19. W. D. Henner, I. Kleber, R. Benzinger, *J. Virol.* **12**, 741 (1973).
20. Strain NS428, N100 (λ*A*am 1 1*b*2*red*3*c*I-857*S*am7), was a gift from N. Sternberg and L. Enquist. Strain dg805, W3350 (λdgal805*c*I-857*S*am7) was the gift from A. Becker. Strain NS428 was grown with shaking at 30° to 33°C in 600 ml NZY broth cultures (*7*) to A_{575} = 0.2, as measured on a Bausch and Lomb spectrophotometer (Spec 20). An additional 600 ml of 64°C NZY broth was added to the cultures to bring the temperature to 45°C, and the flask was incubated at 39°C with shaking for 1 hour (New Brunswick model G25). To test for successful induction a few milliliters of culture were shaken with chloroform and checked for the expected clearing. Freeze thaw lysate (FTL) was prepared by resuspending each pellet in 0.4 ml of cold 10 percent sucrose in 0.05*M* tris-HCl, *p*H 7.4, and transferring it to a 10-ml Oak Ridge type polycarbonate centrifuge tube. Egg white lysozyme (20 μl) (2 mg/ml, freshly dissolved in 0.25*M* tris-HCl, *p*H 7.4) was added with mixing, and the tubes were frozen in liquid nitrogen. Sonic extract (SE) was prepared by resuspending the NS428 pellet in 1.8 ml of buffer A (20 m*M* tris-HCl, *p*H 8, 3 m*M* $MgCl_2$, 0.05 percent 2-mercaptoethanol, 1 m*M* EDTA-KOH, *p*H 7), and was completely sonicated with cooling in an ice-salt bath until an opalescent solution free of intact cells was obtained. The preparation was centrifuged for 6 minutes at 4500*g*, and the supernatant was either used immediately or stored in liquid nitrogen. Protein A (pA) was prepared (A. Becker, personal communication) from six induced cultures of λdg805 prepared as described above for NS428, except that the heating was initiated at an absorbancy of 0.8. The induced cells were centrifuged and resuspended in 100 ml of buffer A and sonicated with cooling; the debris was removed by centrifugation. Polyethyleneimine (Miles, code 23-444) was dissolved to a final concentration of 10 percent in buffer A, and the solution was adjusted to *p*H 8 with HCl. This polyethyleneimine solution was added with stirring in the cold until no further precipitate formed (approximately 10 ml). The precipitate was collected by centrifugation and resuspended with an electric tissue homogenizer in 500 ml of 0.05*M* ammonium succinate (*p*H 6) containing 0.035 percent (by volume) 2-mercaptoethanol and centrifuged again; the pellet was resuspended in 500 ml of 0.1*M* ammonium succinate (as above) and centrifuged again. The pellet was resuspended in 0.2*M* ammonium succinate. The solution was again centrifuged, and the supernatant was saved. Neutralized ammonium sulfate (NAS) was prepared by adjusting saturated ammonium sulfate containing 0.1 percent 2-mercaptoethanol to *p*H 6.5; then 220 ml of NAS was added to the supernatant with stirring in the cold for 20 minutes. The precipitate was removed by centrifugation and 290 ml of NAS plus 15 g of solid ammonium sulfate was added to the supernatant. The precipitate was stirred for 30 minutes in the cold, collected, and resuspended in 5.5 ml of buffer A without $MgCl_2$ and dialyzed against the same buffer. One volume of glycerol was added, and the preparation was subdivided and stored in liquid nitrogen. Immediately before packaging, the FTL was thawed (<5°C) and held on ice for 45 minutes. Next, 0.05 ml of buffer M1 [6 m*M* tris-HCl (*p*H 7.4), 30 m*M* spermidine, 60 m*M* putrescine, 18 m*M* $MgCl_2$, 15 m*M* ATP (from a 0.1*M* stock at *p*H 7), 0.2 percent (by volume) 2-mercaptoethanol] was added, mixed, and centrifuged at 35,000 rev/min at 4°C for 25 minutes in a precooled type 65 rotor. Buffer A (30 μl) was mixed with 4 μl DNA (<1 μg), 4 μl buffer Ml, 20 μl thawed sonic extract, and 1 to 2 μl pA and incubated 15 minutes at room temperature. FTL (150 μl) was added and the mixture was incubated for 60 minutes at room temperature. The resulting phages were then plated.
21. To verify that the safety features of the Charon 3A vector were not compromised by the use of in vitro packaging, two tests were done on each batch of extracts as required by the NIH. (i) To determine whether endogenous DNA from the extracts was packaged, sham experiments were done in which no exogenous DNA was added to a 10× scale packaging mixture (2 ml final volume). The entire output was plated on 20 plates with lawns of bacteria permissive for the prophage in NS428. In five such experiments no phages were seen ($<1.5 \times 10^{-7}$ to $<7.5 \times 10^{-9}$ of the level that would be produced if vector DNA had been added). (ii) To determine whether in vitro packaging could result in loss of the amber mutations of the vector, 10× scale packaging experiments were done with vector DNA alone. In this case, the entire output was plated on a *su*° bacterial lawn to determine levels of amber$^+$ revertants. In five such experiments the highest level of reversion seen was 2×10^{-7}.
