Hemoglobin Switching in Sheep

SYNTHESES, CLONING, AND CHARACTERIZATION OF DNA SEQUENCES CODING FOR THE \( \beta^A \)-, \( \beta^C \)-, AND \( \gamma \)-GLOBIN mRNA*Rs

(Received for publication, November 28, 1978)

Edward J. Benz, Jr., † Peter J. Kretschmer, Craig E. Geist,§ Judith A. Kantor, Patricia H. Turner, and Arthur W. Nienhuis††

From the Clinical Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20205

Synthetic double-stranded DNAs (sDNAs) were prepared from sheep globin mRNA templates isolated from reticulocytes producing either hemoglobin B (HbB) (\( \alpha_2\beta^B \)), HbC (\( \alpha_2\beta^C \)), or HbF (\( \alpha_2\gamma \)). These sDNAs were inserted into the Eco RI site of plasmid pMB9 by the homopolymer tailing method and used to transform \textit{Escherichia coli} X1776 to tetracycline resistance. Recombinant clones were identified by colony hybridization and further characterized by molecular hybridization and restriction endonuclease analysis. All plasmids analyzed thus far contained either \( \beta \)- or \( \gamma \)-globin DNA sequences. Moreover, sDNAs used for cloning yielded restriction endonuclease fragments consistent with the presence of predominantly \( \beta \)- or \( \gamma \)-sDNA, indicating that formation of double-stranded a-sDNA proceeds much less efficiently under our conditions than the formation of non-a-sDNAs. Three recombinant plasmids, pS\( \beta^B \)2, pS\( \beta^C \)69, and pS\( \gamma \)56, were selected for detailed study. These were shown to contain, respectively, \( \beta^B \), \( \beta^C \), and \( \gamma \)-DNA sequences by molecular hybridization and by protection of the appropriate cDNAs from S1 nuclease digestion. Each contained all of the restriction endonuclease sites defined for the synthetic DNAs and protected at least 90% of the sequence length of homologous cDNA. Restriction endonuclease maps of the \( \beta^B \)- and \( \beta^C \)-globin genes were identical at all 12 sites that were mapped, whereas four differences were identified in the \( \gamma \) gene compared to the two others; three of these corresponded to differences in amino acid sequence of the globins. A method was developed to isolate the anti-mRNA strand of the insert for use as a specific molecular hybridization probe analogous to complementsy DNA.

Hemoglobin synthesis in sheep is controlled by the regulated expression of at least five genes coding for the \( \alpha \)-, \( \beta^B \)-, \( \beta^C \)-, and \( \gamma \)-globins (1, 2). During gestation, synthesis of HbF (\( \alpha_2\gamma \)) predominates, whereas in the adult, the allelic genes for \( \beta^B \) and \( \beta^C \) globin determine the preferential synthesis of the normal adult hemoglobins, HbA (\( \alpha_2\beta^A \)) and HbB (\( \alpha_2\beta^B \)). In addition, sheep and a few closely related species are biologically unique among mammals in possessing an erythropoietin-dependent "switch" to the synthesis of HbC (\( \alpha_2\beta^C \)) during periods of erythropoietic stress. Sheep hemoglobin synthesis is thus a useful system for the study of a perinatal switch from HbF to HbA analogous to that which occurs in humans, and for investigation of the molecular basis of erythropoietin action on early erythroid progenitors.

We have previously described the isolation of partially purified synthetic DNAs (cDNAs) complementary to each of the sheep globin mRNA sequences (3, 4). By using these cDNAs as molecular hybridization probes, we have demonstrated that the HbF \( \rightarrow \) HbA switch and the HbA \( \rightarrow \) HbC switch in sheep are regulated at the level of the de novo mRNA accumulation (3-5). In addition, we have previously suggested that the sheep globin mRNAs possess an extremely high degree of sequence homology, requiring unusually stringent conditions of hybridization in order to achieve specificity (3, 4, 6).

In order to facilitate the structural analysis of the sheep globin genes and to obtain hybridization probes of sufficient quantity and purity for analysis of globin mRNA transcription, processing, and stability during hemoglobin switching, we have modified methods recently used successfully in several laboratories (7-10) for the synthesis of double-stranded DNA containing the sequences of globin genes and their insertion into the bacterial plasmid, pMB9 (11). The recombinant plasmids were used to transform the \textit{E. coli} strain X1776 to tetracycline resistance (12), wherein colonies containing globin genes were identified in a colony hybridization assay (13). Utilizing synthetic \( \beta^B \)-, \( \beta^C \)-, and \( \gamma \)-DNAs and the globin gene sequences isolated from the recombinant plasmids, we have constructed and compared restriction endonuclease maps for each of the genes. Single-stranded fragments containing the inserted globin cDNA sequences were also isolated and demonstrated to be useful as molecular hybridization probes.

