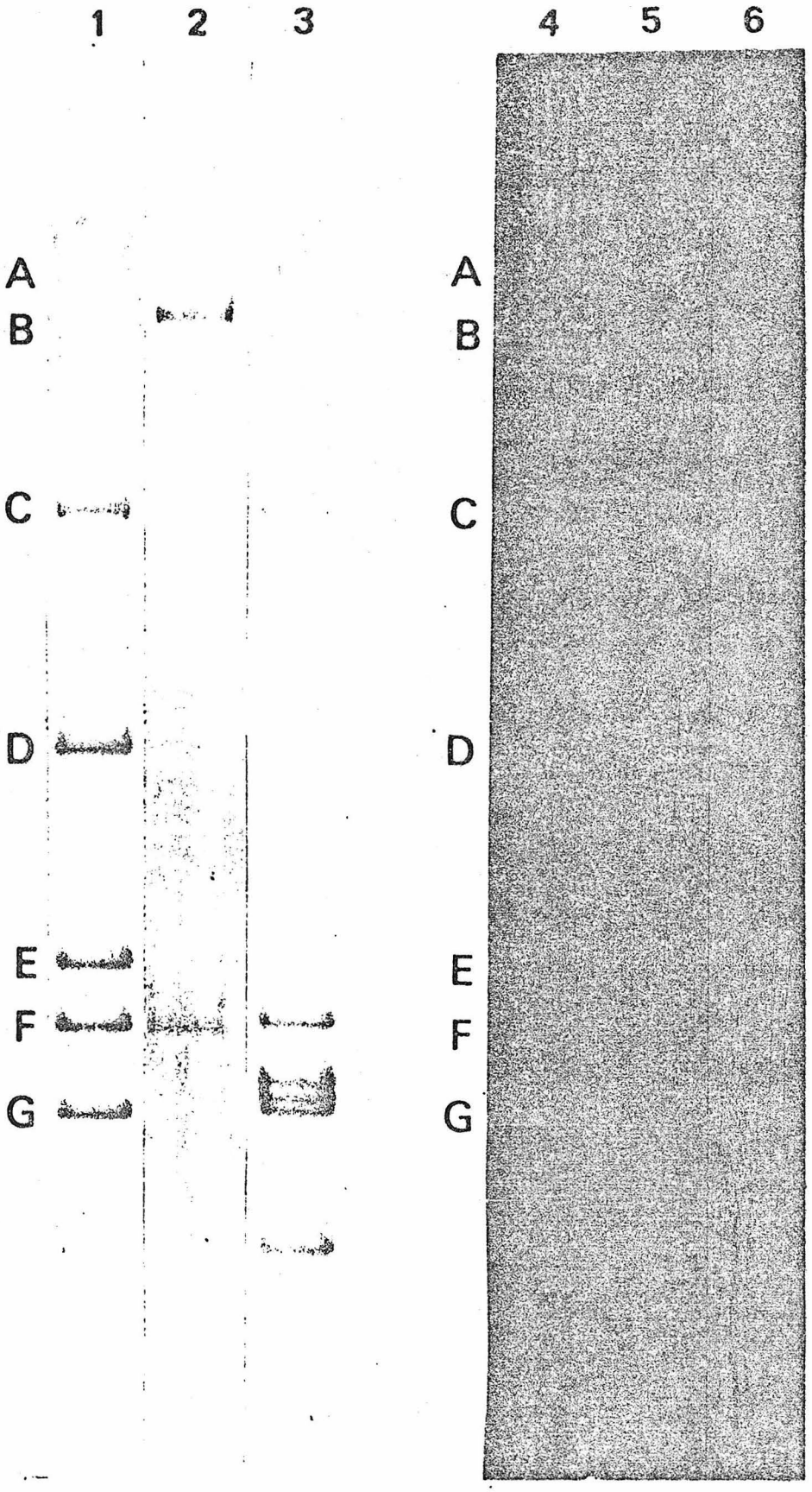

1
2
3
4
5
6
A
B
C
D
E
F
G
A
B
C
D
E
F
G

Figure 3. Mapping Bam HI Sites in HβG2 DNA

The Hind III A and B fragments of HβG2 DNA were isolated from SeaPlaque agarose and digested with Bam HI as described in the legend to Figure 2.

(1) Hind III A fragment digested with Bam HI. Digestion of the Hind III A fragment with Bam HI also produced small products, 0.7 kb (Bam F) and 0.2 kb in length, which migrated off the gel.

(2) HβG2 DNA digested with Bam HI. Two small Bam HI digestion products (0.7 kb [Bam F] and 0.5 kb [Bam I]) were run off the gel.

(3) Hind III B fragment digested with Bam HI.

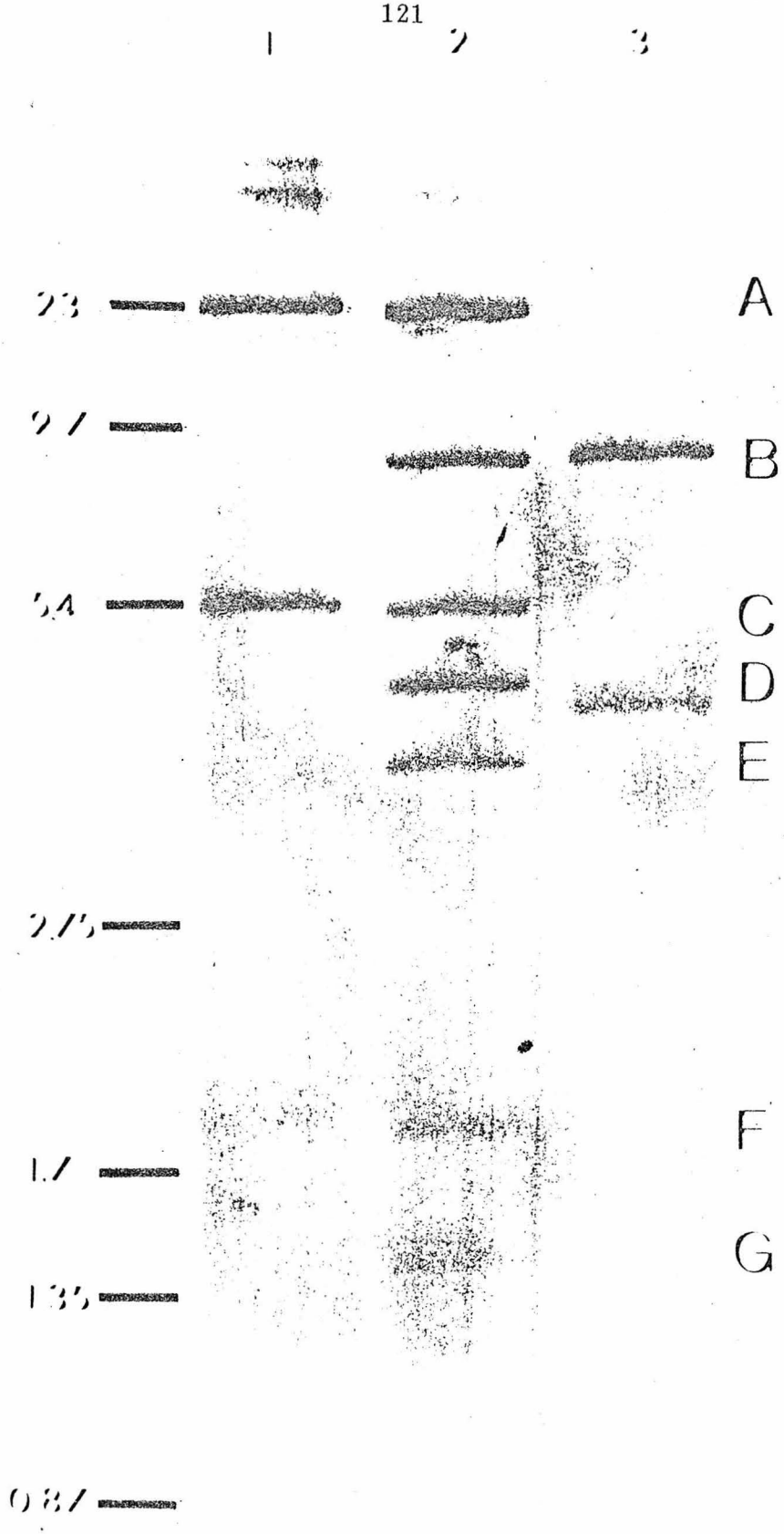

1
2
3
A
B
C
D
E
F
G

Figure 4. Restriction Endonuclease Cleavage Sites in HβG1 and HβG2 DNA

The locations of cleavage sites of restriction endonucleases Eco RI, Bam HI, Xba I, and Hind III in HβG2 DNA are presented. The Eco RI map of HβG1 DNA is also shown for comparison. Restriction fragments are delineated by vertical bars, and are lettered, in order of size, as referred to in the text. Fragment sizes are given in kilobase pairs. Circled letters denote fragments that hybridize to β-globin cDNA plasmid probes. The large vertical arrows mark the junction between the inserted human DNA and the λ Ch4A arms. The direction of transcription is indicated by the arrow at the top of the diagram.

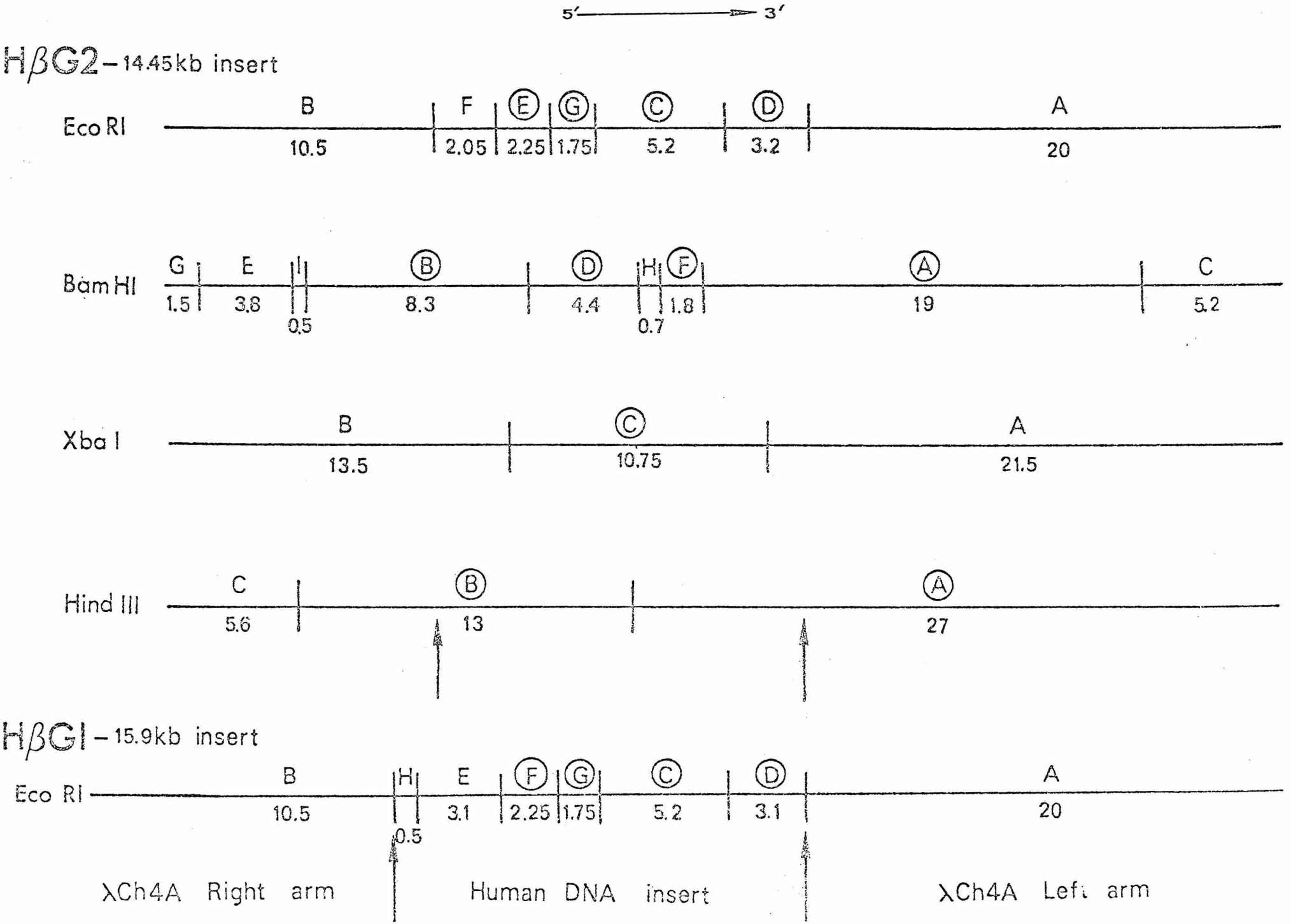

5′ 3′
HβG2 – 14.45kb insert
Eco RI
B F E G C D A
10.5 2.05 2.25 1.75 5.2 3.2 20
Bam HI
G E B D H F A C
1.5 3.8 0.5 8.3 4.4 0.7 1.8 19 5.2
Xba I
B C A
13.5 10.75 21.5
Hind III
C B A
5.6 13 27
HβG1 – 15.9kb insert
Eco RI
B H E F G C D A
10.5 0.5 3.1 2.25 1.75 5.2 3.1 20
λCh4A Right arm
Human DNA insert
λCh4A Left arm

