This is to certify that the

dissertation entitled

CONSTITUTIVE EXPRESSION OF
CHICKEN ADULT ALPHA-GLOBIN GENES

presented by

Jiing-Dwan Lee

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Microbiology

Major professor

Date 11/29/89

CONSTITUTIVE EXPRESSION OF

CHICKEN ADULT α-GLOBIN GENES

By

Jiing-Dwan Lee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Microbiology
and Public Health

1989

ABSTRACT

CONSTITUTIVE EXPRESSION OF
CHICKEN ADULT α-GLOBIN GENES

By

Jiing-Dwan Lee

The chicken adult α-globin genes, α^A and α^D, are closely linked in chicken chromosomal DNA. These two genes are expressed coordinately in primitive and definitive erythroid cells in a 3:1 ratio. The constitutive expression of these two genes was studied in QT6 quail fibroblast cells. Clones containing the α-globin gene(s) and hybrid genes were transiently transfected into the QT6 cell line and RNA was isolated after 48 hrs from the transfected cells. The expression of the transfected gene(s) was assessed using a RNase protection assay. These experiments demonstrated that the promoter region of the α^A-globin gene (but not that of the α^D-globin gene) was sufficient to promote a detectable level of constitutive transcription from both the α^A- and α^D-globin gene coding regions. Mutants with deletions in the α^A-globin promoter region were then transfected into the QT6 cell line, and subsequent RNase and DNase protection analyses suggested the existence of two activator regions and one repressor region directly upstream of this gene. Other regions flanking the α^A-globin gene have also been tested by similar techniques for their effect on the transcription level of the α^A-globin

gene. Orientation-specific activating or inhibitory effects have been observed within these regions. Several protein-binding domains have been identified in an 84 bp fragment which showed a directional activating effect. This result suggests that DNA-binding factors may be involved in this phenomenon.

To Mom and Ling-Ling

ACKNOWLEDGEMENTS

I would like to thank Dr. Jerry Dodgson for his intellectual and long financial support.

Thanks to Dr. Wynne Lewis and Dr. Richard Schwartz for their scientific ideas and great companionship.

Thanks to Dr. Paul Boyer and Dr. Paul Bates for their scientific advice and friendship.

Thanks to Moriko Ito and Sue Kalvonjian for their patience and friendship.

And finally, thanks to my labmates; Dave Browne, S-Y Son, GTP and Natalie Moore for their constant help and endurance.

TABLE OF CONTENTS

LIST OF TABLES.....vii

LIST OF FIGURES.....viii

CHAPTER 1

Introduction.....1
References.....4

CHAPTER 2

Literature Review
Eucaryotic Transcription Regulatory Elements.....6
Erythropoiesis.....12
Globin Protein and Genes.....18
References.....34

CHAPTER 3

Materials and Methods.....41
References.....52

CHAPTER 4

Results and Discussion.....53
References.....100

SUMMARY.....101

APPENDIX I.....102

APPENDIX II.....120

LIST OF TABLE

CHAPTER 4

TABLE #		PAGE
1	Densitometric analysis of mutants containing deletions in the promoter region of the α^A-globin gene.	74
2	The effects of various fragments on the transcription level of the chicken α^A-globin gene.	88

LIST OF FIGURES

FIGURE #		PAGE
	CHAPTER 2	
1	Pathway of Chicken Erythropoiesis	13
2	Chromosomal Organization of the Globin Gene Clusters in Various Species.	20
3	Arrangement of Hemoglobin Genes Along Two Human Chromosomes.	24
4	Organization of the Chicken α-type Globin Genes.	26
5	Organization of Chicken ß-type Globin Genes.	28
	CHAPTER 3	
1	Scheme for Deletion Mutant Construction.	43
	CHAPTER 4	
1	Restriction Enzyme Maps of Three Subclones of Chicken α-Globin Genes.	55
2	Scheme for Testing Expression Level of Clones in QT6 Quail Fibroblasts.	57
3	Scheme for Making Riboprobe for α^A-and α^D-globin Genes and CAT Gene Transcripts.	59
4	Transfection of Clones Containing the α-Globin Gene(s) and Hybrid Genes into QT6 Fibroblast Cells.	61
5	Maps of Chicken α-Globin Hybrid Genes.	64
6	Size and Location of the Deletion Mutants.	70
7	Transfection of Deletion Mutants into QT6 Fibroblast Cells.	72

LIST OF FIGURES

FIGURE #		PAGE
8	DNase Protection Analysis of the Chicken α^A-globin Promoter Using Crude Nuclear Extracts from QT6 Quail Fibroblast.	76
9	Summary of Protein-DNA Interactions of the Chicken α^A-globin Promoter Region.	78
10	Subcloning of Fragments Around the Chicken α^A-globin Gene.	81
11	Transfection of Clones Containing the Flanking Regions of the α^A-Globin Gene into QT6 Fibroblast Cells.	84
12	Transfection of Clones Containing the Flanking Regions of the α^A-Globin Gene into QT6 Fibroblast Cells.	86
13	DNase Protection Analysis of the 84 bp (556-638) Region in the Chicken α^D-globin Gene Using Crude Nuclear Extracts from QT6 Quail Fibroblast.	91
14	Summary of Protein-DNA Interactions on the 84 bp Fragment (556-538) from the Chicken α^D-globin Gene Second Intron (Fragment B).	93
15	The Positions of the GAGGTC Motif in and around the α^A- and α^D-Globin Gene.	97

CHAPTER 1

INTRODUCTION

Globin genes, along with the expression, function and abnormalities thereof have been extensively reviewed (1,2,3,). All normal hemoglobins of higher vertebrates are tetramers consisting of two α and two ß chains. In mammals and birds, the α-type and ß-type genes are organized in separate gene clusters located on different chromosomes (1,3,4). Each cluster contains different globin variants, the expression of which is highly regulated throughout embryonic, fetal and adult stages of development.

Since the expression of globin genes is restricted to one type of cell and is tightly controlled throughout development, these genes serve as a very useful system to study gene regulation. Regulatory elements of transcription in and around α- and ß-globin genes have been identified by transfecting cloned wild type or mutant globin genes into cultured cells and analyzing their expression. Two types of *cis*-acting regulatory elements have been shown by transfection experiments: erythroid-specific elements which only function in erythroid cells (5-12) and constitutive elements which can function in both erythroid and non-erythroid cells (5,13-17). Most of these elements have a positive effect on the

transcription of globin genes and have the properties of transcriptional enhancers. A few of these elements, located in the 5' flanking regions of some globin genes, act in a negative manner and have the properties of transcriptional silencers (18,19).

The chicken α globin gene locus consists of three closely linked genes; two of them (α^A and α^D) are expressed in both primitive and definitive erythroid cells and the other one (π) is expressed only in primitive cells. Primitive cells are the first hemoglobin-containing cells to appear during erythropoiesis in the blood islands at about 35 hrs of development. At approximately 5 days of development, this primitive group of erythroid cells is rapidly replaced by definitive red cells, which become the predominant red cell species of older embryos and adults (20-22). All three genes are transcribed in the same direction and are arranged 5'-π-α^D-α^A-3'(23,24), relative to the direction of transcription. Unlike most mammals which contain two almost identical adult α-globin genes, the two adult chicken α-globin genes are very dissimilar in sequence, which suggests that they arose from an ancient globin gene duplication (23,24). Although α^A- and α^D-globin genes have lower expression levels in primitive red cells than in definitive red cells, they appear to maintain a relative expression ratio of about 3:1 in both cell types (25).

Transient transfection studies in chicken erythrocytes

have identified erythroid-specific enhancers in the promoter of the chicken π and α^D-globin genes and in the 3' flanking region of the chicken α^A-globin gene (26-28). Similar tissue-specific enhancers also have been identified in the chicken ß-globin gene family (8,26,29,30). Moreover, three regions within the promoter of the mouse and rabbit ß-globin genes have been recognized to be essential for maintaining the basal constitutive transcription levels of these genes. These are the CACCC box, CCAAT box and TATA box (31,32). However, less attention has been focused on the constitutive expression of α globin genes. Furthermore, the chicken α^A-and α^D-gene promoters appear to lack consensus CCAAT sequences found in other globin gene promoters, including those of π and the various chicken ß-type globin genes. We transiently transfected QT6 cells, a chemically transformed quail fibroblast cell line, with cloned mutant α^A-and α^D-genes in order to identify constitutive transcription-regulatory elements.

Our studies indicate that there are positive and negative regulatory elements within the 5' flanking region of the α^A-globin gene. Analysis was performed on mutants in which flanking region fragments were inserted upstream and downstream of the α^A-globin gene. These experiments showed an orientation-specific enhancement or repression of RNA levels by some DNA fragments.

References:

(1) Efstratiadis, A. et al., *Cell* **21**:653, 1980.

(2) Maniatis, T. et al., *Annu. Rev. Genet.* **14**:145, 1980.

(3) Collins, F. S., Weissman, S. M. *PNAS* **31**:315, 1984.

(4) Dolan, M. et al., *Cell* **24**:669, 1984.

(5) Antonion, M. et al., *EMBO J.* **7**:377, 1988.

(6) Chao, M. V. et al., *Cell* **32**:483, 1983.

(7) Charnay, P. et al., *Cell* **38**:251, 1984.

(8) Choi, O.-R., Engel, J. D., *Nature* **323**:731, 1986.

(9) Choi, O.-R., Engel, J. D., *Cell* **55**:17, 1988.

(10) Cowie, A., Myers, R. M., *Mol. Cell. Biol* **8**:3122. 1988.

(11) Hess, J. E. et al., *PNAS* **83**:4312, 1986.

(12) Wright, S. et al., *Cell* **38**:265, 1984.

(13) Banerji, J., Rusconi, S., Schaffner, W., *Cell* **27**:299, 1981.

(14) Charnay, P., Mellon, P., Maniatis. T., *Mol. Cell. Biol.* **5**:1498, 1985.

(15) Grosveld, G. C. et al., *Nature* **295**:120, 1982.

(16) Myers, R. M., Tilly, K., Maniatis, T., *Science* **232**:613, 1986.

(17) Treisman, R., Green, M. R., Maniatis, T., *PNAS* **80**:7428, 1983.

(18) Atweh, G. F. et al., *Mol. Cell. Biol.* **8**:5047, 1988.

(19) Cao, S. X. et al., *PNAS* **86**:5306, 1989.

(20) Bruns, G. A., Ingram, V. M., *Trans. Roy. Soc. Lond. Ser. B. Biol. Sci.* **266:**225, 1973.

(21) Mahoney, K. A., Hyer, B. J., Chan, L. N. L., *Devlop. biol.* **56:**412, 1977.

(22) Wilt, F. J., *Advan. Morphogen.* **6:**89, 1967.

(23) Dodgson, J. B. et al., *PNAS* **78:**5998, 1981.

(24) Dodgson, J. B., Engel, J. D., *J. Biol. Chem.* **258:**4623, 1983.

(25) Brown, J. L., Ingram, V. M., *J. Biol. Chem.* **249:**3960,1974.

(26) Evans, T., Reitman, M., Felsenfeld, G., *PNAS* **85:**5976, 1988.

(27) Kemper, B., Jackson, P. D., Felsenfeld. G., *Mol. Cell. Biol.* **7:**2059, 1987.

(28) Knezetic, J. A., Felsenfeld. G., *Mol. Cell. biol.* **9:**893, 1989.

(29) Emerson, B. M. et al., *PNAS* **84:**4786, 1987.

(30) Hesse, J. E. et al., *PNAS* **83:**4312, 1986.

(31) Dierks, P. et al. *Cell* **32:**695, 1983.

(32) Myers, R. M., Tilly, K., Maniatis, T., *Science* **232:**613, 1986.

CHAPTER 2

LITERATURE REVIEW

1. Eucaryotic transcription regulatory elements:

Gene expression can be regulated at each step in the pathway from DNA to protein (1). While transcriptional control is the most common form of regulation in gene expression, regulation can also occur during RNA processing (removal of intervening sequences, addition of the CAP structure and polyadenylation), nuclear transport, mRNA turnover, translation initiation and post-translational modification.

Promoters and enhancers are two well-characterized DNA sequence components required for proper transcriptional regulation in many higher eucaryotic genes. Both are defined in the functional sense, generally using transfection studies similar to those described herein. Promoters are located immediately upstream from the initiation site of transcription (CAP site) and are normally comprised of about 100 base pairs (bp)(2,3). Enhancers on the other hand can be located several kilobase pairs (kb) 3' or 5' away from the gene or in some cases within the structural gene itself (4). The promoter is required for accurate and efficient transcription and operates in a unidirectional manner, while the enhancer is generally

involved in regulating the overal level of transcription. Such modulation may be in response to environmental stimuli or to a tissue-specific or developmental stage specific signal. By definition, enhancers function in both possible orientations (2-4).

A typical promoter contains an AT-rich sequence (TATA box) (5), a CCAAT sequence (CAT box) (11) and other upstream promoter elements (UPEs) (6-8). The TATA box is located 25-30 base pairs upstream from the CAP site. Elimination of the TATA box often results in the usage of multiple CAP sites, but does not always decrease the rate of transcription from the region (9,10). The CAT box, which exists in many eucaryotic promoters, is usually found 80-100 base pairs upstream from the CAP site (11). Omitting the CAT box doesn't affect the CAP site but reduces the transcriptional efficiency of the gene (12-14). Another conserved element is the GC box which has the consensus sequence GGGCGG and usually can be found in the area between 40-200 base pairs upstream from the CAP site (15). The GC box functions in either orientation and is often present in multiple copies. Like the CAT box , the GC box is also responsible for the efficient transcription of the gene (11,16,17).

DNA footprint (18) and gel retardation studies (8) have indicated that there are specific proteins which bind to regulatory elements in promoters and enhancers. In some cases these proteins have been purified and studied biochemically

(15,19,20). TFIID which is partially purified from Hela cells and has an apparent molecular weight of 120-140 kilo daltons, binds to the TATA box and forms a complex with the TFIIA factor which is required for transcription initiation (21). CP1, CP2 and NF-1 are members of the CCAAT-binding protein family, but each of them has a different binding efficiency with the CCAAT sequences in different promoters (22). SP1 is another transcription factor isolated from Hela cell extracts which can bind to the GC box and then activate transcription by RNA polymerase II (23-26).

The first enhancer was discovered in a 200 bp-long DNA segment of the simian virus 40 (SV40). In the following years more enhancers were found in many different viruses, and several of these enhancers were isolated and characterized by an "enhancer trap" selection assay (27,28). Some of these viral enhancers appear to be involved in the induction of certain types of cancer. During a viral infection an enhancer-containing segment of the virus can be inserted in the vicinity of a cellular gene that controls cell proliferation. Because of this insertion, the gene (now referred to as an oncogene) is deregulated, leading to uncontrolled cell growth and to malignant transformation (29).

The first tissue-specific enhancer was discovered in the mouse immunoglobin heavy chain (IgH) gene (30). This shows that not only viral but also cellular genes use enhancers for gene regulation. The IgH enhancer stimulates transcription in

a tissue-specific manner and is the first genetic element described to have such a property. Since then a variety of tissue-specific enhancers have been detected in other cellular genes of higher organisms (23,31). Examples include the pancreas-specific enhancer from the rat insulin II gene (32) and the erythroid-specific enhancers from chicken α- and ß-globin genes (33-35).

Another class of enhancing elements confers inducible transcription, for example, in response to steroid hormones (36,37), heavy metals (38,39), growth factors (40), heat shock (41) and virus infection (42-44). Some transcription factors that bind to the above enhancer sequences have been identified. For example, the heat-shock transcription factor (HSTF) has an apparent molecular weight of 110 kilo daltons (in *Drosophila*), and is also present in uninduced cells, but it is not bound to its cognate heat-shock response element (HSEs) until the cells have been subjected to heat shock (45-47). It has been suggested that in normal cells, HSTF exists in a non-binding form and is converted upon heat shock to a high affinity, sequence-specific binding form by a post-translational modification. Serum response factor (SRF), on the other hand, can interact with serum response elements (SREs) of c-fos and other genes and functions as a constitutive transcription activator *in vitro*. The role of SRF *in vitro* in response to growth factor stimulation is not clear,

but studies suggest that post-translational modification and/or interaction with other protein factors may be involved (48).

The cellular enhancers can be roughly classified as tissue-specific and signal-inducible enhancers as discussed above; however, some enhancers can interact with ubiquitous protein factors and enhance the transcription of the target gene constitutively. For example, the octamer sequence (ATTTGCAT) in the IgH enhancer is not only functional in B cell-specific genes but also in non-B-cell-specific genes, such as histone H2b genes (49-51) and various U snRNA genes (52). Two protein factors, Oct-1 (ubiquitous in mammalian cells) and Oct-2 (expressed mainly in B and T lymphocytes), both can bind to this octamer sequence (53-59). The binding of ubiquitous Oct-1 may explain why the octamer sequence can be functional in non-B-cell environments. The serum response element (SRE) is another example of a constitutive enhancer. SRE, in addition to its serum-inducible activity as we discussed previously, also has a basal constitutive activity which can be eliminated by mutations that block the binding of serum response factor (SRF), a ubiquitous protein factor that binds to the SRE. (60)

Recently a group of negative enhancers (silencers) have been discovered. Silencers have similar properties to enhancers except that they reduce rather than increase the transcriptional efficiency of the target promoter. The first

silencer was found in the MAT locus of *Saccharomyces cerevisiae* which determines the mating type (61). Other elements with analogous properties have been found in the 5' upstream region of the promoters of the following genes; human α-(62) and ϵ-globin (63), chicken lysozyme (64), c-myc (65), human ß-interferon (66) and p53 (67). The general belief is that most cellular genes , perhaps with some exceptions among the housekeeping genes (68), are regulated by a combination of both enhancer(s) and silencer(s) acting cooperatively.

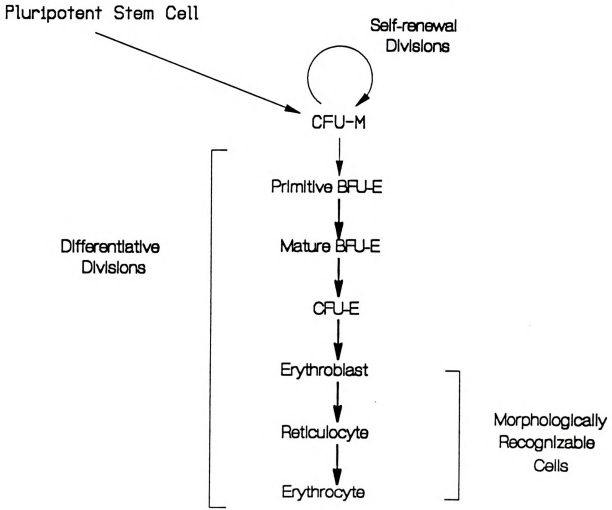
2. Erythropoiesis:

Erythropoiesis, the developmental process by which a pluripotent hemopoietic stem cell is converted to a mature erythrocyte, has been well studied in chicken, mouse and human systems (figure 1). During normal erythropoiesis in the developing chicken embryo, two morphologically distinct erythroid cell lineages appear in sequence. The first lineage is that of the primitive cells which arise in embryonic blood islands, and are the first erythroid cells to appear in the embryonic circulation. They reach a maximum in cell number after 8 days of incubation, followed by a decrease until the cells are no longer detectable in the 16 day old embryo. The definitive cells arise in the yolk sac and then move to the bone marrow (the erythropoietic organ in the adult). They begin to appear in the embryonic circulation at about the fifth day of incubation and increase exponentially to become the predominant red cell species of older embryos and adults (69-71).

There are several types of erythroid precursor cells that can be identified by morphological characteristics, biochemical properties and the surface antigens they present (72-76). The first identifiable erythroid cell type is the colony forming unit-marrow cell (CFU-M) which is defined by its ability to self-renew and to develop into erythrocytic clones when injected into bone marrow of irradiated chickens

Figure 1: Pathway of chicken erythropoiesis

The CFU-M stage may either continue to self-replicate itself or irreversibly commit to further differentiation.
Abbreviations: CFU-M → colony forming unit-marrow, BFU-E → burst forming unit-erythroid, CFU-E → colony forming unit-erythroid.
(72,77,78,81,86,91,92)


Pluripotent Stem Cell
Self-renewal Divisions
CFU-M
Primitive BFU-E
Mature BFU-E
CFU-E
Erythroblast
Reticulocyte
Erythrocyte
Differentiative Divisions
Morphologically Recognizable Cells

(77).

After the CFU-M stage, the progenitors of large, late-appearing bursts (primitive burst forming unit-erythroid, pre BFU-E) arise, and they are succeeded by the progenitors of smaller, earlier-appearing bursts (mature burst forming unit-erythroid, mature BFU-E). These cell types are thus defined by their colonizing ability in *in vitro* erythroid cell cultures. In such *in vitro* semi-solid culture assays, pre BFU-E cells of the mouse first give rise to erythroid cells after 8-14 days of culture, while mature BFU-E cells of the mouse give rise to erythroid colonies in day 3 of culture. Furthermore, in the same assay above, human pre BFU-E cells can form more than 8 clusters of erythroblasts per cell, whereas human mature BFU-E cells can only form 3-8 clusters per cell. Bursts comprise 3-8+ small discrete clusters, and each cluster includes 8-50+ tightly associated erythroblast (60,61,132). The mature BFU-E is the immediate precursor of the colony forming unit-erythroid (CFU-E) cell which can form only one cluster (8-150 cells in chicken) of erythroblasts (81,82) *in vitro*.

Erythropoietin seems to regulate the generation of the new CFU-E from a mature BFU-E and the survival of the CFU-E to carry on differentiation toward the erythroblast stage (83). Although CFU-M and primitive BFU-E show apparent insensitivity to erythropoietin (84,85), there are

implications from *in vitro* studies that primitive BFU-E proliferation is controlled by a leukocyte-derived glycoprotein species (interleukin-3) (86). In addition, an erythroid-potentiating activity (EPA), a glycoprotein with a relative molecular weight of 28,000 has been purified (87,88). Unlike murine interleukin-3 (IL-3) which stimulates precursor cells from all hematopoietic lineages (89), purified EPA specifically stimulates human and murine BFU-E cells (90).

In the closing stages of erythropoiesis, when the size of the cell is reduced and the nucleus begins to condense (86), erythroblasts, reticulocytes and erythrocytes can be distinguished by morphological variation and hemoglobin synthesis. Hemoglobin synthesis can be detected by benzidine staining (91). The following characteristics are observed for each cell type: erythrocytes, strongly benzidine positive oval cells with a nuclear diameter less than half of the diameter of the cell; late reticulocytes, similarly stained, round cells with a large nucleus; early reticulocytes, weakly benzidine positive cells with a large nucleus; and erythroblasts, large benzidine negative cells (92).

The erythroblast is the final erythroid cell type competent to divide, and hemoglobin synthesis begins at this stage. Reticulocytes maintain mRNA and protein but not DNA synthesis capabilities, whereas no noticeable mRNA and protein synthesis occurs in the terminal erythrocyte stage.

Interestingly chicken red cells don't enucleate like mammalian ones.

The mature erythrocyte is a highly specialized cell whose function is to maintain an adequate tissue oxygen supply. Within a few months after its formation, it will be phagocytosed by the reticuloendothelial system, so there is a need for continuous production of new red blood cells.