MATERIALS AND METHODS

Reagents—Reverse transcriptase was obtained through Dr. J. W. Beard (Life Sciences, Inc., Gulfport, Fla.) through the Office of Program Resources and Logistics, Viral Oncology Branch, National Cancer Institute. \textit{E. coli} DNA polymerase I was purchased from Boehringer Mannheim; \( 5' \) nuclease from Miles; DNase I and RNase A from Worthington Biochemicals; and restriction endonucleases from Bethesda Research Laboratories, New England BioLabs, and Boehringer Mannheim. \( \phi X 174 \) Hae III DNA fragments, utilized as molecular weight markers for gel electrophoresis, were purchased from Bethesda Research Laboratories. Agarose was obtained from Bio-Rad Laboratories. Eosin methylene blue agar, MacConkey agar, and nutrient agar were obtained from Difco. Galactose, sodium citrate, and lyso-
zyme were purchased from Sigma; chloramphenicol and diaminopimelic acid from Calbiochem; and tetracycline from ICN. Terminal deoxynucleotidyltransferase (terminal transferase) was a kind gift from Dr. Bernard Forget and Thomas Maniatis, respectively. Recombinant plasmids JW-102 and pPG1 were kindly supplied by Dr. Roy Curtis III and his associates (12). Transformation of this strain by plasmid DNA and extraction of plasmid DNA following growth in L-broth (18) and chloramphenicol amplification (12.5 pg/ml, 4 to 5 h growth) were essentially as described by Curtis et al. (11, 12). For long term storage, log phase cultures in L-broth (20) were diluted in an equal volume of 2~ freezing medium: 10 g of glucose, 12.6 g of KH2PO4, 3.6 g of K2HPO4, 2.8 g of KH2PO4, 0.9 g of sodium citrate, 0.18 g of MgSO4·7 H2O, 1.8 g of (NH4)2SO4, and 86 g of glycerol.1

Biohazard Containment and Precautions—All cloning experiments involving sheep globin DNAs and subsequent manipulations involving recombinant strains were performed under P3EK2 containment procedures as described in a Memoranda of Understanding and Agreement approved by the Biohazards Committee at the National Institutes of Health. Plasmids nG1 and JW-102 were handled under P2 E. coli conditions as certified by the Recombinant DNA Committee of the National Institutes of Health. E. coli X1776 strains were tested before and after cloning experiments to confirm the presence of phenotypic traits relevant to its EK2 status (requirement for thymine and diaminopimelic acid, galactose negative phenotype (white colony) on MacConkey agar), as outlined in a protocol accompanying the strain (12).

Preparation of Globin mRNA and Single-stranded Globin cDNAs—Globin mRNA was obtained from sheep reticulocytes synthesizing predominantly HbB, Hbc, or HbF as described in earlier reports (3-6, 21). Single stranded globin cDNAs were synthesized from the mRNA templates under conditions described previously (3, 4).

Synthesis of Double-stranded sDNA—Each globin cDNA preparation was rendered double-stranded by a second reaction with reverse transcriptase utilizing the self-priming hairpin loop generated during the initial cDNA synthesis (8). Incubation mixtures included the following reagents in the specified concentrations: globin cDNA, 8 to 10 µg/ml; 1 mM concentration each of dATP, dCTP, dGTP, dTTP; 100 mM Tris-HCl, pH = 8.3; 8 mM MgCl2; 60 mM KCl; reverse transcriptase, 300 units/ml, and [α-32P]dCTP (300 Ci/mmol), 250 µCi/ml. Incubation was for 2 h at 37°C, after which E. coli tRNA was added (50 µg/ml), and the mixture was desalted by passage over Sephadex G-50. Radioactive DNA eluting in the void volume was then lyophilized. The complementary strands of DNA prepared in this way are joined by a hairpin loop and are designated as “closed” sDNA.

Electrophoresis mixtures containing closed sDNA (5 µg/ml), 30 mM sodium acetate (pH = 4.6), 280 mM NaCl, 1 mM zine acetate, and S. nucleas (100 units/µg of DNA) (8) were incubated for 1 h at room temperature, and the reactions were terminated by addition of EDTA, pH 7.0, to a final concentration of 25 mM, followed by phenol extraction. This treatment excised the hairpin loop, yielding “open” sDNA. The DNA in the aqueous phase was concentrated by ethanol precipitation, desalted by passage through Sephadex G-25 (coarse), and stored in sterile water in liquid nitrogen at a concentration of 1 to 3 µg/ml.

Addition of Poly(dA) or Poly(dT) Homopolymers and Annealing of Globin and Plasmid DNAs—Homopolymer tails containing approximately 75–100 poly(dA) residues (globin DNA) or 75 to 100 poly(dT) residues (plasmid DNA) were added with terminal transferase according to the methods of Salser and co-workers (8), utilizing cobalt chloride buffers. Approximately equimolar concentrations of each DNA were annealed for 1 h at 50°C followed by slow cooling to room temperature over a period of 3 to 4 h and were used immediately for transformation of E. coli X1776 as described above.

Identification of Recombinant Clones—Denhardt’s solution containing globin DNA sequences were identified by a modification of the colony hybridization procedure of Gristm and Hogness (13). Nitrocellulose filters (Millipore HAWP, 83 µm in diameter, 0.45-µm pore size) were boiled three times for 5 min in sterile deionized water, blotted dry, autoclaved for 10 min, dried, and laid on top of nutrient agar containing tetracycline. A small sample of 2°C for 24 h in fresh Denhardt’s solution, containing globin cDNA (3 x 106 cpm/µg) and [α-32P]dCTP (300 Ci/mmol) and buffer (100 µg/ml) was used in running the 0.5% agarose, 3.5% bisacrylamide slab gels. Electrophoresis was performed in a Bio-Rad slab gel apparatus (model 220 (11 x 14 cm) or model 221 (11 x 28 cm)); running time and voltage were as described in the figure legends. DNA fragments were located after staining with ethidium bromide, 0.5 µg/ml, by viewing with an ultraviolet light. These electrophoretic procedures were essentially those of Maniatis and co-workers (23). Nondenaturing acrylamide gel electrophoresis was performed exactly as described by Dingman et al. (24), and electrophoresis in 385C formamide under denaturing conditions was as described previously by us (3).