22. Cafeteria trays (35 by 45 cm) were lined with aluminum foil, sterilized with ethanol and ultraviolet and filled with 1.5 liters of bottom agar (*7*). After hardening in a still air hood, they were covered with sterile glass and transferred to P3 containment. Approximately 5×10^5 phages were mixed with 8 ml of stationary phase bacteria plus 8 ml of 0.01*M* $MgCl_2$ and 0.01*M* $CaCl_2$ and incubated at 37°C for 10 minutes. The mixture was then added to 150 ml of melted top agar (*7*), mixed, spread evenly over the warm, level surface of the megaplate and allowed to harden for 30 minutes before incubation at 37°C overnight. After development of the lawn the plate was chilled at least 12 hours in a refrigerator. The glass tops were lined with sheets of paper to reduce condensation. Nitrocellulose sheets type BA85 (33 by 45 cm) or rolls (33 cm by 3 m) (Schleicher and Schuell) were cut to size, rinsed in distilled water and 6× SSC (1× SSC is 0.15*M* NaCl, 0.015*M* sodium citrate, *p*H 7.2) and blotted on 3MM paper. These were placed on the surface of the chilled plate for 4 minutes. To aid in subsequent alignment, a series of marks spaced every 3 inches were made by penetration with a syringe needle dipped in India ink. The filter was then transferred to a bed of 3-MM paper saturated with 1.5*M* NaCl, 0.2*N* NaOH for 4 minutes and then neutralized for 4 minutes on a bed of the same paper saturated with 0.5*M* tris-HCl, *p*H 7.2, 3*M* NaCl, blotted and allowed to dry. Up to five filters can be lifted from a single plate with chilling between transfers. The nitrocellulose was then incubated between 3MM sheets for 2 hours at 80°C, and washed for 1 hour with gentle shaking at 68°C in 6× SSD (6× SSC, 0.02 percent Ficoll, 0.02 percent polyvinylpyrrolidone, 0.02 percent bovine serum albumin, 0.5 percent sodium dodecyl sulfate) [see D. T. Denhardt, *Biochem. Biophys. Res. Commun.* **23**, 641 (1966)].
23. The hybridization mixture contained poly(rA) (30 μg/ml; Sigma code P-8878), denaturated sonicated *E. coli* DNA (10 μg/ml) in 6× SSD. For each megaplate filter, 1×10^6 count/min of probe (*24*) was added to 300 ml of the mixture in a flat-bottomed photography tray. The tray was covered with a glass sheet, sealed within a plastic garbage bag, and incubated for 24 hours, with gentle shaking at 68°C. The filter was washed two or three times at 68°C (1 hour each) with shaking in 1 liter of 3 × SSC containing 0.5 percent sodium dodecyl sulfate, blotted, dried in air, and mounted for autoradiography beneath Kodak XR-1 film and a DuPont Quanta III intensifier screen. Exposure was for 2 days at −90°C. For Southern transfers of mini lysate gels (*27*), the hybridization procedure was the same, except that the concentration of probe was about 3×10^4 count/min per milliliter. For Southern transfers of genomic DNA gels (*26*), about 5×10^5 count/min per milliliter probe was used and hybridization was for 40 hours.
24. Human globin mRNA was provided by Drs. Arthur Bank and Francesco Ramirez and mouse globin mRNA was purified from mouse reticulocytes and provided by Dr. Jeffrey Ross. Mouse or human unfractionated mRNA was converted to [α-^{32}P]-labeled cDNA essentially by the method described [D. L. Kacian and J. C. Myers, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2191 (1976)]. The plasmids pRGH-1 [P. H. Seeburg, J. Shine, J. A. Martial, J. D. Baxter, H. M. Goodman, *Nature (London)* **270**, 486 (1977)] and pAU-1 [A. Ullrich, J. Shine, J. Chirgwin, R. Q. Pictet, E. Tischer, W. J. Rutter, H. M. Goodman, *Science* **196**, 1313 (1977)], containing rat growth hormone and rat insulin sequences, respectively, were a gift of Dr. Howard Goodman. The DNA's were made radioactive essentially according to the procedure described by T. Maniatis, A. Jeffrey, and D. G. Kleid [*Proc. Natl. Acad. Sci., U.S.A.* **72**, 1184 (1975)]. Double-stranded probes were denatured with NaOH.
25. J. N. Mol, R. A. Flavell, P. Borst, *Nucleic Acids Res.* **3**, 2367 (1967).
26. A. J. Jeffreys and R. A. Flavell, *Cell* **12**, 429 (1977).
27. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
28. *Note added in proof*: Similar work has just been published by T. Maniatis, R. C. Hardison, E. Lacey, C. O'Connell, D. Quon, G. K. Sim, A. Efstratiadis, *Cell* **15**, 703 (1978).
29. For rapid screening of clones by agarose gel electrophoresis of restriction digests, DNA was prepared directly from lysates. 0.4 ml SDS mix (0.25*M* EDTA, 0.5*M* tris-HCl, *p*H 9, 2.5 percent recrystallized sodium dodecyl sulfate) was added to 2 ml of clarified phage lysate and heated for 30 minutes at 70°C. At this point the samples contained no live phage and could be removed from P3 containment. Then 0.5 ml of 8*M* potassium acetate was added and the mixture was placed on ice for 15 minutes and centrifuged for 20 minutes at 12,000*g*. The supernatant was precipitated with two volumes of ethanol and centrifuged for 30 minutes at 27,000*g*. After the traces of ethanol were removed with a cotton-tipped swab, the pellet was dissolved in 0.4 ml of 0.3*M* sodium acetate, and precipitated again with ethanol. The pellet was dissolved in 50 μl of 1m*M* EDTA, *p*H 8; 1 to 5 μl of the preparation was digested with 2 units of restriction enzyme before agarose gel electrophoresis [T. M. Shinnick, E. Lund, O. Smithies, F. R. Blattner, *Nucleic Acids Res.* **2**, 1911 (1975)].
30. This is paper 2281 from the Laboratory of Genetics at the University of Wisconsin and paper 8 in the series, "Charon Phages for DNA Cloning." Supported by NIH grants GM21812 (F.R.B.), AM20120 and GM20069 (O.S.), CA09075 (K.D.T.), GM06526 and GM07131 (P.W.T.), and GM07133 (J.E.R. and A.E.B.); and an NIH research career development award (to F.R.B). We thank Howard Goodman for his collaboration with us in developing these procedures while visiting in our laboratory; Andrew Becker for hospitality in his laboratory; and D. O. Kiefer, N. Borenstein, T. C. Szeto, E. Kopetsky, and J. Kucera for technical assistance. This work was done under the NIH guidelines which require EK2, P3 containment.