Figure 5. Hybridization of HβG2 DNA Fragments to Probes Specific for the 5' and 3' Regions of β-globin mRNA

HβG2 DNA was digested with Eco RI or Bam HI and the products were electrophoresed in a 0.75% agarose gel, stained with EtdBr, and transferred to nitrocellulose paper as described in the legend to Figure 1. 5' and 3' specific probes were prepared by digesting pJW102 DNA with Hha I and Bam HI and purifying the fragments containing β-globin sequences 5' (5' - Bam) and 3' (3' - Bam) to the Bam HI site (see Figure 9). Similarly, Hha I and Eco RI were used to obtain a fragment representing sequences 3' (3' - RI) to the Eco RI site. Fragments were purified on a 5% polyacrylamide gel and recovered as described in Experimental Procedures. The probes were labeled with ^{32}P by nick translation and hybridized to the gel lanes as described below. Unlabeled 5' and 3' fragments were included as competing DNA. Cross contamination of 5' and 3' probes resulted in some hybridization of 3' specific bands in lanes 3 and 6.

Lane 1: EtdBr-stained gel of HβG2 DNA digested with Bam HI.

Lane 2: Hybridization of 3' - Bam to HβG2 DNA digested with Bam HI.

Lane 3: Hybridization of 5' - Bam to HβG2 DNA digested with Bam HI.

Lane 4: EtdBr-stained gel of HβG2 DNA digested with Eco RI.

Lane 5: Hybridization of 3' - RI to HβG2 DNA digested with Eco RI.

Lane 6: Hybridization of 5' - Bam to HβG2 DNA digested with Eco RI.

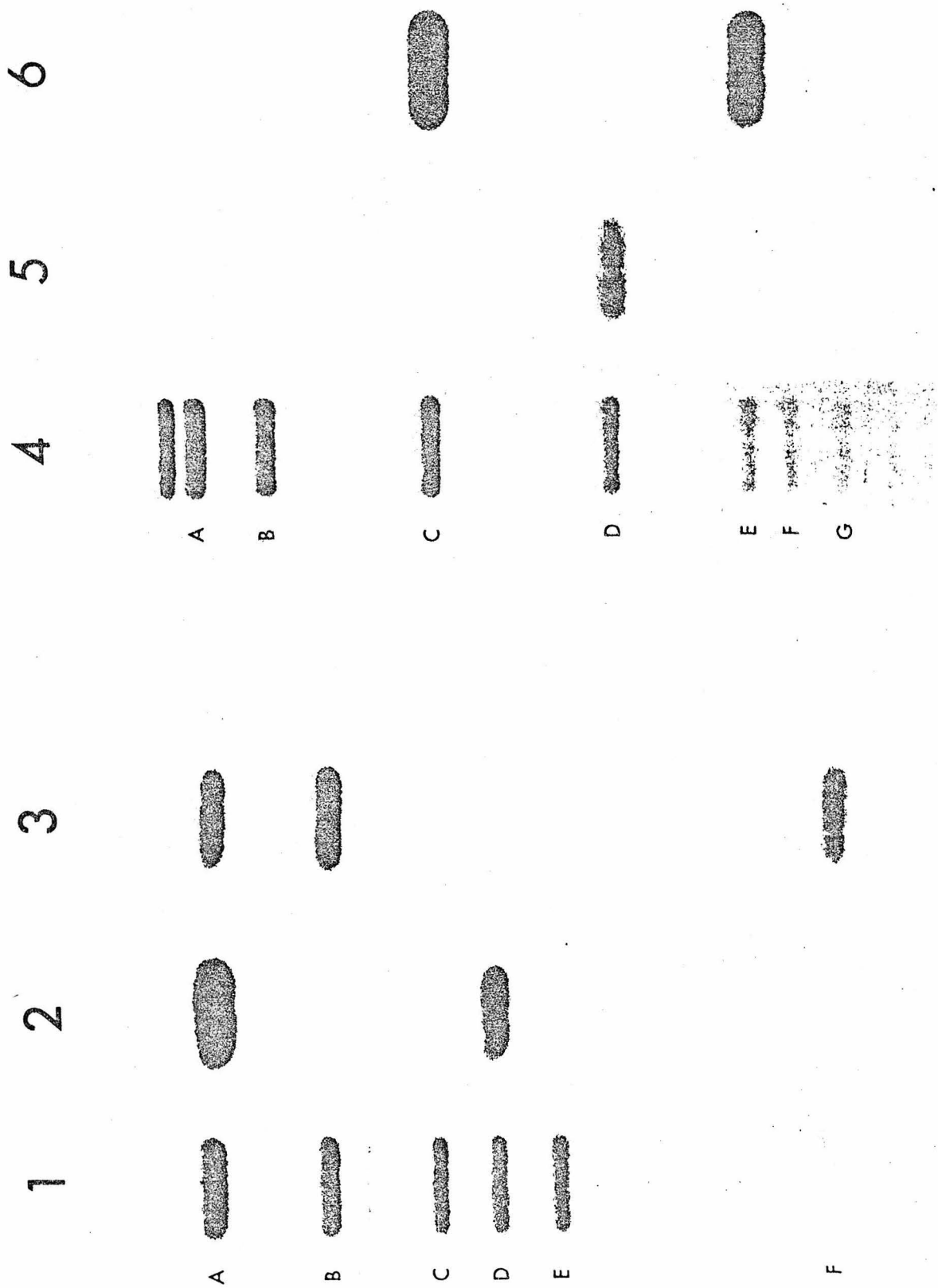

1
2
3
4
5
6
A
B
C
D
E
F
G

Figure 6. Identification of δ- and β-Globin Genes by DNA Sequence Analysis

The 1000 base pair (5') and 950 base pair (3') Eco RI/Bam HI fragments were labeled at the Eco RI cleavage site, purified as described in Experimental Procedures, and subjected to five different base specific cleavage reactions (Maxam and Gilbert, 1977). Aliquots of each reaction were electrophoresed in 20% and 15% polyacrylamide gels containing 7 M urea. Corresponding nucleotide positions in the two gels are marked "O". Within the mRNA coding regions (first 53 sequenced nucleotides), differences between the β and δ gene sequences are denoted by "-". The non-coding intervening sequence regions, beginning at the 54th nucleotide (marked "●"), contain many sequence differences which have not been denoted. Comparison of the derived DNA sequences with mRNA and protein sequences is shown in Figure 7.

(A) Sequence of nucleotides 5-38 from the Eco RI site of the 1000 bp (δ-globin) Eco RI/Bam HI fragment (20% polyacrylamide/7 M urea gel).

(B) Sequence of nucleotides 5-38 from the Eco RI site of the 950 bp (β-globin) Eco RI/Bam HI fragment (20% polyacrylamide/7 M urea gel).

(C) Sequence of nucleotides 28-81 from the Eco RI site of the δ-globin fragment (15% polyacrylamide/7 M urea gel).

(D) Sequence of nucleotides 28-81 from the Eco RI site of the β-globin fragment (15% polyacrylamide/7 M urea gel).

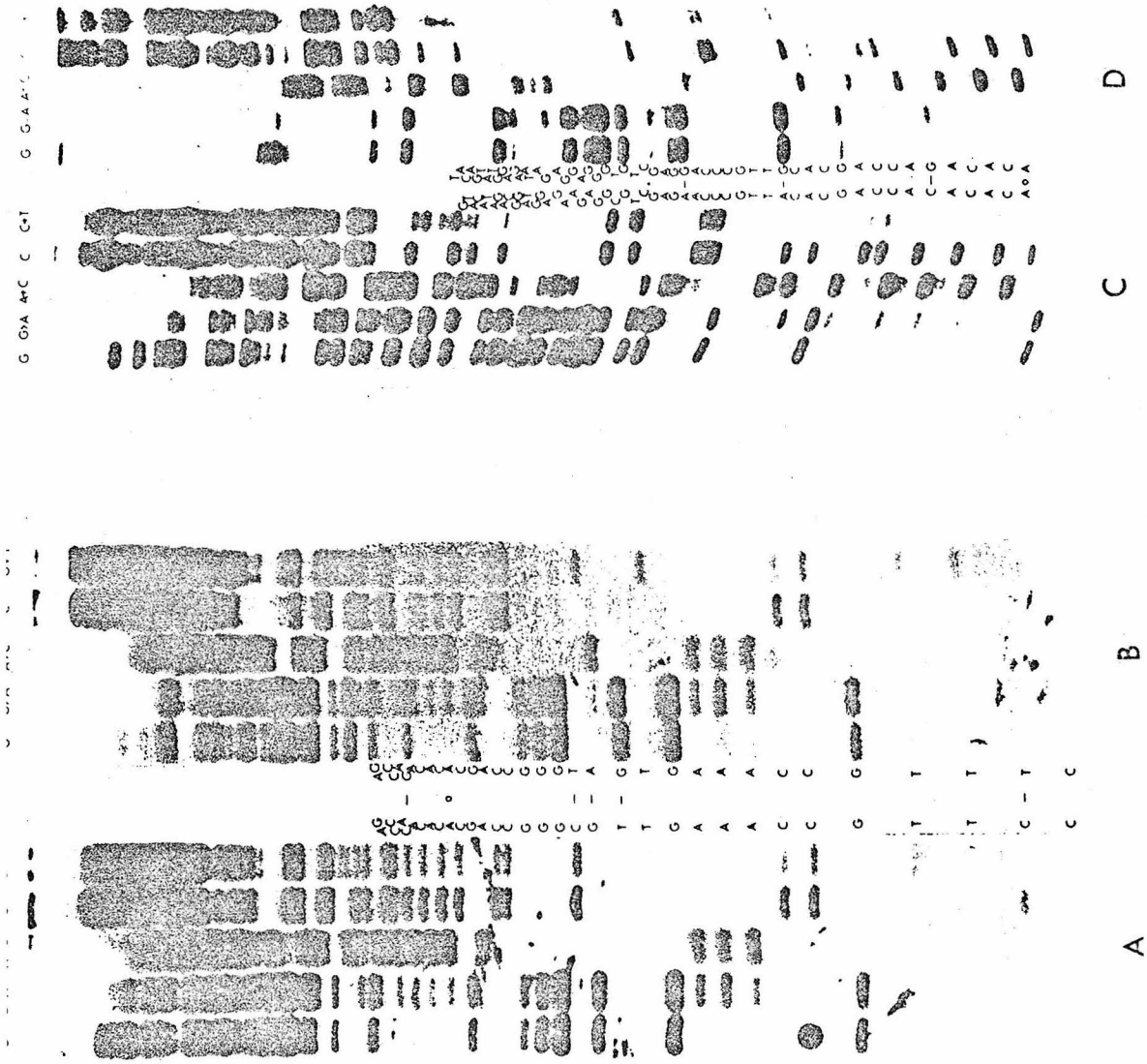

D
C
B
A

Figure 7. Comparison of DNA Sequences of β-Globin Related Genes

Coding Strand: The DNA sequence of the first 82 bases of the 950 or 1000 base pair fragment as determined in Figure 6. The ^{32}P-labeled strand was identified as the coding strand by comparison with the known β-globin mRNA sequence and the β- or δ-globin protein sequence.