Cloning of Sheep Globin Genes 6881

1 W. Salser, personal communication.
EDTA. The peak void was concentrated after addition of NaOAc, pH 5.0, to 50 mm and 100 µg of carrier E. coli RNA, by precipitation with 2 volumes of ethanol. The specific activity of the DNA varied from 2 to 10 x 10^6 cpm/µg.

Preparation of the Anti-mRNA ("cDNA") Strand—One microgram of [32P]Hha I fragment was annealed to 10 to 25 µg of the partially purified homologous reticulocyte mRNA; e.g. the fragment from pSB2 was annealed to HbB mRNA (α + β') in a 100-µl reaction in 50% formamide under standard conditions (3). Annealing was at 66°C for 18 h; this temperature was chosen to prevent formation of DNA duplex. Digestion with S1 nuclease (1200 units) for 30 min at 45°C in a 2-ml reaction was as previously described (3). The mixture was made 0.1 M NaOH, 0.9 M NaCl, and 10 mM EDTA, and a size fraction corresponding to 400 to 500 nucleotides was recovered by alkaline sucrose gradient centrifugation (3, 4). From 60 to 100% of the isolated "cDNA" was protected from S1 nuclease by annealing to globin mRNA; less pure preparations were freed of contaminating plasmid DNA sequences by repeating the above series of reactions.

RESULTS

Synthesis of Double-stranded Globin Genes

We have previously described the properties of single-stranded cDNAs transcribed from sheep globin reticulocyte mRNAs by the methods used for these experiments (3, 4). As shown in Fig. 1, Channel 1, the cDNA was 600 to 640 bases long on denaturing formamide acrylamide gels. Closed sDNA migrated with a chain length of approximately 1250 bases (Fig. 1, Channel 2), while after S1 nuclease digestion the open form of sDNA consisted of two co-migrating single-stranded molecules each 600 to 640 bases long (Fig. 1, Channel 3), corresponding roughly to the initial length of the cDNA. Incubation of sheep globin cDNAs with reverse transcriptase routinely converted about 30 to 60% of the cDNA to sDNA; as shown in Fig. 1, Channel 3, the bulk of the sDNA formed appeared to be full length double-stranded material. Although the efficiency of conversion was somewhat higher with E. coli DNA polymerase I (50 to 80%), the products obtained were more heterogenous, and apparently contained partially double-stranded molecules (data not shown).

Restriction Endonuclease Analysis of sDNAs

Restriction endonuclease analysis of radioactive closed and open sDNAs was performed by electrophoresis of the digests on denaturing acrylamide gels prepared in 98% formamide. For enzymes that cut the sDNAs only once, the fragment common to both open and closed DNA is derived from the 3' end (by convention, the 3' and 5' ends correspond to the polarity of the mRNA), whereas the fragment that differs in size is derived from the 5' end. Fragments common to both the open and closed HbB sDNA were 240 base pairs with Eco RI (Fig. 1), 310 base pairs with Bam HI (Fig. 1), 345 base pairs with Pvu II (Fig. 2) and 235 base pairs with Pst I (data not shown). The corresponding restriction endonuclease sites are approximately these distances from the 3' end of the sDNA. Based on the amino acid sequence (27) and the genetic code, a Bam HI or Eco RI site would not be expected in α-cDNA, yet there is very little DNA that remains undigested when HbB sDNA is cleaved with these enzymes. We surmise therefore that α-cDNA is very inefficiently converted into sDNA. Analysis of HbF and HbC sDNA with Eco RI, Bam HI, and Pst I indicated single sites in these sDNAs in similar positions to those in HbB sDNA (data not shown). Pvu II does not cut HbF sDNA (Fig. 2), whereas HbC sDNA has a site for this enzyme that is approximately 345 base pairs from the 3' end. Again, there was a very little sDNA in these preparations that remained undigested with Bam HI or Eco RI, suggesting a very low content of α-DNA.

![Fig. 1. Synthesis and restriction endonuclease analysis of the synthetic sheep β' globin gene. Slot 1, single-stranded cDNA synthesized using HbB mRNA (α + β') as a template for reverse transcriptase. Slot 2, HbB cDNA rendered double-stranded in a second reaction with reverse transcriptase (closed sDNA). As discussed in the text, these reactions favor the synthesis of β'-sDNA rather than α-cDNA. Slot 3, double-stranded cDNA digested with S1 nuclease to excise the hairpin loop (open sDNA). Slot 4, closed sDNA digested with Eco RI. Slot 5, open sDNA cut with Eco RI. Slot 6, closed sDNA cut with Hae III. Fragment sizes are approximately 300, 280, and 240 base pairs. Slot 7, open sDNA cut with Hae III. Fragment sizes are approximately 350, 310, and 240 base pairs. Slot 8 and 9, closed and open sDNA, respectively, digested with HindIII. Electrophoresis in a 5% polyacrylamide gel run in 98% formamide containing 20 mM sodium phosphate, pH = 7.0 for 3 h. α-XL74 DNA digested with Hae III was run in a parallel slot to define the 600- to 640-base-pair length of the sDNAs. Slot 3, HbB closed sDNA cut with Pvu II; major fragments are 630, 560, and 345 base pairs in length. Slot 4, HbB open sDNA cut with Pvu II. Major fragments are 345, 315, and 280 base pairs in length. The two smaller fragments are thought to be a consequence of the discrete variation in size of the sDNA. Slot 5, HbF open sDNA digested with Pvu II; no reduction in size occurred. Slots 6 and 7, HbB (6) or Hb F (7) open sDNA digested with Hha I. Electrophoresis in a 5% polyacrylamide gel in 98% formamide for 3 h. The buffer was 20 mM sodium phosphate, pH = 7.0.](http://www.jbc.org/)