12 October 1978

Cloning Human Fetal γ Globin and Mouse α-Type Globin DNA: Characterization and Partial Sequencing

Abstract. *Two globin-related clones isolated from collections of bacteriophages containing unfractionated Eco RI fragments of human and mouse DNA were characterized. Charon3AHs51.1Hbγ includes 2.7 kilobase pairs of human DNA containing a large part of a fetal γ globin chain structural gene; Charon 3AMm30.5 includes 4.7 kilobase pairs of mouse DNA related to α globin. The human fetal γ globin gene has within its coding region two intervening sequences of noncoding DNA, IVS 1 and IVS 2, of approximately 100 and 900 base pairs. Sequence IVS 1 is located at the position of one of the two intervening sequences occurring in adult globin genes; IVS 2 is located at the position of the other.*

Mammals utilize various α-type and β-type globin chains to make different hemoglobins at different stages of their lives. In humans, there are two α-type chains: the embryonic chain ζ, and the adult chain α; the β-type chains are the embryonic chain ε, the fetal chains ^{A}γ and ^{G}γ, the major adult chain β, and the minor adult chain δ (*1*). Various α- and β-type chains also occur at different stages in the development of mice (*2*). We hope to learn something about the control of

0036-8075/78/1222-1284$00.50/0

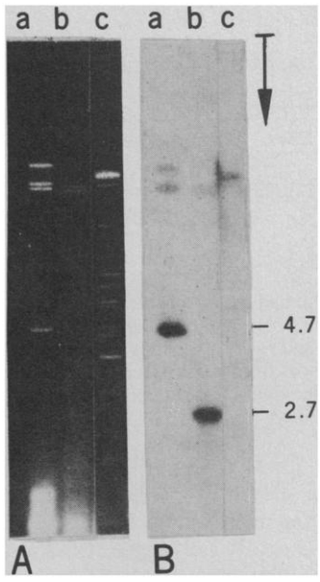

Fig. 1. Agarose gel electrophoresis (5) of Eco RI restriction enzyme digests of DNA from two small-scale phage lysates (3). (A) Stained gel; (B) autoradiogram of its Southern transfer (6) after hybridization to ^{32}P-labeled human globin cDNA. The gel samples were Eco RI digests of DNA from (sample a) the mouse globin clone Mm30.5, (sample b) the human globin clone Hs51.1. Sample c contains DNA fragments of known size (8). The Eco RI fragments of 2.7 and 4.7 kbp length, which hybridize to the cDNA probe, are indicated.

the differential expression of the genes for these several globin chains by studying the organization and sequences of their structural genes and of neighboring stretches of DNA as they occur in the genome. Therefore, we have started experiments to clone and characterize globin genes from the DNA of humans and mice.

The construction and characterization of "shotgun" collections of λ bacteriophages containing unfractionated Eco RI restriction enzyme fragments of human and mouse genomic DNA have been described (3). These collections are presumed to contain random fragments of the genome, except for the nonrandomness imposed by the choice of restriction enzyme and by the range of fragment size which the vector phage accepts (4). We also described (3) the screening of these human and mouse shotgun collections, which yielded two phages, Hs51.1 and Mm30.5; these hybridize to unfractionated mixtures of human and mouse globin complementary DNA (cDNA), respectively. We now show that these phages contain portions of the structural gene for a human fetal γ globin chain and DNA related to mouse α-type globin. We also show that the human fetal γ globin gene has intervening DNA sequences within its coding region at the same positions as they occur in adult globin genes.

Restriction endonuclease studies were carried out as an initial step in characterizing the fragments in the two globin-related phages, Hs51.1 and Mm30.5. Small lysates were prepared from the phages (3, reference 29). The DNA from these lysates was digested with restriction enzymes and examined by electrophoresis in 0.7 percent agarose gels (5). The DNA was transferred to nitrocellulose filters by the Southern method (6) and the filters were hybridized (7) to globin cDNA probes (3, reference 24).

Eco RI digests of the mouse globin clone Mm30.5 (sample a) and the human clone Hs51.1 (sample b) were compared to marker DNA fragments of known size (8) (sample c) (Fig. 1). Mm30.5 (sample a) contains an Eco RI fragment of mouse DNA, which hybridizes strongly to human globin cDNA (and to mouse globin cDNA in other experiments). The size of the fragment is 4.7 kilobase pairs (kbp), as determined by comparison with the marker DNA fragments. Hs51.1 (sample b) contains an Eco RI fragment of human DNA, of size 2.7 kbp, which hybridizes to human globin cDNA. Thus both clones contain globin-related Eco RI fragments.

As shown in other experiments, Eco RI digestion of Hs51.1 yields a small human DNA fragment (0.465 kbp) in addition to the major 2.7-kbp fragment illustrated (Fig. 1). The smaller fragment does not hybridize to globin cDNA. It is probably an Eco RI fragment of unrelated human DNA (see below) adventitiously cloned into the same phage as the globin-related larger fragment. On statistical grounds, phages with two unrelated fragments would be expected to occur reasonably frequently in our shotgun collections (3).

Both Hs51.1 and Mm30.5 were isolated by the use of probes made from mixtures of α- and β-type globin messenger RNA's (mRNA). To determine whether Hs51.1 and Mm30.5 contain α- or β-type globin DNA, we compared the hybridization of their inserts to DNA from phages carrying only α or only β globin cDNA. The 2.7 kbp and 4.7 kbp DNA inserts in the genomic clones Hs51.1 and Mm30.5 were first partially purified by agarose gel electrophoresis, and were then converted into radioactive probes by nick translation (9). DNA from a mouse α globin cDNA-containing phage, 117Hbα (10), from a mouse β globin cDNA-containing phage, 4Hbβ (10), and from their parent phage, Charon3AΔlac (4) were digested with Hpa I restriction enzyme and subjected to electrophoresis

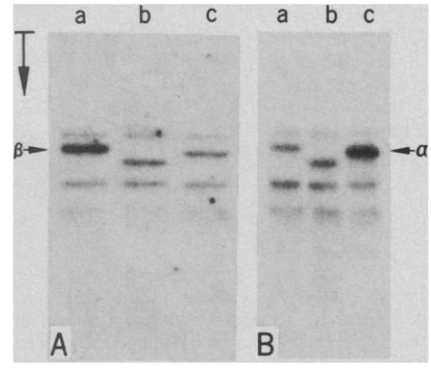

Fig. 2. Autoradiograms of filter transfers (6) made from two electrophoresis gels. One filter (A), was hybridized to a probe made by nick translating the partially purified 2.7-kbp DNA insert from the human clone Hs51.1; the other filter (B) was hybridized to the nick translated 4.7-kbp insert from the mouse clone, Mm30.5. The gels from which the filters were made both contained Hpa I digests of DNA from: (sample a) a mouse β globin cDNA phage, 4Hbβ (10); (sample b), the parent phage, Charon3AΔlac (4); (sample c), a mouse α globin cDNA phage, 117Hbα (10). The arrows labeled α and β point to bands which contain α and β mouse globin cDNA. Hs51.1 hybridizes preferentially to β, Mm30.5 to α.

in two 0.7 percent agarose gels. Nitrocellulose filter transfers of the gels were made, and individually hybridized to the two probes. Autoradiographs made from the filters are shown in Fig. 2A (for the probe made from Hs51.1) and Fig. 2B (for the probe made from Mm30.5). Because the two partially purified probes, in addition to their ^{32}P-labeled inserts, also contain ^{32}P-labeled phage DNA as a contaminant, and because all the fragments in the Hpa I digests contain phage DNA, hybridization occurs to all the bands on the filters. However, only one band (the second from the top of the gel in the Hpa I digests of 117Hbα and 4Hbβ) contains globin cDNA in addition to phage DNA. The probe derived from Hs51.1 hybridizes much more strongly to this cDNA-containing band from 4Hbβ (Fig. 2A, slot a, band labeled β) than to any band from the parent phage Charon3AΔlac (slot b), or from 117Hbα (slot c); the human DNA in clone Hs51.1 is consequently of a β-type. Similarly the probe made from Mm30.5 hybridizes most strongly to the cDNA-containing band from 117Hbα (Fig. 2B, slot c, band labeled α); the mouse DNA in clone Mm30.5 is consequently of an α-type.