3. Globin Protein and Genes:

Hemoglobin (M.W. 64,500), the oxygen carrier in the blood, exists predominantly in red blood cells. It was among the first proteins to have its amino acid sequence determined and its structure worked out by X-ray crystallography (93,94).

Hemoglobin contains four polypeptide chains and four heme prosthetic groups, in which the iron atoms are in the ferrous [Fe(II)] state. The component polypeptides, called globins, consist of two α chains (141 residues each) and two ß chains (146 residues each) (93,94).

X-ray crystallography analysis showed that the hemoglobin molecule is roughly spherical, with a diameter of about 5.5 nm. Each of the four polypeptide chains has a characteristic tertiary structure, in which the chain is tightly folded. There is one heme group bound to each polypeptide chain. With the sixth coordination bond of the iron atom, each heme group is capable of binding one molecule of oxygen (93).

The major task of hemoglobin is to carry oxygen from lungs to respiring tissues. Hemoglobin is also able to bind protons which are released by the ionization of carbonic acid, and this represents the most important buffer system for maintaining neutrality within the red cell.

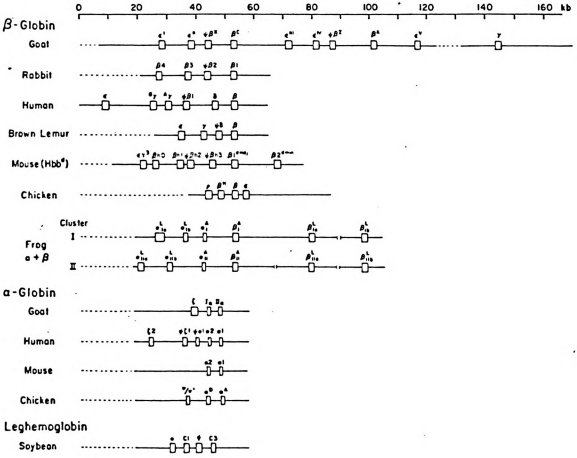
Initial advances in globin gene mapping and isolation were made possible by the development of procedures for synthesizing and cloning double-stranded DNA copies of poly

A+ mRNAs. The successful introduction of double-stranded mouse, rabbit and human ß- and α-globin cDNAs into bacterial plasmids provided homogeneous hybridization probes for gene mapping experiments. A very precise picture of the chromosomal organization of the α- and ß-type (of some mammals and birds) globin gene clusters, with respect to the number of structural loci and intergene distances, has been obtained through the use of the technique of restriction endonuclease mapping using the gel blotting procedure of Southern as well as through the actual isolation and characterization by recombinant DNA technology of large fragments of DNA containing the various globin genes. Sets of overlapping genomic DNA fragments spanning the entire human α and ß globin gene clusters have been obtained by gene cloning, first in bacteriophage lambda (95-97) and later as larger fragments in cosmid vectors (98,99). Detailed analysis of these recombinant DNA clones has led to the determination of the gene organization as shown in Figure 2.

It has been suggested the α and ß globin genes are derived from a single ancestral gene by a duplication event that occured about 450 million years ago (100,101). Since then α-type and ß-type globin gene clusters have evolved and the two clusters have subsequently been separated on two different chromosomes (figure 2). This is true in most vertebrates except the frog, the lowest vertebrate studied thus far, which still has the α and ß globin genes linked on the same

Figure 2: Chromosomal organization of the globin gene clusters in various species.

Figure from Collin (110)


0
20
40
60
80
100
120
140
160 kb
β-Globin
Goat
Rabbit
Human
Brown Lemur
Mouse (Hbb^d)
Chicken
Frog
α + β
Cluster
I
II
α-Globin
Goat
Human
Mouse
Chicken
Leghemoglobin
Soybean

chromosome. In some species of frogs, there is a single such cluster (102), whereas in other species, such as *Xenopus laevis*, there are two similar unlinked clusters that probably arose because of tetraploidization.

Globin pseudogenes have been found in the human, rabbit, mouse, goat and sheep but not in the chicken. For example, there exist $\psi\zeta 1$ and $\psi\alpha 1$ in the human α globin gene cluster and $\psi\beta 1$ in the human ß globin gene cluster as well (103-107). The majority of globin pseudogenes appear to have come from gene duplication within globin gene clusters followed by mutation and inactivation of one of the duplicated genes.

A general feature of the chromosomal organization of globin gene clusters in mammals and birds is the arrangement of the genes in the general order (5'→3') in which they are expressed during development: embryonic → fetal → adult. However, the frog larval ß-type genes (expressed in tadpoles) are located 3' to the adult ß gene, and in the chicken there is a second embryonic gene located 3' to the adult ß gene (108,109). These exceptions indicate that the precise arrangement of the globin genes in order of their developmental expression has not been strictly conserved throughout evolution and therefore is not an absolute requirement for the normal pattern of regulation of globin gene expression during development. Another noteworthy constant feature of the mammalian ß gene cluster is the

presence of one (or more) pseudogenes between the fetal or embryonic genes and the adult ß globin genes. The functional significance of this finding with regard to normal globin gene expression or switching is unclear.

The coding region of the globin genes of human (figure 3) and chicken (figure 4,5) is interrupted at two positions. In the ß-globin genes, the intervening sequences interrupt the sequence between codons 30 and 31 and between codons 104 and 105; in the α-globin gene family, the intervening sequences interrupt the coding sequence between codons 31 and 32 and between codons 99 and 100. The first intervening sequence (intron) is shorter than the second intron in both α-an ß-globin genes, but the second intron of human and chicken α-globin genes is considerably shorter than that of the ß-globin genes. In the case of the ζ-like globin genes, the structure and pattern of intron sizes are rather different from those of the adult α-type and other globin genes. The introns in the α and $\psi\alpha$ genes are rather small (<150 bp). Those of the ζ and $\psi\zeta$ genes are considerably larger ,and the first intron is much larger than second intron, being 8 to 10 times larger than the first intron of any other globin gene. The embryonic α-type π-globin gene of chicken has the same pattern as the ζ and $\psi\zeta$ genes with regard to a first intron (577 bp) being larger than the second intron (294 bp).

The state of methylation of the DNA of a gene usually influences its ability to be expressed (110). In general,

Figure 3: Arrangement of hemoglobin genes along two human chromosomes.

The human α-globin family of genes is located on chromosome 16, and the ß family is on chromosome 11. The α-globin family contains two embryonic (ζ1,ζ2) and two adult (α1,α2) α-like globin genes and one α-like pseudogene (ψα1). The ß family contains one embryonic (ε), two fetal ($^{G}\gamma$,$^{A}\gamma$) and two adult (δ,ß) ß-like globin genes and one ß-like pseudogene (ψß1). Rectangular boxes are the transcribed genes, and horizontal lines are the intervening untranscribed DNA. Exons, destined for translation into amino acid sequences, are represented by vertical black bands across the boxes, and intron sequences that are edited out from the messenger RNA are in white. Figure from Dickerson (93)

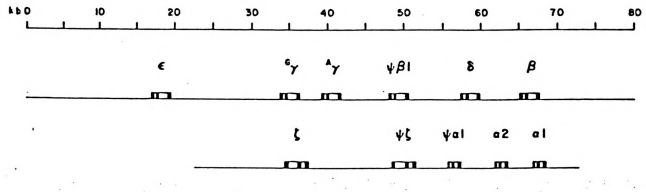

kbp
0
10
20
30
40
50
60
70
80
ε
$^{G}\gamma$
$^{A}\gamma$
ψβ1
δ
β
ζ
ψζ
ψα1
α2
α1

Figure 4: Organization of the chicken α-type globin genes.

The chicken α-globin gene family contains one embryonic (π) and two adult (α^D, α^A) genes. The exons and the introns are defined and represented as in figure 3.

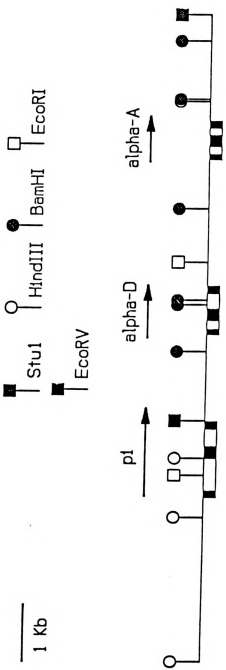
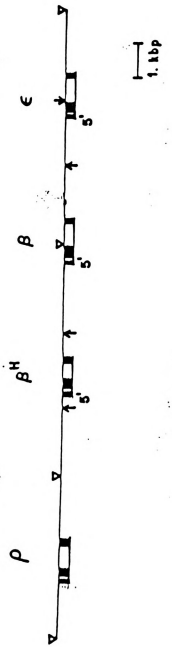

1 Kb
StuI
HindIII
BamHI
EcoRI
EcoRV
p1
alpha-D
alpha-A

Figure 5: Organization of chicken ß-type globin genes.

The chicken ß-globin gene family contains a single adult gene (ß), two embryonic genes (ρ, ϵ) and a gene expressed in definitive cells around the time of hatching ($ß^H$). The exons and the introns are defined and represented as in figure 3. Symbols for the sites of restriction endonucleases: ↓ EcoRI, ↑ BamHI, HindIII.


ρ
βH
β
ε
5'
5'
5'
1.1kbp

genes which are efficiently transcribed often are hypomethylated, whereas inactive genes are usually, but not necessarily, hypermethylated. Methylation of specific sites in DNA can be analyzed by using different restriction endonucleases that have specificities influenced by DNA methylation. For example, the enzyme HpaII will cleave DNA at the sequence CCGG but not at C^{m}CGG; on the other hand, the enzyme MspI will cleave the sequence CCGG whether the second C is methylated or not (except in the case of the sequence GGC^{m}CGG). Globin gene clusters are among the most intensively studied in terms of their chromatin structures and their relation to gene expression. For example, a HpaII site at the 5' end of the α^{D} globin gene and another just 3' to the α^{D} gene are unmethylated in embryonic and in adult erythrocytes where the α^{D} gene is expressed, but are completely or partially methylated in brain and sperm where the α^{D} gene is not expressed (111).

The DNase I sensitivity of a gene in isolated nuclei can be related to the transcriptional activity of the gene. Chicken globin genes have been extensively investigated using DNase I sensitivity assays as well as sensitivity to other nucleases. Both the adult and embryonic ß-type globin gene regions are very sensitive to DNase I digestion in primitive erythroid cells which only express embryonic ß-type (ρ and ϵ) globin genes. The adult ß globin gene remains highly sensitive in definitive erythroid cells where this is the only ß-type

globin gene to be expressed, while the DNase I sensitivity of the now inactive embryonic ß-type globin genes is reduced to a moderate level. All the globin genes are relatively insensitive to DNase I in nuclei from non-erythroid tissues.

DNase I hypersensitive sites are often found in the 5' flanking region of actively transcribed genes. Such sites have been shown in the 5' flanking region of α^A-and α^D-globin genes in definitive erythroid cells and of the π-globin gene in primitive erythroid cells (112). Similar sites also have been found in embryonic and adult ß-type globin genes in appropriate cell types (113-117). Thus, there is a close relationship between DNase I hypersensitive sites at the 5' end of the chicken globin genes and the activity of these genes.

Chicken adult α-globin genes (α^A and α^D) possess some transcriptional activity in primitive erythroid cells ,unlike the adult ß-globin gene which expresses strictly in definitive erythroid cells (69,118). This indicates that the transcription regulatory mechanisms may differ for the chicken adult α globin genes and the adult ß globin gene.

Eryf1, an erythroid-specific DNA-binding factor, recognizes a regulatory sequence which is capable of confering tissue specificity within the 3' enhancer of the chicken ß-globin gene (33,34,119-122). The regulatory sequence to which Eryf1 binds also can be found in the 3' flanking region of the chicken α^A-globin gene, upstream of the TATA sequence of the

α^D gene, and in the π gene promoter (35,120). Moreover, the binding sites within the 3' flanking region of the chicken α^A-globin gene have been shown to be capable of exerting tissue-specific enhancer activity (35).

The mechanism of transcriptional regulation of the chicken ß-type globin gene family has been well explored. There is evidence to suggest that the developmental regulation of the chicken ϵ- and ß-globin genes involves at least two DNA elements. One of these is responsible for tissue specificity (the enhancer in the 3' flanking region of ß globin gene) and another is responsible for definitive erythroid cell stage specificity (the ß-globin developmental stage selector element within the ß-globin gene promoter) (33,120,123,124). Recently, a transcriptional repressor (PAL) and a transcriptional activator (CON) which bind to the promoter region of the chicken ß-globin gene have been identified by *an* *in vitro* transcription assay. The balance between PAL and CON during erythropoiesis has proven to be partially responsible for the developmental regulation of the chicken ß globin gene (125).

Transcriptional silencers of globin genes have been identified in the 5' flanking region of human α- and human ϵ-globin genes (62,63). Short areas within the silencer element of the human ϵ-globin gene have some homology with sequences from other known negative regulatory elements (64). In addition, the ϵ silencer has a stronger effect in Hela cells

(a non-erythroid cell line) than in K-562 cells (erythroid cells) which suggests that the ε silencer may contribute to tissue specificity (63). DNA footprint and deletion mutant studies also suggest that the 3' flanking region of the chicken ß globin gene may contain negative regulatory elements also acting in a tissue specific manner (34).

References:

(1) Darnell, J. E., Jr. *Nature* **297**:365, 1982.

(2) Dynan, W. S., Tjian, R., *Nature* **316**:774, 1985.

(3) McKnight, S., Tjian, R., *Cell* **46**:795, 1986.

(4) Serfling, E. et al., *EMBO J* **4**:3851, 1985.

(5) Corden, J., Wasylyk, B., Buchwalder, A., Chambon, P., *Science* **209**:1406, 1980.

(6) Baty, D., Barrera-Saldana, A. et al., *NAR* 12:915, 1984.

(7) McKnight, S. L., Kingsbury, R. C., Spence, A., Smith, M., *Cell* **37**:253, 1984.

(8) Myers, R. M., Tilly, K., Maniatis, T., *Science* **232**:613, 1986.

(9) Grosschedl, R., Birnstiel, M., *PNAS* 77:1432, 1980.

(10) Wasylyk, B., Derbyshire, R., Guy, A. et al., *Cell* **32**:695, 1983.

(11) Benoist, C., O'Hare, K., Breathnach, R., Chambon, P., *NAR* 8:127, 1980.

(12) Dierks, P. et al. *Cell* **32**:695, 1983.

(13) McKnight, S. L., *Cell* **37**:403, 1984.

(14) Miyamoto, N., Moncollin, V., Wintzerith, M. et al., *NAR* **12**:8779, 1984.

(15) Kadonaga, J. T., Tjian, R., *PNAS* **83**:5889, 1986.

(16) Charnay, P., Mellon, P., Maniatis, T., *Mol. Cell. Biol.* 5:3261, 1985.

(17) Everett, R., *NAR* **11:** 6647, 1983.

(18) Galas, D. J., Schmitz, A., *NAR* **5:**3156, 1978.

(19) Briggs, M. R., Kadonaga, J. T., Bell, S. P., Tjian, R., *Science* **234:**47, 1986.

(20) Jones, K. A. et al., *Cell* **48:**79, 1986.

(21) Reinberg, D., Horikoshi, M., Reeder, R. G., *J. Biol. Chem* **262:**3322, 1987.

(22) Chodosh, L. A. et al., *Cell* **53:**25, 1988.

(23) Conrey, A., Tjian, R., *Cell* **55:**887, 1988.

(24) Jackson, S. P., Tjian, R., *Cell* **55:**125, 1988.

(25) Kadonaga, J. T. et al., *Cell* **51:**1079, 1987.

(26) Kadonaga, J. T. et al., *Science* **242:**1566, 1988.

(27) Weber, F., de Billiers, J., Schaffner, W., *Cell* **36:**983, 1984.

(28) Boshart, M. et al., *Cell* **41:**521, 1985.

(29) Lang, J. C., Spandidos, D. A., *Anticancer Res.* **6:**437, 1986.

(30) Banerji, J., Olson, L., Schaffner, W., *Cell* **33:**729, 1983.

(31) Maniatis, T., Goodbourn, s., Fischer, J. A., *Science* **236:**1237, 1987.

(32) Hanahan, D., *Nature* **315:**115, 1985

(33) Choi, O.-R., Engel, J. D., *Cell* **55:**17, 1988.

(34) Emerson, B. M. et al., *PNAS* **84:**4786, 1987.

(35) Knezetic, J. A., Felsenfeld. G., *Mol. Cell. biol.* **9:**893, 1989.

(36) Chandler, V. L., Maler, B. A., Yamamoto, K. R., *Cell* 33:489, 1983.

(37) Jantzen, H. M. et al., *Cell* 49:29, 1987.

(38) Karin, M. et al., *Nature* 308: 513, 1984.

(39) Serfling, E. et al., *EMBO J.* 4:3851, 1985.

(40) Goodbourn, S., Zinn, K., Maniatis, T., *Cell* 41:509, 1985.

(41) Bienz, M., Pelham, H. R. B., *Cell* 45:753, 1986.

(42) Ryals, J. et al., *Cell* 41:497, 1985.

(43) Fujita, T. et al., *Cell* 41:489, 1985.

(44) Goodbourn, S., Zinn, K., Maniatis, T., *Cell* 41:509, 1985.

(45) Rougvie, A. E., Lis, J. T., *Cell* 54:795, 1988.

(46) Sorger, P. K., Pelham, H. R. B., *Cell* 54:855, 1988.

(47) Wiederrecht, G., Shuey, D. J., Kibbe, W. A., Parker, C. S., *Cell* 48:507, 1987.

(48) Norman, C. et al., *Cell* 55:989, 1988.

(49) Fletcher, C. et al., *Cell* 51:773, 1987.

(50) Harrey, R. P. et al., *Nucleic Acids Res.* 10:7851, 1982.

(51) Barberis, A. et al., *Cell* 50:347, 1987.

(52) Carbon, P. et al., *Cell* 51:71, 1987.

(53) Ingraham, H. A. et al., *Cell* 55:519, 1988.

(54) Robertson, M., *Nature* 336:522, 1988.

(55) Schaffner, W., *Trends Genet.* 5:37, 1989.

(56) Sturm, R., Das, G., Herr, W., *Genes Dev.* 2:1582, 1988.

(57) Ko, H. S. et al. *Cell* 55:135, 1988.

(58) Landschulz, W. H. et al., *Genes Dev.* **2**:786, 1988.

(59) Müller, M. M. et al., *Nature* **336**:544, 1988.

(60) Treisman, R. H., *EMBO J.* **6**:2711, 1987.

(61) Brand, A. H. et al., *Cell* **41**:41, 1985.

(62) Atweh, G. F. et al., *Mol. Cell. Biol.* **8**:5047, 1988.

(63) Cao, S. X. et al., *PNAS* **86**:5306, 1989.

(64) Baniahmad, A. et al., *EMBO J* **6**:2297, 1987.

(65) Remmers, E. F., Yang, J.-G., Marcu, K. B., *EMBO J* **5**:899, 1986.

(66) Goodbourn, S., Burstein, H., Maniatis, T., *Cell* **45**:601, 1986.

(67) Bienz-Tadmor, B. et al., *EMBO J* **4**:3209, 1986.

(68) Sodroski, J. C. et al., *Science* **227**:171, 1985.

(69) Bruns, G. A., Ingram, V. M., *Trans. Roy. Soc. Lond. Ser. B. Biol. Sci.* **266**:225, 1973.

(70) Mahoney, K. A., Hyer, B. J., Chan, L. N. L., *Devlop. biol.* **56**:412, 1977.

(71) Wilt, F. J., *Advan. Morphogen.* **6**:89, 1967.

(72) Cline, M. J., Golde, D. W., *Nature* **277**:177, 1979.

(73) Samarut, J., Jurdic, J. P., Nigon, V. J., *J. Embryol. Exp. Morph.* **50**:1, 1979.

(74) Samarut, J., Blanchet, J. P., Nigon, V., *Develop. Biol.* **72**:155, 1979.

(75) Samarut, J., Bouabdelli, M., *J. Cell. Physio.* **105**:553, 1980.

(76) Samarut, J., Gazzolo, L., *Cell* **28**:921, 1982.

(77) Samarut, J., Nigon, V., *J. Embryol. Exp. Morphol.* **33**:259, 1975.

(78) Gregory, C. J., *J. Cell. Physiol.* **89**:289, 1976.

(79) Gregory, C. J., Eares, A. C., *Blood* **49**:855, 1977.

(80) Stephenson, J. R., Axelrad, A. A., McLeod, D. L.,

(81) Coll, J., Ingram, V. M., *J. Cell Biol.* **76**:184, 1978.

(82) Samarut, J., Nigon, V., *Exp. Cell Res.* **100**:245, 1976.

(83) Eliason, J. F., VanZant, G., Goldwasser, E., *Blood* **53**:935, 1979.

(84) Schooley, J. C., *J. Cell. Physiol.* **68**:249, 1966.

(85) Till, J. E., Siminovitch, L., McCulloch, E. A., *Blood* **29**:102, 1967.

(86) Wagemaker, G., In *in vitro Aspects of Erythropoiesis*, (Murphy, M. J. Jr. ed.), Springer-Verlag, New York, p44, 1978.

(87) Golde, D. W., Bersch, N., Quan, S. G., Lusis, A. J., *PNAS* **77**:593, 1980.

(88) Ihle, J. N. et al., J. Immun. **131**:282, 1983.

(89) Elder, J. H., Alexander, S., *PNAS* **79**:4540, 1982.

(90) Gasson, J. C. et al., *Nature* **315**:768, 1985.

(91) Orkin, S. H., Harosi, F. I., Leder, P., *PNAS* **72**:98, 1975.

(92) Berg, H. et al., *Cell* **28**:907, 1982.

(93) Dickerson, R., Geis, I., *Hemoglobin*, Benjamin/Cumming Publishing Company, Inc., California, pp19-61, 1983.

(94) Lehninger, A., *Principles of Biochemistry*, Worth

Publishers, Inc., New York, pp169-206, 1982.

(95) Lawn, R. M. et al., *Cell* **15**:1157, 1978.

(96) Fritsch, E. F. et al., *Cell* **19**:959, 1980.

(97) Lauer, J. et al., *Cell* **20**:119, 1980.

(98) Grosveld, F. G. et al., *Gene* **13**:227, 1981.

(99) Gollins, F. S. et al., *PNAS* **81**:4894, 1984.

(100) Czelusniak, J. et al., *Nature* **298**:297, 1982.

(101) Hunt, T. L., Hurst-Calderone, S., Dayhoff, M. O., In Dayhoff, M. O. *ed.*: *Atlas of protein Sequence and Structure*, Washington D. C., National Biomedical Research Foundation, pp229-251,1978.

(102) Jeffreys, A. J., *Cell* **21**:555, 1980.

(103) Fritsch, E. F., Lawn, R. M., Maniatis, T., *Cell* **15**:1157, 1978.

(104) Lauer, J., Shen, C. K., Maniatis, T., *Cell* **20**:119, 1980.

(105) Little, P. F. R., *Cell* **28**:683, 1982.

(106) Proudfoot, N. J., Maniatis, T., *Cell* **21**:537, 1980.

(107) Shen, S.-H., Smithies, O., *NAR* **10**:7809, 1982.

(108) Dolan, M. et al., *Cell* **24**:669, 1981.