Anti-coding strand: The DNA sequence complementary to the coding strand.

mRNA: The sequence of β-globin mRNA.

Protein sequence: The amino acid sequence of the β- or δ-globin protein (Dayhoff, 1972).

The entire Eco RI recognition sequence ($^{3'}$ CTTAAG $^{5'}$ / $^{5'}$ GAATTC $_{3'}$) is shown although the ^{32}P label was added at the 5'-adenosine of the coding strand. The first four nucleotides (A-A-T-T) were run off the sequencing gel shown in Figure 6. The five non-sequenced bases of the Eco RI recognition site are shown in lower case letters.

The first 53 nucleotides of the coding strand from the 950 base pair fragment agree precisely with those predicted from the β-globin mRNA sequence. These 53 nucleotides code for amino acids at positions 122-105 in the β-globin protein.

The first 53 nucleotides of the coding strand from the 1000 base pair fragment contain the codons which specify the amino acids found at positions 122-105 in the δ-globin protein. Nucleotides and amino acids which differ between the β- and δ-globin genes are underlined (see text).

Nucleotides 54-81 of the coding strand are not complementary to the known β-globin mRNA sequence nor can they code for the amino acids found

Fig. 7 - continued

in δ-globin. The arrow (v) marks the position in the coding strand where the large intervening sequences begin in both genes (see text). Note that since only one DNA strand has been sequenced, certain base assignments beyond the 65th nucleotide should be considered tentative.

950 bp fragment (β globin)

coding strand: 3'...T ACA GTA TGG AGA ATA GAG GAC GGT GTC ▼ GAG GAC CCG TTG CAC GAC CAG ACA CAC GAC CGG GTA GTG AAA CCG TTT Ctt aag - ^{32}P

anti-coding strand: 5'...A TGT CAT ACC TCT TAT CTC CTC CCA CAG CTC CTG GGC AAC GTG CTG GTC TGT GTG CTG GCC CAT CAC TTT GGC AAA Gaa ttc

mRNA: 5'...G CUG CAC GUG GAU CCU GAG AAC UUC AGG CUC CUG GGC AAC GUG CUG GUC UGU GUG CUG GCC CAU CAC UUU GGC AAA GAA UUC

Leu Leu Gly Asn Val Leu Val Cys Val Leu Ala His His Phe Gly Lys Glu Phe
105 110 115 120

1000 bp fragment (δ globin)

coding strand: 3'...G TAC ATA GAC GGA TGG AGA AGA GGC GTC ▼ GAG AAC CCG TTA CAC GAC CAC ACA CAC GAC CGG GCG TTG AAA CCG TTC Ctt aag - ^{32}P

anti-coding strand: 5'...C ATG TAT CTG CCT ACC TCT TCT CCG CAG CTC TTG GGC AAT GTG CTG GTG TGT GTG CTG GCC CGC AAC TTT GGC AAG Gaa ttc

Leu Leu Gly Asn Val Leu Val Cys Val Leu Ala Arg Asn Phe Gly Lys Glu Phe
105 110 115 120

Figure 8. Evidence for a Second, Small Intervening Sequence in the β-Globin Gene

The Bam HI/Hha I fragment containing the 5' segment of the β-globin gene in the cDNA plasmid JW102 and the Bam HI fragment F containing the 5' segment of the cloned genomic β-globin gene in HβG2 were isolated from a polyacrylamide gel, digested with restriction endonuclease Hae III, and electrophoresed on a 5% polyacrylamide gel. The Hae III digestion products of ØX174 DNA were coelectrophoresed.

Lane 1: ØX174 DNA digested with Hae III. Sizes of fragments are given in base pairs.

Lane 2: Bam HI/Hha I fragment from JW102 digested with Hae III. The A, B, and C fragments correspond, respectively, to the Hha I/Hae III fragment from the Hha I site in PMB9 to codon position 26 in the β-globin message coding sequence, the Hae III/Hae III fragment from codon positions 26 to 75, and the Hae III/Bam HI fragment from codon positions 75 to 100 (see Figure 9B). These assignments were confirmed by additional restriction endonuclease digestions (unpublished results). The faint bands between the A and B fragments are due to contaminating DNA fragments not observed in other gels.

Lane 3: Bam HI fragment F of HβG2 digested with Hae III.


1
2
3
1078
602
A
281
194
B
118
72
C

Figure 9. Location of δ- and β-Globin Genes in HβG2 DNA

(A) The globin gene containing region of HβG2 DNA is shown in this expanded restriction endonuclease map (compare to Figure 4). Arrows pointing down indicate restriction endonuclease sites of Eco RI (O) and Bam HI (o). Sizes of the fragments are given in base pairs. The boxed regions denote the δ- and β-globin genes. The filled boxes represent the mRNA sequence and open boxes represent non-coding intervening sequences. The size of the intervening sequence in the 5' segment of the β-globin gene is not known. The arrows pointing up delineate the distance between the linked genes.

(B) The positions of Eco RI, Bam HI, and Hae III cleavage sites in the β-globin cDNA sequence (Marrota et al., 1977). Only those Hae III cleavage sites 5' to the Bam HI site are shown. The positions of the codons containing each of the restriction endonuclease sites are also given.

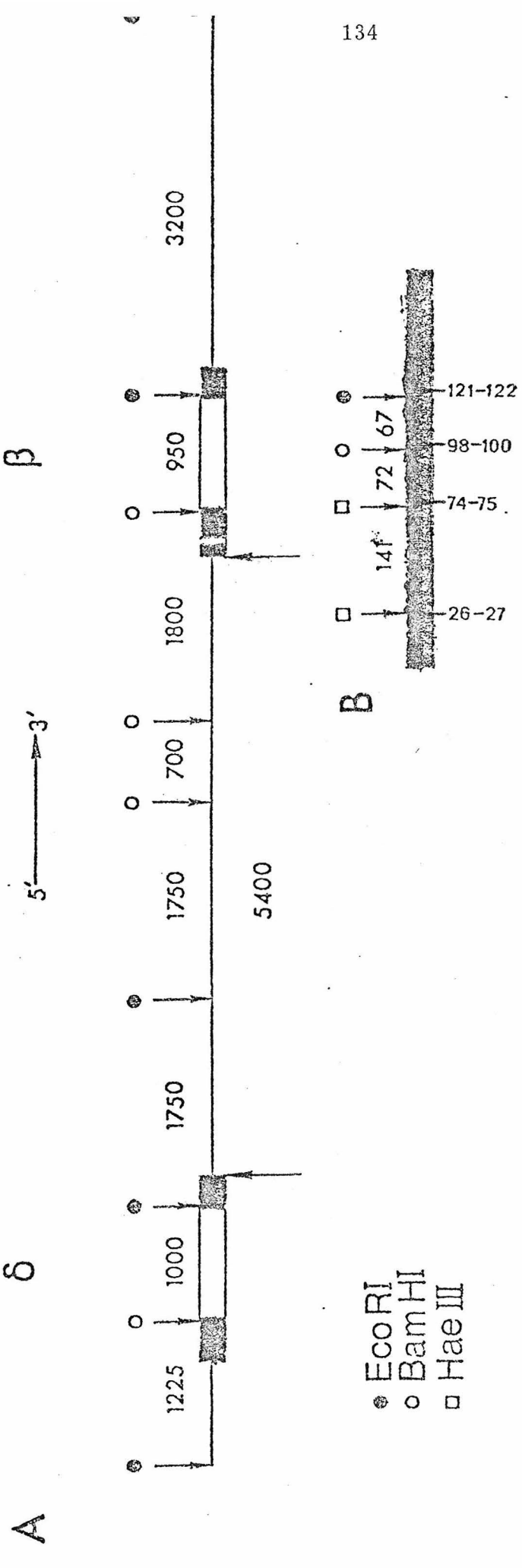

A
δ
1225
1000
1750
5400
1750
700
1800
950
3200
B
121-122
67
98-100
72
74-75
141
26-27
B
5'
3'
Eco RI
Bam HI
Hae III

Figure 10. Agarose Gel Electrophoresis of Pst I Fragments of HβG1 and HβG2 DNA

DNA from clones HβG1 and HβG2 was digested with Pst I, electrophoresed, stained with EtdBr (left panel), transferred to nitrocellulose filter paper, and hybridized with human β-globin cDNA plasmid (right panel) as described in the legend to Figure 1. Sizes of the Pst I fragments containing human DNA are given in Figure 11. Fragment B represents an unresolved doublet, each fragment of which contains human DNA and only one of which contains human β-globin DNA.

Lanes 1 and 3: HβG1 DNA.

Lanes 2 and 4: HβG2 DNA.

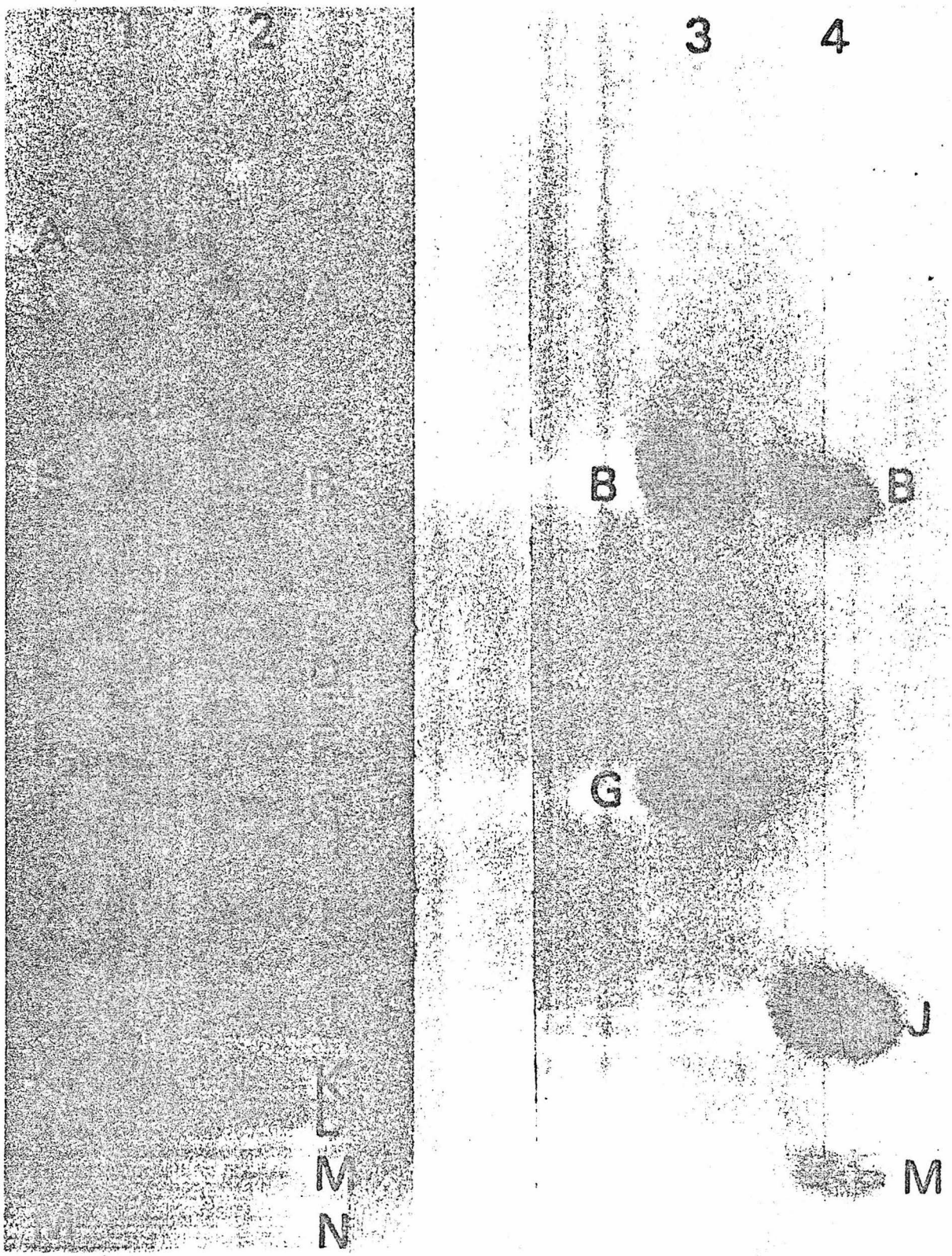

1
2
3
4
A
B
B
G
J
K
L
M
N
M

Figure 11. A Map of Pst I Sites in HβG2 DNA

A Pst I map of HβG2 DNA was constructed on the basis of double enzyme digestions and digestion of restriction fragments isolated by SeaPlaque gel electrophoresis (data not shown). The region of HβG2 DNA containing the human DNA insert is shown. The Pst I site between fragments J and M of HβG2 (marked by an arrow) is absent in HβG1. A fragment 2.3 kb in length (Pst fragment G of HβG1) replaces the Pst J and M fragments of HβG2. The wavy lines represent the junction between the λ Ch4A DNA and inserted human DNA. Filled boxes represent the message coding region of each gene and open boxes represent the large intervening sequence in each gene. The small intervening sequence in the 5' end of the β-globin gene is not shown.

