Proc. Natl. Acad. Sci. USA
Vol. 78, No. 10, pp. 5998–6002, October 1981
Biochemistry

Adult chicken α-globin genes, α^A and α^D: No anemic shock α-globin exists in domestic chickens

(nonanemic cDNA clones/chromosomal clones/DNA sequence)

JERRY B. DODGSON*, KEVIN C. MCCUNE†, DREW J. RUSLING†, ANDREE KRUST‡, AND JAMES DOUGLAS ENGEL†

*Departments of Microbiology and Public Health and Biochemistry, Michigan State University, E. Lansing, Michigan 48824; †Department of Biochemistry, Molecular and Cell Biology, Northwestern University, Evanston, Illinois 60201; and ‡Laboratoire de Genetique Moleculaire des Eucaryotes du Centre National de la Recherche Scientifique, Unite 184 de Biologie Moleculaire et de Genie Genetique de l' Institut National de la Santé et de la Recherche Médicale, Institut de Chimie Biologique, Faculté de Medecine, 67805 Strasbourg Cedex, France

Communicated by Irving M. Klotz, June 12, 1981

ABSTRACT **Three α-type globin genes have been identified in the α-globin linkage group of chickens. No other α-type genes have been directly shown to be within 10 kilobase pairs of any of these three closely linked genes. These three genes have been conclusively identified by DNA sequence analysis. The gene at the 5′ end of the linkage group is an embryonic α-type globin gene, π or π′, and the central gene corresponds to the minor adult α-globin, α^D. The 3′-terminal gene sequence corresponds to the sequence of cDNA clones previously described as "α^S", a presumed anemic shock-induced α-globin gene [Salser, W. A., Cummings, I., Liu, A., Strommer, J., Padayatty, J. & Clarke, P. (1979) in *Cellular and Molecular Regulation of Hemoglobin Switching*, eds. Stamatoyannopoulos, G. & Nienheis, A. (Grune and Stratton, New York), pp. 621–643; Richards, R. I. & Wells, J. R. E. (1980) *J. Biol. Chem.* 255, 9306–9311]. Several groups of workers have isolated α^S-type cDNA clones but no one has identified a cDNA clone corresponding to the published amino acid sequence of the major chicken α-globin, α^A. We have identified the α^S-type sequence as the only abundant α-like globin sequence in cDNA clones made from reticulocyte mRNA isolated from nonanemic chickens. Therefore, we suggest that the α^S-type sequence corresponds to the true α^A-globin species.**

Adult chicken erythrocytes contain two quite different α-globin polypeptides, α^A and α^D (1). The amino acid sequences of both of these proteins have been determined (2, 3). Several groups of workers have prepared cDNA clones from globin mRNA isolated from the reticulocytes present in the blood of phenylhydrazine-treated adult chickens (4–7), but none of them have been able to isolate a cDNA clone whose nucleotide sequence corresponds to the α^A amino acid sequence. Instead, a different α-type sequence has routinely been cloned that has been assumed to arise from the phenylhydrazine-induced anemia (4–6). This sequence has been assigned to an anemic shock-induced α-globin gene, "α^S" (5). The α^S sequence of Richards and Wells (5) appears to be similar to both the pHb1003 and the pHb1008 sequences of Salser *et al.* (4). The differences observed between these and other α^S-cDNA cloned sequences may arise from artifactual sequence variability generated in the cDNA cloning itself (5, 7).

Analysis of chicken chromosomal α-type globin regions has identified only three globin genes (8–10). One of these genes is clearly embryonic, and the other two have been identified as α^A and α^D by their hybridization to cDNA made to adult chicken reticulocyte mRNA and by the greater similarity of the α^A globin to mammalian adult α-globins (8, 10). We now extend the isolated region of the chicken α-globin gene linkage group a considerable distance in the 5′ direction. We have not yet demonstrated the presence of two embryonic α-globin genes (π and π′; ref. 1); however, we believe from previous chromosomal blotting data (8) and recent mapping data on these recombinants (not shown) that another embryonic α-globin locus does exist.