The mouse α-type globin-related clone, Mm30.5, has been further characterized by DNA sequencing. To date we have sequenced about 500 base pairs of DNA from its 4.7-kbp insert in five stretches of 200, 100, 75, 75, and 50 base pairs. We have not yet been able to cor-

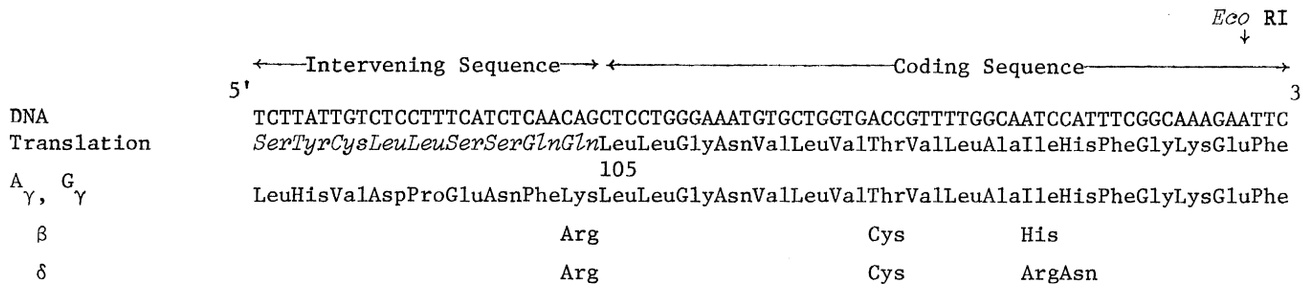

Fig. 3. The DNA sequence at the 3′ end (in the transcriptional sense) of the 2.7-kbp Eco RI fragment of human DNA in Hs51.1 is given in the top part of the figure together with a hypothetical amino acid sequence that could be translated from this DNA. Normal type indicates agreement of this hypothetical amino acid sequence with the actual amino acid sequence of the human fetal chains, $^{A}\gamma$ and $^{G}\gamma$ (*13*), which is shown in the figure; italics show disagreement. Positions at which the human β and δ globin chains differ in amino acid sequences from those of the γ globin chains are also shown. The intervening DNA sequence preceding Leu at position 105 in the coding sequence, and the Eco RI site at codons 121 and 122 are indicated.

relate these DNA sequences unequivocally with any published α-type globin amino acid sequences. Consequently, our classification of Mm30.5 as containing DNA related to α globin rests on its hybridization characteristics. We cannot at present choose between the possibilities that Mm30.5 contains DNA from an adult α globin gene, from an embryonic α-type globin gene, or from some part of the genome that has sufficient homology to a globin cDNA to cause the observed hybridization.

Specific identification of the human globin gene from which Hs51.1 is derived has been accomplished by partial DNA sequencing (*11*) by the method of Maxam and Gilbert (*12*). Figure 3 shows the resulting DNA sequence data from one end of the 2.7-kbp fragment. The figure also compares a hypothetical amino acid sequence which could be translated from this DNA with the actual amino acid sequences from the human β-type globins: $^{A}\gamma$, $^{G}\gamma$, β, and δ (*13*). The DNA-derived amino acid sequence and that of the two fetal γ globin chains are in complete agreement from the leucine (Leu) residue at position 105 to the phenylalanine (Phe) at 122, whereas the amino acid sequences of the β and δ globin chains disagree at several positions. We consequently conclude that Hs51.1 includes part of the structural gene for a human fetal γ globin chain, and that the 3′ end (in the transcriptional sense of the globin gene) of the 2.7-kbp piece is an Eco RI site formed by the codons for glutamic acid (Glu) and Phe at positions 121 and 122 of the structural gene. Further sequence (data not shown) from the Bam HI site at codons 98 to 100 (Fig. 5), when translated into a hypothetical amino acid sequence, is also in complete agreement with the amino acid sequence of the fetal γ globin chain from codons 98 through 104 (*13*).

We have also been able to make a direct comparison of the base sequence of our genomic fragment with the base sequence of a clone of human fetal $^{G}\gamma$ globin cDNA (*14*). The sequences are in complete agreement from codon 105 through codon 122, which confirms our conclusion that Hs51.1 includes part of a fetal γ globin chain gene. There are 14 differences in this region when a comparison is made with the comparable base sequence either of human β globin mRNA or its cDNA (*15*), or of the human δ globin gene (*16*).

Further inspection of Fig. 3 shows that the sequence of the DNA from Hs51.1 which is immediately 5′ to the Leu codon at 105 diverges completely from that expected from the sequence of the fetal globin chains that is amino terminal to position 105, or from any of the other β-type globin sequences. Further sequence (data not shown) of the DNA immediately 3′ to codon 104 (lysine) also shows complete divergence from the sequence expected for β-type globin chains.

The presence in eukaryotic genes of intervening sequences not represented in the translated mRNA has been described (*17*). Two intervening sequences are present in the structural gene for mouse adult β globin major; the larger of the two, 646 base pairs in length, occurs between the DNA coding for the amino acids at positions 104 and 105 (*18*), which is exactly at the place where we observe the divergent sequences. Thus, our DNA sequence data show that the human fetal γ globin chain gene in Hs51.1 also contains an intervening DNA sequence at the same position as the larger one in the mouse adult β globin chain gene.

Further evidence supporting the existence of this intervening DNA sequence in the coding region of the fetal γ globin structural gene cloned in Hs51.1 was obtained by electron microscopy and by restriction enzyme mapping studies.

Heteroduplexes (*19*) between DNA from phage Hs51.1 and DNA from the well-characterized cDNA-containing phage 4Hbβ (*10*) were studied in the electron microscope. Phage 4Hbβ contains a nearly full-length cDNA copy of mouse β globin mRNA. The orientation of the cDNA in 4Hbβ, designated u (*10*), is such that its 5′ end, in the sense of the corresponding globin mRNA, is toward the right arm of the vector bacteriophage (5′ below refers to the sense of globin mRNA). The total length of DNA inserted in 4Hbβ is 1092 ± 41 base pairs.