![Fig. 2. Characterization of HbB (slot 1) and HbF (slot 2) sDNA that had been treated with S1 nuclease to excise the hairpin loop. Standard markers (α-XL74 DNA digested with Hae III) were run in a parallel slot and used to define the 600- to 640-base-pair length of the sDNAs. Slot 3, HbB closed sDNA cut with Pvu II; major fragments are 630, 560, and 345 base pairs in length. Slot 4, HbB open sDNA cut with Pvu II. Major fragments are 345, 315, and 280 base pairs in length. The two smaller fragments are thought to be a consequence of the discrete variation in size of the sDNA. Slot 5, HbF open sDNA digested with Pvu II; no reduction in size occurred. Slots 6 and 7, HbB (6) or Hb F (7) open sDNA digested with Hha I. Electrophoresis in a 5% polyacrylamide gel in 98% formamide for 3 h. The buffer was 20 mM sodium phosphate, pH = 7.0.](http://www.jbc.org/)
Enzymes that did not cut HbC, HbB, or HbF sDNAs included HindIII (Fig. 1), Hha I (Fig. 2), Ava I, and Hpa I. HbB and HbC sDNA were also shown to lack sites for Xho I, Xho I and Ava II. Digestion of HbB and HbC sDNAs with Hinf I yielded 395- and 90-base-pair fragments common to the open and closed forms of sDNA, whereas digestion of HbF sDNA yielded a single 480-base-pair fragment common to both forms of sDNA (data not shown). Digestion with Hae III yielded several fragments with each sDNA.

Hpa II cut HbB and HbC sDNA twice, yielding fragments common to the open and closed sDNAs of 230 and 340 base pairs (Table I). HbF sDNA had only a single Hpa II site; the 230-base-pair fragment was present in the digest of the open HbF sDNA. We surmised, therefore, that the 340-base-pair fragment seen in HbB and HbC sDNA was an internal fragment, a supposition verified by subsequent analysis of recombinant plasmids (see below).

Insertion of Globin sDNAs into E. coli X1776

Sheep globin sDNAs were cloned in two separate experiments. In the first experiment, full length HbB sDNA was size-selected by isolation on nondenaturing gels prior to incubation with terminal transferase. This preparation was annealed to plasmid pMB9 DNA that had been digested twice with Eco RI prior to addition of poly(dT) to reduce the number of tetracycline-resistant colonies containing only pMB9. Of 13 transformants obtained in this manner, 12 proved to be authentic recombinants containing β-globin DNA sequences. In the second experiment, HbC and HbF sDNAs were not size-selected prior to addition of poly(dT), nor was pMB9 digested twice with Eco RI. Both the background obtained with poly(dT)pMB9 alone and the number of transformants obtained with poly(dT)pMB9 plus poly(dA)-sDNA were increased relative to the first experiment. However, only 19 of 94 colonies prepared with HbC sDNA gave positive colony hybridization signals; of these, 16 were further analyzed by restriction endonuclease analysis, and only 8 proved to be true recombinants. Similarly, 21 of 91 colonies from the HbF sDNA experiment gave positive colony hybridization signals, and of the 14 subjected to further analysis, only 10 were positive. Thus no change in the yield of true recombinants was observed when size selection was omitted. The HbB sDNA recombinants appeared to contain uniformly large inserts (600 to 750 base pairs), whereas a number of the colonies obtained with HbC and HbF sDNAs yielded plasmids containing small inserts.

| Table 1 |

Fragments obtained by Hpa II digestion of sheep globin synthetic DNAs

As discussed in the text, the sDNA obtained with mixed α- and β-(or γ-) mRNA is thought to be derived almost exclusively from the non-α component.

<table>
<thead>
<tr>
<th></th>
<th>Closed a</th>
<th>Open b</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbB</td>
<td>340</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
</tr>
<tr>
<td>HbC</td>
<td>340</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
</tr>
<tr>
<td>HbF</td>
<td>415</td>
<td>230</td>
</tr>
</tbody>
</table>

a Closed sDNA is that which results from synthesis of the second strand by reverse transcriptase.
b Open sDNA is that obtained after the hairpin loop is excised with S. nuclease.

Lengths are in base pairs as estimated by comparison to φX174 Hae III fragments.