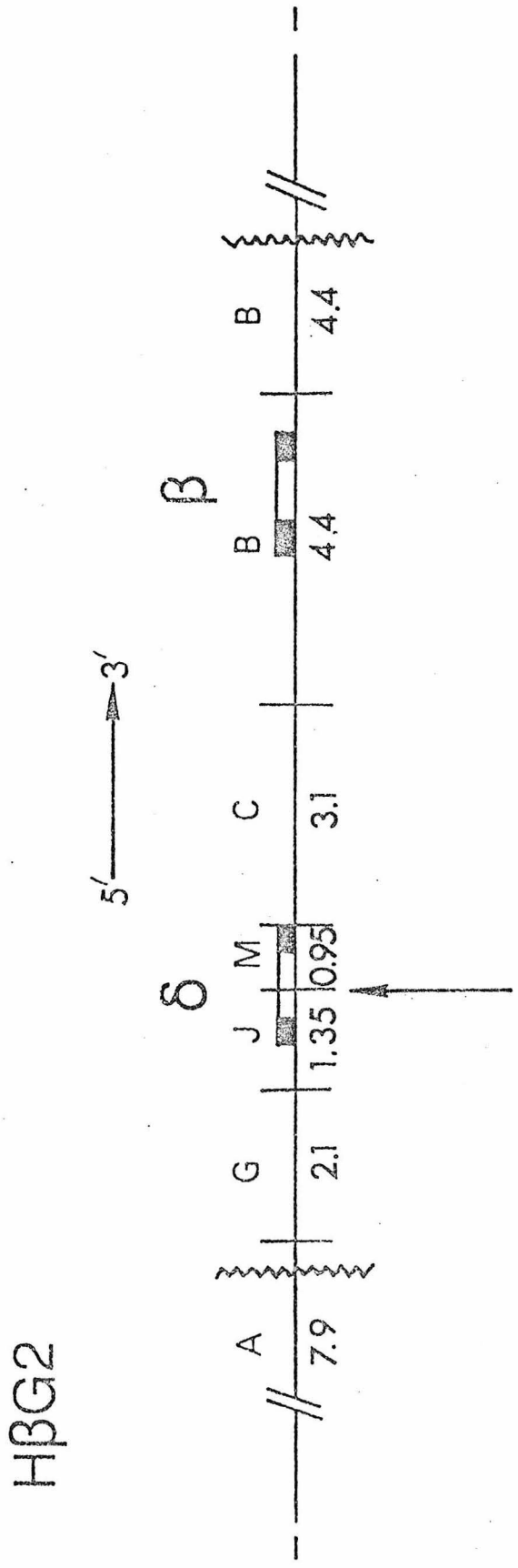

HBG2
A
7.9
G
2.1
J
1.35
M
0.95
δ
C
3.1
5′
3′
B
4.4
B
4.4
B
4.4

Figure 12. Pst I Digestion of Bam HI/Eco RI Fragments Containing the δ-Globin Intervening Sequence

The 1000 bp Bam HI/Eco RI fragment containing the δ-globin intervening sequence was isolated from HβG1 and HβG2 DNAs following preparative acrylamide gel electrophoresis. Samples of each DNA were electrophoresed on a 1% agarose gel with and without digestion by Pst I, and DNA from the gel was transferred to nitrocellulose filter paper. A portion of the isolated Bam HI/Eco RI fragment from HβG1 DNA was nick-translated, hybridized to the filter-bound DNA, and autoradiographed.

Lane 1: Bam HI/Eco RI fragment from HβG1.

Lane 2: Bam HI/Eco RI fragment from HβG1 digested with Pst I.

Lane 3: Bam HI/Eco RI fragment from HβG2.

Lane 4. Bam HI/Eco RI fragment from HβG2 digested with Pst I.

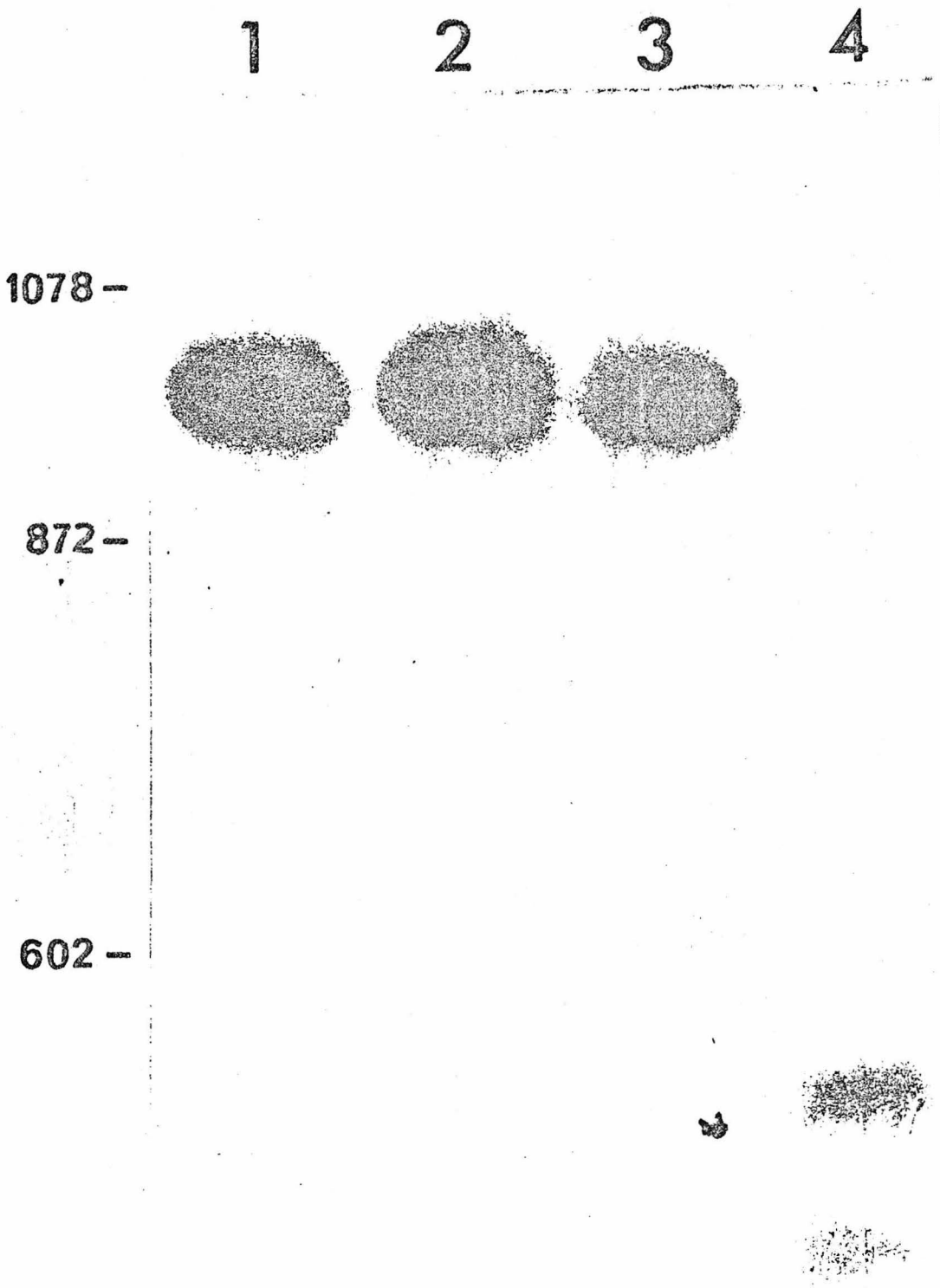
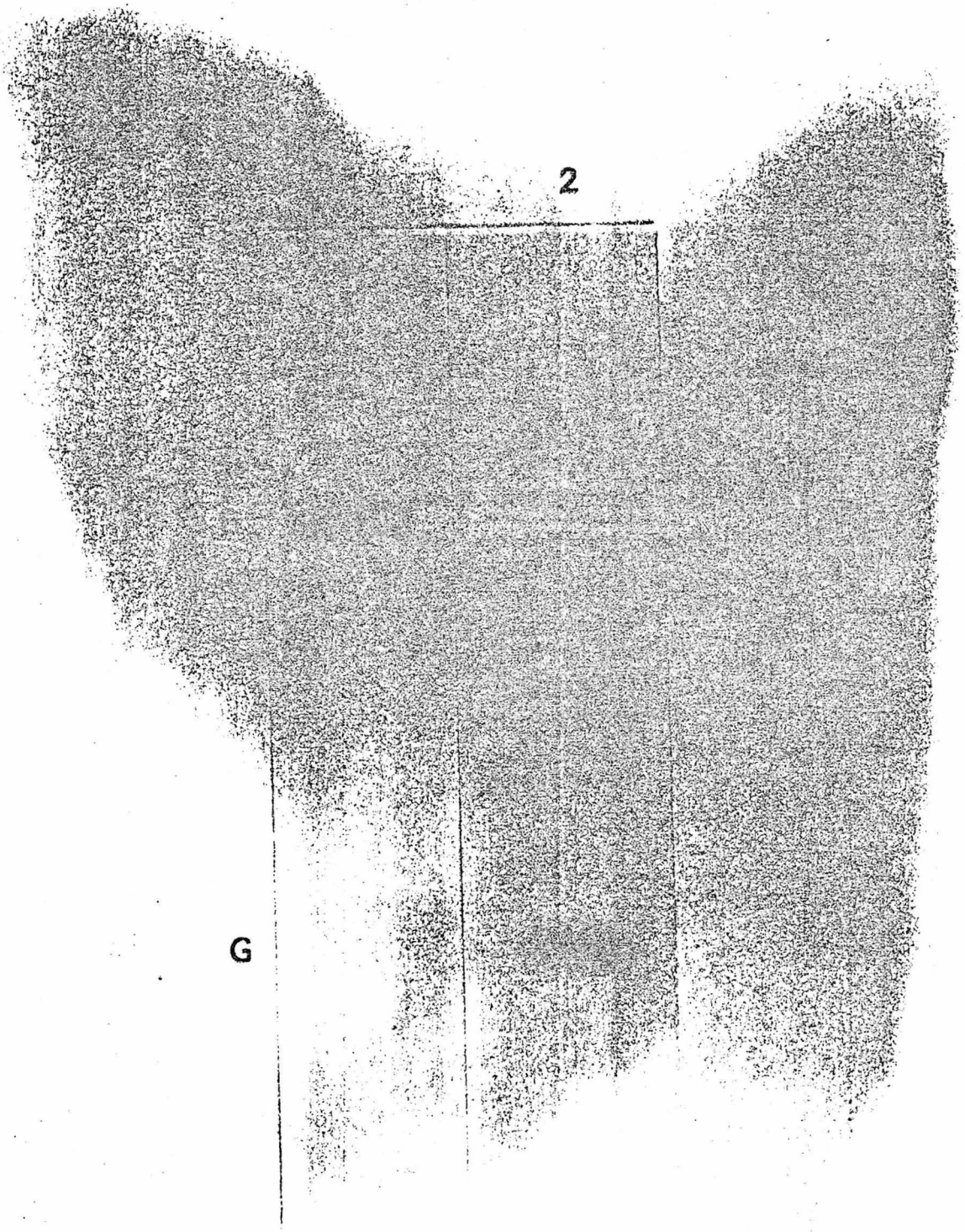