In this paper, we show that the chromosomal gene that we have designated α^A (10) corresponds to the α^S-cDNA cloned sequence. Finally, sequence analysis of cDNA clones prepared from nonanemic reticulocyte mRNA also shows that α^S-type sequences are the most abundant α-globin mRNAs present in normal (untreated) red cells. No cDNA clones corresponding to the published α^A amino acid sequence could be found among several such clones analyzed. These results lead us to suggest that the α^S-type sequence corresponds to the true α^A-globin gene sequence, at least for the standard White Leghorn chicken used in most laboratory studies.

MATERIALS AND METHODS

Preparation of Recombinant DNA Clones. Recombinant cDNA clones were prepared from total reticulocyte mRNA isolated from nonanemic chickens as described (11). The isolation and characterization of chromosomal chicken globin gene-containing bacteriophage has been described (10, 12). Restriction analyses were done by single and multiple digestions of phage DNA with commercial restriction enzymes followed by analysis on agarose gels and Southern (13) hybridization of the blots from these gels to various labeled nucleic acids (10, 12). Subcloning of DNA fragments isolated from these phage DNAs into pBR322 plasmid DNAs was performed as described (14–16).

DNA Sequence Determination. DNA sequence analysis was performed by the Maxam and Gilbert partial chemical degradation technique (17); in later cases, the modifications of Smith and Calvo (18) were used. Singly end-labeled restriction enzyme fragments were isolated from ethidium bromide-stained agarose gels as described (19).

RESULTS

Characterization of the Complete α-Globin Gene Linkage Group. Several recombinant DNA bacteriophage that contain α-type globin genes have been isolated from a complete library of chicken chromosomal DNA (refs. 9 and 10; Fig. 1). As seen in the map (Fig. 1A), these clones define an overlapping segment of chicken chromosomal DNA ≈35 kilobase pairs (kbp)

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

Abbreviation: kbp, kilobase pair(s).