(109) Villeponteau, B., Martinson, H., *Nucleic Acids Res.* **9**:3731, 1981.

(110) Collins, F. S., Weissman, S. M., *Prog. Nucleic Acid Res. Mol. Biol.* **31**:315, 1984.

(111) Haigh, L. S., *PNAS* **79**:5332, 1982.

(112) Weintraub, H., Larsen, A., Groudine, M., *Cell* **24**:333, 1981.

(113) Dolan, M., Dodgson, J. B., Engel, J. D., *J. Biol. Chem.* **258**:3983, 1983.

(114) Groudine, M., Weintraub, H., *Cell* **24**:393, 1981.

(115) Larsen, A., Weintraub, H., *Cell* **29**:609, 1982.

(116) McGhee, J. D. et al., *Cell* **27**:45, 1981.

(117) Stalder, J. et al., *Cell* **20**:451, 1980.

(118) Brown, J. L., Ingram, V. M., *J. Biol. Chem.* **249**:3960, 1974.

(119) Emerson, B. M., Lewis, C. D., Felsenfeld, G., *Cell* **41**:21, 1985.

(120) Evans, T., Reitman, M., Felsenfeld, G., *PNAS* **85**:5976, 1988.

(121) Lewis, C. D. et al., *Genes Dev.* **2**:863, 1988.

(122) Plumb, M. A. et al., *NAR* **14**:7675, 1986.

(123) Choi, O.-R., Engel, J. D., *Nature* **323**:731, 1986.

(124) Hesse, J. E. et al., *PNAS* **83**:4312, 1986.

(125) Emerson, B. M., Nickol, J. M., Fong, T. C., *Cell* **57**:1189, 1989.

CHAPTER 3

MATERIALS AND METHODS

Plasmid Construction:

Subclones of the α-globin gene, pBRα7-1.7 and pHRα5-4.3 have been described (1). Plasmid pAT48D-HV5.6 contains both adult genes on a 5.6 Kb fragment from an EcoRV site upstream of the α^D-globin gene to the Hind III site just downstream of the α^A gene. Subclone pdBRα7-1.7 is identical to pBRα7-1.7 except that its vector (pATdT) contains a deletion from nucleotide 401 (relative to EcoRI) to 1283 in order to make further construction easier. Various fragments from flanking regions of the α^A-globin gene have been cloned into BamHI or BglII sites in pBRα7-1.7; detailed descriptions will be provided in Results and Discussion.

Bal 31 Digestion:

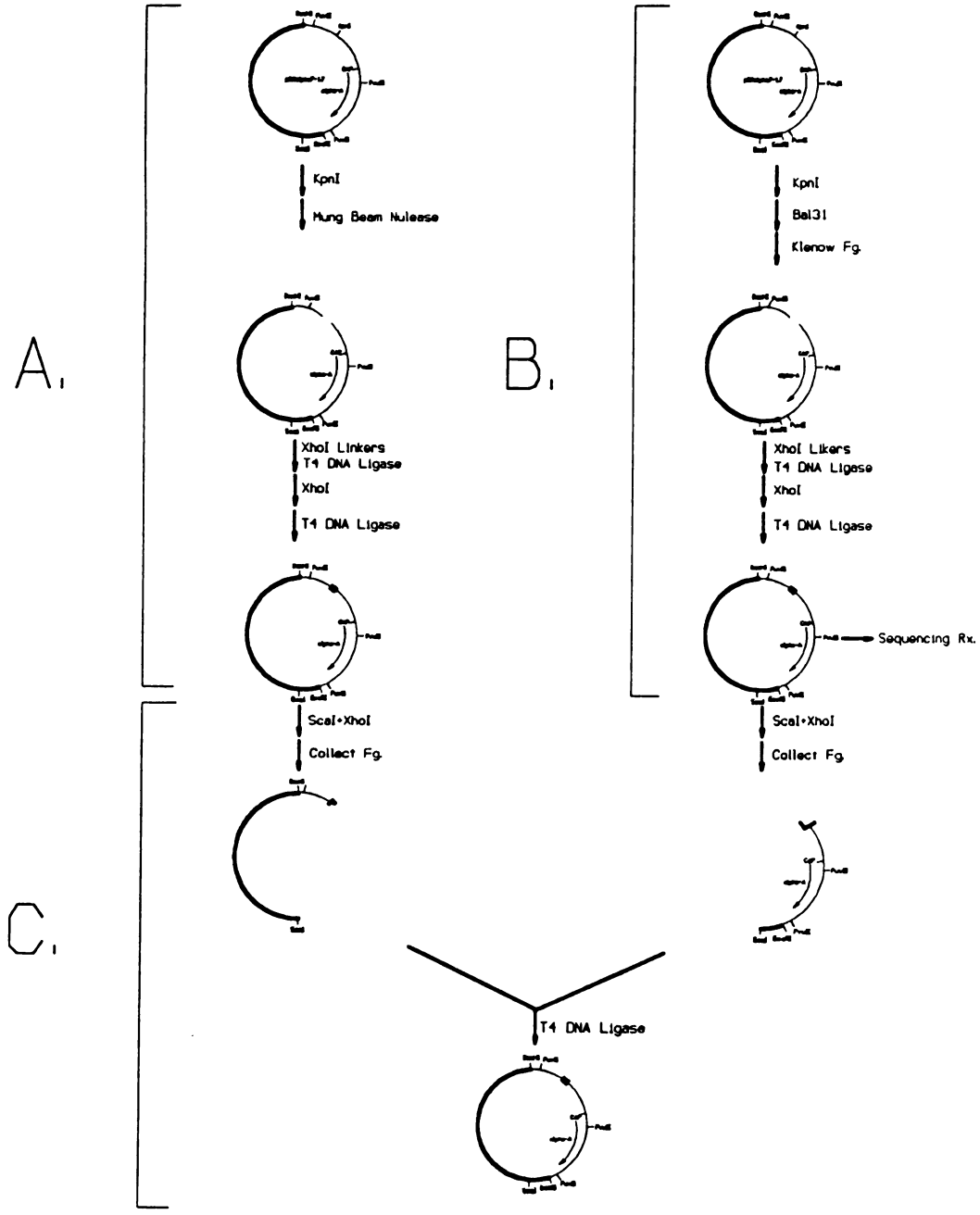
Bal31 exonuclease was used as described by Maniatis et al. (2). 5 μg of pBRα7-1.7 was first linearized using KpnI and then subjected to 2-5 units of Bal31 digestion at 30^0 for 2-10 min. The digestion with Bal31 was stopped by adding 1/10 reaction volume of 0.25 M EGTA followed by phenol/chloroform extraction. These linearized plasmids were then blunt-ended

by 0.5 unit of DNA polymerase I large fragment (Klenow) with four deoxynucleotides (dATP, dTTP, dGTP, dCTP; 0.25mM each) in the reaction mix. 0.3 μg Kinased XhoI linkers were ligated to these blunt-ended DNA fragments using a high concentration (0.1 u/μl, final conc.) of T4 DNA ligase at 16^0 overnight. The excess linkers were removed afterward by XhoI digestion and EtOH precipitation. The final step was to recircularize these treated fragments using a low concentration (0.01-0.001 u/μl, final conc.) of T4 DNA ligase at 16^0 overnight. The end product was then transformed into the competent cells of *E. coli* DH5 prepared as described (2). Individual colonies were picked and used to prepare plasmid DNA for restriction enzyme analysis as described in (2). Those bacterial clones with a XhoI linker and a deletion of interest were stored as glycerol stocks at -70^0.

Deletion Mutant Construction:

A scheme for making deletion mutants is shown in figure 1. A series of deletions was made by Bal 31 digestion as described in the previous section and shown in figure 1B. Mutant K→X is constructed as shown in figure 1A. The construction procedure is almost identical to that used to make the deletions (figure 1B) except that mung bean nuclease was substituted for Bal31 in order to covert the KpnI site to

Figure 1: Scheme for Deletion Mutant Construction.
The thick line represents pBR322 sequence and chicken globin gene region DNA is indicated by a thin line. Details of the construction are described in the text (Materials and Methods).


A.
KpnI
Mung Bean Nulease
XhoI Linkers
T4 DNA Ligase
XhoI
T4 DNA Ligase
ScaI+XhoI
Collect Fg.
B.
KpnI
Bal31
Klenow Fg.
XhoI Likers
T4 DNA Ligase
XhoI
T4 DNA Ligase
Sequencing Rx.
ScaI+XhoI
Collect Fg.
C.
T4 DNA Ligase

a blunt end. The mung bean nuclease digestion is performed at 37^0 in buffer containing 30 mM sodium acetate (pH 5.0), 100 mM NaCl, 2 mM $ZnCl_2$, 10 % glycerol. The end products from A and B were both subjected to ScaI and XhoI digestion and followed by fragment collection and ligation as shown in figure 1C.

Hybrid Genes Construction:

Hybrid genes which contain the α^A-(or α^D-) globin gene promoter linked to the CAP site and the body of the α^D-(or α^A-) globin gene were constructed using the linker scanner mutagenesis (3). Most of the details are described in (4) except for the 3' deletion mutants of the α^A-globin gene. The 3' deletion mutants were constructed identically to the 5'deletion mutants except that plasmid pdBRα7-1.7 was substituted for pBRα7-1.7, and the first enzyme digestion was changed from KpnI to NaeI. All the hybrid genes we constructed were made by joining deletion clones with an 8 bp XhoI linker at bp -15 (-14 for α^D promoter) to -7 between the TATA sequence and the CAP site (+1) of the relevant promoter. The resulting hybrid genes are shown in Results and Discussion.

Sequencing:

The deletion clones were analyzed by sequencing in order

to precisely locate the end point of deletion. The chemical degradation method of Maxam and Gilbert (5) as modified by Smith and Calvo (6), was utilized to conduct the above analysis. To end label the DNA fragments, two methods were employed: in one T4 polynucleotide kinase is used to transfer the γ-phosphate of [γ-^{32}P]ATP to the 5'-OH termini of DNA fragments (2), the other involves the use of the Klenow fragment of DNA polymerase to fill in the 3' recessed end of these fragments with [α-^{32}P]dCTP (2). The unwanted labeled ends were removed by a second restriction enzyme digestion. Four of the standard Maxam-Gilbert reactions (A, A+G, T+C, C) were performed and the reaction products were run on a 20% denaturing polyacrylamide gel in order to determine the end of given deletion.

Cell Culture:

Chemically transformed quail fibroblasts (7) were grown in Dulbecco's modified Eagle medium (Gibco Laboratories) supplemented with 4% fetal calf serum, 1% chicken serum , 1% DMSO (Fisher scientific), and the antibiotics, penicillin and streptomycin (50 u/ml each).

DNA Transfection:

The calcium phosphate procedure (8) was used to transiently transfect QT6 cells. QT6 cells were plated to a density of 1-2 X 10^6 cells per 100mm tissue culture plate 16-20 hrs before transfection. 2-4 hrs before transfection, old media was replace by 7 ml of fresh media. 20 μg of test plasmid and 20 μg of internal control plasmid (pRSVCAT) were coprecipitated beforehand. The DNA/calcium phosphate precipitate was made by mixing 418 μl ddH_2O, 20 μl DNA solution, 62 μl 2M $CaCl_2$ and 500 μl 2X Hepes buffered saline (280 mM NaCl, 50 mM Hepes, 1.5 mM Na_2HPO_4, pH 7.10 ± 0.05) followed by incubation at the room temperature for 30 min. The above mixture was added to the QT6 cells and then removed and replaced by fresh QT6 media after 4 hrs of incubation at 37^0. RNA was isolated from the cells 48 hrs after the DNA-calcium phosphate mixture was removed.

Poly A+ RNA Isolation:

Two 100mm plates of QT6 cells were lysed with 3mls of lysis buffer (0.5 M NaCl, 10 mM Tris HCl pH 7.5, 1 mM EDTA, 1% SDS, 200 μg/ml proteinase K) and the high molecular weight DNA was sheared by passing through a 22 gauge needle. This homogenate was incubated at 37^0C for 1 hr. Oligo-dT cellulose was swelled in RNase-free ddH2O and mixed with lysis buffer to remove residual RNase activity. The cell homogenate was then added to this treated oligo-dT cellulose and rocked at

room temperature for one hour. The oligo-dT cellulose was subsequently spun from the solution and washed twice with high salt buffer (0.5 M NaCl, 10 mM Tris HCl pH 7.5, 1 mM EDTA), and twice with low salt buffer (0.1 M NaCl, 10 mM Tris HCl pH 7.5, 1 mM EDTA). The poly A+ RNA was removed from the cellulose by adding 2 mls RNase-free ddH_2O followed by precipitation with the addition of 200 μl 3M NaOAc(7.0) and 2.5 volumes of EtOH.

RNase Protection Assays:

Vector pT7-1 (U. S. Biochemical Corp.) was used to clone the 342 bp Sau3A fragment from the α^A-globin gene and the 329 bp MspI fragment from the α^D-globin gene. Furthermore, another vector pT7/T3-mp18 (Bethesda Research Laboratories) was used to clone the 148 bp HindIII/ PvuII fragment of the CAT gene. Both vectors are able to template *in vitro* transcription from the T7 RNA polymerase promoter which is positioned directly upstream from each cloned fragment. In all cases, the plasmid DNAs were first linearized by HindIII digestion, so they would template RNA probes of specific defined length. T7 RNA polymerase was used to generate labeled RNA transcripts as described in the directions from the supplier (U. S. Biochemical Corp.). These riboprobes (5 X 10^5 cpm) were then hybridized with poly A+ RNA samples in hybridization buffer

overnight at 55^{0}C, and later treated with RNase A and T1 as described by Melton et al. (9). The final products were afterward run on 6% denaturing polyacrylamide gels.

Crude Nuclear Extract Preparation:

Six 100 mm plates of 80% confluent QT6 cells were first washed twice with 1X PBS (0.1 M NaCl, 2.68 mM KCl, 2.68 mM KH_2PO_4, 8.7 mM Na_2HPO_4) and then incubated with 5 ml of solution A (10 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT) on ice for 5 mins. All the following steps were carried out at 0-4^0. Swelled QT6 cells were scraped into a 30 ml Corex tube and pelleted at 2000 rpm for 5 mins in the SS34 rotor. The pellet was resuspended in 5 volumes (of pellet) of solution A and incubated for 10 mins. The cell suspension was pelleted in an SS34 rotor as above. These QT6 cells were then resuspended in 2 volumes of solution A and lysed by 10 strokes of an all-glass dounce homogenizer (B type pestle). The resultant crude nuclei were pelleted in an Eppendorf microfuge at 4000 rpm for 5 mins. The pellet was resuspended in buffer C (20 mM Hepes, pH 7.9, 25% v/v glycerol, 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT) with 10 more strokes of a homogenizer, and incubated on ice for 30 mins with constant gentle mixing. This nuclear suspension was then centrifuged in the microfuge as above, the resulting pellet was discarded, and the supernatant was dialyzed against 50

volumes of buffer D (20mM Hepes, pH 7.9, 20% v/v glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT) overnight. The dialyzed supernatant was centrifuged as above,the precipitate was discarded, and the supernatant (crude nuclear extract) was aliquoted and stored at -80^0.

DNase Protection Analysis:

10 ng of labeled DNA fragments were mixed with 0-40 μg of nuclear extract in binding buffer (0.1 μg/μl poly dI-dC, 5 mM $MgCl_2$, 100 mM KCl, 10 mM Hepes, pH 7.9, 0.1 mM EDTA, 10% v/v glycerol, 0.25 mM DTT) and subsequently incubated for 30 mins at 37^0. After binding, complexes were transferred to 21^0 for 1 min, and 0.1 unit of DNase was then added for 1 min at 21^0. The reactions were terminated by the addition of 25 mM EDTA, 300 mM NH_4OAc and 5 μg of yeast RNA. Resulting DNA fragments were purified by phenol/chloroform extraction and EtOH precipitation. The final products were suspended in 90% formamide loading buffer and run on 6% denaturing polyacrylamide gels.

Densitometry Analysis:

The signal intensities resulting from RNase protection analysis were quantitated with a LKR 2222-010 UltraScan XL Laser Densitometer (Bromma, Sweden). Peak areas were recorded

and later compared.

Miscellaneous:

Restriction enzyme digestion, plasmid DNA preparations, DNA labeling, and agarose and polyacrylamide gel electrophoresis were essentially as described by Maniatis et al.(2).

References:

(1) Dodgson, J. D., Engel, J. D., *J. Biol. Chem.* **258**:4623, 1983.

(2) Maniatis, T. et al., *Molecular Cloning: A Laboratory Mannual* (Cold Spring Harbor Laboratory, NY) 1982.

(3) Keshet, E., Cedar, H., *NAR* **11**:3571, 1983.

(4) Bates, P., *Ph. D. Dissertation* (Michigan State University) 1984.

(5) Maxam, A., Gilbert, W., *Mithods Enzymol.* **65**: 499, 1980.

(6) Smith, D., Calvo, J., *NAR* **8**:2255, 1980.

(7) Moscovici, C. et al., *Cell* **11**:95, 1977.

(8) Wigler, M. et al., *Cell* **14**:725, 1978.

(9) Melton, D. A. et al., *NAR* **12**:7035, 1984.

CHAPTER 4

RESULTS AND DISCUSSION

A. α-Globin Gene Expression from Clones Containing the α^A- and/or α^D-Globin Genes.

A variety of chicken α-globin gene clones were initially transiently transfected into a quail fibroblast cell line (QT6), a chicken erythroid-precursor cell line (HD3) and into chicken erythroid cells isolated from 9-day embryos. No measurable expression of the exogenous α-globin genes was observed in transfected HD3 cells and embryonic cells (results not shown). The expression of the endogenous globin mRNA showed that this was not a defect in the assay. The inability to measure expression of exogenous α-globin genes may be due to low transfection efficiency, inability to transcribe transfected DNA or some other unknown reason. However, an exogenous α^A-globin transcript can be detected in QT6 cells transfected with the chicken α^A-globin gene, and this RNA is initiated at the normal α^A-gene CAP site (figure 2). Since the transfected quail fibroblast cell line is the only system we have found to date in which detectable transcription of the transfected α^A-globin gene could be demonstrated, this system was used in all of the studies reported herein.

Plasmids containing the chicken α^A-(pBRα7-1.7), and α^D-

(pHRα5-4.3) globin genes separately or together (pAT48D-HV5.6), as shown in figure 1, were transiently transfected into QT6 quail fibroblasts to study their constitutive expression. Poly A+ RNA was isolated 48 hrs after transfection and assayed by RNase protection analysis. In each transfection experiment, 20 μg of pRSVCAT (1) DNA, a plasmid containing a chloramphenicol acetyl transferase gene using the long terminal repeat (LTR) of Rous Sarcoma Virus as a promoter, was cotransfected with 20 μg of the test clone. Half of the RNA isolated from the transfectants was hybridized with a riboprobe complementary to CAT RNA followed by RNase protection analysis to serve as an internal control for the relative tranfection efficiency in each experiment. The detailed procedure used is shown in figure 2.

A 342 bp Sau3A fragment from the α^A-globin gene was inserted downstream of the T7 RNA polymerase promoter in the pT7-1 vector such that RNA transcribed *in vitro* would be complementary to α^A-globin mRNA. This was used to prepare a labeled α^A-specific antisense riboprobe which was hybridized to RNA prepared from transfected cells. RNAs present in these cells which arose from transcription beginning at the normal α^A-globin mRNA initiation site (CAP site) would generate a 131 bp protected fragment after hybridization to the probe and RNase A and T1 digestion. The riboprobe for the α^D-globin gene and the CAT gene were made similarly and should result in

Figure 1: Restriction Enzyme Maps of Three Subclones of Chicken α-globin genes.

The two adult (α^A, α^D) and the embryonic (π) genes are shown in A. The arrows indicate direction of transcription. Exons are shown as black boxes and introns are closed boxes. B, C and D show three subclones containing various regions of the chicken α-globin gene cluster.

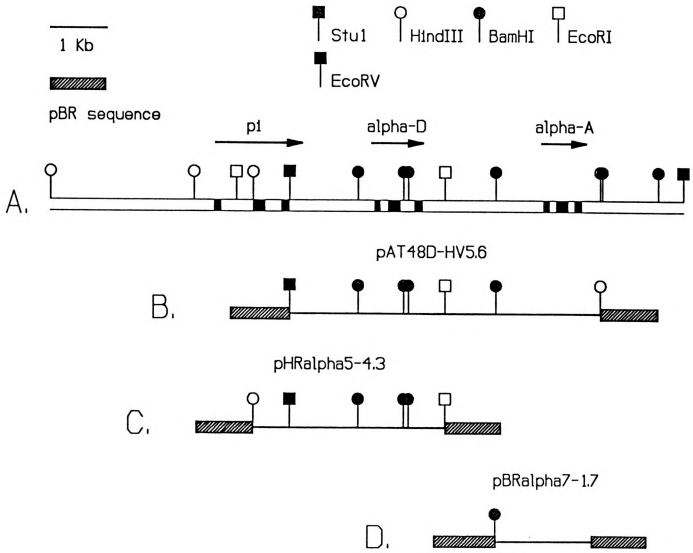

1 Kb
pBR sequence
Stu1
HindIII
BamHI
EcoRI
EcoRV
p1
alpha-D
alpha-A
A.
pAT48D-HV5.6
B.
pHRalpha5-4.3
C.
pBRalpha7-1.7
D.

Fig 2: Scheme for Testing Expression Level of Clones in QT6 Quail fibroblast.

Two plates (100 mm) of QT6 cells were used in transfection experiment for each test clone. Poly A+ RNA was isolated 48 hrs after transfection. Half of the RNA isolated from two plates was used to examine the expression of the α-globin genes, and the other half was used to examine the expression of the CAT gene.

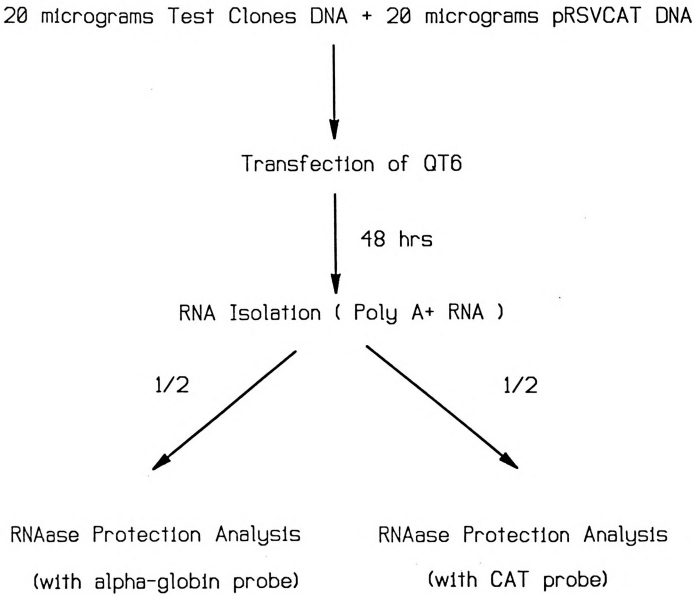

20 micrograms Test Clones DNA + 20 micrograms pRSVCAT DNA
Transfection of QT6
48 hrs
RNA Isolation (Poly A+ RNA)
1/2
1/2
RNAase Protection Analysis
(with alpha-globin probe)
RNAase Protection Analysis
(with CAT probe)

Figure 3: Scheme for Making the Riboprobe for α^A and α^D-Globin Genes and CAT Gene Transcripts.