A typical heteroduplex between Hs51.1 and 4Hbβ is shown in Fig. 4A, together with an accompanying line drawing interpretation in Fig. 4B. The photograph is printed so that the 5′ to 3′ orientation of the globin sequence in 4Hbβ is from left to right. The heteroduplex consists of three regions of double-stranded DNA separated by two regions where the single strands do not pair. Regions 1 and 7 are double-stranded DNA corresponding, respectively, to the right and left arms of the phages. They were identified by their lengths (*20*), and they end at the Eco RI site of the phage Charon3AΔlac used to construct both Hs51.1 and 4Hbβ. The single-stranded region 2 is longer (1343 ± 68 bases) than the total insert in the phage 4Hbβ (1092 ± 41 base pairs) and so must be human DNA; the human DNA in region 2 does not form a duplex with the single-stranded region 3 of 4Hbβ. The length of region 3 (192 ± 30 bases), its position in 4Hbβ, and the known structure of 4Hbβ show that region 3 must include the polythymidylate [poly(T)] or polydeoxyadenylate [poly (dA)] tails of 4Hbβ and possibly a small length of transcribed sequences 5′ to the coding region of the mouse β globin gene. Region 4 is a duplex region between Hs51.1 and 4Hbβ. The existence of this duplex indicates that Hs51.1 is homologous with the 5′ part of the coding region of 4Hbβ and that Hs51.1 has the same orientation, u, in Charon3AΔlac as does 4Hbβ (*10*). The length of the region

(324 ± 29 base pairs) and its position relative to 4Hbβ show that Hs51.1 must contain about 100 codons of coding sequence from the 5′ half of the fetal γ globin structural gene. The 3′ end of the duplex region is more than 1000 bases from the 3′ end of the human DNA in Hs51.1, as indicated by the length of region 5 (1365 ± 55 bases). [Region 5 must derive from Hs51.1 because it is too long to have derived from 4Hbβ.] The failure of the human DNA in region 5 to form a duplex with the mouse β globin cDNA in region 6 shows the presence of a large length of DNA in the human globin gene, on the 3′ side of the first 100 codons, which is not homologous to mouse globin cDNA. The 3′ end of region 5 in the heteroduplex also contains the 465 base pair Eco RI fragment of nonglobin DNA (Fig. 4B). [This fragment is presumed to be nonglobin DNA because it fails to hybridize to globin cDNA and because of its base sequence (*21*); its location in Hs51.1 was determined by the restriction enzyme mapping data presented below.]

The DNA sequence data from Hs51.1 showed that the cloned human insert contains DNA coding for amino acids 98 through 104 and 105 through 121 of the γ globin gene, and that this latter coding DNA is at the 3′ end of the 2.7-kbp Eco RI fragment. The heteroduplexes show that Hs51.1 also has coding DNA corresponding approximately to amino acids 1 through 104, but that this DNA is not located close to the 3′ end of the clone. Together, the sequence data and heteroduplexes provide the basis for the interpretation of the heteroduplexes presented in Fig. 4B in which a large intervening DNA sequence (shown by a broken line in the figure) is placed between codons 104 and 105. The length of this intervening sequence can be estimated from the length of region 5 in the heteroduplex minus the lengths of the nonglobin fragment and minus the DNA for codons 105 through 121. This estimate is 849 ± 57 base pairs [that is, (1365 ± 55) − (465 ± 10) − 3(121 − 104)].

Confirmation of the existence of the large intervening sequence and an independent estimate of its length were obtained by restriction enzyme mapping. A single Bam HI restriction enzyme site occurs at the codons 98 to 100 for valine (Val), aspartic acid (Asp), and proline (Pro) in the coding DNA of human fetal γ globins (*14*) and in Hs51.1. If Hs51.1 contains all the coding region of fetal γ globin up to amino acid 121, then Bam HI should cut the 2.7-kbp Eco RI fragment of Hs51.1 at this site, and the coding sequence from codon 1 to codon 99 should be on one piece of DNA, which should hybridize strongly to globin cDNA probe. If there is no intervening sequence in the coding region of Hs51.1 between codons 100 and 121, then there should also be a single, second weakly hybridizing piece, 67 base pairs in length, containing these codons. If there is an intervening sequence in the coding region between codons 104 and 105, then there should be a weakly hybridizing piece, or pieces, if there is a Bam HI site in the intervening sequence, of substantially greater length.

We find that Bam HI cuts the 2.7-kbp fragment once to give two pieces. The strongly hybridizing one is 1740 ± 50 base pairs in length, and restriction mapping experiments show that it includes the 5′ end of the 2.7-kbp fragment. The weakly hybridizing one is 960 ± 15 base pairs in length, and it is from the 3′ end of the 2.7-kbp fragment. This is much greater than the 66 base pairs length expected if there is no intervening sequence. These restriction enzyme data consequently provide further evidence for the existence of a large intervening sequence in the fetal γ globin structural gene between codons 104 and 105 and provide a second measure of its length as 894 ± 15 base pairs [that is, (960 ± 15) − 3(121 − 99)]. This length agrees with that obtained from the heteroduplexes (849 ± 57 base pairs).

Experiments were performed, on Hs51.1 end-labeled with ^{32}P at the single Bam HI site in the human DNA, to determine whether there is a second intervening sequence in the human fetal γ glo-