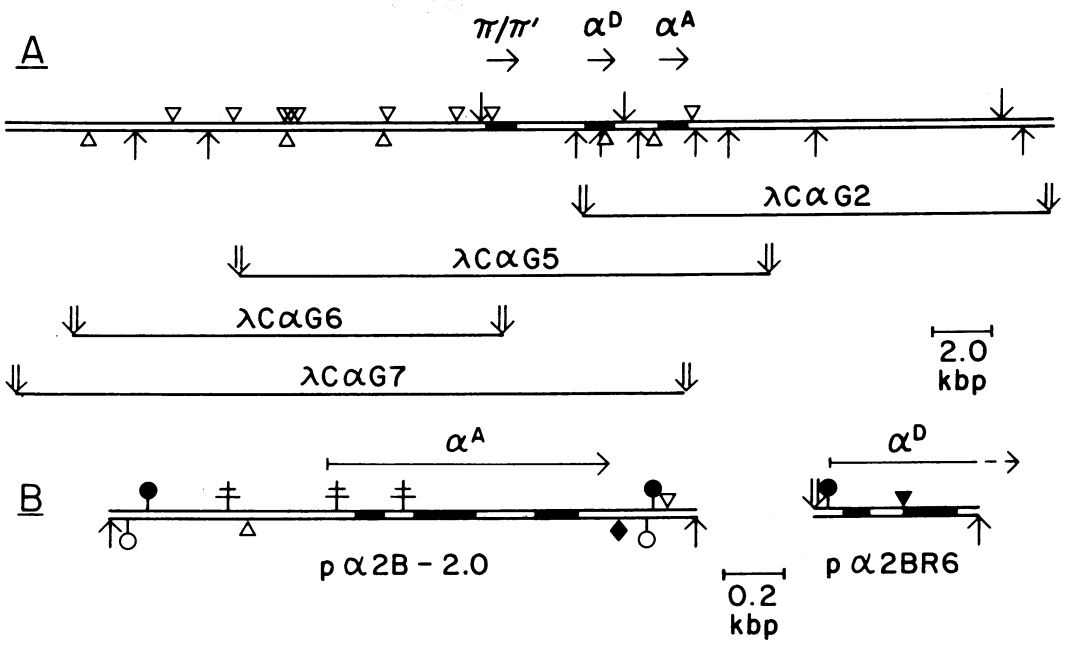

FIG. 1. (*A*) Chromosomal recombinants in the chicken α-globin complex. The overlapping recombinants λCαG2 and λCαG5 (10), λCαG6, and λCαG7 are shown below the contiguous portion of the chicken chromosome (≈35 kbp) containing the α-globin genes (8, 10). ■, Positions of the adult α^A and α^D genes (10) and of the π or π' (embryonic α-globin) gene-coding regions (unpublished observations); →, transcriptional direction of the three identified genes (5' to 3'). (*B*) Higher resolution restriction enzyme maps of subclones prepared from λCαG2 for DNA sequence analysis—pα2B-2.0 (containing the 2.0-kbp *Bam*HI chromosomal DNA fragment including the entire α^A gene) and pα2BR6 [containing the 0.56-kbp *Bam*HI/*Eco*RI (linker) insert fragment from the 5' end of the α^D gene]. ■, Positions of exons within each structural gene; → approximate extent and direction of transcription within these subclones as deduced from the cDNA clones described here and elsewhere (4–6). ⇓, *Eco*RI linker. Cleavage sites: ↓, *Eco*RI; ↑, *Bam*HI, ▽, *Hin*dIII; △, *Kpn* I; ⏚, *Sac* I; 🎈, *Pst* I; ⫩, *Hin*fI; ◆, *Bgl* I; ▼, *Msp* I.

long. Three regions that hybridize to adult or embryonic (or both) globin cDNA have been identified in this region. These three genes are transcribed in the same direction, left to right.

The 5'-terminal gene of this linkage group has been shown to hybridize to embryonic cDNA only (10). DNA sequence analysis has now confirmed that this gene corresponds to an embryonic α-type globin gene, π or π' (ref. 1; unpublished observations). The central hybridizing region in this area corresponds to the minor α-type globin present in both adult and embryonic erythrocytes, α^D. This is shown in Fig. 2 in which DNA sequence data from the chromosomal gene are compared with the sequence of a cloned cDNA copy of α^D mRNA and with the amino acid sequence of the α^D polypeptide (3).

The 3'-terminal hybridizing region (Fig. 1*A*) was identified as the α^A gene because it hybridized strongly to cDNA made to adult chicken globin mRNA and also to mammalian α-globin probes (8, 10). The amino acid sequence reported for α^A (2) is much more closely related to mammalian adult α-globins than is that of α^D (3). The 2.0-kbp *Bam*HI fragment of λCαG2 and the 1.7-kbp *Bam*HI/*Eco*RI (linker) fragment of λCαG7, which contain this gene, have been subcloned into pBR322. Detailed restriction mapping of this region is shown in Fig. 1*B*. DNA sequence data in a region to the right of the right-hand *Hin*fI site in Fig. 1*B* are shown in Fig. 3. It can be seen that the chromosomal sequence covers all of the central exon of this α-globin gene. In this region, the chromosomal sequence and our cDNA clone sequence (see below) show complete agreement with the α^S sequence of Richards and Wells (5). This is also true for the 3'-untranslated region of the α-globin gene as determined by DNA sequence analysis to the left of the *Eco*RI site generated by the insertion of a synthetic linker near the 3' end of the gene in λCαG7 (Fig. 1*A*; unpublished observations) and by comparison of the cDNA clone sequence (pCcαG5) 3'-untranslated region (data not shown). The sequence around the central exon identifies the junction of this exon with the two introns present in this gene (10) at the expected positions for α-globin genes (20). Also, the dinucleotides G-T and A-G at the 5' and 3' ends of introns (21) appear in the expected consensus positions of the chromosomal sequence. We have aligned these consensus sequences to join (in the mature mRNA) separated 5' and 3' exons within codon 31 and between codons 99 and 100, respectively, for both the α^D- and α^A-chromosomal sequences (Figs. 2 and 3).