A. A Sau3A fragment (342 bp) containing the α^A-globin gene first exon and its flanking region was cloned into the pT7-1 vector.

B. A MspI fragment (340 bp) containing the α^D-globin gene first exon and its flanking region was cloned into the pT7-1 vector.

C. A PvuII/HindIII fragment (147 bp) containing the coding region of the CAT gene was cloned into the pT7/T3-mp18 vector. Both pT7-1 and pT7/T3-mp18 vectors are able to template *in vitro* transcription from the T7 RNA polymerase promoter which is positioned directly upstream from each cloned fragment. The procedures for making anti-sense riboprobe and the RNase protection assay are described in Materials and Methods. After RNase digestion , three protected fragments of 131 bp, 138 bp and 147 bp are expected, respectively, from the RNA transcripts of α^A, α^D and CAT genes.

CAP: Starting site for mRNA transcription.

I: Exon 1.

II: Exon 2.

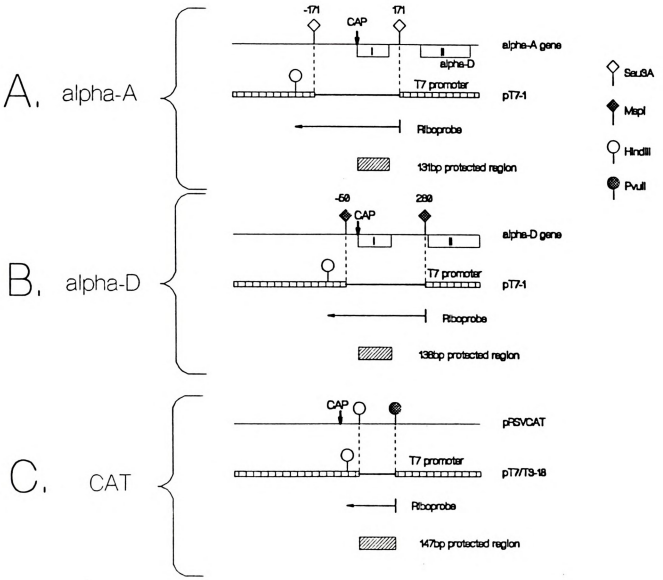

A. alpha-A
-171
CAP
171
alphe-A gene
alpha-D
T7 promoter
pT7-1
Riboprobe
131bp protected region
B. alpha-D
-58
CAP
289
alphe-D gene
T7 promoter
pT7-1
Riboprobe
138bp protected region
C. CAT
CAP
pRSVCAT
T7 promoter
pT7/T3-18
Riboprobe
147bp protected region
Sau3A
MspI
HindIII
PvuII

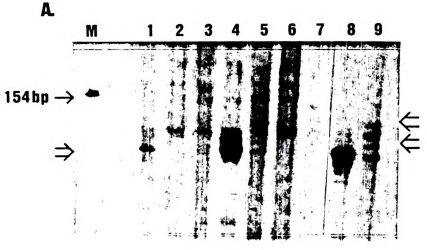
Figure 4: Transfection of Clones Containing the α-Globin Genes and Hybrid Genes into QT6 Fibroblast Cells:

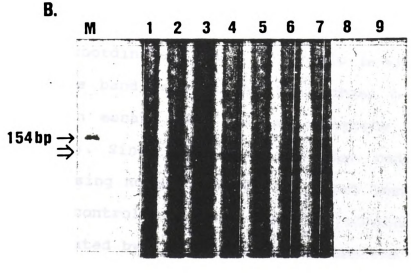
An RNase protection assay was performed on poly A+ RNA isolated from QT6 cells transfected with (Lane 1) pBRα7-1.7; (lane 2) pHRα5-4.3; (lane 3) pAT48D-HV5.6; (lane 4) A/A; (lane 5) A/D; (lane 6) D/D; (lane 7) D/A and 60 ng of total RNA from anemic chicken reticulocytes (lane 8 and 9). The riboprobe for the α^A-globin gene was used in lanes 1,4,7 and 8. That for the α^D-globin gene was used in lanes 2,3,5,6 and 9.

(B) RNA assayed in all the lanes used are as in panel A. The riboprobe for the CAT gene was used in all lanes.

Procedures for transfection, RNA isolation, RNase protection assay and gel electrophoresis are described in Materials and Methods.

→ indicates the marker size. ⇒ indicates the size of the protected fragment for the α^A-globin gene. ⇐ indicates the size of the protected fragments for the α^D-globin gene.


A.
M
1
2
3
4
5
6
7
8
9
154bp

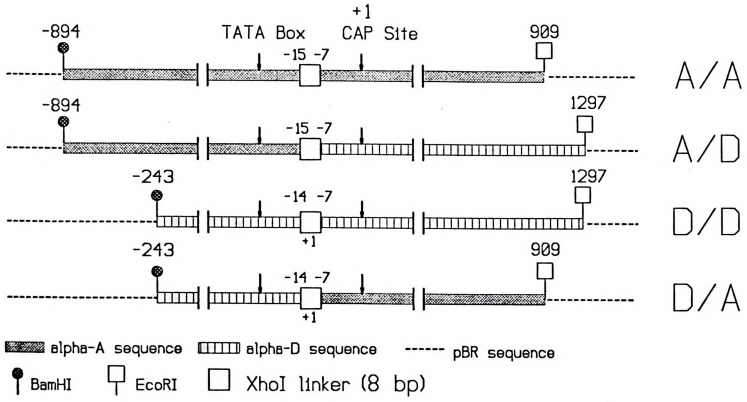

B.
M
1
2
3
4
5
6
7
8
9
154bp

138 bp and 147 bp protected fragments, respectively, after RNase digestion. More details are shown in figure 3.

The results of such a RNase protection analysis are shown in figure 4. The α^A-globin gene in pBRα7-1.7 (figure 4, lane 1) and in pAT48D-HV5.6 (data not shown) shows a detectable level of RNA initiated from its normal CAP site (figure 4 lane 1), whereas the α^D-globin gene transcript was not detectable for both pHRα5-4.3 and pAT48D-HV5.6 plasmid (figure 4, lanes 2 and 3). We estimate that a band with about 1/10 the intensity of the α^A-globin band could be detected by these experiments. Therefore, in our transfected QT6 cell system, RNA initiating from the α^A promoter is at least 10 fold higher than that from α^D promoter, whereas α^A and α^D-globin genes express coordinately in a 3:1 ratio *in vivo*(2). There is a noticeable band between the two arrows (⇐) in every RNase protection assay using the α^D riboprobe (figure 5, lanes 2,3,5,6,9). Since this band was also found in a negative control using normal QT6 RNA (data not shown) and since the positive control (figure 5, lane 9) contains two sharp bands (as indicated by the arrows ⇐), we believe that the band is an artifact resulting from incomplete RNase A+T1 digestion of this riboprobe, and that it should be ignored. In this and subsequent RNase protection analyses, the poly A+ RNA of untransfected QT6 cells was used as a negative control. No cross-hybridization was found between the riboprobes and QT6

Figure 5: Maps of Chicken α-Globin Hybrid Genes.

The construction of hybrid genes is described in Materials and Methods. Hybrid genes with the α^D-globin gene promoter have one extra base pair between the TATA box and the CAP site.


+1
-894
TATA Box
CAP Site
909
-15 -7
A/A
-894
1297
-15 -7
A/D
-243
1297
-14 -7
+1
D/D
-243
909
-14 -7
+1
D/A
alpha-A sequence
alpha-D sequence
pBR sequence
BamHI
EcoRI
XhoI linker (8 bp)

RNA (figure 11C, lane 7) as expected, since the QT6 fibroblast cells do not express endogenous globin RNAs. Total RNA from anemic chicken reticulocytes was used as a positive control to identify the normal CAP sites of α^A- and α^D-globin gene transcripts (figure 4A, lanes 8 and 9), and to compare relative expression levels in different assays.

In order to explore the reason(s) for this difference in relative expression levels from the α^A and α^D gene promoters as compared to their *in vivo* behavior, chimeric α-globin genes, consisting of one α-globin gene's promoter linked to the other gene's body, were constructed by linking two deletion mutants (see below) at a 8 bp XhoI linker site inserted between the TATA box and the CAP site (figure 5). Since an 8 bp region between the TATA box and CAP site (7 bp for the α^D promoter) has been replaced by an 8 bp XhoI linker (figure 5) in these hybrid genes, two control genes, A/A and D/D (figure 5), were constructed to monitor any possible variation in transcription efficiency caused by the XhoI linker replacement. To our surprise, the expression level of the α^A-gene in the A/A clone (figure 4, lane 4) is much higher than in pBRα7-1.7 (figure 4, lane 1). This phenomenon is presumably due to the XhoI linker replacement, but the exact reason for the change remains obscure. However, a comparison among the expression levels of these hybrid genes should still be meaningful since all of these genes possess this XhoI linker replacement. The

1 extra bp added directly upstream from the Xho I linker in D/D and D/A (figure 5) was not expected to have a substantial effect on the expression level of these genes and on the positions of the cap site. For example, in a similar study on the expression of the SV40 early gene, mutants containing deletions directly downstream of the TATA box gave similar expression levels to wild type, and the 5' ends of RNAs from the mutants began approximately the same distance downstream from the TATA box as for the wild type gene (4).

The results of transfecting these constructs into QT6 cells are shown in figure 4. The α^A promoter linked to the α^D body (A/D) expresses RNA at a level about 1/3 that of the control α^A gene promoter-α^A gene body construct (A/A) (figure 4, lane 4 and 5), a ratio consistent with the expression ratio of these coding sequences *in vivo* (figure 4 lane 8 and 9). However, the chimeric genes with the α^D-globin promoter (attached either to the α^A or α^D gene body, figure 4, lanes 6 and 7, respectively) showed no detectable signals. These results suggest that in QT6 quail fibroblasts, the α^D-globin gene has a much weaker promoter activity than does the α^A-globin gene and that it may require an additional activator(s) or/and relief from repressor(s) to express at a detectable level. It should be noted that absence of expression from the α^D-globin gene in a fibroblast cell is, in fact, the behavior expected from a normally erythroid cell-specific gene such as

a globin gene. We know that the endogenous α^A- and α^D-globin genes are not expressed in QT6 cells (3). However, expression of transfected genes including globin genes, in analogous mammalian systems, has often been observed in cell types where the corresponding endogenous genes are inactive. These results do suggest that the body of α^D-globin gene does not exert a significant inhibitory effect which blocks expression from the α^D-globin gene, since the chimeric A/D gene shows detectable expression of α^D-globin gene sequences (figure 4, lane 5). All the clones tested in figure 4. have been examined in at least two separate transfections, and results consistent with those shown were obtained.

The above results are similar to expression data obtained in cells stably transfected with these clones, in this case using G418 resistance expressed from a co-transfected pSV2neo plasmid to select transfectants. However, in this case no transcription from the A/D mutant was detected (appendix II). The relatively insensitive S1 protection assay used in these studies and the low level of A/D expression observed by RNase protection in our transient transfections may explain the difference in these results.

B. α^A-Globin Gene Expression from Clones Containing Deletions in the Upstream Promoter Region.

Although the promoter elements important for maintaining

basal levels of the mouse and rabbit ß-globin genes have been well characterized (5,6), there is little known about promoter elements of this kind in the chicken α^A-globin gene. In order to further understand the promoter elements controlling constitutive expression of the α^A-globin gene, a nested set of deletion mutants of α^A-globin gene 5' flanking sequence was constructed. Each deletion began at the KpnI site (-340) and ended upstream of the TATA box. The construction of this deletion series is described in Materials and Methods. The resulting mutants all have a deletion starting at -340 and ending at a point between -325 to -84 as shown in figure 6. The naming of these clones is described in the legend of figure 6.

Transient transfection of the above mutant DNAs into QT6 quail fibroblasts was performed, followed by RNase protection analysis as described above (figure 2). The expression level of the α^A-globin gene varied little with deletion of sequences from -340 to -169 (figure 7A, lanes 2-10; figure 7C, lane 4), on wild type pBRα7-1.7 (figure 7C, lane 1). However, once the deleted region extended to -156, a major drop in expression was observed (figure 7C, lane 5). Δ156 expressed the α^A-gene transcript about 20 fold less than those mutants with shorter deletions. Surprisingly, however, deletions of increasing extent toward the α^A-globin gene (Δ147, Δ144, Δ114) show a gradual increase in the expression level (figure 7C, lanes 6,7,8) compared to the reduced level of Δ156. However, when

Figure 6: Size and Location of the Deletion Mutants.
For each deletion mutant except for ΔB/X, the deletion starts from the KpnI site (-340) and ends at the base pair for which it is named (Each clone's name is given by Δ followed by the absolute value of the 3' end of the deletion.). Mutant ΔB/X has a deletion from -894 to -340.

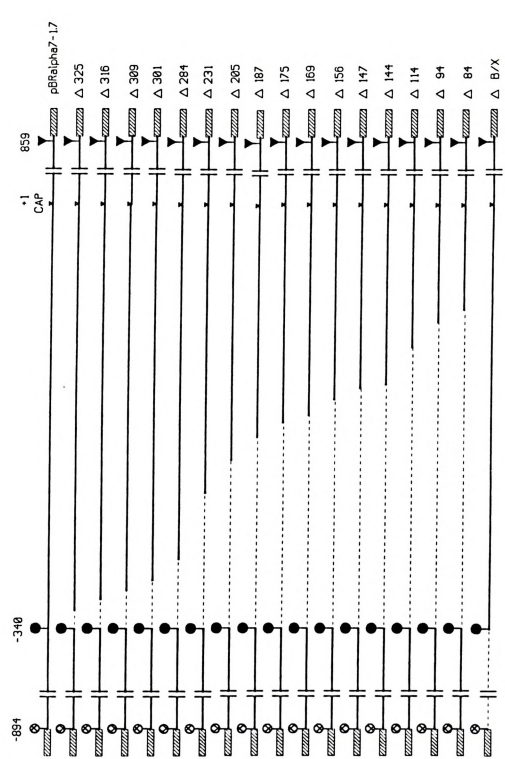

-894
-340
+1
CAP
859
pBRalpha7-17
Δ 325
Δ 316
Δ 309
Δ 301
Δ 284
Δ 231
Δ 205
Δ 187
Δ 175
Δ 169
Δ 156
Δ 147
Δ 144
Δ 114
Δ 94
Δ 84
Δ B/X
pBR sequence
Deleted Region
BamHI
KpnI
BglII

Figure 7: Transfection of deletion mutants into QT6 fibroblast cells.

(A, B) RNase protection assays were performed on poly A+ RNA from QT6 cells transfected with (lane 1) K→X; (lane 2) Δ325; (lane 3) Δ316; (lane 4) Δ309; (lane 5) Δ301; (lane 6) Δ284; (lane 7) Δ231; (lane 8) Δ205; (lane 9) Δ187; (lane 10) Δ175.

(C,D) RNase protection assays were performed on poly A+ RNA from QT6 cells transfected with (lane 1) pBRα7-1.7; (lane 2) ΔB/X; (lane 3) K→X; (lane 4) Δ169; (lane 5) Δ156; (lane 6) Δ147; (lane 7) Δ144; (lane 8) Δ114; (lane 9) Δ94; and on 20ng of total RNA from anemic chicken reticulocytes (lane 10).

Riboprobe for the α^A-globin gene was used in panels A and C. Riboprobe for the CAT gene was used in panels B and D.

Procedures for transfection, RNA isolation, RNase protection assay and gel electrophoresis are described in Materials and Methods.

→ indicates the marker size. ⇒ indicates the size of protected fragment for α^A-globin gene in panel A and C. ⇒ indicates the size of protected fragment for CAT gene in panel B and D. Mutant K→X is almost identical to pBRα7-1.7 except that the KpnI site (-340) has been converted to XhoI site. The construction of K→X is described in Materials an Methods.

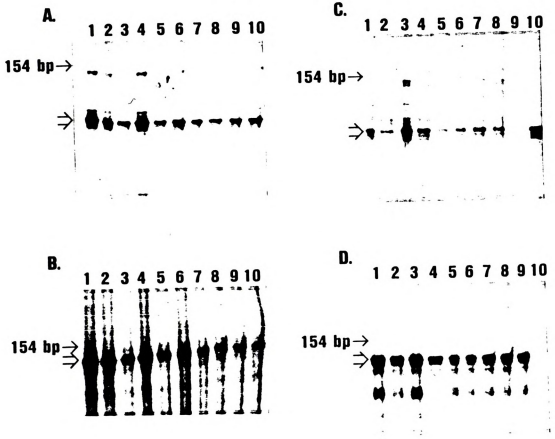

A.
1 2 3 4 5 6 7 8 9 10
154 bp→
B.
1 2 3 4 5 6 7 8 9 10
154 bp→
C.
1 2 3 4 5 6 7 8 9 10
154 bp→
D.
1 2 3 4 5 6 7 8 9 10
154 bp→

Table 1: Densitometric analysis of mutants containing deletions in the promoter region of the α^A- globin gene.

	experiment 1[a]	experiment 2[a]	experiment 3[b]
pBRα7-1.7	1	1	
K→X	8		8
Δ284	10.4	3.6	
Δ231	9.6		
Δ205	3.2	7.4	
Δ187	14.4		
Δ175	24.8		1.76
Δ169	7.6	1.2	2.16
Δ156	0.24		0.016
Δ147	1.44		0.04
Δ144	3.6		0.4
Δ114	1.76	2.24	0.064
Δ94	0.016	0.16	2×10^{-4}
Δ84		0.16	0.008

a: In experiments 1 and 2, the expression level of the chicken α^A-globin gene has been normalized relative to internal control (CAT gene). The expression from pBRα7-1.7 was arbitrarily assigned a value of 1.0.

b: Due to the low specific activity of the CAT riboprobe, internal controls were unable to be used in experiment 3. An equal transfection efficiency for each test clone was assumed in this experiment. Since pBRα7-1.7 was not included in this experiment, the value for K→X was arbitrarily set equal to that found in experiment 1.

the deletion extended to -94, the expression of the α^A-globin gene became almost undetectable (figure 7C, lane 9). All the deletion mutants have been tested at least twice with consistent results as shown in table 1.

An analysis of α^A-globin gene 5' flanking sequences protected from DNase I by QT6 cell nuclear extracts was carried out to investigate protein-binding activities within this DNA region. The source of binding protein was a crude nuclear extract from QT6 quail fibroblasts. Extracts were incubated with the EcoRI(+859)/XhoI (-205 and -169) fragments end-labeled at the XhoI site containing the α^A-globin 5' flanking region. These fragments were derived from the Δ205 and Δ169 constructs. By comparison of band intensities between DNase I digestion without added QT6 nuclear extract (figure 8 A and B, lane 1) and with increasing amounts of QT6 nuclear extract (figure 8 A and B, lane 2-4), several regions (defined in figure 9) were observed which give rise to bands of decreasing intensity as nuclear extract levels increase. Therefore we suspect that one or more DNA-binding proteins from the QT6 nuclear extract interact with these regions. Since the size of the region protected usually exceeds the actual protein binding site, further *in vitro* binding experiments using oligonucleotide competitors would need to be performed to identify the specific binding site sequence elements of these proteins. Some of the protected regions

Figure 8: DNase Protection Analysis of the Chicken α^A-Globin Gene Promoter Using Crude Nuclear Extracts from QT6 Quail Fibroblasts.

(A) (Lane M) HinfI-digested pBR322 as size markers; (lane G and A+G) DNA sequencing ladders; (lane 1) no N.E. added; (lane 2, 3, 4) 5, 10, 15 μg N.E. added.

(B) (lane G and A+G) DNA sequencing ladders; (lane 1) no N.E. added; (lane 2, 3, 4) 10, 20,40 μg N.E. added.

A 1,114 bp EcoRI-XhoI end-labeled fragment from +909 to -205 was used in panel A. A 1,078 bp EcoRI-XhoI end-labeled fragment from +909 to -169 was used in panel B. For each lane (1-4), 10 ng of labeled fragments was used.

The numbers to the left of panel A and B indicate the nucleotide positions with respect to the α^A-globin promoter CAP site as +1.

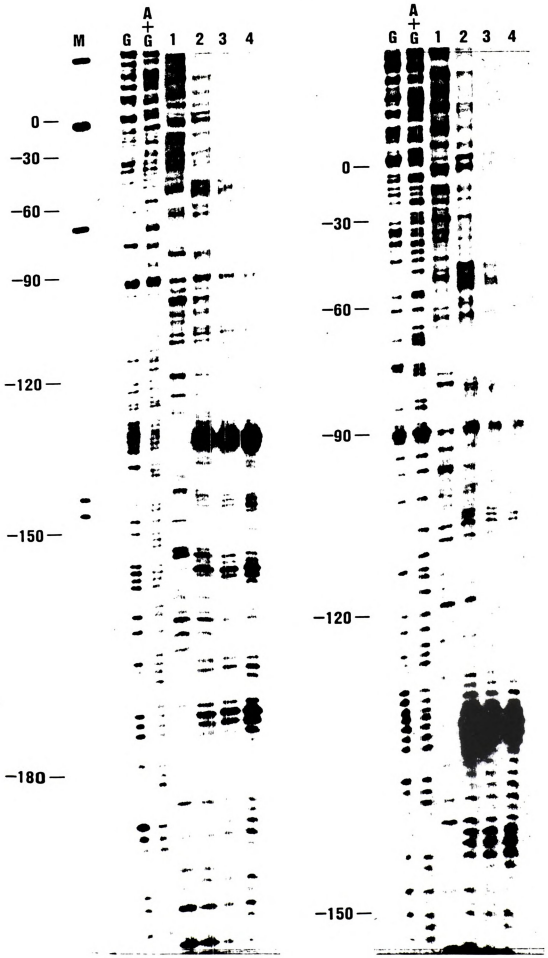
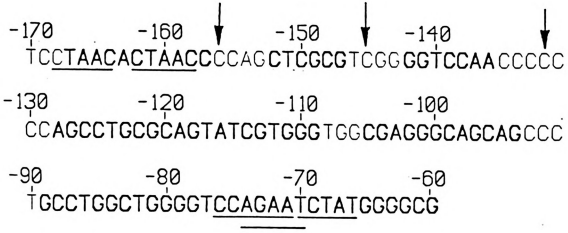

M
G
A+G
1
2
3
4
0
−30
−60
−90
−120
−150
−180
G
A+G
1
2
3
4
0
−30
−60
−90
−120
−150

Figure 9: Summary of Protein-DNA Interactions on the Chicken α^A-Globin Promoter Regions.