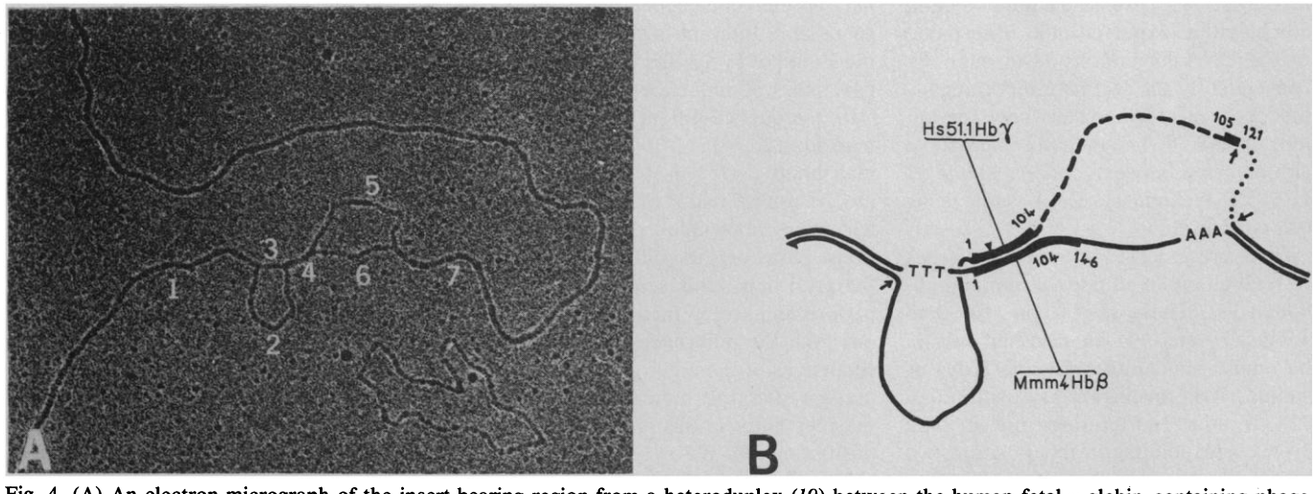

Fig. 4. (A) An electron micrograph of the insert-bearing region from a heteroduplex (*19*) between the human fetal γ globin-containing phage Hs51.1Hbγ, and the mouse β globin cDNA phage, Mmm4Hbβ (*10*). [DNA from a single-stranded phage (G4) and a double-stranded plasmid (pColE1) were included as length standards.] Double-stranded regions 1 and 7, from their measured lengths, are the phage right and left arms (*20*). The single-stranded region 2 contains human DNA (see text); its length, measured in 22 heteroduplexes, is 1343 ± S.E.M. of 68 bases. Region 3 (192 ± 30 bases) is single-stranded DNA from 4Hbβ. Region 4 (324 ± 29 base pairs) is a duplex between the human DNA in Hs51.1 and part of the transcribed sequences in the mouse β globin gene. The single-stranded region 5 (1365 ± 55 bases) of the human DNA cloned in Hs51.1 does not form a stable duplex with the single-stranded region 6 (649 ± 42 bases) of 4Hbβ. (B) Interpretative diagram of the heteroduplex, compiled from the known structure of Mmm4Hbβ (*10*) and from the data presented in Fig. 3 and (*21*). Double-stranded regions are shown by parallel solid lines. The Eco RI sites in Hs51.1 are indicated by arrows. The heavy black bars show DNA regions coding for the numbered amino acid sequences. The broken line indicates a large intervening sequence. The dotted line shows the 465 base pair fragment (*21*) of nonglobin human DNA in Hs51.1. The tails of 4Hbβ are shown by AAA and TTT. No duplex would be expected between the mouse β globin and human γ globin DNA for codons 105 through 121 since their sequences differ in 17 out of 51 bases. [See (*18*) for the mouse sequence data.] The small filled triangle above the duplex region of Hs51.1 shows the position of a second small intervening sequence (see text) which could not be detected in the electron micrographs.

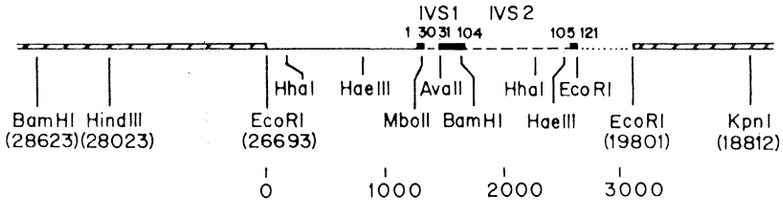

Fig. 5. Summary of the restriction enzyme data used in the structural analysis of Hs51.1. Cross-hatching shows DNA derived from the vector phage; the coordinates, in base pairs from the left end of the phage, of the restriction enzyme sites in the phage DNA are given in parentheses (*20*). The figure is oriented with the 5′ end of the globin sequences to the left, so that the phage left arm is to the right. The scale at the bottom of the figure is in base pairs starting at the 5′ end of the human insert. IVS 1 and IVS 2 are the two intervening sequences (shown by a broken line) which occur in the coding DNA (shown by the heavy black bars). The numbers above the black bars show the positions of IVS 1 and IVS 2. The dotted line indicates the small Eco RI fragment of human nonglobin DNA [see text and (*21*)].

bin structural gene analogous to that present between codons 30 and 31 in the adult mouse β globin major gene (*18*) (Fig. 5). Human $^{G}\gamma$ globin cDNA has restriction enzyme recognition sites for Mbo II, Ava II, and Bam HI at codons 21 to 22, 37 to 38, and 98 to 100, respectively (*14*). The lengths of the corresponding Mbo II–Bam HI fragment (obtained by digestion with both enzymes), allowing for the displaced cutting site of Mbo II, should be 225 base pairs if there is not a second intervening sequence between codons 30 and 31. We have determined its length to be 340 base pairs. Consequently the human γ globin structural gene cloned in Hs51.1 has a second intervening sequence of about 115 base pairs in the coding region to the 3′ side of codon 28. We find the Ava II–Bam HI fragment to be 181 ± 10 base pairs in length, which is that expected (186 base pairs) for the DNA for codons 38 through 99. Consequently the second intervening sequence occurs somewhere between codons 28 and 38. The data are compatible with its being between codons 30 and 31 (Fig. 5). Preliminary DNA sequencing data confirm this.

There are at least two nonallelic genes for fetal globin in all normal human individuals (*1*). These genes code for fetal globins, $^{G}\gamma$ and $^{A}\gamma$, which differ only in the amino acid, either glycine (Gly) or alanine (Ala), occurring at position 136 (*22*). In that Hs51.1 does not include DNA corresponding to this position, we cannot at present determine whether Hs51.1 is derived from a $^{G}\gamma$ or an $^{A}\gamma$ fetal globin gene. However, since we have demonstrated unequivocally that Hs51.1 includes a large part of the structural gene for a fetal γ globin chain, we propose to give it the name Hs51.1Hbγ. An additional G (Gly) or A (Ala) can be added at a later date, if appropriate.

Our successful isolation and partial characterization of a fetal globin structural gene opens the way to looking for DNA sequences that may control the expression of globins during development. We can, for example, ask whether the occurrence and location of intervening sequences are important factors. Since IVS 1 and IVS 2 occur in the fetal γ globin gene at the same positions as they occur in adult globin genes (*16, 18, 23*), we conclude that neither their absence nor a change in their positions can be the means whereby the fetal globin genes are controlled differently from the adult globin genes. Indeed, the occurrence and positions of IVS 1 and IVS 2 appear to be conserved in all globin genes so far described (*16, 18, 23*).

We can also ask whether the specific base sequence of the noncoding intervening DNA is important in the context of fetal versus adult expression by looking for elements common to the adult genes and different in the fetal gene. At the 3′ end of IVS 2 (for which we have 27 base pairs of sequence) the human fetal IVS 2 sequence differs from that of human adult β (*16*) in 23 positions, from human adult δ (*16*) in 14 positions, from mouse adult β major (*18*) in 15 positions, and from rabbit adult β (*23*) in 14 positions. However, the differences between the fetal and adult sequences show no pattern and are, in fact, no more numerous than the differences among the IVS sequences of the adult genes themselves (except that the mouse and rabbit sequences differ in this region at only nine positions) (*23*). We are unable to find any clearly fetal-specific sequences or possible secondary structures in this region (*24*).