The data of Fig. 3 identify the 3'-terminal gene of the cluster as that which corresponds to the α^S sequence observed in cDNA cloning experiments (4–6). We have examined the rest of the cloned 35-kbp α-globin gene region for other globin genes by hybridization to adult globin cDNA at both high and low criteria (0.05 M Na^+, 65°C and 0.3 M Na^+ 60°C, respectively). No regions that hybridize to adult globin cDNA other than the genes described above could be identified in any of the clones in Fig. 1*A* or in chromosomal DNA (8) when digested by a variety of multiple restriction enzyme combinations (results not shown).

DNA Sequence of cDNA Clones Prepared from Normal Chicken Reticulocyte mRNA. One of the major problems in the analysis of globin cDNA (and therefore indirectly, mRNA) populations in chickens has been that the globin message has been isolated from animals treated with phenylhydrazine to increase the relative population of circulating immature red cells, wherein globin mRNA is still very abundant. This led to the speculation (4–6) that the disparity between the amino acid sequence predicted from cDNA cloning of anemic α-globin mRNAs and the previously reported α^A polypeptide sequence (2) was due to the presence of an abundant anemic-shock α-globin polypeptide in treated birds.

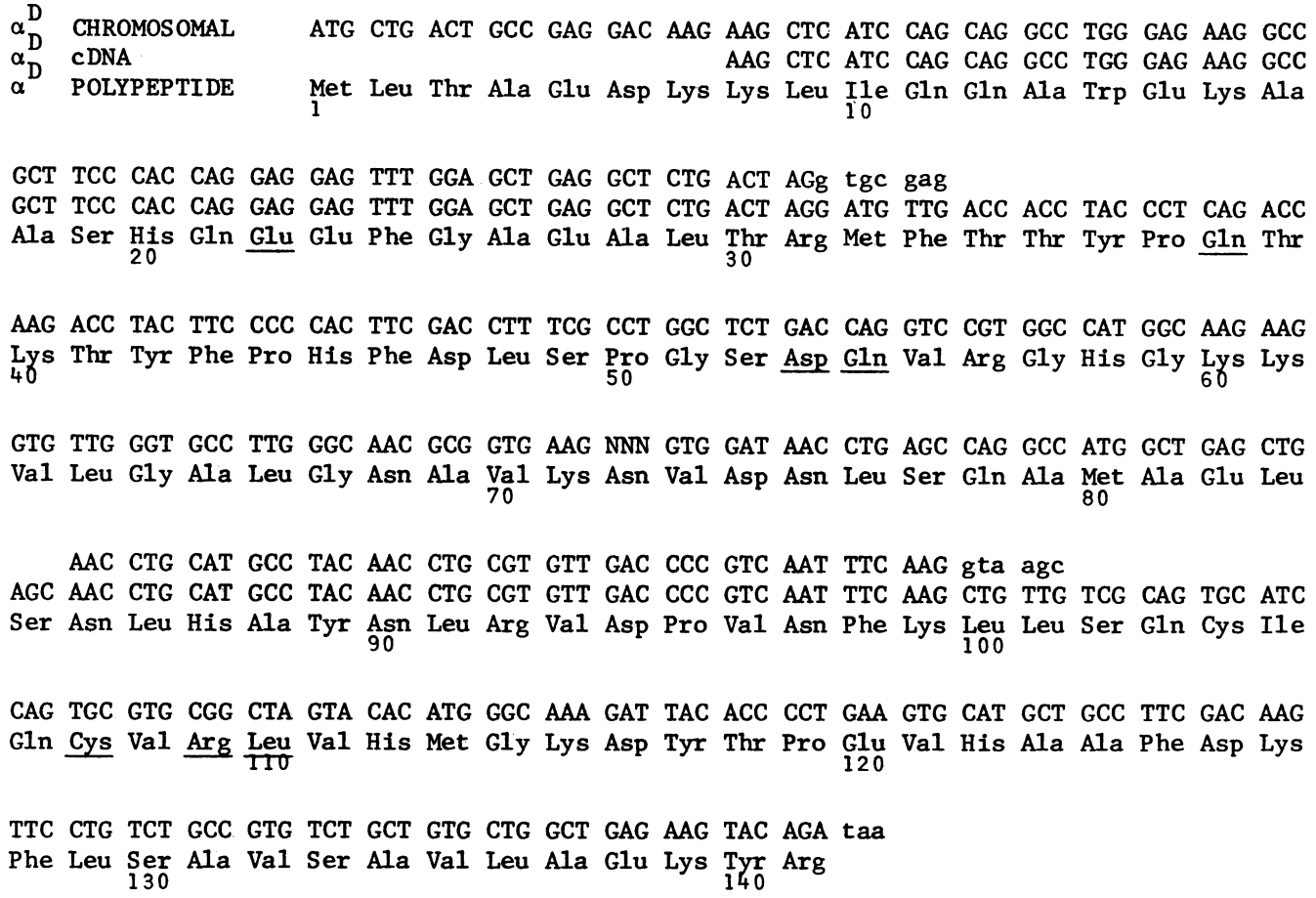

FIG. 2. DNA sequence determination of the α^D gene in chromosomal and cDNA recombinants: Comparison of DNA sequence of pCcαG24 (α^D, cDNA) with that of pα2BR6 (α^D chromosomal) and the α^D polypeptide sequence. The polypeptide sequence is primarily from Takei *et al.* (3) but includes seven changes (underlined) as deduced from the cDNA and chromosomal DNA sequences. N, positions of abnormalities in sequencing gels (or of divergent results obtained by comparison of two strands sequenced in opposite directions). Noncoding chromosomal DNA sequences are indicated by lower-case letters.

We therefore prepared mRNA (and subsequently, cDNA clones) from untreated adult hens for analysis of globin mRNA species in normal chicken red cells. Of ≈35 recombinants isolated, 5 were demonstrated to be non-β-globin genes by selective screening using an adult chromosomal β-globin probe (15) versus adult α-globin probes (10). Of the five non-β cDNA clones isolated, four