Sequences in boldface indicate regions of protein binding, underlined sequences denote putative CCAAT box regions and the arrows designate enhanced cleavage by DNAase I in response to binding.



revealed by this experiment corresponded with those regions affecting the expression of the α^A-globin gene in QT6 cells as revealed by the deletion analysis discussed above. For example, Δ94 has a deletion extending from -340 to -94, which deletes an apparent factor binding site at -106 to -93 (figures 8 and 9). The above results suggest the possibility that an activating protein binds to the region between -106 to -93. Another transcription rate reduction occurred when the 5' deletion was extended from -169 to -156, whereas further deletion, from -156 to -114, increased the transcription rate of α^A-globin gene relative to the Δ156 construct. There are two possible factor binding sites at -167 to -157 and at -152 to -147, in the region affected by the above deletions. The deletion of the first protected site (-167 to -157) resulted in a decrease in transcription , whereas the deletion of the second site (-152 to -147) appears to enhance the transcription rate. These observations could be explained by an activating protein binding to the first site and a repressor or negative regulatory protein binding to the second site. Of course, a variety of more complex scenarios is also possible. This phenomenon is somewhat similar to that observed in studies of the chicken ß-globin gene promoter, which possesses a binding site for a transcriptional repressor (PAL) which is side by side with another binding site for a transcriptional activator (CON), although, in this case, these

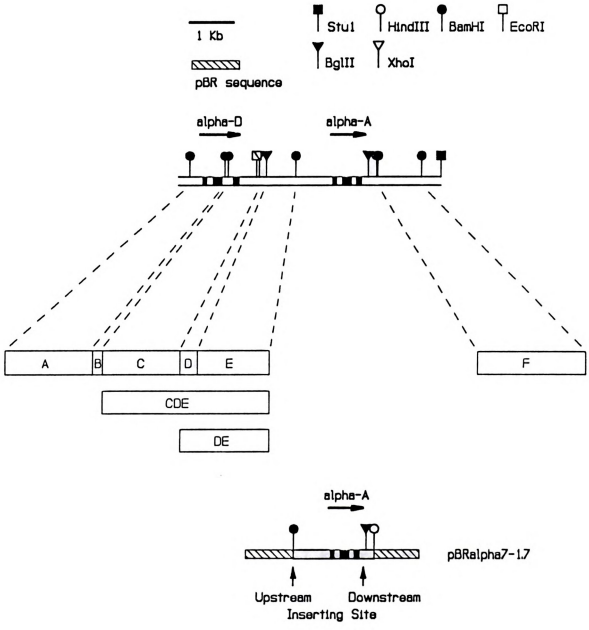
were observed in an *in vitro* transcription system using a chicken red cell extract (7). Thus these ß-globin results were obtained in an erythroid-specific system. Five putative CCAAT box regions (8) have been located in the two protected regions (-170 to -160 and -90 to -60), as shown in figure 9. Therefore the activator activity between -170 to -160 may involve CCAAT-binding protein(s). However, as noted previously, these putative CCAAT box regions deviate considerably from the normal consensus sequence. Since deletion of sequences to -94 leads to a very low expression level of the α^A-globin gene, our deletion series experiments were unable to detect specific effects that might result from interaction with the -90 to -60 region (data not shown). Further experiments using linker scanning or site-directed mutagenesis would be required to assess such effects.

C. α^A-Globin Gene Expression from Clones Containing Flanking Regions of the α^A-Globin Gene.

The ability of various DNA fragments from the chicken α-globin locus to regulate transcription from the α^A-globin promoter has been analyzed in chicken erythroid cells from 9-day embryos. A tissue-specific enhancer region has been located 3' to the α^A gene (9). However, little was known about the effects of these fragments on the constitutive expression

Fig 10: Subcloning of Fragments Around the Chicken α^A-Globin Gene.

Upstream and downstream inserting sites are indicated by arrows. Construction details are provided in the text.


1 Kb
pBR sequence
StuI
HindIII
BamHI
EcoRI
BglII
XhoI
alpha-D
alpha-A
A
B
C
D
E
F
CDE
DE
alpha-A
pBRalpha7-1.7
Upstream
Downstream
Inserting Site

of the α^A gene. To study the effects of neighboring regions of the α^A-globin gene on its expression, we have inserted various fragments from the flanking region of the α^A-globin gene, in both orientations, into the upstream BamHI site or downstream BglII site of the α^A-globin gene in plasmid pBRα7-1.7 (figure 10). The name of these subclones starts with a letter which specifies the inserted fragment, followed by two letters; the first one indicating the insertion site (U, upstream-BamHI or D, downstream-BglII), and the second one representing the fragment orientation (S indicating that the fragment has its same orientation with respect to transcription as in the chromosomal locus and O indicating an opposite orientation). For example, A UO is the name for a clone with the fragment A inserted at upstream BamHI site in an orientation opposite to its normal direction in the chicken α-globin gene locus, and DE DS is the name for a clone which has the fragment DE inserted at the downstream BglII site in its normal orientation. As a further example, note that clones DE US and CDE US are equivalent to just cloning a larger fragment of the α-globin gene locus than the 1.7 kb fragment in pBRα7-1.7.

The effects of these fragments on the transcription level of the chicken α^A-globin gene are summarized in Table 2. As shown in Table 2, column US, several fragments (A, B, C, DE and F) exhibit various levels of a positive effect on α^A-globin gene expression when inserted at the upstream (BamHI) site (figure 11A,lane 9; figure 11C, lanes 2 and 4; figure

Figure 11: Transfection of Clones Containing the Flanking Regions of the α^A-Globin Gene into QT6 Fibroblast Cells.

(A,B) RNase protection assays were performed on poly A+ RNA from QT6 cells transfected with (lane 1) pBRα7-1.7; (lane 2) CDE US; (lane 3) CDE UO; (lane 4) E US; (lane 5) E UO; (lane 6) D US; (lane 7) D UO; (lane 8) C US; (lane 9) C UO and on 200 ng of total RNA from anemic chicken reticulocytes (lane 10). Poly A+ RNA from untransfected QT6 cell was used in lane 11 as a negative control.

(C,D) RNase protection assays were performed on poly A+ RNA from QT6 cells transfected with (Lane 1) pBRα7-1.7; (lane 2) A US; (lane 3) A UO; (lane 4) B US; (lane 5) B UO and on 200 ng of total RNA from anemic chicken reticulocytes (lane 6). Poly A+ RNA from untransfected QT6 was used in lane 7 as a negative control.

(E,F) RNase protection assays were performed on poly A+ RNA from QT6 cells transfected with (lane 2) pBRα7-1.7; (lane 3) DE US; (lane 4) DE UO; (lane 5) F US; (lane 6) F UO; and on 20 ng of total RNA from anemic chicken reticulocytes (lane 7). Poly A+ RNA from untransfected QT6 cell was used in lane 1 as a negative control.

Riboprobe for the α^A-globin gene was used in panels A, C and E. Riboprobe for the CAT gene was used in panels B, D and F. → indicates the marker size. ⇒ indicates the size of the protected fragment for the α^A-globin gene in panels A, C and E. ⇒ indicates the size of protected fragment for the CAT gene in panels B, D and F.

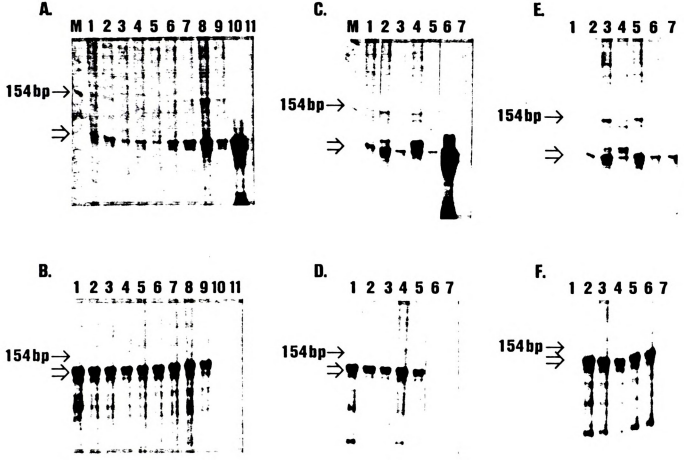

A.
M 1 2 3 4 5 6 7 8 9 10 11
154bp→
⇒
C.
M 1 2 3 4 5 6 7
154bp→
⇒
E.
1 2 3 4 5 6 7
154bp→
⇒
B.
1 2 3 4 5 6 7 8 9 10 11
154bp→
⇒
D.
1 2 3 4 5 6 7
154bp→
⇒
F.
1 2 3 4 5 6 7
154bp⇒
⇒

Figure 12: Transfection of Clones Containing the Flanking Regions of the α^A-Globin Gene into QT6 Fibroblast Cells.

(A,B) RNase protection assays were performed on poly A+ RNA from QT6 cells transfected with (lane 2) pBRα7-1.7; (lane 3) DE DS; (lane 4) DE DO; (lane 5) D DS; (lane 6) D DO and on 20 ng of total RNA from anemic chicken reticulocytes (lane 1). Poly A+ RNA from untransfected QT6 cells was used in lane 7 as a negative control.

(C,D) RNase protection assays were performed on poly A+ RNA from QT6 cells transfected with (lane 2) pBRα7-1.7; (lane 3) F DS; (lane 4) F DO; (lane 5) E DS; (lane 6) E DO; (lane 7) C DS; (lane 8) C DO and on 20 ng of total RNA from anemic chicken reticulocytes (lane 9). Poly A+ RNA from untransfected QT6 cells was used in lane 1 as a negative control.

Riboprobe for the α^A-globin gene was used in panels A and C. Riboprobe for the CAT gene was used in panels B and D.

→ indicates the marker size. ⇒ indicates the size of the protected fragment for the α^A-globin gene in panels A and C. ⇒ indicates the size of the protected fragment for the CAT gene in panels B and D.

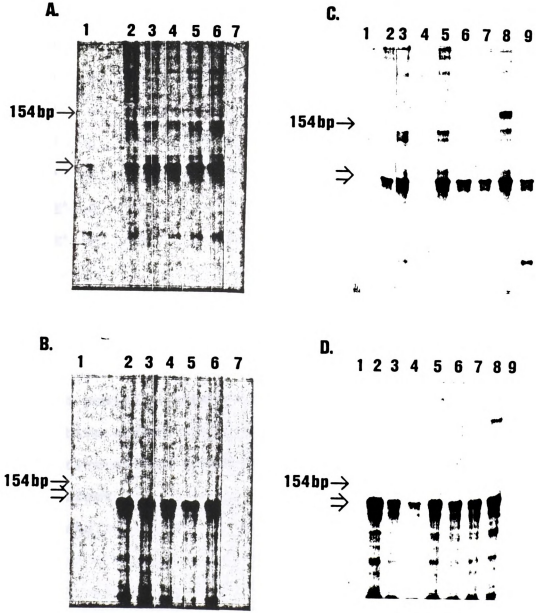

A.
1 2 3 4 5 6 7
154bp
B.
1 2 3 4 5 6 7
154bp
C.
1 2 3 4 5 6 7 8 9
154bp
D.
1 2 3 4 5 6 7 8 9
154bp

Table 1: The effects of various fragments on the transcription level of the chicken α^A-globin gene.

	US[b]	UO[b]	DS[b]	DO[b]
A[a]	++	0	*	*
B[a]	+	0	*	*
C[a]	++	0	0	0
D[a]	0	0	0	0
E[a]	0	-	0	0
F[a]	+	0	0	-
CDE[a]	0	-	*	*
DE[a]	+	0	0	o

a. The fragments are described in figure 10.

b. The orientations and the insertion sites are defined in the text.

++: large positive effect (10-20 fold increase)

+: positive effect (5-10 fold increase)

-: negative effect (5-10 fold decrease)

0: no effect.

*: no data.

11E, lanes 3 and 5). However, as shown in table 2, column DS, when these fragments are inserted at the downstream (BglII) site, no significant positive effect is observed (figure 12A, lanes 3 and 5; figure 12C, lanes 3,5 and 7). Thus, these DNA fragments do not behave in the manner expected by classical enhancers. It's hard to explain why fragments C (in C US) and DE (in DE US) alone can enhance the transcription of the α^A-globin gene but do not do this when they are linked together (fragment CDE in CDE US). One possible reason is that the separation of fragment C from DE could result in the destruction of a repressor binding site. This hypothesis could be tested by introducing fragments with different breakpoints or by inserting foreign DNA at the C to DE junction.

As shown in table 2, column US and UO, two fragments (CDE and E) showed an inhibitory effect on the transcription level of pBRα7-1.7 when they were inserted in the opposite orientation at the upstream site (figure 11A, lanes 3 and 5), but no effect in the positive orientation. Several other clones (fragments A, B, C, DE and F) showed no effect in the opposite orientation but strong positive effects when inserted in their normal orientation with respect to transcription. When these fragments are inserted at upstream the BamHI site, with a positive orientation, they lead to the expression of 5-20 fold higher levels of α^A-globin RNA than they do in a negative orientation (figure 11A, lanes 8 and 9; figure 11C, lanes 2,3,4 and 5; figure 11E, lanes 3,4,5 and 6). However,

all these fragments (except for fragment F) lose this directional enhancement or inhibition effect when they are inserted at downstream BglII site (see table 2, columns DS and DO).

Fragment F has been shown in another lab to have an erythroid-specific enhancer domain containing three binding sites for Eryf1 (10), Eryf1 is an erythroid-specific binding protein that also plays an important role in the activity of the chicken ß-globin enhancer (11). Note, however, that Eryf1 should not be present in QT 6 fibroblasts and thus should not exert an effect via fragment F binding. Unlike other fragments, the F fragment has a directional-specific effect at both the upstream and downstream insertion sites as shown in table 2, row F (but F increases transcript levels in the US position and decrease transcript levels in the DO position). Fragment F is also the only fragment isolated from the 3' flanking region of the α^A-globin gene. It may be significant that DNA regions normally upstream from the α^A gene no longer exert a direction-specific effect when placed downstream of the gene, but the normally 3' fragment F continue to have some direction-specific effect in the downstream position. In any case, note that none of the fragments which confer a positive effect on α^A gene expression show both the location and orientation independence of classical enhancer sequences.

In order to further examine this orientation effect,

Figure 13: DNase Protection Analysis of the 84 bp (556-638) Region in the Chicken α^D-Globin Gene Using Crude Nuclear Extract from QT6 Quail Fibroblasts.

(A) (Lane M) HinfI-digested pBR322 as size markers; (lane G and A+G) DNA sequencing ladders; (lane 1) no N.E. added; (lane 2, 3, 4) 20, 40, 60 μg N.E. added.

A 1,382 bp PstI-EcoRI end-labeled fragment from clone B DS was used as a probe and 10 ng of this fragment was used in each lane (1-4). The numbers to the left indicate the positions in α^D-globin gene sequence.

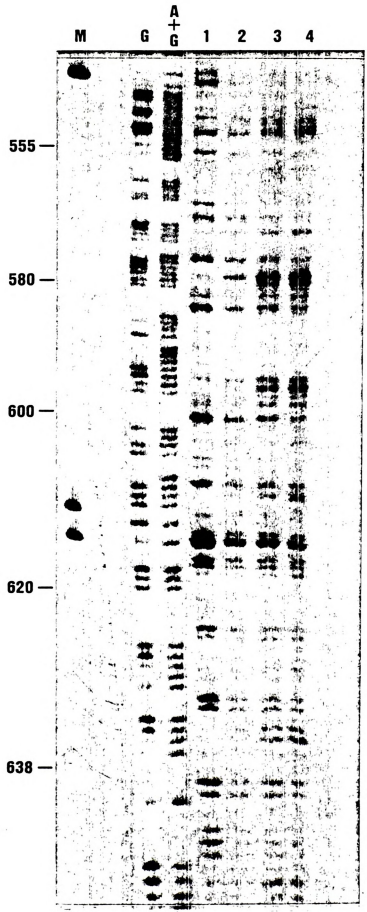
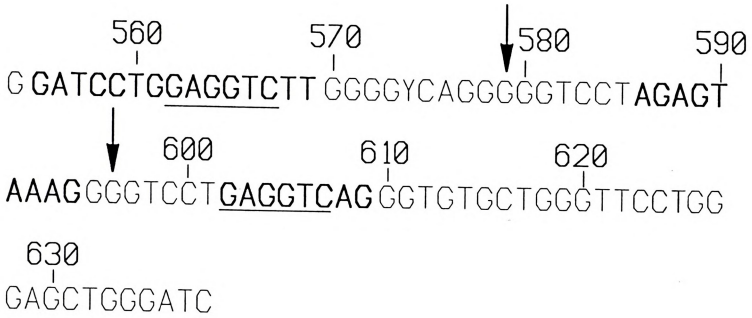

M
G
A+G
1
2
3
4
555
580
600
620
638

Figure 14: Summary of Protein-DNA Interactions on the 84 bp Fragment (556-538) from the Chicken α^D-Globin Gene Second Intron (Fragment B).

Sequences in boldface indicate regions of protein binding, underlined sequences denote a GAGGTC motif and the arrows designate enhanced cleavage by DNAase I in response to binding.


560
570
580
590
G GATCCTGGAGGTCTT GGGGYCAGGGGGTCCT AGAGT
600
610
620
AAAG GGGTCCT GAGGTCAG GGTGTGCTGGGTTCCTGG
630
GAGCTGGGATC

fragment B, the shortest fragment (84 bp) which demonstrated this effect, was used for DNase I protection experiments. This fragment is in fact completely contained within the second intron of the upstream α^D-globin gene. By comparison of band intensities between DNase I digestion without added QT6 nuclear extract (figure 13, lane 1) and with increasing amounts of QT6 nuclear extract (figure 13, lane 2-4), three protected regions (defined in figure 14) were observed, which give rise to bands of decreasing intensity as nuclear extract levels increase. Therefore we suspect that one or more DNA-binding proteins from the QT6 nuclear extract interact with these regions. Since this 84 bp fragment can enhance the α^A-globin RNA levels in QT6 cells in an orientation-specific manner, one might postulate that this property is due to: 1. binding an orientation-specific activating protein(s); 2. Binding both a bidirectional activating protein and an orientation-specific negative regulatory factor which can reduce the transcription efficiency (or block enhancing effects) in a certain orientation relative to the promoter; or 3. a combination of orientation-specific positive and negative effects. Of course, at this stage other, more complex explanations are also possible; for example, effects on topology and long range spacing of elements.

Most of these fragments from the chicken α-globin locus seem to have an effect that increases the transcript levels of the α^A-globin gene when they are tested in their wild-type

orientation and decrease these levels or show no effect in the opposite orientation. This effect could serve a function by selectively enhancing transcription in the direction which produces the wild-type mRNA and/or inhibiting the transcription from the opposite direction which would produce anti-sense mRNA. However, since we have been unable to observe significant α-globin gene transcription in any transfected erythroid cell system, it remains unclear whether these flanking DNA segments exert a similar effect in cells which normally transcribe globin mRNA. The development of new methods of transfection, new avian erythroid cell types or viable *in vitro* erythroid transcription systems will be required to answer this question.

Three regions of protection by QT6 nuclear extracts within the 84 bp Bam HI fragment were observed as described above (figure 14). The sequences of these three protected regions have been searched for homologies with various known DNA sequences that bind regulatory protein factors. No striking sequence similarities to known DNA-binding motifs were found among the three regions. However, we observed that two out of these three regions share a common 6-bp GAGGTC motif (figure 15). The sequence of the chicken α-globin gene cluster was searched for similar motifs. Ten identical copies were located in and around the α^A- and α^D-globin genes, but none were found within the embryonic π gene (figure 15). In

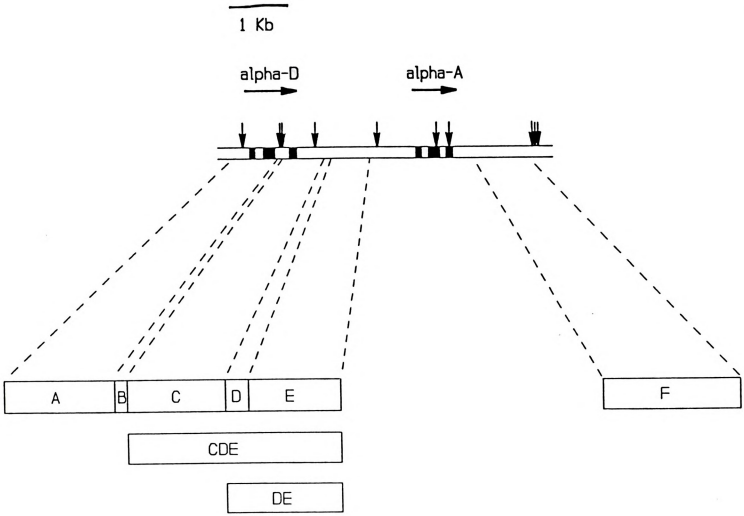
Figure 15: The Positions of the GAGGTC Motif In and Aroud the α^A- and α^D-Globin Genes.

↓ indicates the positions of the GAGGTC motif.

→ indicates the transcription direction of the genes.

Exons are shown as black boxes and introns are closed boxed.

Fragments A, B, C, D, E, F, CDE and DE are defined as in figure 10.


1 Kb
alpha-D
alpha-A
A
B
C
D
E
F
CDE
DE

addition, this motif also occurs twice within the adult ß-globin gene but not within the embryonic ρ- and ϵ-globin genes. However, since this 6-bp sequence doesn't exist on all of the fragments that show the orientation-specific effect, it alone cannot account for this effect. Note that three of these motifs exist within the cloned α^A gene fragment in pBRα7-1.7. Thus, it is unclear how the addition of further such elements in the flanking fragments tested might alter the control level of α^A-globin gene expression. In order to understand the effect, if any, of the GAGGTC sequence, further *in vitro* protein binding experiments with oligonucleotide competitors would need to be performed and new, more specific mutant α-globin gene constructs would have to be tested.

It is interesting that this GAGGTC sequence only occurs in the chicken adult α- and ß-globin genes but not in the embryonic ones. This suggests that this sequence could be important for the expression of the adult chicken globin genes in definitive erythroid cells.

References:

(1) Gorman, C. et al., *PNAS* **79:** 6777, 1982.

(2) Brown, J. L., Ingram, V. M., *J. Biol. Chem.* **249:**3960, 1974.

(3) Moscovici, C. et al., *Cell* **11:**95, 1977.

(4) Gluzman, Y. et al., *PNAS* **77:**3898, 1980.

(5) Maniatis, T. et al., *Science* **236:**1237, 1987.

(6) Hanahan, D., *Nature* **315:**115, 1985.

(7) Emerson, B. M. et al., *Cell* **57:**1189, 1989.

(8) Dodgson, J. B., Engel, J. D., *J. Biol. Chem.* **258:**4623, 1983.

(9) Knezetic, J. A., Felsenfeld, G., *Mol. Cell. Biol.* **9:**893, 1989.

(10) Evans, T. et al., *PNAS* **85:**5976, 1988.

(11) Emerson, B. M. et al., *PNAS* **84:**4786, 1987.

SUMMARY

A study of the regulation of constitutive expression of the chicken α-globin genes in QT6 fibroblast cells has demonstrated that:

(1) The promoter region of α^A-globin gene (but not α^D) was sufficient to promote a measurable expression from both the α^A- and α^D-globin gene coding regions.

(2) Two activating and one repressing regions appear to exist in the promoter region of the α^A-globin gene.

(3) Orientation-specific activating and inhibitory effects were observed to result from insertions of various regions of the α-globin gene locus into sites upstream and downstream of the α^A-globin gene.

Our studies have shown that α^A-globin gene expression is subject to a complex series of regulatory influences in this constitutive system. It seems likely that similarly complex interactions hold true in erythroid cells although this can't at present be proven. Part of the process of globin gene activation in such erythroid cells may entail release from negative regulatory effects such as those described herein.

Appendix I

The Chicken Carbonic Anhydrase II Gene: Evidence for A Recent Shift in Intron Position

Note: My involvement in this project included:

(1) isolation of chicken carbonic anhydrase II cDNA clone from a chicken reticulocyte cDNA library.