Preliminary Analysis of Recombinant Plasmids

To obtain an approximate estimate of the size of the inserted sequence, each recombinant plasmid was analyzed with Hha I. Hha I cleaves pMB9 several times, yielding a series of fragments, the largest of which is 1030 base pairs (Fig. 3, Slots 4 and 8) and contains the single Eco RI site present in pMB9 (7). The β, β', and γ-sDNAs are not cut by Hha I (Fig. 2); hence, recombinants, when digested with this enzyme, should yield a large fragment that is 1030 base pairs long plus the length of the insert. The results obtained with three recombinants, pSy56, pSβ' 69, and pSβ' 2, are shown in Fig. 3, Slots 5 to 7. The insets in these plasmids were among the largest obtained; hence they were chosen for detailed study.

Fig. 4 shows the results obtained by digestion of pSy56 and pSβ' 69 with Hpa II. pSy56 yielded two novel fragments 1050 and 780 base pairs in length, which replace a 1110-base-pair fragment in pMB9 (Fig. 4, Slot 4); these data indicate that pSy56 contains approximately 720 base pairs of inserted sequence. pSβ' 69 yielded three novel fragments of 785, 740, and 340 base pairs in length, which replace the 1110-base-pair fragment in pMB9; the estimated insert is therefore 755 base pairs. Analyses of pSβ' 2 gave results similar to those obtained with pSβ' 69. These data verify that the 340-base fragment observed when HbB and HbC sDNAs are digested with Hpa II (Table I) is an internal fragment.

Size and Identity of Cloned Sheep Globin sDNAs

Further characterization of the inserted sequences was facilitated by the observation that the largest Hha I fragment that contains the entire globin insert of each recombinant plasmids was easily separable from the remainder of the plasmid fragments by alkaline sucrose density gradient centrifugation. Hha I fragments isolated in this manner migrated as a single band on agarose gels (data not shown). In order to verify the identity of each of the three plasmids, and to determine the length of the inserted sequence that actually represented globin gene DNA, Hha I fragments from each plasmid were denatured and allowed to reanneal in the pres-
Cloning of Sheep Globin Genes

FIG. 4. Analysis of recombinant plasmids pSy56 and pSy69 with Hpa II. The plasmid vector, pMB9, was digested with Hpa II and run in Slot 4 for comparison. Slot 2, pSy56. The 1110-base-pair fragments seen in pMB9 are replaced by two fragments of 1050 and 780 base pairs in length, indicating an inserted sequence of approximately 720 base pairs. Slot 3, pSy69. The large fragment present in pMB9 is replaced by three unique fragments of size 785, 740, and 340 base pairs in length, reflecting a net insert size of 750 base pairs. Slot 1, X174 Hae III fragments. Electrophoresis was in a 0.5% agarose, 3.5% polyacrylamide gel run in Tris/borate buffer, pH = 8.0, for 15 h at 30 V. The gel was subsequently stained in ethidium bromide to visualize the DNA fragments.

Further Restriction Endonuclease Analysis of Cloned Sheep Globin Genes and Determination of the Orientation of theInserted Sequences

HinfI—The isolated Hha I fragments from pSy69, pSy69, and pSy56 were labeled with 32P-dCTP by nick translation, digested with HinfI, and analyzed in a neutral polyacrylamide gel. The large Hha I fragment from pMB9 contained a single HinfI site 345 base pairs from one end and 65 base pairs from the Eco RI site into which our globin genes had been inserted (7); as expected, each Hha I fragment therefore yielded a 345-base-pair piece of DNA. The Hha I fragments from pSy69 yielded additional fragments of DNA of 790, 510, and 90 base pairs in length, indicating that there are two HinfI sites in the cloned sequences. Digestion of the Hha I fragment from pSy69 gave fragments of 790, 470, and 90 base pairs. The internal 340-base-pair Hpa II fragment (Fig. 5) from pSy56 (a recombinant derived from cloning of HbB sDNA, which has an insert size of 32P-labeled HbB (α + β), HbC (α + β), or HbF (α + γ) cDNA. The annealed preparations were then digested with S1 nuclease in 48% formamide (see details in the legend to Fig. 5) and analyzed by polyacrylamide gel electrophoresis as shown in Fig. 5. pSy69 protected only HbB cDNA; the length of the protected sequence corresponded to approximately full length cDNA (620 bases). Similarly, pSy69 protected only HbC cDNA; the protected cDNA was 580 bases in length. pSy56 protected only HbF cDNA, and the protected cDNA was approximately 580 bases long. If any of these recombinants had contained α-globin DNA sequences, its Hha I fragment should have protected a fraction of the molecules in each mixed α + non-α cDNA. We conclude, therefore, that these three plasmids contain virtually full length β, β', and γ-DNA sequences coding for their respective mRNAs.