hybridized specifically to the chromosomal locus previously identified as the α^A gene on mapping blots and the fifth (pCcαG24) hybridized to the α^D gene (ref. 10; mapping blots not shown). The recombinant containing the largest inserted double-stranded cDNA among the four α^A-hybridizing segments (pCcαG5) and also the single α^D-cDNA clone were sequenced directly (Figs. 2 and 3). pCcαG5 differed from its three α^A siblings only in length (unpublished observations). As shown in Figs. 2 and 3, the cDNA and chromosomal sequences *are identical* in the regions of overlap, demonstrating that, in normal healthy French White Leghorn chickens (from which the nonanemic globin mRNA was prepared), the common abundant adult α-globin mRNA is of the α^S type and that the presence of the α^S sequence is not due to phenylhydrazine-induced anemia in these animals.

DISCUSSION

Our results, along with previous cDNA cloning studies (4–7) suggest two possible explanations for the inability of previous investigators to isolate cDNA clones corresponding to the published α^A-globin amino acid sequence. First, there could be a gene corresponding to the α^A amino acid sequence (2) that is yet to be discovered. We think this improbable for several reasons. Although the α^A-globin protein is the major α-type globin in definitive chicken erythrocytes (1), no one has been able to prepare a cDNA clone to a mRNA that corresponds to the published protein sequence. We have shown that even in nonanemic chicken reticulocytes the abundant α-type cDNA observed is the same as that found in anemic birds, the α^S type. An argument for the presence of both α^A and α^S genes in the chicken genome would require that the chromosomal gene corresponding to the α^A-amino acid sequence either be unlinked to the rest of the chicken α-globin genes or be so different in nucleotide sequence as to be unable to hybridize to α^D- or α^S-type probes. The latter possibility seems unlikely because there are long regions of close correspondence between α^A- and α^S-amino acid sequences (2, 5), and the former possibility contradicts all other globin gene linkage groups (14, 22–24). The chicken globin gene linkage groups, in fact, seem to exhibit extremely close linkage properties, with adjacent genes rarely more than 3 kbp apart (10, 15).