(2) sequence determination of an isolated chicken carbonic anhydrase II cDNA clone.

Volume 15 Number 2 1987 Nucleic Acids Research

The chicken carbonic anhydrase II gene: evidence for a recent shift in intron position

Corinne M.Yoshihara+, Jiing-Dwan Lee and Jerry B.Dodgson

Departments of Microbiology and Biochemistry and Genetics Program, Michigan State University, East Lansing, MI 48824, USA

Received September 19, 1986; Revised and Accepted December 5, 1986

ABSTRACT

The complete nucleotide sequence of the coding region of the chicken carbonic anhydrase II (CA II) gene has been determined from clones isolated from a chicken genomic library. The sequence of a nearly full length chicken CA II cDNA clone has also been obtained. The gene is approximately 17 kilobase pairs (kb) in size and codes for a protein that is comprised of 259 amino acid residues. The 5' flanking region contains consensus sequences commonly associated with eucaryotic genes transcribed by RNA polymerase II. Six introns ranging in size from 0.3 to 10.2 kb interrupt the gene. The number of introns as well as five of the six intron locations are conserved between the chicken and mouse CA II genes. The site of the fourth intron is shifted by 14 base pairs further 3' in the chicken and thus falls between codons 147 and 148 rather than within codon 143 as in the mouse gene. Measurements of CA II RNA levels in various cell types suggest that CA II RNA increases in parallel with globin RNA during erythropoiesis and exists only at low levels, if at all, in non-erythroid cells.

INTRODUCTION

The carbonic anhydrases constitute an ancient family of proteins. Five different carbonic anhydrase (CA) isozymes (1-9) (possibly more, 10-12) have so far been identified, each of which is believed to be coded for by a separate genetic locus. Together the CA isozymes are extensive in their distribution. They are found in practically all organisms and in most tissues of any higher organism. Although the most obvious and well-studied role of CA is the hydration of CO_2 in red blood cells, other CA functions have been elucidated. A few of these roles include involvement in ion fluxes in neurons (13), avian eggshell formation (14), and eye morphogenesis (15). The CA gene family is variable in its expression pattern. For example, CA II is the only isozyme expressed in avian red blood cells, whereas both CA I and CA II isozymes are expressed in most mammalian red blood cells (16).

Only recently has CA gene structure been examined at the DNA level. DNA sequence data of the mouse CA II gene (17,18) and a rabbit CA I cDNA

clone (19) have been reported. Analysis of the mouse CA II gene showed that it was composed of seven exons that were stretched over 16 kb of DNA (18). The human CA II gene has been partially sequenced, and a comparison of human and mouse CA II promoters has been made (20). We previously described the characterization of a partial chicken CA II cDNA clone (21). In this paper the isolation and structural analysis of the complete chicken CA II gene is given. The chicken gene shows interesting similarities to and differences from the analogous mouse gene. In addition, CA II mRNA levels in various cells and tissues have been measured.

MATERIALS AND METHODS

Isolation of cDNA and Genomic Clones

A λgt10 cDNA library prepared from chicken red cell poly(A)$^+$ mRNA (22) was screened at a 99% representation of the library. The plaques were transferred to nitrocellulose filters and processed as described (23). CA II cDNA was identified by hybridization to both a 5'-end chicken CA II cDNA clone probe (21) and a full length mouse CA II cDNA clone probe (17, generous gift of Dr. P. Curtis, Wistar Institute). Probes were prepared by nick translation (24). CA II cDNA fragments were isolated from positive λgt10 clones by digestion at the Eco RI linker sites and inserted into pBR325 plasmid DNA. A fine structure restriction map was derived for the subclone with the largest insert (pCA-1.2) and used for the sequence analysis of the cDNA. A λCharon 4A chicken genomic DNA library (23) was also screened with the 5'-end chicken CA II cDNA clone as described above. Restriction maps of the unique CA II-containing phage, designated λcaIII and λcaXVI, were prepared by standard multiple restriction digestion (24). Subclones of restriction fragments of phage DNAs were prepared by standard techniques (23,24) following their isolation from agarose gels (25). The resultant plasmid DNAs were further analyzed by restriction enzyme digestion and blot hybridization (26).

DNA Sequence Analysis

Subcloned CA II DNA was restriction enzyme digested, gel fractionated and the appropriate fragment isolated (25). Fragments were treated with calf alkaline phosphatase, 5'-end labeled with (γ-^{32}P)ATP, recut with the appropriate secondary enzyme and the resultant singly-labeled fragments isolated. Chemical degradation and gel electrophoresis were as described previously (27-29).

S1 Analysis

Restriction fragments were isolated from restriction enzyme digested

pBBca-2.8 by cleavage within exon 1 (RsaI site at +57, Fig. 3) and upstream beyond the likely start site (RsaI, -600). This fragment was labeled with polynucleotide kinase as described above, recut with SinI (-133) and the resultant 190 base pair (bp) fragment isolated. The end-labeled fragment (40 ng at 5 x 10^6 dpm/µg) was hybridized to 50 µg of each RNA to be tested at 55° in 80% formamide hybridization buffer (30) for 15 hours. Samples were quenched by dilution into 0.3 ml of S1 digestion buffer (30 mM NaOAc, 250 mM NaCl, 4 mM Zn $(OAc)_2$, 200 µg/ml denatured salmon sperm DNA, pH 4.5). S1 nuclease (1200 U/ml) was added, digestion was allowed to proceed for 15 min at 37°, and the reaction was stopped by phenol/chloroform extraction. (S1 digestion at 300 U/ml gave equivalent results.) Labeled DNA was run on a sequencing gel as described (27).

Miscellaneous

Restriction enzymes were obtained from International Biotechnologies, Inc.; New England Biolabs, Inc. and Bethesda Research Labs, Inc. Polynucleotide kinase was from Amersham or International Biotechnologies, Inc. Enzymes were used according to manufacturers specifications. Other materials and bacterial strains were as previously described (23,24,30). RNA was isolated as described (23,24). Manipulations of recombinant DNA were done according to the appropriate NIH Guidelines.

RESULTS

Isolation of Larger Chicken CA II cDNA Clones

The partial chicken CA II cDNA clone previously described (21) was used to obtain more nearly complete clones from a λgt10 red cell poly(A)$^+$ cDNA library (22). The largest CA II cDNA insert (1.2 kb) subcloned into pBR325 is termed pCA-1.2. The cDNA clone was restriction mapped and its complete sequence was determined. It was found to code for sequences from codon 9 to a point about 440 bp 3' to the TAA stop codon. The cDNA sequence will be described in more detail below in comparison to the genomic sequence.

Isolation of the Chicken CA II Gene

A λCharon 4A chicken genomic library (23) was screened with the chicken CA II cDNA clone (21) containing the 5'-end of the cDNA. Two of the resulting clones, λcaIII and λcaXVI, have been mapped and characterized in detail. The restriction maps of the two phage are shown in Fig. 1. Subclones of appropriate phage restriction fragments were constructed, and their fine structure restriction maps are given in Fig. 2

Sequence of the Chicken CA II Gene

The subclone maps were used to develop a sequencing strategy for each

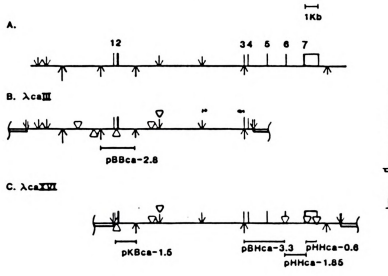


Figure 1. Restriction map of the chicken CA II gene locus. (A) Restriction map of chromosomal DNA contained within clones λcaIII and λcaXVI. (B) Restriction map of clone λcaIII which contains exons 1 to 4. (C) Restriction map of clone λcaXVI which contains exons 2 to 7. The solid boxes represent the coding regions and the open boxes represent the untranslated regions of the exons. The numbers above the boxes indicate the exons and the horizontal lines below the boxes represent the subclones. The arrows above the line indicate the direction and extent of DNA sequence determination. Only the BamHI and EcoRI sites are shown in line A.

of the CA II exons. The direction and extent of the regions sequenced are indicated by the arrows in Figs. 1 and 2. Except for that sequence 5' to codon 9, all the coding sequences were also sequenced on one or both strands of the cDNA clone, pCA-1.2.

The DNA sequence of the chicken CA II gene is given in Fig. 3. The genomic sequence shown encodes 259 amino acid residues and approximately 0.8 kb of untranslated region. The coding sequence of the gene was identified by comparison to the chicken CA II cDNA (pCA-1.2) sequence. The amino acid sequence predicted from the nucleotide sequence is identical to that predicted from the cDNA clone sequence. However, there are five differences in the cDNA and genomic sequence (Fig. 3). These include the third nucleotide of codon 140 (C in the genomic and T in the cDNA) and codon 152 (A in the genomic and G in the cDNA). There are also three changes in the 3' untranslated region: at nucleotide 1001 a G in the cDNA is an A in the genomic DNA; at nucleotide 1177 a C in the cDNA is an A in the genomic DNA;

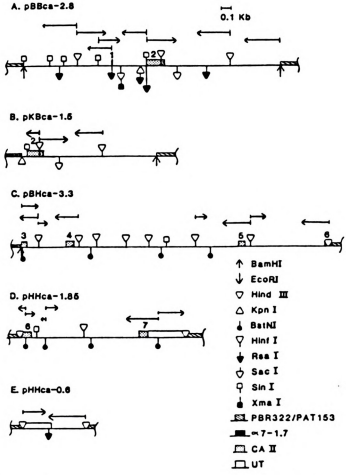


Figure 2. Restriction maps of the subcloned DNA fragments of the chicken CA II gene. Partial restriction maps of (A) subclone pBBca-2.8 which contains exons 1 and 2, (B) subclone pKBca-1.5 which overlaps with pBBca-2.8 and also contains exon 2, (C) subclone pBHca-3.3 which contains exons 3,4,5, and a small portion of exon 6, (D) subclone pHHca-1.85 which contains the greater portion of exon 6, all of exon 7, and a portion of the 3'-untranslated region, (E) subclone pHHca-0.6 which contains the remainder of the 3'-untranslated region. The filled boxes represent the exons which are identified by numbers. Open boxes show 3' untranslated sequences. The region designated as α7-1.7 (B) is a portion of chicken globin gene DNA used to provide the KpnI site for this subclone. The arrows above the boxes show the direction and extent of DNA sequence analysis.

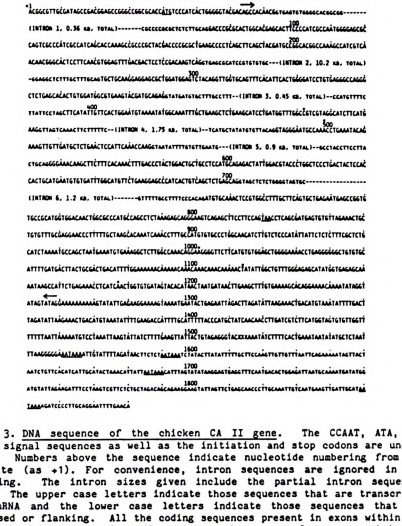


Figure 3. DNA sequence of the chicken CA II gene. The CCAAT, ATA, and AATAAA signal sequences as well as the initiation and stop codons are underlined. Numbers above the sequence indicate nucleotide numbering from the cap site (as +1). For convenience, intron sequences are ignored in the numbering. The intron sizes given include the partial intron sequences shown. The upper case letters indicate those sequences that are transcribed into mRNA and the lower case letters indicate those sequences that are processed or flanking. All the coding sequences present in exons within the left and right arrows above the sequence shown were also sequenced in the cDNA clone, pCA-1.2. Asterisks indicate sites where the cDNA and genomic clone sequences differ (see text). X indicates a nucleotide whose identity could not be resolved. Y indicates C or T.

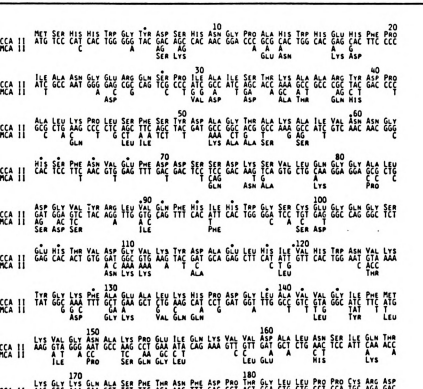


Figure 4. Amino acid sequence comparison of chicken and mouse CA II genes. The predicted amino acid sequence of chicken CA II is compared with homologous amino acid sequences of mouse CA II (17,18). The amino acid sequence predicted by the nucleotide sequence is given above the coding regions along with its numbering. Only those nucleotides in the mouse sequence that differ from that of the chicken sequence are given along with the resulting amino acid change, if any. Asterisks indicate those amino acid residues that are located in the active site region of the CA II protein.

TABLE I. DNA sequence of intron donor and acceptor sites.

Intron	Donor	Acceptor
1	ACG/GTGAGT	CGCGCTCTCTTGCAG/G
2	CAG/GTGAGC	CTCTTTGCTTTGCAG/T
3	GAG/GTATGA	TTTTCTTATTCCTAG/C
4	AAG/GTTAGT	CTATATGTGTTACAG/G
5	AAG/GTAAT	CCTTCCTTACTGCAG/G
6	CAG/GTAGCT	TGCCTTTTCCCACAG/A
consensus (32)	$^{C}_{A}$AG/GT$^{A}_{G}$AGT	$(^{T}_{C})_{11}$ N$^{C}_{T}$AG/G

and at nucleotide 1113 the cDNA contains an extra A that is not present in the genomic clone. These changes presumably reflect genetic diversity in the chickens used to prepare the cDNA and genomic libraries. It cannot be determined at present whether the changes seen are actually differences in the germ line of the chickens used or whether any or all of the changes could have occurred during the cloning procedures. Similar changes were seen by Venta et al. (18) between YBR and BALB/c mouse strains.

The chicken CA II amino acid sequence (Fig. 4) has 65% homology to the mouse CA II sequence (18). There are 69 base changes that result in silent substitutions, and there are 164 base substitutions that result in amino acid changes. Overall, the nucleotide divergence between the mouse and chicken CA II genes is fairly evenly spread across all seven exons. The active site residues as well as the unique and invariant residues (31) are fairly well conserved. The chicken amino acid sequence is considered in more detail in the Discussion.

Chicken CA II Gene Intron/Exon Organization

The chicken CA II gene is interrupted by six introns. Introns 1 and 2 interrupt the codons for Gly-11 and Val-77, respectively. Introns 3, 4, 5, and 6 fall between the codons for Glu-116/Leu-117, Lys-147/Val-148, Lys-168/Gly-169, and Gln-220/Met-221, respectively (Figs. 2 and 3). The locations of five of the six introns relative to the amino acid sequence are conserved between the chicken and mouse CA II genes. Surprisingly, the location of one of the introns, intron 4, is different in the chicken gene, falling between codons 147 and 148 rather than within codon 143 as in the mouse. The chicken intron 4 location, however, is also observed in the human CA I and CA Z genes (12).

The 5' and 3' boundaries of the six introns of the chicken CA II gene

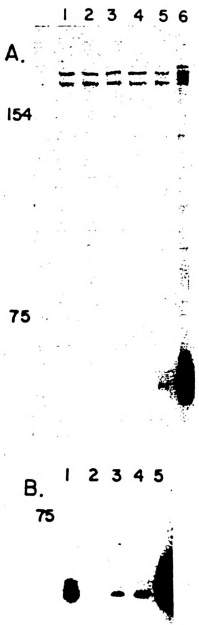


Figure 5. S1 analysis of CA II RNA levels. 50 µg of the various RNA samples were hybridized to the CA II 5' probe, digested, and the products electrophoresed as described in Materials and Methods. RNA samples were isolated from: breast muscle, lane 1; chicken embryo fibroblasts, lane 2; oviduct, lane 3; liver, lane 4; HD3 cells, lane 5; anemic red cell cytoplasm, lane 6. The dried gel was exposed for (A) 17 hr with no intensifying screen and (B) 40 hr with one intensifying screen. The position of co-electrophoresed labeled markers is shown by numbers to the left of the figure.

all fit the consensus donor and acceptor sequences (32) seen for most eucaryotic introns (Table I). This fit to the consensus sequence holds true for the donor and acceptor sites for intron 4, the intron whose position is

altered in the chicken gene relative to the mouse CA II gene.

The sizes of the six chicken CA II introns are given in Fig. 3. In general, most of the intron sizes in the chicken gene are roughly similar to the corresponding mouse CA II introns (18). However, intron 1 of chicken is 0.35 kb which is approximately one-third the size of mouse intron 1 (18). Intron 2, on the other hand, is about 3 kb larger than the corresponding intron in mouse. The remaining four introns of chicken differ from the corresponding introns of mouse by anywhere from 0.1 to 0.8 kb but the difference is not as striking as in the first two introns.

5' and 3' Ends of the Chicken CA II mRNA

The 5' end of the CA II mRNA has been determined by nuclease protection experiments (Fig. 5A, lane 6). A DNA fragment labeled at the RsaI site in exon 1 was hybridized to total chicken reticulocyte cytoplasmic RNA, digested with S1 nuclease and run on a 6% sequencing gel. The major protected fragment is about 58 bases in size which places the RNA start site (cap site) in the CCACG sequence about 39 bp upstream from the ATG initiation codon (Fig. 3). When the protected fragment is run next to a Maxam-Gilbert sequencing ladder of the S1 probe fragment, the major band corresponds to initiation at the A in the CCACG (results not shown).

We have been unable to definitively locate the 3'-end of the chicken CA II mRNA, apparently due to the very A:T rich character of this region. (S1 nuclease treatment of labeled DNA:RNA hybrids shows preferential cleavage at a site around 1460 in Fig. 3 probably due to the 17 contiguous A:T bp in this region as there are no poly(A) signal sequences upstream from this site.) The cDNA clone, pCA-1.2, terminates at 1259 (Fig. 3) apparently due to the oligo(dT) primer (used in the cDNA preparation) binding to the 10 contiguous transcribed A residues from 1260 to 1270. There are four potential poly(A) addition signal sequences (33) in the 3' region of the CA II gene (Fig. 3). We have arbitrarily assumed that termination occurs shortly 3' to the first of these although we can't rule out that any or all of the other three signals may be used. Blots of chicken reticulocyte poly-(A)$^+$ RNA run on denaturing gels and hybridized to the CA II cDNA show a band about 1650 nucleotides in size in agreement with use of one of the first three AATAAA signal sequences to position the polyadenylation site (M. Federspiel and J. Dodgson, unpublished results).

CA II RNA Levels

We have used the S1 nuclease protection assay to measure CA II RNA levels in several cell types. Levels of CA II RNA, as expected, are quite high in total cytoplasmic RNA isolated from anemic hen reticulocytes

(Fig. 5A, lane 6). A lower, but still significant, level of CA II RNA is observed in total cellular RNA isolated from uninduced HD3 cells (Fig. 5A, lane 5). The HD3 cell line is an erythroid progenitor cell line transformed by temperature-sensitive avian erythroblastosis virus, and its uninduced state has been shown to correspond roughly to the CFU-E (erythroid colony forming unit) stage of erythroid development (34-36). Comparison of several different exposure times of the gel shown in Fig. 5 indicates that anemic reticulocytes contain about 50-fold more CA II than uninduced HD3 cells. This difference is comparable to that seen for αA-, αD-, and β-globin RNAs (36; Wynne Lewis and J. Dodgson, unpublished results). The bands observed above 154 bases in Fig. 5 are probably due to undigested probe DNA, both free single strand and renatured double strand. The faint band at about 90 bases in all lanes is due to slight contamination of the 190 bp probe with the SinI/RsaI fragment from the other end of the originally labeled RsaI fragment (see Materials and Methods). A large excess of labeled probe was used in all lanes to insure that the observed signal was proportional to CA II RNA. We can't rule out that there is a small constant level of CA II RNA initiated at upstream sites in all samples tested. Clearly, however, the major CA II initiation site in reticulocytes is at the proposed cap site.

A longer exposure of this protection experiment (Fig. 5B) shows that low but measurable levels of CA II RNA are observed in total RNA isolated from adult liver (lane 4), oviduct (lane 3) and breast muscle (lane 1) whereas CA II RNA levels were not detectable in chicken embryo fibroblast RNA (lane 2) from cells grown in culture. The three adult tissues show about 1/10 the level of CA II RNA as do the HD3 cells. Control experiments which measure red cell contamination in these tissues by S1 protection of an αA-globin gene probe demonstrated that much, if not all, of the CA II RNA in liver may arise from red blood cell contamination (results not shown). Red cell contamination in the oviduct RNA was not detectable in this assay, and breast muscle RNA was not tested.

DISCUSSION

5' and 3' Flanking Sequences of the Chicken CA II Gene

Over 700 bp upstream from the CA II coding region have been sequenced. Putative signal sequences that are common to eucaryotic genes transcribed by RNA polymerase II are found in the chicken gene (Fig. 3). A Goldberg-Hogness (37) or ATA block is located 23 to 30 bp upstream from the cap site.

```
                    -120                    -100
CCAII   CCGCCCCCGAGCGAAGTCTCCCTCCGCCCCCGCC
MCAII   A CT  GTCC  C CC  CA GGT T  T  C T
HCAII   A CT  G CC TC CC  C-------- T  C T

                -80 _____               -60
CCAII   CG-CGC-TCCCCACCC---CTTCCTCC----GGCCGCGGAGAAGGGCAT
MCAII   T---T AGGT    T --GG C     CCTG       CC---  A   G
HCAII   TTCGCTAGGT      GAG CC     CCCG       CC--CC A   C

            -40          _____ -20                  +1
CCAII   GGAGTTCGCGGGAGCCTATAAAAGCCCCTGACAGCCCGCCGAGGCCACG
MCAII   CA   G                    GGACGGT   AC   C -A   A
HCAII    A   G                    GGGCCGGC  GAC  C  A   A
```

Figure 6. Comparison of the 5' flanking region of the chicken, mouse, and human CA II genes. The upper line gives the sequence of the chicken CA II gene for 120 bp upstream of the cap site. The first mRNA nucleotide is numbered as +1 in this figure. The second line gives those nucleotides of the mouse gene that differ from the chicken gene and the third line gives those nucleotides of the human gene which differ from the chicken gene. Dashes indicate a deletion in one line relative to the others. The putative TATA and CCAAT sequences are indicated by a line over the chicken sequence.

Although a sequence that corresponds accurately to the consensus sequence CCAAT (38) cannot be identified, a region (at -74 in Fig. 3) that has limited homology (CCACC) to the CCAAT site can be found. The absence of good consensus CCAAT sequences has also been noted for the adult chicken α-globin genes (39). Fig. 6 compares the -120 to +1 region of the chicken CA II gene to the corresponding regions of the mouse and human genes (18,40). It can be seen that the ATA and putative CCAAT sequences are very similar among the three genes both in terms of actual sequence and approximate spacing relative to the transcription start site. (The actual cap sites of the two mammalian CA II genes have not been determined experimentally, but are estimated from their sequence.) The 5' untranslated regions of the three CA II genes are different in both sequence and length being 39, 59, and 73 nucleotides long in chicken, mouse, and human, respectively.