FIG. 5. Determination of the size of the inserted globin gene sequences and demonstration of hybridization specificity of the globin gene sequences contained in the recombinant plasmids. Slot 1, HbB cDNA. Slot 2, HbB cDNA digested with S1 nuclease. Slots 3 to 5, HbB cDNA annealed to the Hha I fragment from pSy69 (slot 3), pSy69 (slot 4), or pSy69 (slot 5) and digested with S1 nuclease. Slots 6 to 8, HbC cDNA annealed to the Hha I fragment from pSy69 (slot 6), pSy69 (slot 7), or pSy69 (slot 8) and digested with S1 nuclease. Slots 9 to 11, HbF cDNA annealed to the Hha I fragment from pSy69 (slot 9), pSy69 (slot 10), or pSy69 (slot 11) and treated with S1 nuclease. From 1 to 5 ng of [32P]-labeled cDNA (specific activity = 2 x 10^6 cpm/ng) were annealed to 50 ng of the respective plasmid fragment under standard hybridization conditions at 50°C (see Methods). The S1 nuclease digests were performed in 0.5 M NaOAc, 1 mM ZnSO4, and 0.2 M NaCl containing 5% formamide. Fifty units of S1 were added to each 40-μl reaction, and incubation was at 50°C for 30 min, after which EDTA was added to a final concentration of 20 mM. Gel electrophoresis was performed in a 1% polyacrylamide gel containing 5% glycerol for 8 h. The buffer was 0.5 M Tris/borate, pH = 7.8, containing 2.5 mM EDTA. The HbB cDNA was estimated to be 640 base pairs in length by comparison to φX174 Hae III fragments run simultaneously. The estimated size of the globin insert in pSy69 is 620 base pairs, and the inserts in pSy69 and pSy56 are 580 base pairs based on this analysis.
comparable to pSβ69) was isolated from the plasmid after it had been labeled with [32P]dCTP by nick translation. This 340-base-pair Hpa II fragment was then digested with HindII, and the resulting fragments were displayed by radioautography after electrophoresis in a polyacrylamide gel under denaturing conditions. Three fragments were observed, indicating that the two HindII sites present in HbB (and HbC) sDNA were between the two Hpa II sites. Because the opened and closed HbB sDNA yielded identical 395-base-pair fragments (see results of analysis of sDNA), we can infer that one HindII site is 395 base pairs from the 3' end and the second is 90 base pairs away toward the 5' end.

The data obtained with HindII digestion of the isolated Hha I fragment permits us to orient the globin DNA inserts in pSβ69 and pSβ69 (Fig. 6B). The 790-base-pair fragment must consist of 610 base pairs of pMB9 sequence, a homopolymer dA-dT region, and 180 to 150 base pairs of globin gene DNA from the 5' end. The 610-base-pair fragment includes the 3' part of the coding sequence of the gene, the homopolymer dA-dT region, and 65 base pairs of pMB9.

Digestion of the Hha I fragment from pSy56 with HindII yielded three fragments—the 345-base-pair piece of pMB9 DNA and two other fragments 790 and 595 base pairs in length; thus the inserted sequence has only one HindII site. Because analysis of HbF sDNA had demonstrated a 480-base-pair fragment common to all of the genes (Fig. 7), digestion of the isolated Hha I fragment from pSy56 with HindII and HpaII digestion of the isolated Hha I fragment yielded, among others, a fragment approximately 240 base pairs in length (Fig. 7B). Hae III digestion of the isolated Hha I fragment yielded, among others, a fragment approximately 240 base pairs in length (Fig. 7B, Slot 2). A double digestion of the isolated Hha I with Hae III and Eco RI resulted in loss of a 240-base-pair piece and yielded a fragment of 180 base pairs (Fig. 7B, Slot 4), indicating that there is a Hae III site 180 base pairs from the Eco RI site and another approximately 60 base pairs away. The Eco RI site is approximately 240 base pairs from the 3' end; a Hae III site 180 base pairs further toward the 5' end of the gene would correspond to one of two Hae III sites found in the isolated Hpa II fragment, whereas part of the 5' end of the globin gene, whereas the 595-base-pair piece includes most of the globin gene insert (Fig. 6A).

Hpa II—The isolated Hha I fragments of pSy56, pSβ69, and pSβ69 were digested with Hpa II and subjected to electrophoresis in a neutral polyacrylamide gel followed by radioautography (data not shown). The size of the fragments obtained from the Hha I fragment of pSy56 and pSβ69 are indicated in Fig. 6, A and B, respectively. As predicted from the analysis of sDNA (Table I), the insert of pSy56 contains one Hpa II site, whereas two sites separated by approximately 340 base pairs of DNA were found in pSβ69 and pSβ69 (see also Fig. 4). The fragments obtained with Hpa II digestion support the orientation of the inserted globin gene sequences deduced from the HindII data (Fig. 6).

Hae III—Several Hae III sites were found in each of the globin genes. Two sites separated by approximately 110 base pairs of DNA were common to all of the genes (Fig. 7). Digestion of the isolated Hpa II fragment from pSβ69 with Hae III (see above) indicated that these two sites are between the Hpa II sites (of the βγ- and ββ-globin genes), since a 110-base-pair fragment was observed. The position of a third Hae III site in the βγ gene can be deduced from the data shown in Fig. 7B. Hae III digestion of the isolated Hha I fragment yielded, among others, a fragment approximately 240 base pairs in length (Fig. 7B, Slot 2). A double digestion of the isolated Hha I with Hae III and Eco RI resulted in loss of a 240-base-pair piece and yielded a fragment of 180 base pairs (Fig. 7B, Slot 4), indicating that there is a Hae III site 180 base pairs from the Eco RI site and another approximately 60 base pairs away. The Eco RI site is approximately 240 base pairs from the 3' end; a Hae III site 180 base pairs further toward the 5' end of the gene would correspond to one of two Hae III sites found in the isolated Hpa II fragment, whereas