The more probable explanation for the data presented here and elsewhere (4–6) is that the published amino acid sequence for α^A-globin (2) is in fact the sequence of an unusual protein variant of chicken α^A and that, in fact, the amino acid sequence predicted by the α^S-cDNA clone (5) is the more common α^A-

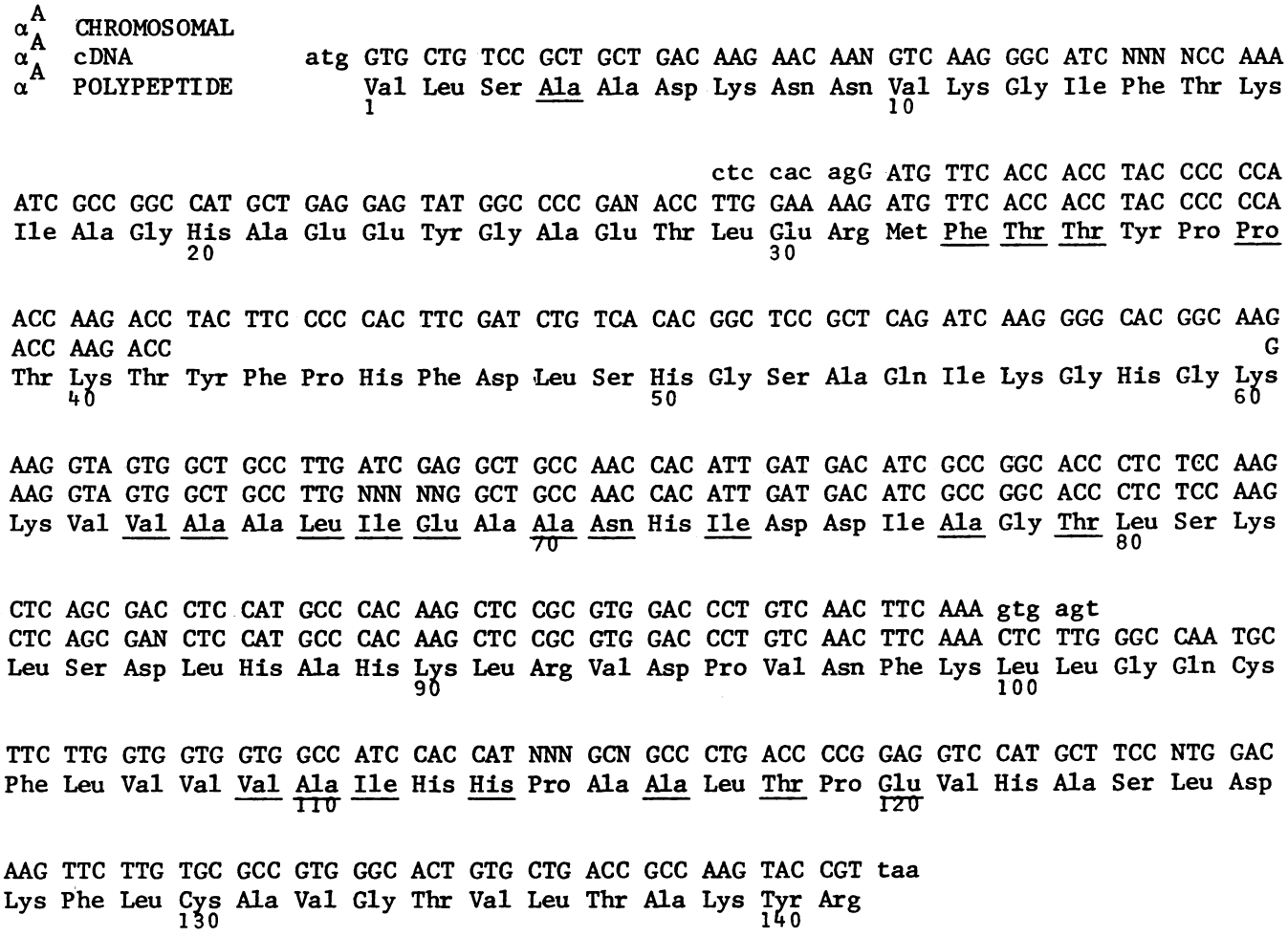

FIG. 3. DNA sequence determination of the α^A gene in chromosomal and cDNA recombinants. DNA sequences generated from pCcαG5 (α^A cDNA) and pα2B-2.0 (α^A chromosomal) are compared. In the region of overlap, the chromosomal and nonanemic α-globin sequences are identical, and both are also identical to the sequence of the α^S globin of Richards and Wells (5). The α^A polypeptide sequence is primarily that of Matsuda *et al.* (2), with inferred changes underlined. N, internal discrepancies in DNA sequence. Chromosomal DNA sequences flanking the central exon are indicated by lower case letters.