As shown in Fig. 6, the region from 45 to 22 bp upstream of the mRNA start site (which includes the TATAAA sequence) is over 95% homologous between the chicken and mouse sequences. The homology in this 23 bp region between the CA II genes exceeds that seen between analogous chicken and mammalian globin gene promoter regions (30,39,41,42), suggesting that these

sequences may have an important role in the regulation of CA II gene expression. The position of this sequence block just 5' to and including the ATA region may indicate a role in controlling the initiation of transcription of these genes. The overall homology between the chicken and mammalian CA II genes in the -80 to +1 region is about 60% which is also unusually high in comparison with the same region in globin genes.

McKnight and Kingsbury (43) identified GC-rich sequences in the thymidine kinase gene of Herpes simplex virus that appear to be involved in efficient transcription, possibly by acting as binding sites for the Sp1 transcription factor (44). The core consensus Sp1 transcription factor binding site is 5' GGGCGG 3' or its complement 5' CCGCCC 3'. Two such sequences exist 7 bp 5' to the CCACC box in the chicken CA II gene (Fig. 6). The mouse and human CA II genes contain several such sequences both 5' to CCAAT and between CCAAT and ATA (18,40). While no exact match to the consensus sequence is seen between CCACC and ATA in chicken CA II, this gene does contain a partially homologous GC-rich sequence, GGCCGCGG, in the appropriate position which may function in the same manner. This latter GC-rich region (-69 to -59, Fig. 6) constitutes the one other region of high homology between the chicken and mammalian CA II gene promoters besides the CCACC and -45 to -22 regions discussed above.

An imperfect tandem repeat of 14 to 15 bp that is thought to function as an upstream promoter element and is found in five mammalian and one avian β-globin genes and in several rat pancreatic genes (40,45) is also found in the human and mouse CA II genes as CCNGTCACCTCCGC (40). In the β-globin genes this tandem repeat is 9 to 25 bp upstream from the CCAAT sequence in mammals and 55 bp upstream in chickens. In the human and mouse CA II genes these repeat elements have been found 15 and 22 bp upstream, respectively, from the CCAAT boxes. There are similar repeat elements in the chicken CA II gene: CAAAGCACCTCCCC and AGGACCACCACAGC. However, these elements are located very far upstream of the CA II promoter at -427 and -572 in the chicken CA II gene, and they are not closely linked to each other. Furthermore, these two elements show a rather poor match to the consensus. The function of all of these elements, if any, remains unknown and whether the homologues near the chicken CA II gene are functional or merely coincidental is also unclear.

Amino Acid Sequence of Chicken CA II

There are 30 amino acid residues that are postulated to occur in the active site regions of CA isozymes (1). When chicken CA II amino acid

residues are compared to the analogous active site residues of mammals most of the chicken residues are found to be conserved (Fig. 4). The 15 residues that are invariant in all of the CA I, II and III proteins of all species sequenced to date are also invariant in chicken. At position 90 the residue that is present in most mammals is Ile except for ox which has Val (31). Chicken is similar to ox in that it too has Val at this position. Chicken does differ from mammals at active site residue 202. Mammalian CA IIs that have been sequenced all have Leu at this position whereas in chicken the residue is His. When the overall chicken CA II amino acid sequence is compared to those of mouse and human there is 65% sequence homology to the mouse sequence and 70% homology to the human sequence.

Hewett-Emmett et al. (31) have compiled those amino acid residues that, to date, are invariant among the known examples of a specific CA isozyme but unique to that isozyme. The chicken CA II gene codes for 9 of the 15 previously unique and invariant residues for CA II including both of those (Asn at 66 and Glu at 68) in the active site. The gene possesses only 1 of 18 unique and invariant residues for CA I and 8 of the 39 for CA III. These results, along with the extensive nucleotide sequence homology of the chicken gene to the mouse CA II gene (Fig. 4) confirm its assignment as a CA II isozyme gene.

Exon/Intron Organization

In comparing the exon/intron organization between the chicken and mouse CA II genes, the most surprising result is the different locations of intron 4 relative to the respective coding sequences. In almost all cases studied to date, intron positions within coding regions are conserved between homologous mammalian and avian genes. Even in genes such as the α- and β-actin genes where intron positions are known to vary considerably between different species, the rat and chicken genes retain identical intron locations (46). The change in intron 4 position between the two CA II genes shifts this intron 14 bp toward the 3' end in the chicken relative to the mouse CA II gene (Fig. 7). This "new" intron position has also recently been found in the human CA I and CA Z genes (12). The similarity of the chicken CA II intron position to that of the human CA I and CA Z introns suggests that the chicken CA II gene structure is the more ancient form. Given the similarities observed in the rest of the two CA II genes, it appears that the intron shift occurred via a small number of mutational events as opposed to the possibility that the two genes are the product of two lines of CA II gene evolution that have been separate for longer than

```
          141                                                  145      148
          VAL LEU G                                        LY TYR PHE LEU LYS ILE GLY
MOUSE     GTT TTG G/GT ATT TTT TC. ... TGC CCT GCA G/GC TAT TTT TTG AAG ATT GGA

CHICKEN   GTC GTA GGC  ATC TTC ATG AAG/GTT AGT ... CTA  TAT GTG TTA CAG/GTA GGG
          VAL VAL GLY  ILE PHE MET LYS                                 VAL GLY
          141          145                                             148
```

Figure 7. Comparison of intron 4 location in chicken and mouse CA II genes. The upper half of the figure shows the nucleotide and amino acid sequence of the mouse CA II gene around intron 4 (18). The corresponding region of the chicken CA II gene is shown on the lower half of the figure. Slashes indicate the boundaries of the intron region in both genes.

most mammalian/avian homologues. The mutational shift may have occurred after divergence of birds and mammals. Fig. 7 shows that the intron boundary sequence at the intron 4 acceptor site retains a considerable level of homology to the corresponding sequence of the mouse gene which in mouse is used to code for amino acids 144 to 147. It seems likely that a mutation in the mammalian evolutionary line resulted in a shift to previously cryptic splice donor and acceptor sites. For example, if the ancestral gene to the mouse CA II gene had the chicken arrangement, and the donor site at codon 147 was partially inactivated, a cryptic donor site 14 bp upstream might have become active followed by a similar shift of the acceptor to the present intron 4 acceptor site in mouse. Note that the codon 147 to 148 junction in the mouse gene shows a good fit to the consensus intron acceptor sequence even though it is apparently not used *in vivo*.

Other than the difference in position of intron 4, the organization of the chicken and mouse CA II genes is quite similar. All other introns have identical locations within the coding sequence, both genes have relatively long 3' untranslated regions, and except for a few nucleotide differences in the donor/acceptor sites, intron junctions are fairly well conserved. There are some differences in the size of the analogous introns, but the effect of these changes is likely to be minimal.

CA II RNA Levels

Our preliminary assays of CA II RNA levels present in different chicken tissues and cells (Fig. 5) show that CA II is induced over 100-fold during erythroid differentiation. Comparison of CA II and globin RNA levels in HD3 erythroid progenitor cells versus anemic hen reticulocytes suggests that CA II and the adult globin genes are induced approximately in parallel in agreement with the protein studies of Weil *et al.* (47) for human and mouse

erythroid progenitor cells (but not for mouse erythroleukemia cells). As mentioned previously, chickens differ from most mammals in that they appear to express only CA II in their red cells as opposed to both CA I and CA II (16). The late induction of CA II and globin RNAs in chick red cell maturation is in contrast to the red cell-specific H5 histone gene whose RNA levels in HD3 cells are similar to those in mature reticulocytes (34, Paul Boyer and J. Dodgson, unpublished results). Despite the possible requirement for carbonic anhydrase activity in several, if not all, non-erythroid tissues, the levels of CA II RNA in liver, oviduct and muscle appear to be no more than 0.2% that seen in anemic reticulocytes, and CA II RNA was undetectable in chicken embryo fibroblasts grown in culture. Further, more sensitive, experiments will be required to accurately assess RNA levels in adult (and embryonic) tissue of CA II and other CA isozymes.

ACKNOWLEDGMENTS

We are grateful to Dr. P. J. Curtis for providing the mouse carbonic anhydrase cDNA clone. We also thank Drs. R. E. Tashian, D. Hewett-Emmett, J. C. Montgomery, and P. J. Venta for providing a human CA II exon 1 clone, and for communication of results prior to publication. We thank Mark Federspiel and Wynne Lewis for some of the RNA samples, the cDNA library, and the αA-globin probe used. This research was supported by Grants GM28837 and a Research Career Development Award to J.B.D. from N.I.H. This is Journal Article 12103 from the Michigan Agricultural Experiment Station.

+Present address: Department of Biology, 16-820 Massachusetts Institute of Technology, Cambridge, MA 02139, USA

REFERENCES

1. Tashian, R.E., Hewett-Emmett, D. and Goodman, M. (1983) In Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (eds), Isozymes: Current Topics in Biological and Medical Research. Alan R. Liss, Inc., New York. pp. 79-100.
2. Sapirstein, V.S., Strocchi, P. and Gilbert, J.M. (1984) Ann. N.Y. Acad. Sci. 429, 481-493.
3. Whitney, P.L. and Briggle, T.V. (1982) J. Biol. Chem. 257, 12056-12059.
4. Henry, R.P. and Cameron, J.N. (1983) J. Exp. Biol. 103, 205-223.
5. McKinley, D.N. and Whitney, P.L. (1976) Biochem. Biophys. Acta 445, 780-790.
6. Sanyal, G., Pessah, N.I. and Maren, T.H. (1981) Biochem. Biophys. Acta 657, 128-137.
7. Wistrand, P.J. (1984) Ann. N.Y. Acad. Sci. 429, 195-206.
8. Dodgson, S. J., Forster, R.E., II, Storey, B.T. and Mela, L. (1980) Proc. Natl. Acad. Sci. USA 77, 5562-5566.

C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.
39. Dodgson, J.B. and Engel, J.D. (1983) J. Biol. Chem. 258, 4623-4629.
40. Venta, P.J., Montgomery, J.C., Hewett-Emmett, D. and Tashian, R.E. (1985) Biochem. Biophys. Acta 826, 195-201.
41. Dolan, M., Dodgson, J.B. and Engel, J.D. (1983) J. Biol. Chem. 258, 3983-3990.
42. Engel, J.D., Rusling, D.J., McCune, K.C. and Dodgson, J.B. (1983) Proc. Natl. Acad. Sci. USA 80, 1392-1396.
43. McKnight, S.L. and Kingsbury, R. (1982) Science 217, 316-324.
44. Gidoni, D., Dynan, W.S. and Tjian, R. (1984) Nature 312, 409-413.
45. Dierks, P., Van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. and Weismann, C. (1983) Cell 32, 695-706.
46. Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) Nucleic Acids Res. 10, 3861-3876.
47. Weil, S.C., Walloch, J., Frankel, S.R. and Hirata, R.K. (1984) Ann. N.Y. Acad. Sci. 429, 335-337.

9. Vincent, S.H. and Silverman, D.N. (1982) J. Biol. Chem. 257, 6850-6855.
10. Feldstein, J.B. and Silverman, D.N. (1984) Ann. N.Y. Acad. Sci. 429, 214-215.
11. Montgomery, J.C., Venta, P.J. and Tashian, R.E. (1986) Isozyme Bull. 19, 12.
12. Venta, P.J., Montgomery, J.C. and Tashian, R.E. (1986) In Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (eds), Isozymes: Current Topics in Biological and Medical Research. Alan R. Liss, Inc., New York, in press.
13. Cammer, W.T., Fredman, T., Rose, A.L. and Norton, W.T. (1976) J. Neurochem. 27, 165-171.
14. Benesch, R.N., Barron, N.S. and Mawson, C.A. (1944) Nature 153, 138-139.
15. Linser, P. and Moscona, A.A. (1984) Ann. N.Y. Acad. Sci. 429, 430-446.
16. Tashian, R.E. (1977) In Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (eds) Isozymes: Current Topics in Biological and Medical Research. Alan R. Liss, Inc., New York. pp. 21-62.
17. Curtis, P.J., Withers, E., Demuth, D., Watt, R., Venta, P. J. and Tashian, R.E. (1983) Gene 25, 325-332.
18. Venta, P.J., Montgomery, J.C., Hewett-Emmett, D., Wiebauer, K. and Tashian, R.E. (1985) J. Biol. Chem. 260, 12130-12135.
19. Boyer, S.H., Ostrer, H., Smith, K.D., Young, K.E. and Noyes, A.N. (1984) Ann. N.Y. Acad. Sci. 429, 324-331.
20. Venta, P.J., Montgomery, J.C., Wiebauer, K., Hewett-Emmett, D. and Tashian, R.E. (1984) Ann. N.Y. Acad. Sci. 429, 309-323.
21. Yoshihara, C.M., Federspiel, M. and Dodgson, J. B. (1984) Ann. N.Y. Acad. Sci. 429, 332-334.
22. Yamamoto, M., Yew, N.S., Federspiel, M., Dodgson, J.B., Hayashi, N. and Engel, J.D. (1985) Proc. Natl. Acad. Sci. USA 82, 3702-3706.
23. Dodgson, J. B., Strommer, J. and Engel, J.D. (1979) Cell 17, 879-887.
24. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. Girvitz, S.C., Bacchetti, S., Rainbow, A.J. and Graham, F. L. (1980) Anal. Biochem. 106, 492-496.
26. Southern, E. (1975) J. Mol. Biol. 98, 503-518.
27. Maxam, A.M. and Gilbert W. (1980) Methods Enzymol. 65, 499-560.
28. Smith, D.R. and Calvo, J.M. (1980) Nucleic Acids Res. 8, 2255-2274.
29. Simoncsits, A. and Torok, I. (1982) Nucleic Acids Res. 10, 7959-7964.
30. Dodgson, J.D., Stadt, S.J., Choi, O.R., Dolan, M., Fischer, H.D. and Engel, J.D. (1983) J. Biol. Chem. 259, 12685-12692.
31. Hewett-Emmett, D., Hopkins, P.J. and Tashian, R.E. (1984) Ann. N.Y. Acad. Sci. 429, 338-350.
32. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
33. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
34. Beug, H., Palmieri, S., Freudenstein, C., Zentgraf, H. and Graf, T. (1982) Cell 28, 907-919.
35. Samarut, J. and Gazzolo, L. (1982) Cell 28, 921-929.
36. Weintraub, H., Beug, H., Groudine, M. and Graf, T. (1982) Cell 28, 931-940.
37. Goldberg, M. (1979) Ph.D. thesis, Stanford University, Palo Alto, California.
38. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R.M., O'Connell,

Appendix II

Adult Chicken α-Globin Gene Expression in Transfected QT6 Quail Cells: Evidence for a Negative Regulatory Element in the α^D gene region

This manuscript is currently being revised for resubmission.

Note: My involvement in this project included:

(1) construction of hybrid genes.

(2) optimization of the RNase protection assay to measure mRNA levels in this system.

Adult Chicken α-Globin Gene Expression in Transfected QT6 Quail Cells: Evidence for a Negative Regulatory Element in the αD Gene Region

Wynne Lewis, Jiing-Dwan Lee, and Jerry B. Dodgson*

Departments of Microbiology and Biochemistry

Michigan State University

East Lansing, MI 48824

Running title: CHICKEN α-GLOBIN GENE EXPRESSION IN QT6 CELLS

*Corresponding Author; Tel. No. (517) 353-5024

Abstract

The chicken adult α-globin genes, αA and αD, are closely linked in chromosomal DNA and are coordinately expressed *in vivo* in about a 3:1 ratio, respectively. When subcloned DNAs containing one or the other gene are stably transfected into QT6 quail fibroblasts, the αA-globin gene is expressed at measurable RNA levels, but the αD gene is not. The αA gene expression can be considerably increased by the presence of a linked Rous sarcoma virus long terminal repeat enhancer, but that of the αD gene remains undetectable. Transfection with subclones containing both genes, either in *cis* or in *trans*, leads to considerably reduced αA RNA levels and still no observable αD gene expression. Transfection with deleted subclones suggests that maximal expression levels in this system require the αA-globin gene promoter, as opposed to that of the αD gene, but that such expression is greatly reduced by one or more DNA sequences which lie approximately 2,000 base pairs upstream of the αA promoter, within the body of the αD-globin gene.

Globin gene expression has served as a useful model system in which to study transcriptional regulation in higher eukaryotes. Regulatory elements present in and around mammalian α- and ß-globin genes (1,9-11,14,16,26,34,39,42) and, to a lesser extent, the chicken ß-globin gene (12,13,27) have been identified primarily by transfecting cloned globin gene DNA (in vitro-mutated or wild type) into cultured cells (or transgenic animals; 4,29,38,40) and examining the levels of expression of the exogenous globin gene(s). Transient and stable transfection experiments of this type have provided evidence for two types of cis-acting DNA sequences: constitutive elements consisting of sequences capable of functioning in both nonerythroid and erythroid cells (1,2,10,26,34,39) and erythroid-specific sequences that function solely in erythroid lineage cells (1,9,11-14,27,42). Most such elements appear to act in a positive manner (i.e., deletion or alteration of the element reduces gene expression). The action of several of these is relatively unaffected by orientation or exact location with respect to the gene, and thus they fit the definition of an enhancer (2).

Cis-acting elements of mammalian globin genes that function constitutively have been studied in HeLa (2,10,26,34,39), COS (39), 293 (39), 3T6 (16), and murine erythroleukemia (MEL) (10,14) cells, among others. Erythroid-specific DNA sequences have been identified in mammalian globin genes primarily in MEL cells (1,9-11,14,42) and, to a lesser extent, in transgenic mice (4,29,38,40). Constitutive regulatory elements have not been as carefully delineated in avian globin genes. However, an erythroid-specific enhancer element has been accurately located 3' to the chicken ß-globin gene (12,20,27,36).

The studies reported here were undertaken to investigate constitutive regulatory sequences of the chicken α-globin genes. Avians are unusual in that they have two closely-linked adult α-globin genes, αA and αD, which are very different in sequence, and are therefore presumably the result of an ancient globin-gene duplication (17,18). Both genes are expressed at low levels in primitive avian erythrocytes, but are the only α-globin genes expressed in definitive red cells. Interestingly, both types of red cells contain αA- and αD-globin in about a 3:1 ratio, respectively (7). Our interest in the chicken adult α-globin genes derived from their somewhat unusual promoter structure (17); both promoters lack any obvious CCAAT sequences which are typically found in other active globin genes (19).

This paper describes a system in which the constitutive expression of an exogenous chicken αA-globin gene and in vitro-mutated constructs thereof can be monitored in analogy with studies of human globin gene expression in HeLa and COS cells. The αD-globin gene, however, is not expressed in this or any other tissue culture system we have tested. Interestingly, when a DNA fragment containing both the αA- and αD-globin genes (in their normal arrangement as in chromosomal DNA) is transfected, αA-globin gene expression is substantially reduced and αD expression remains undetectable. This suggests the presence of a negative regulatory element which represses α-globin gene expression in non-erythroid cells.

MATERIALS AND METHODS

Cell culture. The QT6 Japanese quail cell line (33) was grown in Dulbecco's modified Eagle medium (Gibco Laboratories), supplemented with 4% fetal bovine serum, 1% chicken serum, 1% dimethyl sulfoxide (Fisher Scientific), and the antibiotics, penicillin and streptomycin (50 U/ml each).

QT6 transfections. Stable transfections were performed using the calcium phosphate procedure (25). 5×10^5 cells were transfected with test plasmid DNA (20 µg) and pSV2neo (25) plasmid DNA at a 2:1 molar ratio. G418 antibiotic (Gibco Laboratories) was added at 1 mg/ml 48 hr after transfection was begun. G418-resistant colonies were observable within 7-14 days after transfections, and coalescence into mass cultures occurred within the subsequent 1-2 weeks.

RNA isolation. RNA was isolated from transfected cells by the guanidinium isothiocyanate extraction and CsCl gradient centrifugation procedures (31). An average of 150 µg of total RNA per 100 mm plate of cells was isolated, and total RNA was prepared from 2-6 mass cultures for each transfection. RNA levels were assayed 4-6 weeks after transfection. Each RNA preparation was from cells expanded from $>10^3$ resistant colonies.

S1 nuclease protection assays. S1 analysis was done as previously described (23) with S1 digestion performed at 660-1320 U/ml at 25°C for 15 min. The 5' end-labeled DNA fragments used in the assay were the BamHI (-814)-NarI (+112) fragment of pBRα7-1.7 and the 710 base pair (bp) StuI fragment (ca. -630 to +81) of pHRα5-4.3 for αA- and αD-globin RNAs, respectively. Numerical designations of restriction sites are given with the respective transcription initiation or cap-sites designated as +1 (17).

RNase protection assays. RNA probes were prepared by in vitro transcription of pT7-1 (U.S. Biochemical Corp.) clones containing the 342 bp Sau3A (-172 to +170) fragment of the αA-globin gene and the 329 bp MspI (-51 to +278) fragment of the αD-globin gene, respectively. In both cases the plasmid DNA was linearized with HindIII-digestion just downstream of the cloned insert. Labeled RNA transcripts were prepared with T7 RNA polymerase according to the directions of the supplier (U.S. Biochemical Corp.), and hybridized to total RNA samples and RNase treated as described by Melton et al. (32). Protected RNAs were run on 6% sequencing gels as used for S1 analysis.

Plasmid constructions. Subclones of α-globin gene-containing DNAs are as shown in Fig. 1. Plasmids pBRα7-1.7 and pHRα5-4.3 have been described (17). Plasmid pHRα5-2.9 contains the αA-globin gene from the nearest upstream EcoRI site to the downstream HindIII site and pRStα5-1.8 contains the αD-globin gene from the nearest upstream StuI site to the 3' EcoRI site. Plasmid pHV-5.6 contains both adult α-globin genes on a 5.6 kb fragment from an EcoRV site upstream of the αD-globin gene to the HindIII site just downstream of the αA gene. All of these fragments were cloned into the corresponding sites of pBR322 or pAT153 plasmid DNAs (31) except StuI and EcoRV blunt ends for which vector EcoRV and NruI sites were used, respectively. The αA-globin gene subclone pdBRα7-1.7 is identical to pBRα7-1.7 except that its vector (pATdT) contains a deletion from nucleotide 401 (relative to EcoRI) to 1283 in order to make further constructions easier.

The control chicken histone H3.2 gene plasmid DNA, pH3-L3'S, was derived from the EcoRI fragment containing the H3.2 gene of p3dR1 (21) cloned into the pSplink vector (8) followed by the insertion of an EcoRI-permuted Rous sarcoma

virus (RSV) long terminal repeat (LTR) into the EcoRI site 3' to the gene with the LTR in the opposite transcriptional orientation to that of the histone gene.

Subcloning of the RSV LTR into plasmids containing the α-globin genes was performed using permuted forms of the LTR, obtained by either EcoRI or Sau3AI digestion of two tandem LTRs followed by insertion at an EcoRI, BamHI or BglII site as appropriate. Regulatory regions within the 358 bp RSV LTR include both an enhancer and a promoter element (15). In the permuted forms that were employed for subcloning purposes, the enhancer portion of the U3 region remains uninterrupted in order to maintain its potential effect on the α-globin genes, but the normal LTR promoter structure was either disrupted or deleted.

Miscellaneous. Plasmid DNA preparations, Southern blot analysis, DNA labeling, restriction enzyme digestion and bacterial strains used are as described previously (8,17,18).