![Fig. 6. Restriction endonuclease map of the Hha I fragments from recombinant plasmids pSy56 and pSβ69. The vertical dashed line in the middle represents the position of the Eco RI site in the globin gene inserts. The heavy lines on the ends indicate DNA sequences derived from pMB9; the sites in these fragments are those published by Maniatis et al. (7). The vertical arrows indicate the positions of the homopolymer dA-dT regions, whereas the thin line represents the globin gene insert. The derivation of these maps is discussed in detail in the text. The lengths of the fragments within the globin gene sequences have been adjusted to the exact length predicted from the amino acid sequence of the respective globins (see "Discussion").](http://www.jbc.org/)

![Fig. 7. Digestion of the isolated Hha I fragments from pSy56 (A) and pSβ69 (B) with Hae III. A: Slot 1, 5X174 fragments obtained by Hae III digestion; Slot 2, fragments obtained from the Hha I fragment of pSy56 when digested with Hae III. The estimated fragment sizes are 550, 470, 270, 190, and 110 base pairs in length. B: Slot 1, 5X174 Hae III fragments; Slot 2, Hha I fragment of pSβ69 digested with Hae III. Estimated fragment sizes are 630, 420, 270, 240, and 110 base pairs. Slot 3, double digest of the Hha I fragment of pSβ69 with Hae III and Xho II. The 240-base-pair fragment seen in Slot 2 is reduced in size to 165 base pairs. Slot 4, digestion of the Hha I fragment of pSβ69 with Hae III and Eco RI. The 240-base-pair fragment seen in Slot 2 is reduced in size to 180 base pairs. Fragments smaller than 110 base pairs are not visualized on this gel. Electrophoresis was in a 6% polyacrylamide gel in Tris/borate buffer for 8 h. The gel was stained in ethidium bromide to visualize the DNA fragments.](http://www.jbc.org/)
Cloning of Sheep Globin Genes

The amino acid sequences of the several sheep globins are quite similar; only 19 to 22 amino acid differences are found among the sequences (see "Results"). The exact position of the various restriction endonuclease sites, with respect to the amino acid sequence of globins (27), was deduced by searching for the appropriate sequence very close to the position predicted from the restriction endonuclease analysis as detailed above. The amino acid sequences of the several sheep globins are quite similar; only 19 to 22 amino acid differences are found in comparing the $\beta^\alpha$- and $\beta^\gamma$-globins, and only 25 to 27 differences are found between the $\beta^\gamma$- and $\gamma$-globins (3, 27). By study of the thermal stability of the duplexes formed between the purified cDNAs and various mRNAs (3), we have determined that the nucleotide sequences of the sheep globin genes are highly conserved. Very few changes in nucleotide sequence are thought to occur except for those which result in amino acid alterations. The restriction endonuclease sites mapped

Specificity of the Isolated "cDNA" Probes in Annealing Reactions

The radioactive anti-mRNA (cDNA) strands isolated from the large Hha I fragment of pSy56, pSp$^\beta$69, and pSp$^\beta$2 (Fig. 3) were annealed to the homologous (e.g. pSp$^\beta$69 cDNA to pSy56 Hha I) and one heterologous (e.g. pSp$^\beta$69 cDNA to pSp$^\beta$69 Hha I) Hha I fragments (Fig. 8). Complete specificity was achieved. Incorporation of only 30 to 40% of the cDNA probes into $\beta$-resistant duplex is thought to result from the relatively high temperature at which the annealing reaction was conducted, and also because reannealing of the large Hha I strands is likely to be a kinetically more favorable reaction than annealing to the cDNA. From 90 to 100% of each cDNA was incorporated into duplex when incubated with homologous mRNA at 50°C (data not shown).

**DISCUSSION**

We have cloned synthetic DNA sequences corresponding to most of the mRNA coding portions of the sheep $\beta^\alpha$, $\beta^\gamma$, and $\gamma$-globin genes. From the analysis of the sDNAs and subsequently of the cloned sequences, we have constructed restriction endonuclease maps of the three genes (Fig. 9). The derivation of these maps is detailed in the Results but may briefly be summarized as follows. The positions of sites for enzymes that cut only once, namely Eco RI, Pst I, Pvu II, and Bam HI, were determined by analysis of opened and closed sDNAs (Figs. 1 and 2). The positions of the Hpa II sites in the $\beta^\alpha$ and $\beta^\gamma$ synthetic genes are deduced from the data in Table I and the result in Fig. 4, which verifies that the 340-base-pair fragment is present in the inserted sequence and therefore cannot be from either the 3' or 5' end of the sDNA. The position of the two Hinf I sites in the $\beta^\alpha$ and $\beta^\gamma$ synthetic genes is defined by their position in the isolated 340-base-pair Hpa II fragment and the analysis of sDNA, which indicates that the site closest to the 3' end is approximately 385 bases away from that terminus. The positioning of the several Hae III sites was complex and required analysis of the cloned sequences (see "Results"). The exact position of the various restriction endonuclease sites, with respect to the amino acid sequence of globins (27), was deduced by searching for the recognition sequence of each enzyme among the various combinations of codons for the individual amino acids in the globin chains. In all cases it was possible to identify a unique appropriate sequence very close to the position predicted from the restriction endonuclease analysis as detailed above.