globin sequence. A number of observations support this contention. First, cDNA clones prepared to anemic reticulocyte RNA isolated from hens from Texas (4), California (5, 25), and Australia (6), as well as cDNA clones prepared from nonanemic French hens, are all precisely (or very close to) the same in DNA sequence as the chromosomal locus we identify as the α^A gene. This implies that the most universal sequence for α^A is indeed that reported here. Second, the report demonstrating the existence of a new hemoglobin species as a result of either chronic or acute phenylhydrazine-induced anemia in hens (26) also demonstrates that this electrophoretic variant is a minor component present only in the blood of treated animals, reaching a maximum of 5% of total hemoglobin. It would seem to be unlikely then that all characterized α-globin cDNA clones reported to date (from anemic and nonanemic reticulocyte RNA) find only this α-type globin sequence. In fact, by this latter explanation, it should have been impossible for us to isolate the α^S gene product. Although our sequence prediction does not explain the tryptic peptide map of Brown and Ingram (1), neither does the original amino acid sequence reported for α^A (2). As demonstrated previously, however, the greater number of spots observed than predicted could be accounted for either by partial proteolytic cleavage (2), oxidation products (1), or various combinations of these.

Although sequence analysis errors (either in DNA or amino acid analysis) may in part contribute to the differences between the α^A and the α^S sequences, it seems unlikely that such a major effect (22 of 141 amino acids) could be explained in this fashion. In fact, the β-globin polypeptide sequence (25) precisely matches both the cDNA (7) and chromosomally derived recombinant DNA (15) sequence predictions, and we have shown here that the α^D-gene sequence predictions vary by seven amino acids from the polypeptide sequence for α^D (3). Of course, further protein sequence analyses will be required to confirm this explanation. We would predict that analyzing the α component of HbA in American or European White Leghorn chickens would demonstrate the presence of the α^A gene sequence reported here (and by others as α^S).

Although it is well recognized that β-globin structural genes crosshybridize to heterologous β-globin loci (e.g., see ref. 15), it is *prima facie* somewhat odd that α-globin genes crosshybridize only poorly. However, the sequence analysis reported here shows that α^A and α^D have diverged in coding sequence by almost 35% but even more thermodynamically important is the fact that the longest homologous DNA sequence held in common between α^A- and α^D-coding segments is 26 nucleotides (from codons 39–47). At our normal blot-washing stringency (20 mM Na^+ equivalent, 65°C; ref. 15), we would not expect

heterologous hybrids (matched precisely for no more than 26 nucleotides) to be stable.

In summary, we conclude the following. (*i*) The abundant α-globin component in anemic and nonanemic definitive reticulocytes is the same. (We also present negative evidence from chromosomal recombinant DNA studies that no closely linked gene exists that corresponds to an anemic-shock α-globin.) (*ii*) The intron positions in α-globin genes (20) and the consensus nucleotides surrounding exons (21) are evolutionarily conserved. (*iii*) The sequence analyses of the α^D cDNA and chromosomal DNA copies of the minor adult chicken α-globin gene are as shown in Fig. 2. (*iv*) The α-globin genes in chickens maintain the 5′... embryonic–adult ... 3′ temporal and physical relationship observed everywhere else but in the chicken β-globin gene cluster (15). This final observation may mean, therefore, that the physical gene arrangement has no effect on the temporal control of gene expression or that *cis*-acting elements have a different arrangement in the β-globin gene cluster than in all other globin gene families studied to date.