1
2
3
4
1078
872
602
374

Figure 13. Hybridization of δ- or β-Globin Specific Intervening Sequences to Restriction Fragments Containing δ- or β-Globin Genes

HβG1 DNA was digested with restriction endonuclease Pst I, electrophoresed in two lanes of a 1% agarose gel, and transferred to nitrocellulose filter paper. The Bam HI/Eco RI fragments containing the δ-globin (1000 bp) and β-globin (950 bp) intervening sequences were separately prepared by preparative polyacrylamide gel electrohporesis. Each of the fragments also contains 67 nucleotides of δ- or β-globin coding sequence. The isolated fragments were nick-translated and hybridized separately to the filter-bound Pst I DNA fragments of HβG1. The filters were washed in 12 mM NaCl, 2.5 mM Tris (pH 7.4), and 1 mM EDTA at 68°C.

Lane 1: Pst I digest of HβG1 DNA hybridized with Bam HI/Eco RI fragment containing β-globin intervening sequence.

Lane 2: Pst I digest of HβG1 DNA hybridized with Bam HI/Eco RI fragment containing the δ-globin intervening sequence.


2
G

ISOLATION AND CHARACTERIZATION OF MAMMALIAN GLOBIN GENES

Elizabeth Lacy, Richard M. Lawn, Edward Fritsch, Ross C. Hardison,
Richard C. Parker, Tom Maniatis

Division of Biology, California Institute of Technology
Pasadena, California 91125

Key words: linked globin genes/recombinant DNA/gene isolation/
gene mapping/intervening sequences

This work was supported by grants from the National Science Foundation, the National Institutes of Health, and funds from an NIH Biomedical Research Support Grant to the California Institute of Technology. T.M. is a recipient of a Rita Allen Foundation Career Development Award.

Summary

Using a gene isolation procedure which involves screening cloned libraries of large, random fragments of genomic DNA, we have obtained direct evidence for physical linkage between different mammalian globin genes. Restriction endonuclease cleavage analyses and hybridization studies of cloned rabbit and human globin genes revealed two β-globin sequences separated by less than 9 kb of DNA in rabbits and 5.5 kb of DNA in humans. The β-related sequence in rabbit DNA lies to the 5' side of the adult β-globin gene, but the relative direction of transcription of the two linked β sequences is not yet known. Using 5' and 3' specific cloned cDNA hybridization probes, it was possible to show that both human β genes are transcribed in the same direction, with the more weakly hybridizing β-like sequence lying to the 5' side of the presumptive β-globin gene. On the basis of a comparison of detailed restriction maps of the cloned β genes with preliminary genomic mapping data from normal and Hb Lepore DNA, we tentatively identified the β-related sequence as the δ gene. Restriction maps of the cloned DNA also reveal the presence of a large (900-950 bp) non-coding intervening sequence within both human β genes.

Isolation of Mammalian Globin Genes

One approach to understanding the molecular basis of hemoglobin switching is to study the organization and expression of isolated globin genes. Although an individual globin gene represents only one five millionth of the DNA in a mammalian cell, recent advances in the development of molecular cloning techniques have made it possible to obtain a number of globin genes and their associated sequences in the amounts required for detailed biochemical studies. The first globin sequences to be cloned and amplified in bacteria were double stranded DNA copies of globin mRNA (1-3). These cDNA clones have been used to determine the nucleotide sequence of globin mRNA (4-6), to map restriction sites flanking globin genes in genomic DNA (7, 8), and to identify and isolate globin genes and their surrounding sequences (9, 10, 11).

Globin genes were first isolated from genomic DNA using a combination of gene enrichment and recombinant DNA techniques which are described by Leder in this Symposium. We have recently established a procedure for gene isolation which does not require partial purification of the gene prior to cloning (12). The essential features of this procedure are illustrated in Figure 1. High molecular weight DNA is randomly fragmented and then size fractionated on sucrose gradients to obtain molecules of approximately 20 kilobases (kb) in length. These molecules are then adapted for cloning in bacteriophage λ in the following manner. First, the eukaryotic DNA is treated with the enzyme Eco RI methylase which renders Eco RI sites within the eukaryotic DNA resistant to cleavage by Eco RI. Next, the methylated DNA is ligated to a synthetic dodeca-nucleotide duplex DNA molecule bearing an Eco RI site (Eco RI linker) using the blunt end joining activity of T4 polynucleotide kinase. Finally, the methylated eukaryotic DNA is digested with

Eco RI. Because the Eco RI sites in the eukaryotic DNA are blocked by methylation, only the Eco RI sites within the synthetic DNA are cleaved. This produces random, high molecular weight DNA fragments bearing Eco RI cohesive ends which are suitable for insertion into an appropriate cloning vector.

The Charon 4A strain of bacteriophage λ (13) was used as a cloning vector. As shown in Figure 1, λ Charon 4A DNA is cleaved three times by Eco RI to produce two internal fragments and two end fragments. The internal fragments carry genes which are not essential to phage viability and therefore can be removed and replaced with eukaryotic DNA. The internal fragments are separated from the end fragments by annealing the cohesive ends of the molecule, digesting the circular DNA with Eco RI and fractionating the products on a sucrose gradient. The 31 kb end fragments are readily separated from the 7 and 8 kb internal fragments. Recombinant phage molecules can then be obtained by joining the 31 kb end fragments of the vector to the eukaryotic DNA through ligation of their Eco RI cohesive ends.

Because of the enormous complexity of the mammalian genome, a very large number of independent phage recombinants is required if the collection of cloned sequences (cloned library) is to contain all of the sequences present in the genome. For example, the calculated number of independently derived recombinant phage necessary to achieve a 99% probability of having any given DNA sequence in the library (for 20 kb DNA inserts) is nearly 7×10^5 (12). To obtain the necessary number of phage recombinants, we employed a highly efficient procedure for introducing phage DNA into bacteria which involves <u>in vitro</u> packaging of the phage DNA into viable phage particles (14, 15). This procedure is from ten to fifty times more efficient than calcium chloride transfection procedures (see reference 12 for discussion). Once a large number of viable phage recombinants

is obtained, they can be amplified 10^6-fold by low density growth in bacteria on agar plates with no apparent loss in sequence complexity (12). The amplified phage then constitute a "permanent" library of genomic DNA which can be screened repeatedly.

Recently, a rapid in situ plaque hybridization procedure has been developed which makes it possible to screen many hundreds of thousands of phage recombinants at a time (16). The procedure involves plating approximately 10,000 phage on an agar plate, transferring phage DNA from the resulting plaques onto a nitrocellulose filter, fixing the DNA to the filter and hybridizing with ^{32}P-labeled hybridization probes. Plaques which carry the sequence of interest are identified by autoradiography. By aligning the resulting autoradiogram with the agar plates, the plaques which hybridize can be identified, picked and purified. Using this procedure, it is possible to screen a mammalian DNA library (7×10^5 independent plaques) in 1-2 weeks.

The library approach to gene isolation offers the following advantages: 1) several different genes can be isolated in a single step by screening a library with a mixture of gene probes; 2) isolation of a set of overlapping clones, all of which contain a given gene, permits study of sequences extending many kilobases from the gene in the 5' and 3' direction; and 3) distant regions along the chromosome can be obtained by rescreening the library using terminal fragments of the initially selected clones. Rescreening is facilitated by the fact that the entire genome can be rapidly screened.

Using this procedure of gene isolation, we have obtained a number of different globin genes from libraries of rabbit and human DNA. Restriction endonuclease cleavage analyses and hybridization experiments have revealed that a number of these globin genes are closely linked.

Characterization of Linked Globin Genes

Mammalian globin genes constitute a relatively simple gene family comprised of a number of members (α and non-α types) which almost certainly evolved from a single ancestral gene by duplication and subsequent sequence divergence (17). Genetic (18, 19) and cell fusion (20, 21) studies have shown that the α and non-α types of globin genes are located on different chromosomes, but that both types are arranged in closely linked clusters. Models for the mechanism of hemoglobin switching have emphasized the possible relationship between the organization of genetically linked globin genes and the control of their temporal expression, but until now this possibility could not be examined at the molecular level. In fact there was no direct evidence for close physical linkage between globin genes. The availability of large fragments of cloned genomic DNA bearing globin sequences provides the opportunity for studying the detailed organization of the globin gene family, including possible linkage relationships. We present here the current status of our analysis of the organization of linked rabbit and human globin genes.

Linked Rabbit β-Globin Genes

As shown in Table 1, there are at least 4 to 6 different globin genes in rabbit. Two types of globin chains (designated χ and ε) are found in nucleated red blood cells derived from yolk sac islets (22). At least two different ε chains (y and z) with similar but nonidentical amino acid sequences have been detected. After 20 days of the 32-day gestation period, the fetal liver becomes the center for erythropoesis and only the α- and β-globin proteins are synthesized (23). Until recently, there was no evidence for more than one adult β-globin gene. However, a β-like mRNA sequence has been identified in proerythroblasts and basophilic

erythroblasts of bone marrow from anemic rabbits which is clearly different from the adult β-globin gene (24). Although it is possible that the presence of β-like mRNA in immature erythroid cells results from anemia-induced activation of an embryonic gene, the temporal expression of this mRNA during red cell maturation is strikingly similar to that of the δ gene of human (24) and a minor hemoglobin is in fact found in immature erythroid cells from nonanemic rabbits (25).

Although the possibility of genetic linkage between different β-related genes in rabbit has not been studied, such linkage has been demonstrated in other mammals (19, 26-29). For example, in mouse, the adult β-major and β-minor genes are linked to each other and to an embryonic ε gene (27, 28). In the process of analyzing several clones selected from a library of rabbit DNA by hybridization to a β-globin probe, we have identified one or more β-related genes closely linked to the adult β-globin gene.

From a screen of 750,000 recombinant phage, four independent β-globin clones were isolated. DNA from each of the four clones was digested with various restriction enzymes. The resulting fragments were fractionated by agarose gel electrophoresis, transferred from the gel to a nitrocellulose filter (30), and subsequently hybridized to nick translated pβG-1, a rabbit adult β-globin cDNA plasmid (2). (For convenience, we will refer to this procedure as "blotting".) Such experiments revealed that two clones (designated λCh4A-RβG-2 and λCh4A-RβG-5), contain β-globin related sequences linked to the adult β-globin gene. As an example, Figure 2 shows a blotting experiment with the DNA from a clone designated RβG-2, using Kpn I, an enzyme which does not cleave within the double stranded cDNA sequence of the adult rabbit β-globin mRNA.