The amino acid sequences of the several sheep globins are quite similar; only 19 to 22 amino acid differences are found in comparing the $\beta^\alpha$- and $\beta^\gamma$-globins, and only 25 to 27 differences are found between the $\beta^\gamma$- and $\gamma$-globins (3, 27). By study of the thermal stability of the duplexes formed between the purified cDNAs and various mRNAs (3), we have determined that the nucleotide sequences of the sheep globin genes are highly conserved. Very few changes in nucleotide sequence are thought to occur except for those which result in amino acid alterations. The restriction endonuclease sites mapped
for the $\beta^b$ and $\beta^c$ genes are identical, but four differences were found in the map of the $\gamma$ gene (Fig. 9). Because of the predicted conservation of the nucleotide sequence, it is of interest to determine which of these correspond to amino acid differences between the respective globins. As shown in Table II, three differences in restriction endonuclease site between $\gamma$ and $\beta^b$- (or $\beta^c$-) globin genes correspond to amino acid sequence differences, and the other occurs either because of a substitution of asparagine for aspartic acid at position 68 or because of a silent mutation. Further comparison of the sheep globin genes must await determination of their nucleotide sequence, a task now readily feasible because of the presence of these genes in the recombinant plasmids isolated in this study.

To utilize cloned DNA sequences as hybridization probes, it is often desirable to recover them free of the DNA of the plasmid vector. Selection of the insert by virtue of the relative sensitivity of the dA-dT homopolymer duplexes to this nuclease has been used (28), although this technique apparently works reliably only when the homopolymers are quite long. Exonuclease VII also will excise inserts put in by a homopolymer tailing method (29), although this enzyme is not generally available. Both of these methods yield both strands of the insert. The technique we have devised for the recovery of the insert has the advantages that it can potentially be used regardless of the method by which the insert was joined to the vector and that it yields only the anti-sense strand of the inserted DNA sequences. Titration of mRNA sequence concentration in mixtures of RNA species is most frequently and easily done with a single-stranded probe, and thus our method may prove quite useful. We have recently used cDNA probes prepared from pSy56, pS162, and pSp69 to investigate this structure of the individual globin genes in nuclei from sheep erythroid cells (30), and we anticipate that these probes will continue to facilitate our molecular analysis of hemoglobin switching in sheep.

Acknowledgments—We wish to thank Dr. Roy Curtis and his colleagues for providing us the EK 2 certified host X1776, Dr. George Khoury for suggesting the conditions for the nick-translation reaction, and Dr. Robert Goldberger and his colleagues for allowing us to use their P-3 lab. We wish particularly to acknowledge our indebtedness to W. Salser for his assistance in initiating these studies and for providing the terminal transferase. We are also most grateful to Eric Schneider and H. Coon for excellent technical assistance.

REFERENCES

**Table II**
Comparison of differences in restriction endonuclease sites among the sheep globin genes to the amino acid sequences

<table>
<thead>
<tr>
<th>Site Present in $\beta^b$ and $\beta^c$ sDNA But Not Present in $\gamma$</th>
<th>Sites Present in $\beta^b$ and $\beta^c$ sDNA But Not Present in $\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hpa II (Amino Acid Positions 11-12)</td>
<td>2. Hinf I (Amino Acid Positions 68-69)</td>
</tr>
<tr>
<td>$\beta^b$ and $\beta^c$</td>
<td>$\beta^b$ and $\beta^c$</td>
</tr>
<tr>
<td>Asp – Ser</td>
<td>Asp – Ser</td>
</tr>
<tr>
<td>Ser – Leu</td>
<td>Ser – Leu</td>
</tr>
<tr>
<td>(GAC UC)</td>
<td>(GAC UC)</td>
</tr>
<tr>
<td>Or</td>
<td>Or</td>
</tr>
<tr>
<td>3. Pvu II (Amino Acid Positions 66-67)</td>
<td></td>
</tr>
<tr>
<td>$\beta^b$ and $\beta^c$</td>
<td></td>
</tr>
<tr>
<td>Glu – Lys</td>
<td></td>
</tr>
<tr>
<td>(CAG CUG)</td>
<td></td>
</tr>
<tr>
<td>CAG CUG</td>
<td></td>
</tr>
<tr>
<td>AG (U)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Site Present in $\gamma$ Only</td>
<td></td>
</tr>
<tr>
<td>Hae III (Amino Acid Positions 124-125)</td>
<td></td>
</tr>
<tr>
<td>$\beta^b$ and $\beta^c$</td>
<td></td>
</tr>
<tr>
<td>Val – Leu</td>
<td></td>
</tr>
<tr>
<td>Gln – Leu</td>
<td></td>
</tr>
<tr>
<td>(GUUCU)</td>
<td></td>
</tr>
<tr>
<td>GA(CGU)</td>
<td></td>
</tr>
<tr>
<td>GCAG</td>
<td></td>
</tr>
<tr>
<td>LA (G)</td>
<td></td>
</tr>
</tbody>
</table>

1. Among the possible codons for these amino acids, those pairs were chosen which gave the recognition sequence for the endonuclease known to cut in approximately this position. The recognition sequence is indicated in the box.
2. All possible codons for these amino acids are shown.
3. Loss of this restriction site could occur by virtue of replacement of aspartic acid with asparagine in the globin or due to silent nucleotide change.
Cloning of Sheep Globin Genes

U.S.A. 72, 3961–3965


Access the most updated version of this article at http://www.jbc.org/content/254/15/6880

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/254/15/6880.full.html#ref-list-1