The support of Drs. P. Chambon, in whose laboratory the cDNA clones were prepared, and N. Davidson, in whose laboratory the chromosomal clones were initially isolated, is gratefully acknowledged. We thank Dr. B. Chapman for spirited discussions of this work and access to protein sequence data prior to publication and our colleagues B. J. Sugarman and M. Dolan for fruitful discussions. We also thank Mses. Patricia Creatura and Sheryl Niemiec for excellent technical assistance. J.B.D. is supported by National Institutes of Health Grant GM28837, J.D.E. is supported by National Institutes of Health Grants HL24415 and CA15145, and A.K. is supported by Centre National de la Recherche Scientifique Grant ATP 006520/50 to P. Chambon.

1. Brown, J. L. & Ingram, V. M. (1974) *J. Biol. Chem.* **249**, 3960–3972.
2. Matsuda, G., Takei, H., Wu, K. C. & Shiozawa, T. (1971) *Int. J. Prot. Res.* **3**, 173–174.
3. Takei, H., Ota, Y., Wu, K. C., Kiyohara, T. & Matsuda, G. (1975) *J. Biochem. (Tokyo)* **77**, 1345–1347.
4. Salser, W. A., Cummings, I., Liu, A., Strommer, J., Padayatty, J. & Clarke, P. (1979) in *Cellular and Molecular Regulation of Hemoglobin Switching*, eds. Stamatoyannopoulos, G. & Nienheis, A. (Grune and Stratton, New York), pp. 621–643.
5. Richards, R. I. & Wells, J. R. E. (1980) *J. Biol. Chem.* **255**, 9306–9311.
6. Deacon, N. J., Shine, J. & Naora, H. (1980) *Nucleic Acids Res.* **8**, 1187–1199.
7. Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E. & Goodman, H. M. (1979) *Nucleic Acids Res.* **7**, 1137–1146.
8. Engel, J. D. & Dodgson, J. B. (1978) *J. Biol. Chem.* **253**, 8239–8246.
9. Dodgson, J. B., Strommer, J. & Engel, J. D. (1979) in *Eucaryotic Gene Regulation*, Eighth Annual ICN-UCLA Symposium on Molecular and Cellular Biology, eds. Maniatis, T., Axel, R. & Fox, C. F. (Academic, New York), pp. 383–392.
10. Engel, J. D. & Dodgson, J. B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2596–2600.
11. Cochet, M., Perrin, F., Gannon, F., Krust, A., Chambon, P., McKnight, G. S., Lee, D. C., Mayo, K. E. & Palmiter, R. (1979) *Nucleic Acids Res.* **6**, 2435–2452.
12. Dodgson, J. B., Strommer, J. S. & Engel, J. D. (1979) *Cell* **17**, 879–887.
13. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
14. Lacy, E., Hardison, R. C., Quon, D. & Maniatis, T. (1979) *Cell* **18**, 1273–1281.
15. Dolan, M., Sugarman, B. J., Dodgson, J. B. & Engel, J. D. (1981) *Cell* **24**, 669–677.
16. Engel, J. D. & Dodgson, J. B. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2856–2860.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
18. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255–2274.
19. Girvitz, S. C., Bacchetti, S., Rainbow, A. J. & Graham, F. L. (1980) *Anal. Biochem.* **106**, 492–496.
20. Leder, A., Miller, H. I., Hamer, D. H., Seidman, J. G., Norman, B., Sullivan, M. & Leder, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 6187–6191.
21. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
22. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) *Cell* **19**, 959–972.
23. Jahn, C. J., Hutchison, C. A., III, Phillips, S. J., Weaver, S. J., Haigwood, N. L., Voliva, C. F. & Edgell, M. H. (1980) *Cell* **21**, 159–168.
24. Lauer, J., Shen, J. C.-K. & Maniatis, T. (1980) *Cell* **20**, 119–130.
25. Matsuda, G., Maita, T., Misumo, K. & Ota, H. (1973) *Nature (London) New Biol.* **244**, 244.
26. Stino, F. K. R. & Washburn, K. W. (1970) *Poult. Sci.* **49**, 101–114.