Unexpectedly, two Kpn I fragments, approximately 4.7 and 18.8 kb in length hybridize to the β-globin probe. This observation indicates that a β-related sequence is closely linked to the adult β-globin gene or that the adult β-globin gene has an intervening sequence containing a Kpn I recognition site. The latter possibility can be ruled out by the fact that restriction mapping and DNA sequencing data from the cloned 4.7 kb Kpn I fragment indicates that this fragment contains the entire β-globin gene plus intervening sequences (unpublished results). This is consistent with the observation that the 4.7 kb Kpn I fragment is roughly the same size as the only Kpn I fragment which hybridizes to pβG-1 in genomic DNA blotting experiments (7, 8). Thus, the additional Kpn I fragment in RβG-2 which hybridizes to the β-globin probe must contain a β-related globin gene that is physically linked to the adult gene.

Bam HI and Kpn I restriction enzyme maps have been constructed for RβG-2 and are presented in Figure 3. An analysis of the products from a Bam HI and Kpn I double digest indicated that the cloned rabbit DNA included a 9.1 kb Bam HI fragment that contains sequences from both the adult β and the linked β-related globin genes. A 9.9 kb Bam HI fragment that hybridizes to sequences on the 5' side of the Bam HI site in the adult β-globin gene has been detected in genomic blots of rabbit DNA (7). Most likely, this 9.9 kb fragment corresponds to the 9.1 kb fragment seen in RβG-2. On the basis of these observations, we have mapped the β-related gene approximately 9 kb to the 5' side of the adult gene. Preliminary mapping experiments on RβG-5 DNA revealed additional Eco RI fragments which hybridize pβG-1. These fragments map in the 5' direction from the β-related sequence. They could be part of a third linked β-globin gene or may indicate the presence of a large intervening sequence in the β-related gene.

The identities of the β-related sequences in RβG-2 and RβG-5 have not yet been determined. To decide whether the related genes are embryonic β-globin and/or adult β-minor globin genes, we are preparing cDNA clones from mRNA isolated from the blood islands of 12-day rabbit fetuses and from the bone marrow of anemic adult rabbits. Once the β-related genes are identified, it will then be possible to establish the order of linkage of the genes in the rabbit β-globin family. Furthermore, it should also be possible to determine whether there are more linked β-like genes by using terminal fragments from the four β clones to rescreen the library.

The Human Globin Gene Family

As shown in Table 1, there are at least 8 members of the human globin gene family (19). In contrast to the developmental pattern of expression in rabbit, where only one switch in the expression of hemoglobin types occurs (embryo to adult), two switches occur in human (36). In the early embryo, the ζ and ε genes are expressed in erythroid cells thought to be derived from the yolk sac. The human ζ chain is analogous to the embryonic χ chain of rabbit in that the primary structure of the two globins resembles that of their respective adult α-globin chains (37).

When the site of red blood cell production changes from the yolk sac to the fetal liver, the adult α-globin and fetal γ-globin genes are expressed. In some human populations, the α chain locus is thought to be duplicated (19). Evidence presented in this Symposium by S. Orkin verifies this prediction. At least two structural genes for the γ chains exist in humans, coding for polypeptide chains which have either glycine (γ^G) or alanine (γ^A) in position 136 of their amino acid

sequence (38). On the basis of a study of γ-chain variants, a four loci model for γ-globin gene organization has been proposed (39, W. Schroeder, this Symposium). Just prior to birth, the adult β- and δ-globin chains appear and by 6 months HbA ($\alpha_2 \beta_2$) and HbA_2 ($\alpha_2 \delta_2$) represent greater than 98% of the hemoglobin in peripheral blood. There is thought to be only one copy each of the β- and δ-globin structural genes per haploid genome (19).

Linked Human Globin Genes

The linkage relationships and chromosomal locations of the ζ and ε genes are not known (19). Cell fusion experiments have shown that α-globin gene(s) are located on chromosome 16 (20). Chromosomal blotting experiments presented at this Symposium suggest that the α genes are, in fact, duplicated and closely linked (approximately 3-4 kb apart) (S. Orkin, this Symposium). The cell fusion experiments have also shown that non-α genes are located on chromosome 11 (21). Analysis of two structural mutants, hemoglobins Lepore and Kenya (29) suggest that the γ, β and δ genes are closely linked. In hemoglobin Lepore the N-terminal amino acid sequence of δ globin is joined to the C-terminal sequence of β globin, while the γ^A and β-globin genes are fused in a similar fashion in hemoglobin Kenya (29). This information and other genetic data are consistent with either of the following gene arrangements: 5' γ^G - γ^A - δ - β 3' or 5' γ^A - δ - β - γ^G - 3' (29). Until now physical linkage between non-α genes has not been demonstrated.

The human globin gene system presents the possibility of using mutants to establish relationships between the structure and organization of globin genes and the mechanisms of their differential expression. In order to characterize

globin genes and their associated DNA sequences, we constructed a human DNA library by the methods described above using human fetal liver DNA (a gift of B. G. Forget). The library was screened using ^{32}P-labeled nick-translated human α, β and γ globin cDNA plasmids (pJW 101, pJW 102, pJW 151, respectively, a gift of B. G. Forget and co-workers) (36). We have initially studied two clones which hybridize the β-globin probe (designated HβG-1 and HβG-2).

DNA from the two clones was digested with restriction endonucleases which do not recognize cleavage sites within the β-globin message sequence (e.g., Hha I, Bgl I or Xba I) or which recognize a single cleavage site (Bam HI and Eco RI) (37). Digestion of HβG-1 and HβG-2 with Xba I (Figure 4), or Bgl I or Hha I (data not shown) generated a single, large band which hybridized the β-globin cDNA plasmid. Digestion of HβG-1 and HβG-2 with either Eco RI or Bam HI generated four bands which hybridized with the β-globin cDNA plasmid (Fig. 4). Because Eco RI and Bam HI are known to cleave within the β-globin message sequence at only one site, each clone must contain two β-related genes or a single gene with more than one non-coding intervening sequence containing Bam HI and Eco RI recognition sites.

Genetic evidence indicates that the human δ-globin gene (a minor adult globin) is linked to the β-globin gene in the order 5' δ - β 3' (17). The δ-globin amino acid sequence is very similar to the β-globin sequence and therefore the β- and δ-globin genes should cross-hybridize. Evidence presented below indicates that HβG-1 and HβG-2, each contain two closely linked β-globin genes, most likely β and δ. This conclusion is consistent with the results of genomic blotting experiments presented at this Symposium by A. Bank and R. Williamson and similar experiments conducted in our laboratory.

The position of several restriction enzyme cleavage sites around the two regions which hybridize β-globin probe in HβG-2 are shown in Figure 5. To determine the orientation of the two globin sequences with respect to each other, hybridization probes specific for the 5' and 3' portion of the β-globin gene were prepared by restriction enzyme digestion of the β-globin cDNA plasmid JW 102 followed by in vitro labeling of the appropriate fragments. A probe corresponding to sequences 5' to the internal Bam HI site at Codons 98-100 hybridized to the Eco RI fragments designated C and E and the Bam HI fragments B and F. Sequences located 3' to the Bam HI site hybridized to the Bam HI fragments A and D. Similarly, a probe containing sequences 3' to the Eco RI site at Codons 121-122 hybridized to the Eco RI fragments D and G. On the basis of these data and the Bam HI and Eco RI restriction maps of Figure 7, we conclude that two linked β-globin genes are present in HβG-2. The δ gene has been shown to lie 5' to the β by genetic analysis and restriction enzyme analysis of genomic DNA (17, A. Bank, R. Williamson, this Symposium). In both HβG-1 and 2, the 3' end of the more weakly hybridizing β-like sequence lies about 5.5 kb from the 5' end of the presumptive β gene.

Using a cloned γ cDNA sequence (36) to identify genomic clones bearing the γ^A and/or γ^G genes, it should be possible to determine the precise linkage relationships between the γ, δ, β globin gene complex.

Location of Non-coding Intervening Sequences

Large non-coding intervening sequences in the presumptive β and δ globin genes were revealed by comparing the location of Bam HI and Eco RI sites within the regions which hybridize the β-globin probe in HβG-2 (Figure 5) to the corresponding sites in the β-globin mRNA sequence (36). The Eco RI and Bam HI sites in

the β-globin message sequence are separated by 67 base pairs. In HβG-2, the number of base pairs of DNA separating these sites is approximately 900. Similarly, if we assume that Eco RI and Bam HI sites are located at the same positions within the δ globin mRNA sequence, these sites are separated by approximately 920 base pairs in the δ globin sequence of HβG-2. Thus, in both genes the coding sequences in genomic DNA are interrupted by at least 900 base pairs of non-coding sequences. It is interesting to note that an insert of approximately 600 bp is located between the corresponding Bam HI and Eco RI sites of the rabbit β-globin (8), the mouse β-major (10), and the mouse β-minor globin genes (11). A second, smaller non-coding intervening sequence has been located in the mouse (P. Leder, this Symposium) and rabbit (Efstratiadis, Lacy and Maniatis, unpublished results) β-globin genes, located between codons 30 and 31. Although the functional significance of these inserts is not understood, the non-coding sequences are transcribed in the nucleus (11) and presumably removed before the message sequence is transported to the cytoplasm. Experiments are in progress to search for additional inserts in the human β-globin genes and to test for homology between inserts in the β- and δ-globin genes.

The restriction map of HβG-2 is consistent with blot hybridization experiments we have performed with wild-type adult spleen and fetal liver DNA and with DNA from a thalassemic patient with hemoglobin Lepore. Our data are in general agreement with the restriction maps derived from genomic blots presented by others at this meeting (A. Bank and R. Williamson). Thus, no gross rearrangement of DNA has taken place during the cloning and propagation of the recombinant phage.

Discussion

The intergene distance observed for the different pairs of linked globin genes described above is remarkably similar. The number of base pairs separating Bam HI cleavage sites within the two members of linked rabbit β and human β globin genes is approximately 9 and 7 kb, respectively. In the case of the linked human and rabbit β globin genes, the β-related sequence is located to the 5' side of the adult β-globin gene. Although the minor β-globin gene in rabbit has not yet been identified with a particular globin mRNA, it is possible that this gene codes for the β-like mRNA expressed in immature adult erythroid cells of anemic rabbits (24). Clissold et al. (24) have argued that this mRNA may be the product of a gene which is the functional analogue of the human δ-globin gene. Both the human δ gene and the minor β-gene in rabbit (24) have sequences which are similar but not identical to the respective β-globin genes and are expressed primarily during the early stages of adult erythropoeisis. Thus, it is possible that linked β-globin genes in rabbit and human have similar functions and organization. The relationship, if any, between the close linkage and sequential expression of these genes can now be approached using cloned genomic DNAs as hybridization probes to nuclear RNA.

The existence of partially characterized functional deficiencies in human hemoglobin expression (thalassemias) provides the opportunity for studying relationships between globin gene organization and function. The feasibility of applying genomic blotting procedures to the study of globin genes from thalassemic patients has been demonstrated using labeled globin cDNA probes (A. Bank and R. Williamson, this Symposium). The availability of cloned genomic globin DNA with its associated non-coding sequences will extend the utility of this approach by making it possible

to detect structural changes in sequences not accessible to cDNA probes.

Once structural differences have been detected by genomic blotting, the genetic defect can be studied at the level of nucleotide sequence by cloning the affected gene. The gene isolation procedure described above and elsewhere (12) should allow the rapid isolation of globin genes from thalassemic patients. Relatively small amounts of DNA are required and the procedure is not effected by changes in restriction endonuclease cleavage sites which flank the affected genes. This is due to the fact that randomly fragmented DNA, rather than specific restriction fragments, is cloned. Thus, thalassemias resulting from large deletions or sequence rearrangements as well as those resulting from single base changes will be accessible to structural analysis. Studies of this kind should identify the molecular basis of thalassemias and possibly provide insight into human genetic disorders in general.