Another question of general interest is whether there are specific base sequences at the borders of intervening sequences which might serve as recognition sites for RNA processing enzymes. In the three human globin genes considered here (and in the mouse gene also), we find only four base pairs conserved: the last base pair of IVS 2 at its 3′ end, and the three base pairs of codon 105. Consequently, if a unique uninterrupted base sequence is needed for recognition purposes at the 3′ border of all human globin intervening sequences, it cannot be longer than four base pairs.

In conclusion, we have shown that two intervening sequences occur in a human fetal globin chain coding region at the positions at which they occur in adult globin genes. Accordingly, the differential expression of the fetal globin genes during development cannot be the result of the fetal genes lacking these intervening sequences. It remains to be seen whether the specific base sequences of any parts of the intervening sequences are important in this context, although our limited sequence data show no clear fetal characteristics in the DNA at one end of the larger intervening sequence.

OLIVER SMITHIES
ANN E. BLECHL
KATHERINE DENNISTON-THOMPSON
NANNETTE NEWELL
JULIA E. RICHARDS
JERRY L. SLIGHTOM
PHILLIP W. TUCKER
FREDERICK R. BLATTNER
Laboratory of Genetics,
University of Wisconsin, Madison 53706

References and Notes

1. H. F. Bunn, B. G. Forget, H. M. Ranney, *Human Hemoglobins* (Saunders, Philadelphia, 1977).
2. M. L. Craig and E. S. Russell, *Dev. Biol.* **10**, 191 (1964); A. Fantoni, A. Bank, P. A. Marks, *Science* **157**, 1327 (1967).
3. F. R. Blattner, A. E. Blechl, K. Denniston-Thompson, H. E. Faber, J. E. Richards, J. L. Slightom, P. W. Tucker, O. Smithies, *Science* **202**, 1279 (1978).
4. F. R. Blattner, B. G. Williams, A. E. Blechl, K. Denniston-Thompson, H. E. Faber, L. A. Furlong, D. J. Grunwald, D. O. Kiefer, D. D. Moore, J. W. Schumm, E. L. Sheldon, O. Smithies, *Science* **196**, 161 (1977).
5. T. M. Shinnick, E. Lund, O. Smithies, F. R. Blattner, *Nucleic Acids Res.* **10**, 1911 (1975).
6. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
7. Hybridization of filter transfers was overnight at 68°C in 6 × SSC (saline sodium citrate) Denhardt's solution [D. T. Denhardt, *Biochem. Biophys. Res. Commun.* **23**, 641 (1966)] containing 0.5 percent sodium dodecyl sulfate, and 30 μg of poly(rA) per milliliter. Typical hybridizations were with ^{32}P-labeled (100,000 count/min) probe in 3 ml. The final washing was usually for 30 minutes at 68°C in 3 × SSC containing 0.5 percent sodium dodecyl sulfate.
8. The sizes of the marker fragments are 19.8, 12.4, 9.57, 7.56, 6.76, 5.95, 5.69, 4.925, 4.59, 3.9, 3.6, 3.11, 1.99, and 1.32 kilobase pairs (D. L. Daniels, J. deWet, and F. R. Blattner, in preparation).
9. T. Maniatis, A. Jeffrey, D. G. Kleid, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1184 (1975).
10. Mmm117Hbα and Mmm4Hbβ (abbreviated where appropriate, to 117Hbα and 4Hbβ) are cDNA clones that we constructed from *Mus musculus* globin mRNA, using a tailing method (*4*). Each has been verified by DNA sequencing. They are close to full length. The 5′ end in the transcriptional sense of the cloned mRNA in 4Hbβ is toward the right arm of the parent phage, Charon3AΔlac. The orientation of cloned fragments in Charon phages is designted n, for normal, when the numbering system of the fragment (for example, from 5′ to 3′ of the corresponding mRNA) is in the same sense as the numbering system of lambda (that is, left arm to right arm). When the orientation is in the opposite (unnatur-

al) sense it is designated as u. Mmm4Hbβ is accordingly u, as are Hs51.1 and Mmm117Hbα.

11. DNA from Hs51.1 was first digested with Eco RI, and the 5′ ends were labeled with ^{32}P (*12*). The end-labeled DNA was further digested with Hha I to give two Eco RI-Hha I singly labeled fragments of 170 and 350 base pairs and a doubly labeled Eco RI fragment of 465 base pairs (see Fig. 5). These three fragments were purified by polyacrylamide gel electrophoresis, and the doubly labeled fragment was redigested with Hinf I to yield two singly labeled pieces of 310 and 144 base pairs. All four fragments were partially sequenced (*12*). The experiment was repeated with Hae III as the second enzyme, which yielded four singly labeled Eco RI-Hae III fragments of 805, 109, 280, and 160 base pairs.
12. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
13. L. T. Hunt and M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, D.C., 1976), vol. 5, suppl. 2, pp. 191-223.
14. Data made available to us by S. M. Weissman, B. G. Forget, and R. Cavellesco prior to its publication.
15. C. A. Morotta, J. J. Wilson, B. G. Forget, S. M. Weissman, *J. Biol. Chem.* **252**, 5040 (1977); F. E. Baralle, *Cell* **12**, 1085 (1977).
16. R. M. Lawn, E. F. Fritsch, R. C. Parker, G. Blake, T. Maniatis, *Cell*, in press.
17. Reviewed by B. Williamson, *Nature (London)* **270**, 295 (1977).
18. S. M. Tilghman, D. C. Tiemeier, J. G. Seidman, B. M. Peterlin, M. Sullivan, J. V. Maizel, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 725 (1978); D. Konkel, S. M. Tilghman, P. Leder, *Cell*, in press.
19. R. W. Davis, M. Simon, N. Davidson, in *Methods in Enzymology 21D*, L. Grossman and K. Moldave, Eds. (Academic Press, New York, 1971), pp. 413-428.
20. A complete restriction enzyme map of the EK2 vector Charon 3A has been compiled with restriction site coordinates in base pairs starting at the left end of the phage (J. R. deWet, D. L. Daniels, J. L. Schroeder, B. G. Williams, K. Denniston-Thompson, D. D. Moore, F. R. Blattner, in preparation). The length of the left arm, up to the first Eco RI site, is 19,801 base pairs; the length of the right arm, from the second Eco RI site at 26,693, is 21,600 base pairs. Other Charon 3A restriction sites referred to in this report are a Kpn I site at 18,812, a Hind III site at 28,023, and a Bam HI site at 28,623.
21. Partial sequences from the Eco RI ends of the 465 base pair fragment were determined. The six amino acid sequences that could be translated from this DNA show no relationship to human β-type globin sequences. From the Eco RI sites the two ends have the provisional sequences:

5′ GAATTCTCCTTTTGTAAAAATGGGA-
5′ GAATTCTCAATGAGGATGGCAGCCC-

CCTTTATTTTTGTTTTGACCA-
TACCTGGACTTCAAGCCTGAA-

CAAAGTGAACATGGTCAGGCG 3′
GACCAGCCCCATTTTGACATC 3′

The 465-base-pair fragment does not hybridize to globin cDNA.