REFERENCES

1. Rougeon F, Kourilsky P, Mach B: Insertion of a rabbit β-globin gene sequence into an E. coli plasmid. Nucl Acid Res 2:2365, 1975
2. Maniatis T, Sim GK, Efstratiadis A, Kafatos FC: Amplification and characterization of a β-globin gene synthesized in vitro. Cell 8:163-182, 1976
3. Higuchi R, Paddock GV, Wall R, Salser W: A general method for cloning eukaryotic structural gene sequences. Proc Nat Acad Sci USA 73:3146, 1976
4. Efstratiadis A, Kafatos FC, Maniatis T: The primary structure of rabbit β-globin mRNA as determined from cloned DNA. Cell 10:571-585, 19
5. Liu AY, Paddock GV, Heindell HC, Salser W: Nucleotide sequences from a rabbit alpha globin gene inserted in a chimeric plasmid. Science 196:192-194, 1977
6. Browne J, Paddock GV, Liu A, Clarke H, Heindell H, Salser W: Nucleotide sequences from the rabbit β-globin gene inserted into Escherichia coli plasmids. Science 195:389, 1977
7. Jeffreys AJ, Flavell RA: A physical map of the DNA regions flanking the rabbit β globin gene. Cell 12:429-439, 1977a
8. Jeffreys AJ, Flavell RA: The rabbit β-globin gene contains a large insert in the coding sequence. Cell 12:1097-1108, 1977b
9. Tilghman SM, Tiemeier DC, Polsky F, Edgell MH, Seidman JG, Leder A, Enquist LW, Norman B, Leder P: Cloning specific segments of the mammalian genome: Bacteriophage λ containing mouse globin and surrounding gene sequences. Proc Nat Acad Sci USA 74:4406-4410, 1977

10. Tilghman SM, Tiemeier DC, Seidman JG, Peterlin BM, Sullivan M, Maizel J, Leder P: Intervening sequence of DNA identified in the structural portion of a mouse β-globin gene. Proc Nat Acad Sci USA 75:724-729, 1978

11. Tilghman SM, Curtis PJ, Tiemeier DC, Leder P, Weissman C: The intervening sequence of a mouse β-globin gene is transcribed within the 15S β-globin mRNA precursor. Proc Nat Acad Sci USA 75:1309-1313, 1978

12. Maniatis T, Hardison RC, Lacy E, Lauer J, O'Connell C, Quon D, Sim GK, Efstratiadis A: The isolation of structural genes from libraries of eukaryotic DNA. Cell (in press)

13. Blattner FR, Williams BG, Blechl AE, Thompson KD, Faber HE, Furlong, LA, Grunwald DJ, Kiefer DO, Moore DD, Schumm JW, Sheldon EL, Smithies O: Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. Science 196:161-196, 1977

14. Hohn B, Murray K: Packaging recombinant DNA molecules into bacteriophage particles in vitro. Proc Nat Acad Sci USA 74:3259-3263, 1977

15. Sternberg N, Tiemeier D, Enquist L: In vitro packaging of a λ Dam vector containing Eco RI DNA fragments of E. coli and phage P1. Gene 1:255-280, 1977

16. Benton WD, Davis RW: Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196:180-182, 1977

17. Ingram VM: The Hemoglobins in Genetics and Evolution. Columbia University Press, 1963

18. Nienhuis AW, Benz EJ: Regulation of hemoglobin synthesis during development of the red cell. The New England J of Med 297:1318-1328, 1371-1381, 1430-1436, 1977

19. Bonn FH, Forget BG, Ranney HM: Human Hemoglobins. W. B. Saunders Co., Philadelphia, 1977

20. Deisseroth A, Nienhuis A, Turner P, Velez R, Anderson WF, Ruddle F, Lawrence J, Creagen R, Kucherlapati R: Localization of the human alpha globin gene to chromsome 16 in somatic cell hybrids by molecular hybridization assay. Cell 12:205-218, 1977

21. Deisseroth A, Nienhuis A, Lawrence J, Giles R, Turner P, Ruddle F: Chromosomal localization of human beta globin gene on human chromosome 11 in somatic cell hybrids. Proc Nat Acad Sci USA 75:1456-1460, 1978

22. Steinheider G, Melderis H, Ostertag W: Mammalian embryonic hemoglobins, International Symposium on the Synthesis, Structure and Function of Hemoglobin. Edited by H Martin, L Novicke. Munich, Lehmans, 1974, pp 222-235

23. Kitchen H, Brett I: Embryonic and fetal hemoglobin in animals. Ann NY Acad Sci 241:653-670, 1974

24. Clissold PM, Arnstein HRV, Chesterton CJ: Quantitation of globin mRNA levels during erythroid development in the rabbit and discovery of new β-related species in immature erythroblasts. Cell 11:353-361, 1977

25. Borsook H: The change of hemoglobin synthesis from minor to major in the course of erythroblast maturation. Ann NY Acad Sci 149:416-422, 1968

26. Russell ES, McFarland EC: Genetics of mouse hemoglobins. Ann NY Acad Sci 41:25-38, 1974

27. Gilman JG, Smithies O: Fetal hemoglobin variants in mice. Science 160: 885-886, 1968

28. Popp RA: Sequence of amino acids in the major and minor β chains of the diffuse hemoglobin from BALB/c mice. Biochim Biophys Acta 303:61-67, 1973

29. Clegg JB, Weatherall DJ: Molecular basis of thalassaemia. Br Med Bull 32:262-269, 1976

30. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503-517, 1975

31. Wilson JT, Wilson LB, deRiel JK, Villa Komaroff L, Efstratiadis A, Forget BG, Weissman SM: Insertion of synthetic copies of human globin genes into bacterial plasmids. Nucleic Acids Res 5:563-581, 1978

32. Kamuzora H, Lehmann H: Human embryonic haemoglobins including a comparison by hemology of the human ζ and α chains. Nature 256:511-513, 1975

33. Shroeder WA, Shelton JR, Shelton JB: World wide occurrence of non-allelic genes for the γ chains of human fetal haemoglobin in newborns. Nature [New Biology] 240:273, 1972

34. Huisman THJ, Shroeder WA, Bannister WH, Grech JL: Evidence for nonallelic structural genes for the γ chain of human fetal hemoglobin. Biochem Gen 7:131-135, 1972

Table 1. The Globin Gene Families of Rabbit and Human

		Embryonic	Fetal	Adult
Rabbit	α	χ		α
	non-α	ε(y) ε(z)		β, (β minor)
Human	α	ζ		α_1, α_2
	non-α	ε	γ^A γ^G	β, δ

Figure 1 Schematic diagram illustrating the strategy used to isolate globin genes from genomic DNA. (Reproduced from reference 12 by permisson.)

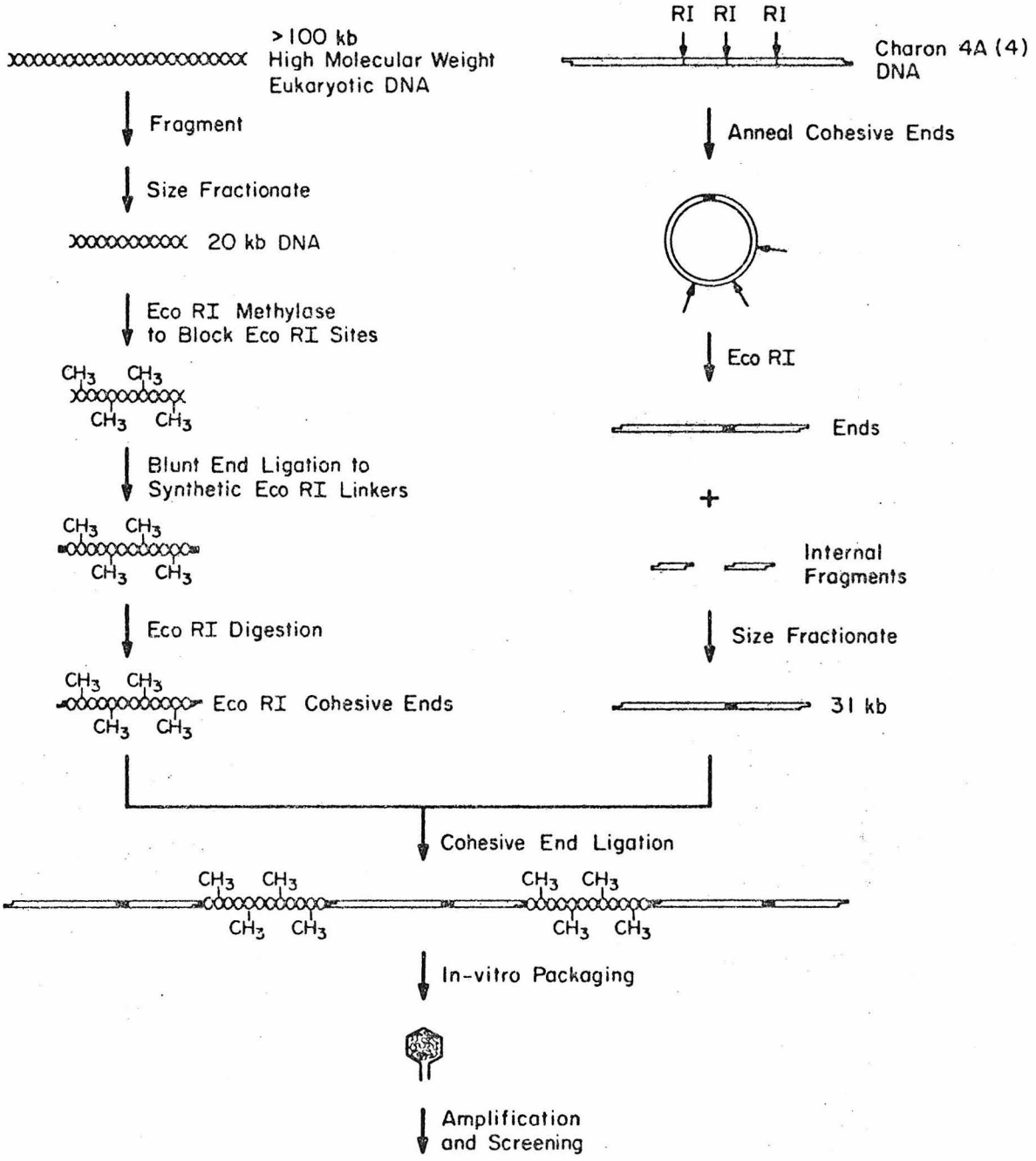

>100 kb
High Molecular Weight
Eukaryotic DNA
Fragment
Size Fractionate
20 kb DNA
Eco RI Methylase
to Block Eco RI Sites
CH3 CH3
CH3 CH3
Blunt End Ligation to
Synthetic Eco RI Linkers
Eco RI Digestion
Eco RI Cohesive Ends
RI RI RI
Charon 4A (4)
DNA
Anneal Cohesive Ends
Eco RI
Ends
+
Internal
Fragments
Size Fractionate
31 kb
Cohesive End Ligation
In-vitro Packaging
Amplification
and Screening

Figure 2 Evidence for the presence of a β-related globin sequence closely linked to the adult rabbit β-globin gene. Cloned rabbit β-globin genomic DNA (RβG-2) was digested with Kpn I, fractionated by electrophoresis on a 1.4% agarose gel, transferred to a nitrocellulose filter (30) and hybridized to ^{32}P-labeled pβG-1 DNA (2). The column on the left shows the ethidium bromide stained gel. The column on the right shows an autoradiogram of the nitrocellulose filter. The size of the globin-containing DNA fragments is indicated in kilobase pairs.

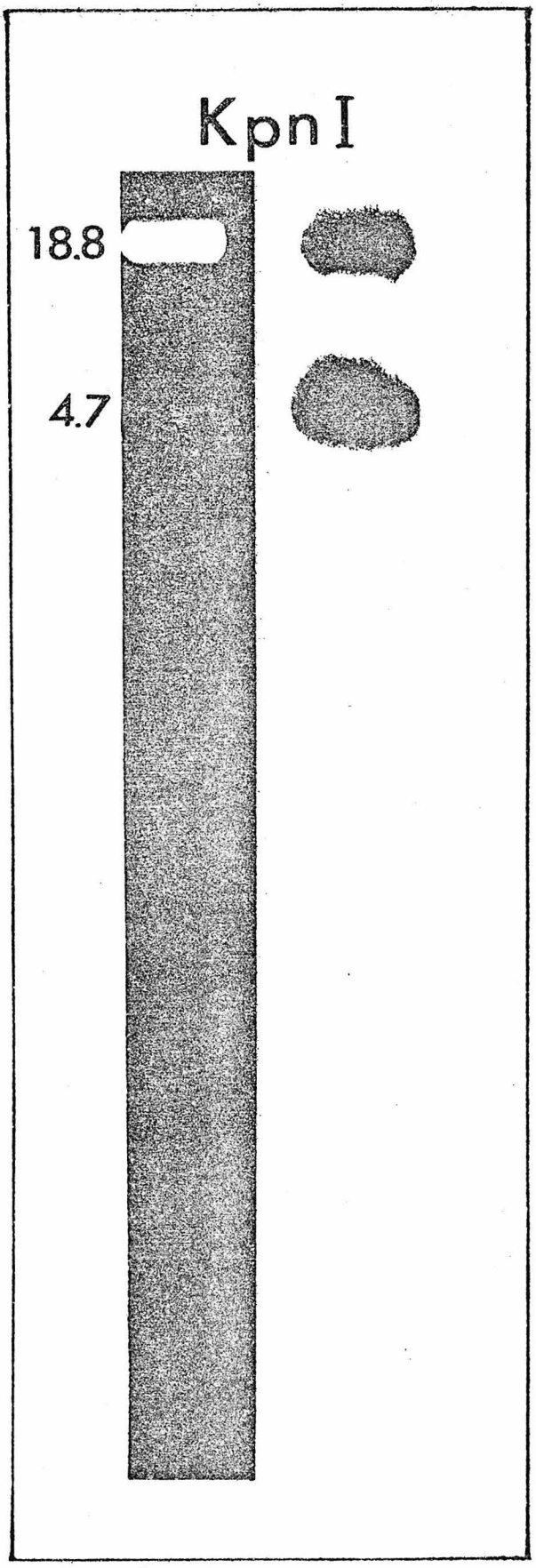

Kpn I
18.8
4.7

Figure 3 Bam HI and Kpn I restriction maps of the rabbit β-globin genomic clone RβG2. The vertical lines indicate the positions of Bam HI and Kpn I recognition sites. The approximate number of kilobase pairs between restriction sites is also indicated. The hatched box represents the approximate position of the β-related sequence in the cloned DNA. The interrupted solid box represents the position of the adult β-globin gene. The clear area indicates the presence of a 600 bp non-coding intervening sequence (8). A smaller (125 bp) intervening sequence located nearer to the 5' end of the coding region (between codons 30 and 31) is not shown (unpublished). Experiments have not yet been performed to determine whether the minor β gene contains an intervening sequence. The horizontal arrow (5' → 3') indicates the orientation of the β-globin gene with respect to the direction of transcription. The orientation of the β-related gene is not yet known. The vertical arrows show the junction between the rabbit DNA insert and λ Ch 4A DNA.

λCh4A RβG2 – 48.5kb

Rabbit DNA insert – 18.9kb

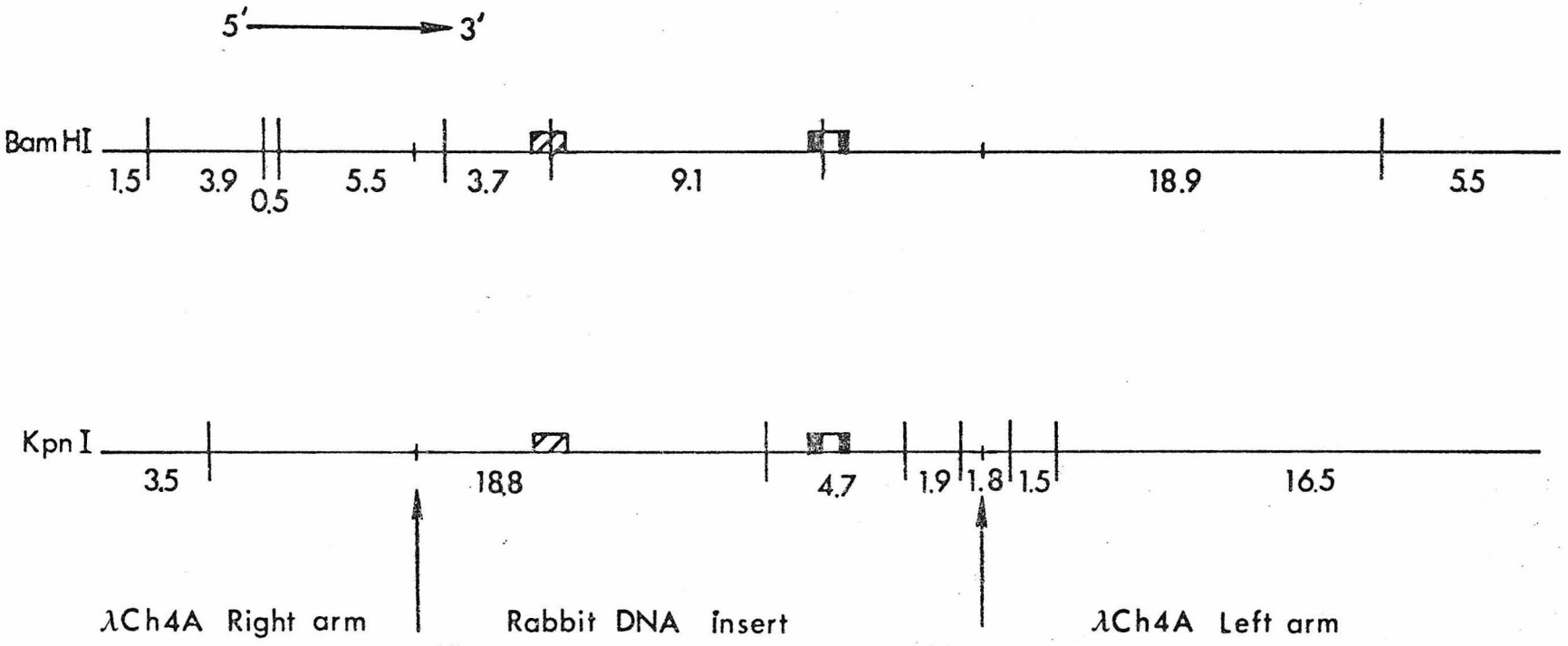


Figure 4 Detection of β-globin related sequences in restriction endonuclease digests of cloned human DNAs. DNA from clones HβG-1 and HβG-2 was digested with restriction endonucleases Bam HI (left lanes), Xba I (center lanes), Eco RI (right lanes), and Bgl I and Hha I (data not shown) and electrophoresed on a 0.7% agarose gel. The separated fragments were transferred to nitrocellulose filter paper (30) and fragments containing β-globin related sequences were detected by hybridization to ^{32}P-labeled nick-translated human β-globin cDNA plasmid and subsequent autoradiography. Included in the gel were hybridization markers consisting of rabbit β-globin cDNA plasmid (pβG-1; ref 2) digested separately with Eco RI and Bam HI. The sizes of the markers are 5.7 kb, 4.8 kb, and 0.9 kb. Digestion of HβG-1 and HβG-2 DNAs with Bgl I or Hha I generated, in each case, a single large hybridizing fragment (data not shown) similar to the pattern observed for Xba I. The sizes of all hybridizing bands were determined from the ethidium bromide stained pattern of the gel which included a large number of molecular weight markers.

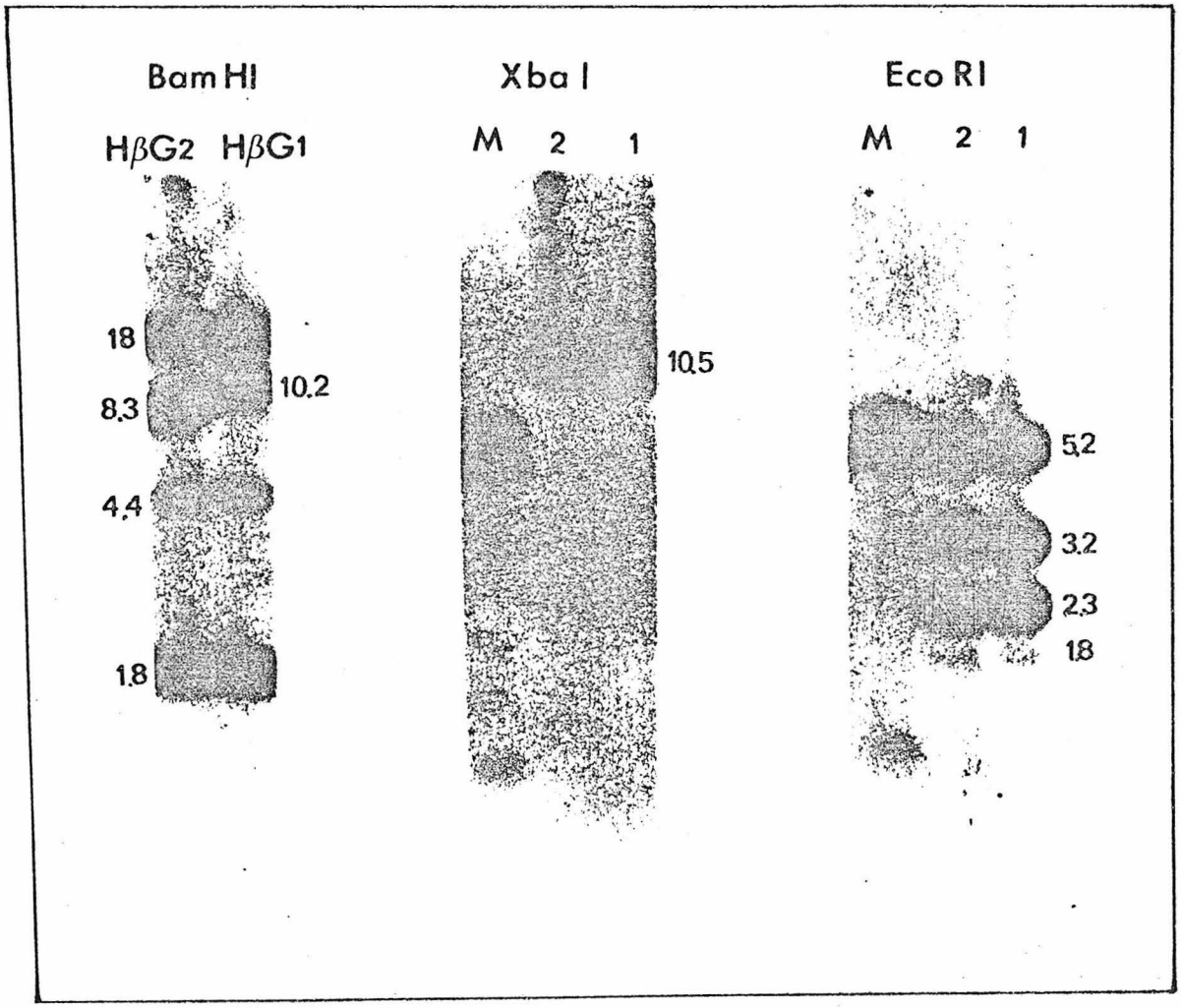

Bam HI
HβG2 HβG1
18
8.3
4.4
1.8
10.2
Xba I
M 2 1
10.5
Eco RI
M 2 1
5.2
3.2
2.3
1.8

Figure 5 Bam HI, Eco RI, Xba I and Hind III restriction maps of the human β-globin genomic clone HβG-2. The vertical lines indicate the position of various restriction enzyme recognition sites. The approximate number of kilobase pairs is also indicated. The interrupted solid boxes represent the position of the presumptive δ and β globin genes. The clear areas indicate the presence of large (900-950 bp) non-coding intervening sequences. The possible existence of smaller intervening sequences has not yet been investigated. The horizontal arrow indicates the direction of transcription for both genes. The vertical arrows show the junction between the rabbit DNA insert and Charon 4A DNA.

λCh4A HβG2 - 45kb

Human DNA insert – 14.5kb