22. W. A. Schroeder, T. H. J. Huisman, J. R. Shelton, J. B. Shelton, E. F. Kleihauer, A. M. Dozy, B. Robberson, *Proc. Natl. Acad. Sci. U.S.A.* **60**, 537 (1968).
23. J. Van den Berg, A. van Ooyen, N. Mantei, A. Chamböck, G. Grosveld, R. A. Flavell, C. Weissmann, *Nature (London)* **276**, 37 (1978).
24. Dr. W. M. Fitch compared available globin gene sequences by computer programs designed to detect sequence homology and to detect possible secondary structures that might be formed from the sequence.
25. The cloning experiments were done under NIH guidelines using P3 physical containment and an EK2 host vector system. We thank N. Borenstein for help and supervision of the P3 facility; J. Devereux, E. Kopetsky, D. Kiefer, J. Kucera, and T. Szeto for technical assistance; A. Bank and J. Ross for gifts of mRNA, A. Biro and R. Cavellesco for help and advice in the restriction mapping of Hs51.1; and S. Weissman, P. Leder, and T. Maniatis for making DNA sequence data available to us prior to their publication. This is paper No. 2299 from the Laboratory of Genetics, University of Wisconsin-Madison, and paper 9 in the series "Charon Phages for DNA Cloning." Supported by NIH grants AM 20120, GM 20069 (O.S.), GM 21812 (F.R.B.), CA 69075 (K.D.T.), GM 06526 (P.W.T.), GM 07133 (A.E.B. and J.E.R.), and GM 07131 (N.N.).

12 October 1978

Human Leukocyte Interferon Purified to Homogeneity

Abstract. *One of the species of human interferon produced by incubation of leukocytes with Newcastle disease virus was purified to homogeneity. It exhibited one peak of activity coinciding with a single protein band on sodium dodecyl sulfate-polyacrylamide gel electrophoresis.*

The purification of interferon has been a continuous challenge since its discovery by Isaacs and Lindenmann (*1*) and Nagano and Kojima (*2*). Numerous attempts have been made to purify it by conventional techniques (*3-6*) as well as by affinity chromatography with antibodies (*5-7*) and other methods (*6, 8*). So far, it has been reported that human fibroblast interferon (*4*) and mouse L-cell interferon were purified to homogeneity, but no characterization of these molecules has yet been provided. Although much information is available on the physical and chemical properties of interferons as estimated by effects on antiviral activity, such indirect results are inherently only approximate. Complete physical and chemical characterization of the molecules depend on obtaining homogeneous interferon in sufficient amounts for characterization.

We have purified one species of human leukocyte interferon to homogeneity in amounts sufficient for physical and chemical characterization. Details of the purification and initial characterization, including amino acid analysis, of the homogeneous species have been described (*9*).

Leukocytes were isolated from blood of normal donors and incubated with Newcastle disease virus in a serum-free medium supplemented with casein (*10*). Interferon titers of 5000 to 20,000 U/ml were obtained. The purification procedure consisted of selective precipitations, gel filtration in 4*M* urea, and several steps of high performance liquid partition chromatography (*9*). Several species of human leukocyte interferon activity were obtained. One of the major species was purified 80,000-fold to a specific activity of 4×10^8 U/mg with bovine serum albumin as a standard (*9*).

A sample of the homogeneous interferon, treated with 2-mercaptoethanol and sodium dodecyl sulfate, was analyzed by sodium dodecyl sulfate-polyacrylamide slab gel electrophoresis (Fig. 1). A single protein band of molecular weight 17,500 was obtained upon staining with Coomassie blue. After being stained, the gel was cut into 1-mm slices, each slice was homogenized in 0.4 ml of a solution containing 0.5*M* $NaHCO_3$ and 0.1 percent sodium dodecyl sulfate, and each fraction was assayed for interferon activity. A single peak of activity coinciding with the protein band was obtained.

In addition, a duplicate of the stained band was excised from the polyacrylamide gel and subjected to hydrolysis. Amino acid analysis was performed with the fluorescamine analyzer (*11*). The

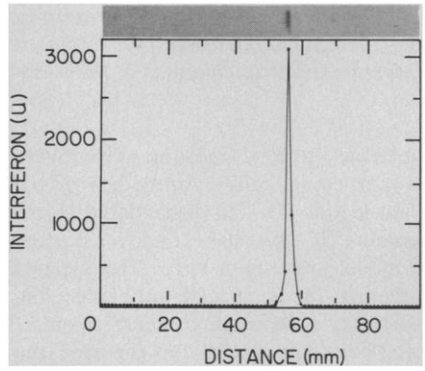

Fig. 1. Sodium dodecyl sulfate-polyacrylamide slab gel electrophoresis of human leukocyte interferon. Two samples (150,000 units each) were dissolved in 7 μl of a solution containing 1 percent (weight to volume) sodium dodecyl sulfate and 2 percent (by volume) 2-mercaptoethanol. After 60 minutes at 25°C, 7 μl of a solution containing 15.6 m*M* tris-HCl (*p*H 6.8), 40 percent (weight to volume) sucrose, and 0.1 percent (weight to volume) bromphenol blue was then added and the samples were applied to a 12.5 percent polyacrylamide slab gel in a tris-glycine (*p*H 8.4) buffer containing 0.1 percent sodium dodecyl sulfate (*12*). After electrophoresis and staining with Coomassie blue, one strip of the polyacrylamide gel was cut into 1-mm slices. Each slice was homogenized in 0.5 ml of a solution of 0.5*M* $NaHCO_3$ and 0.1 percent sodium dodecyl sulfate. Gel fragments were removed by centrifugation at low speed, and the supernatant from each homogenate was assayed for interferon activity (*9*). Interferon titers on the ordinate are expressed in terms of reference units per milliliter, calibrated against the reference standard for human leukocyte interferon (G-023-901-527) (*13*). Low recovery of interferon activity was expected after treatment of human leukocyte interferon with 2-mercaptoethanol and sodium dodecyl sulfate (*14*). The total activity recovered was 4 percent of the activity applied to the gel. The duplicate interferon band (not shown) was used for amino acid analysis and measurement of protein recovery (20 percent). Thus, 80 percent of the interferon protein was lost during electrophoresis or during staining and destaining of the polyacrylamide gel. When adjusted for this factor, the interferon activity recovered represents 20 percent of the material retained in the gel after staining and fixation.

0036-8075/78/1222-1289$00.50/0