RESULTS

Characterization of constitutive globin gene expression in stably-transfected QT6 cells. A variety of adult chicken α-globin gene clones were initially transfected into mammalian cell lines (HeLa, COS7 and 293 cells) that have been shown to constitutively express several exogenous mammalian globin genes (39). In no case did we observe measurable expression of either chicken adult α-globin gene in stable or transient transfections (23 and results not shown). Apparently, transcription of chicken α-globin genes, unlike that of chicken vimentin (35) or histone genes (S. Y. Son and J.B.D., unpublished results), requires one or more activities that are lacking in these mammalian cell expression systems.

When the chicken αA-globin gene (on pdBRα7-1.7, Fig. 1) is stably co-transfected (with pSV2neo) into Japanese quail QT6 cells (33), an exogenous αA-globin transcript can clearly be detected (Fig. 2, lane 6). The band observed corresponds to RNA which starts at the normal cap site of the chicken αA-globin gene (17), so it appears that at least the initiation of transcription is normal in these cells. No transcription of endogenous globin genes in QT6 cells (normal or mock-transfected) is observed (Fig. 2, lane 7). (The endogenous quail αA-globin mRNA, if present, would be expected to cross-hybridize to the chicken probes up to approximately the protein initiation site (+37) producing a band about 35-40 nucleotides smaller than the exogenous gene band.)

QT6 cells are transformed fibroblasts (33), and thus it seemed likely that the level of exogenous αA-globin gene transcription would be increased by inserting an enhancer sequence known to function in fibroblasts into the αA-globin gene plasmid DNA. Fig. 2 (lanes 2,3) shows that the presence of a

permuted RSV LTR enhancer 3' to the αA-globin gene increased expression from the αA promoter by about 100-fold. (The band in lane 5 resulting from loading 1/10 the normal assay sample following transfection with the LTR is about 10 times as intense as that in lane 6 resulting from loading all the sample after transfection without an LTR.) In this case the permuted LTR was inserted at an EcoRI site about 900 bp 3' to the αA cap site (Fig. 1). In agreement with the definition of an enhancer sequence, comparable levels of expression were observed in cells transfected with clones containing the 3' LTR in the same (Fig. 2, lane 11) or opposite (Fig. 2, lane 10) orientation with respect to the direction of gene and LTR transcription. The levels of expression from the clones with an LTR 3' to the αA gene were approximately two-fold greater than in clones containing LTRs 5' to the gene (at the BamHI site at -814; Fig. 2, lanes 8 and 9). The two-fold difference is probably not significant and may relate to the fact that different permutations of the LTR sequence were used (Fig. 1). Note that the enhancing effect specifically increases the level of RNA initiated at the αA-globin gene cap site. We have not attempted to measure the level of transcripts initiating within the LTR itself, but the fact that the LTRs used in these constructions have their 5' promoter-flanking sequences rearranged or deleted and the fact that the enhancement is relatively position and orientation independent argue that this effect is not directly related to the promoter function of the LTR sequences.

In contrast to the above results, no expression of the closely-linked αD-globin gene was observed when either the pHRα5-4.3 or the smaller pRStα5-1.8 (Fig. 1) subclone was used to transfect QT6 cells. Furthermore, even when permuted LTRs were inserted 5' or 3' to the αD-globin gene, that gene was inactive in transfected cells (Fig. 3). This was true using either S1 assays

(Fig. 3) or RNase protection assays (not shown), even though easily detectable signals were obtained in both cases using either total anemic hen reticulocyte RNA or undifferentiated HD3 (5) total RNA. The HD3 cell line is a ts AEV-transformed cell line which consists of predominantly erythroblasts but still expresses low levels of globin mRNAs (and other red cell mRNAs such as carbonic anhydrase II, 43) in the undifferentiated state used here (41). Generally, undifferentiated HD3 cells express about 2% of the levels of globin and carbonic anhydrase II mRNAs as do anemic hen reticulocytes (43). Since it appears that the αD-globin RNA assays are of comparable sensitivity to those used successfully for αA-globin RNA, we estimate that constitutive expression of the αA gene in QT6 cells is at least 10-20 times higher than that of the αD gene in the absence of an RSV LTR enhancer and on the order of 1,000 times higher in the presence of a linked LTR enhancer sequence.

α-Globin gene expression in linked clones.

Since the chicken αA- and αD-globin genes are closely linked in chromosomal DNA and coordinately expressed in vivo, the large difference in their expression levels in transfected QT6 cells was unexpected. Therefore, a larger clone, pHV-5.6 (Fig. 1), was prepared that contained both genes linked just as they are in chicken chromosomal DNA (18). Following stable transfection of this clone into QT6 cells both RNase protection (Fig. 4) and S1 analysis (not shown) demonstrated a low level of αA-globin gene expression, approximately 1/10 of that observed from transfection of the pdBRα7-1.7 clone described previously (compare, Fig. 4, lane 7 to lane 3). Furthermore, insertion of the RSV LTR enhancer directly 3' to the αA globin gene in this clone (BglII site at +850 relative to the αA cap site) in either orientation gives only a slight increase, if any, in αA expression (Fig. 4, lanes 8,9). Since

in this case the exact site and configuration of the LTR differed slightly from the 3' LTRs used previously (in pdBRα7-1.7; Fig. 2 and Fig. 4, lane 4), we also tested the effect of LTR insertion at the identical BglII site in the larger αA gene clone pHRα5-2.9 (Fig. 1). Lane 5 of Fig. 4 shows that the RSV LTR induces high levels of αA gene expression in this subclone just as LTR insertion into the EcoRI site of pdBRα7-1.7 did. Thus, the presence of the upstream region of DNA which contains the αD gene (in pHV-5.6) appears to mask the enhancing effects of LTR insertion which are seen in pHRα5-2.9 (Fig. 4, compare lane 5 to lane 9). When the permuted LTR was located in the BglII site upstream of the αA gene in pHV-5.6 (about 1.5 kb 5' to the cap site, Fig. 1), a slight (2-3-fold) but measurable enhancement was observed (Fig. 4, lanes 10 and 11). However, the level of enhancement in this clone containing both the αA and αD genes was considerably less than that exhibited by the same LTR placement in the pHRα5-2.9 clone (Fig. 4, compare lanes 6 and 10). No αD-globin gene expression was observed in transfections with all seven of the pHV-5.6-derived clones (results shown only for two of them, see Fig. 3, lanes 8, 9).

The constitutive expression of the αA-globin gene in QT6 cells was therefore in some way repressed when linked in cis to the αD-globin gene region (in the same fashion that the genes are linked in normal chicken chromosomal DNA). Furthermore, this repression appears, for the most part at least, to be dominant to the enhancing effect of a linked RSV LTR. In order to verify that this was an effect on gene expression and not on co-transfection frequency, Southern blots of DNA isolated from pooled transfectants were performed. Hybridization with a portion of the transfected αA-globin gene DNA shows that co-transfection with pHV-5.6 results in integration of the exogenous chicken

sequences into the genome of the selected QT6 cells just as does co-transfection with pdBRα7-1.7 (results not shown). No consistent difference was observed in the copy number of integrated transfecting DNA between the pdBRα7-1.7L3'S plasmid which lacks the αD-globin gene region and those plasmids such as pHV-5.6L3'S which contain the αD gene. Thus, transfection efficiency does not appear to account for the difference in αA-globin gene transcript level between these two cases. The studies described below also argue against the reduction in αA gene expression being due to altered transfectability.

Co-transfection of αA- and αD-globin genes in **trans**. Since the expression of the αA-globin gene in QT6 cells was reduced when linked to the αD-globin region DNA, we also wished to test the result of co-transfecting the αA gene and αD gene DNAs on separate plasmids. As a control, the αA-globin gene plasmid was also co-transfected with a plasmid containing the chicken H3.2 histone gene (pH3-L3'S, MATERIALS AND METHODS). Of course, all co-transfections also contained the pSV2neo plasmid used for G418 selection of stable transfectants. Expression of the αA-globin gene DNA with a 3' permuted LTR enhancer sequence (pdBRα7-1.7L3'S) was comparable in cells transfected with the αA-globin gene alone or co-transfected with the histone H3.2 gene DNA containing a 3'-LTR (Fig. 5, compare lanes 1, 2 to 11, 12). In similar co-transfections with αD-globin gene-containing plasmids (with or without an LTR on the αD plasmid), the levels of expression of the αA-globin gene dropped to approximately 10% or less of the singly transfected αA gene level (Fig. 5, compare lanes 1, 2 to lanes 3-10). No αD-globin expression was detectable in cotransfections involving these DNAs (not shown).

While at first glance this might suggest that the repression of αA expression exerted by the αD gene region also occurs in trans, this need not be the case. The transfected DNAs (αA gene, αD gene, and pSV2neo) are expected to form a mixed tandem multimer on integration (37) and thus at least some of the transfected αA genes are likely to be linked to a nearby αD gene region DNA. Southern blotting of pooled co-transfectant DNAs is in agreement with this possibility (not shown). Thus, at this point, we cannot say for sure whether the down regulation of αA-globin gene expression functions in trans or not.

αA-globin gene expression in transfections of hybrid genes. A hybrid α-globin gene was constructed which contained the αD-globin gene promoter linked to the cap site and body of the αA-globin gene. This clone was termed pdBRα7-D/A. Since it was constructed by joining deletion clones with an 8 bp XhoI linker at bp -15 to -7 between the TATA sequence and the cap site (+1), a control clone, pdBRα7-X, was constructed which is identical to the pdBRα7-1.7 used previously except that bp -15 to -7 are replaced by the XhoI linker. Fig. 6 (lanes 3, 4) shows that, as expected, the mere insertion of the XhoI linker in pdBRα7-X had no effect on transcript level in transfected QT6 cells. However, when the αD-globin gene promoter was exchanged for the αA gene promoter, expression was essentially completely abolished (Fig. 6, lane 5). This implies a fundamental difference between the two chicken adult α-globin gene promoters, at least in QT6 cells.

When the reciprocal recombinant was constructed, i.e., by attaching the αA promoter to an αD-globin gene body, again no measurable αD gene expression was observed (not shown). This implies that the promoter of the αA gene but not that of the αD gene is active in this system and suggests that sequences

in the neighborhood of the αD gene lower expression from an otherwise active αA gene promoter.

αA-globin gene expression from clones containing deletions of upstream flanking sequences. The results described above indicate that sequences present in the 5.6 kb insert of pHV-5.6 but absent in pdBRα7-1.7 (Fig. 1) reduce αA-globin gene expression in transfected QT6 cells. In order to localize this effect in more detail, two deletion derivatives of pHV-5.6 have been prepared and assayed. Fig. 7A (lane 3) shows that, as described above, the pHV-5.6 clone gives rise to very little αA gene expression in QT6 cells. Insertion of an RSV LTR enhancer 3' to the gene still gives rise only to low levels of αA expression (Fig. 7B, lane 4). Two BamHI fragments (807 and 80 bp) were deleted from the αD gene region of pHV-5.6 to form pHV-5.6dB1 (Fig. 1). The region deleted extends from -251 to +637 with respect to the αD gene (17) and spans the αD promoter region into intron 2. When pHV-5.6dB1 was used to transfect QT6 cells (Fig. 7A and B, lane 5), αA gene expression increased about 20-fold with respect to pHV-5.6. A second deleted subclone, pHV-5.6dB2 (Fig. 1), lacks a 1.56 kb BamHI fragment, with the deleted region extending from +638 to +2201 with respect to the αD cap site. (This deletion ends at position -814 with respect to the αA cap site.) Expression of the αA gene on pHV-5.6dB2 (Fig. 7A and B, lane6) is nearly 100-fold greater than that on pHV-5.6. Thus, deletion of either of two different portions of the αD-globin gene-containing region results in significantly increased expression from the αA-globin gene promoter, even though the deletion endpoints are 2.4 or 0.8 kb, respectively, upstream of the αA gene cap site.

DISCUSSION

We began using the transfected QT6 cell system in order to study constitutive regulatory sequences in and around the chicken globin genes. The behavior of the chicken αA-globin gene in this system is intermediate to that of the human α- and β-globin genes in HeLa cells (16,39). Like that of the human α-globin gene, αA expression in the absence of an enhancer is clearly detectable, but the 100-fold level of αA enhancement by the RSV LTR enhancer is similar to that of the β-globin gene by the SV40 enhancer rather than to the 5-10-fold enhancement of the human α gene. These quantitative differences are not surprising, given the differences in expression systems being employed. In this regard, there must be at least one avian-specific factor in the QT6 cells that is absent in mammalian globin gene expression systems, since transfected chicken αA-globin genes are not expressed in any mammalian cell system we have tested, either by transient or stable transfection assays.

It should be pointed out that all the results presented employed stable transfection procedures. Therefore, the exogenous DNAs were integrated into QT6 quail chromosomal DNA and presumably organized into normal nucleosomal structures. Over a thousand transfectants were used for each assay in order to average out the effects of integration position on exogenous gene expression. Of course, in all cases the exogenous DNA probably exists as a multimeric cointegrate (37). Chromosomal blotting of pooled transfectant DNAs indicated that exogenous globin gene DNA sequences were present in copy numbers of about two to five on average. Copy numbers of the selected pSV2neo plasmid were similar.

The considerable difference in αA and αD-globin gene expression in QT6 cells was unexpected since, although the two genes have evolved separately for

a long period, their two promoters are fairly similar in general appearance (17). It might be argued that the lack of measurable αD gene expression is more "natural" since QT6 cells are fibroblast in nature and don't express their endogenous α-globin genes. However, transfected genes often are expressed where their endogenous counterparts are not (1,14,39). The αD-globin gene is also very poorly expressed in transient QT6 transfections (our unpublished results) which, along with the results described above, suggests that this trait is a particular aspect of the αD gene (perhaps its promoter, Fig. 6). Whether the transfected αD genes are blocked by a mechanism that acts normally in vivo to prevent their expression in non-erythroid cells and, if so, how this mechanism is relieved during erythropoiesis remain to be determined.

It was also surprising to find that upstream DNA sequences in the neighborhood of the αD-globin gene repress expression of αA-globin RNA both in the presence (Fig. 4) and absence (Fig. 7) of an LTR enhancer. Negative regulatory mechanisms are not uncommon in eucaryotes (3,6,24,28,30). In some cases these effects are local and may be maintained by mechanisms similar to better known repressor mechanisms in procaryotes (3,24,28,30), but at least in some cases the negative regulation is exerted by sequences several kb away from their point of action (6). The term "silencer" (6) has been coined for such a cis-acting sequence in analogy with enhancer. Our data are consistent with the presence of such a silencer sequence in the αD gene region which acts on the αA gene promoter. (We can't tell if it acts on the αD promoter as well, since the αD promoter appears to be inactive in QT6 cells even when joined to an αA-globin gene in the absence of the putative silencer region.) However, it is not yet clear whether the sequences that negatively regulate αA

gene expression in QT6 cells fit all the classical criteria of enhancer/silencer sequences or not. In some aspects the effects observed above are similar to those observed by Choi and Engel in the chicken β-globin gene cluster (13). These investigators identified a locus in the β-globin gene promoter which appears to block expression of the linked ε-globin gene, perhaps by competition for regulatory DNA binding proteins. However, in this case the β gene was actively transcribed when the ε gene was being repressed whereas in our system the αD gene is not expressed. Furthermore, in vivo, β- and ε-globin gene expression are mutually exclusive whereas the αD- and αA-globin genes are expressed together.

Our results are consistent with a model in which the αD-globin gene promoter has a stringent requirement for one or more erythroid-specific factors such as the one described by Evans et al. (22), but with the αA gene promoter being somewhat leaky in its tissue specificity. Additionally, there appears to be one or more sequences within the αD gene region which repress αA gene activity, perhaps by competing for some constitutive regulatory binding protein or enhancer interaction factor. In this regard it should be noted that none of the constructs described contains the putative erythroid-specific enhancer sequence 3' to the αA gene (cited in 22). Whether the negative regulatory sequence(s) that we have suggested exists in or near the αD gene actually functions in the coordinate control of α-globin levels in vivo is unknown. We are beginning to approach this question by examining these constructs in the two available avian erythroid expression systems (12,27).

ACKNOWLEDGMENTS

We thank Paul Bates for advice and assistance in the initial phases of this work and Sue Conrad for comments on the manuscript. This work was supported by NIH Grant GM 28837. This is Journal Article from the Michigan Agricultural Experiment Station.

LITERATURE CITED

1. **Antoniou, M., E. deBoer, G. Habets, and F. Grosveld.** 1988. The human β-globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. EMBO J. 7:377-384.
2. **Banerji, J., S. Rusconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27:299-308.
3. **Barberis, A., S.-F. Giulio, and M. Busslinger.** 1987. Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. Cell 50:347-359.
4. **Behringer, R. R., R. E. Hammer, R. L. Brinster, R. D. Palmiter, and T. M. Townes.** 1987. Two 3' sequences direct adult erythroid-specific expression of human β-globin genes in transgenic mice. Proc. Natl. Acad. Sci. USA 84:7056-7060.
5. **Beug, H., S. Palmieri, C. Freudenstein, H. Zentgraf, and T. Graf.** 1982. Hormone-dependent terminal differentiation in vitro of chicken erythroleukemia cells transformed by ts mutants of avian erythroblastosis virus. Cell 28:907-919.
6. **Brand, A. H., L. Breeden, J. Abraham, R. Sternglanz, and K. Nasmyth.** 1985. Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. Cell 41:41-48.
7. **Brown, J. L., and V. M. Ingram.** 1974. Structural studies on chicken embryonic hemoglobins. J. Biol. Chem. 249:3960-3972.
8. **Brush, D., J. B. Dodgson, O.-R. Choi, P. W. Stevens, and J. D. Engel.** 1985. Replacement variant histone genes contain intervening sequences. Mol. Cell. Biol. 5:1307-1317.

9. **Chao, M. V., P. Mellon, P. Charnay, T. Maniatis, and R. Axel.** 1983. The regulated expression of ß-globin genes introduced into mouse erythroleukemia cells. Cell **32**:483-493.

10. **Charnay, P., P. Mellon, and T. Maniatis.** 1985. Linker scanning mutagenesis of the 5'-flanking region of the mouse ß-major-globin gene: sequence requirements for transcription in erythroid and nonerythroid cells. Mol. Cell. Biol. **5**:1498-1511.

11. **Charnay, P., R. Treisman, P. Mellon, M. Chao, R. Axel, and T. Maniatis.** 1984. Differences in human alpha- and beta-globin gene expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell **38**:251-263.

12. **Choi, O.-R., and J. D. Engel.** 1986. A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult ß-globin gene. Nature (London) **323**:731-734.

13. **Choi, O.-R.B., and J. D. Engel.** 1988. Developmental regulation of ß-globin gene switching. Cell **55**:17-26.

14. **Cowie, A., and R. M. Myers.** 1988. DNA sequences involved in transcriptional regulation of the mouse ß-globin promoter in murine erythroleukemia cells. Mol. Cell. Biol. **8**:3122-3128.

15. **Cullen, B. R., K. Raymond, and G. Ju.** 1985. Functional analysis of the transcription control region located within the avian retroviral long terminal repeat. Mol. Cell. Biol. **5**:438-447.

16. **Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit ß-globin gene in mouse 3T6 cells. Cell **32**:695-706.

17. **Dodgson, J. B., and J. D. Engel.** 1983. The nucleotide sequence of the adult chicken alpha-globin genes. J. Biol. Chem. **258**:4623-4629.

18. **Dodgson, J. B., K. C. McCune, D. J. Rusling, A. Krust, and J. D. Engel.** 1981. Adult chicken α-globin genes, αA and αD: no anemic shock α-globin exists in domestic chickens. Proc. Natl. Acad. Sci. USA **78**:5998-6002.

19. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. De Riel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human ß-globin gene family. Cell **21**:653-668.

20. **Emerson, B. M., J. M. Nickol, P. D. Jackson, and G. Felsenfeld.** 1987. Analysis of the tissue-specific enhancer at the 3' end of the chicken adult ß-globin gene. Proc. Natl. Acad. Sci. USA **84**:4786-4790.

21. **Engel, J. D., B. J. Sugarman, and J. B. Dodgson.** 1982. A chicken histone H3 gene contains intervening sequences. Nature (London) **297**:434-436.

22. **Evans, T., M. Reitman, and G. Felsenfeld.** 1988. An erythroid-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc. Natl. Acad. Sci. USA **85**:5976-5980.

23. **Fischer, H. D., J. B. Dodgson, S. H. Hughes, and J. D. Engel.** 1984. An unusual 5' splice sequence is efficiently utilized in vivo. Proc. Natl. Acad. Sci. USA **81**:2733-2737.

24. **Goodbourn, S., H. Burstein, and T. Maniatis.** 1986. The human ß-interferon gene enhancer is under negative control. Cell **45**:601-610.

25. **Gorman, C.** 1985. High efficiency gene transfer into mammalian cells, p. 143-190. In D. M. Glover (ed.), DNA cloning: a practical approach, vol. 2. IRL Press Ltd., Oxford, U.K.

26. **Grosveld, G. C., E. deBoer, C. K. Shewmaker, and R. A. Flavell.** 1982. DNA sequences necessary for the transcription of the rabbit β-globin gene in vivo. Nature (London) 295:120-126.

27. **Hesse, J. E., J. M. Nickol, M. R. Lieber, and G. Felsenfeld.** 1986. Regulated gene expression in transfected primary chicken erythrocytes. Proc. Natl. Acad. Sci. USA 83:4312-4316.

28. **Johnson, A. D., and I. Herskowitz.** 1985. A repressor (MATα2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell 42:237-247.

29. **Kolias, G., N. Wrighton, J. Hurst, and F. Grosveld.** 1986. Regulated expression of human Aγ-, β-, and hybrid γβ-globin genes in transgenic mice: manipulation of the developmental expression patterns. Cell 46:89-94.

30. **Langer, S. J., and M. Ostrowski.** 1988. Negative regulation of transcription in vitro by a glucocorticoid response element is mediated by a trans-acting factor. Mol. Cell. Biol. 8:3872-3881.

31. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

32. **Melton, D. A., P. A. Krieg, M. R., Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. 12:7035-7056.

33. **Moscovici, C., M. G. Moscovici, H. Jiminez, M.M.C. Lai, M. J. Hayman, and P. K. Vogt.** 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell 11:95-103.

34. **Myers, R. M., K. Tilly, and T. Maniatis.** 1986. Fine structure genetic analysis of a β-globin promoter. Science 232:613-618.

35. **Ngai, J., V. C. Bond, B. J. Wold, and E. Lazarides.** 1987. Expression of transfected vimentin genes in differentiating murine erythroleukemia cells reveals divergent cis-acting regulation of avian and mammalian vimentin sequences. Mol. Cell. Biol. 7:3955-3970.

36. **Nickol, J. M., and G. Felsenfeld.** 1988. Bidirectional control of the chicken β- and ε-globin genes by a shared enhancer. Proc. Natl. Acad. Sci. USA 85:2548-2552.

37. **Robins, D. M., S. Ripley, A. S. Henderson, and R. Axel.** 1981. Transforming DNA integrates into the host chromosome. Cell 23:29-39.

38. **Townes, T. M., J. B. Lingrel, H. Y. Chen, R. L. Brinster, and R. D. Palmiter.** 1985. Erythroid-specific expression of human β-globin genes in transgenic mice. EMBO J. 4:1715-1723.

39. **Treisman, R., M. R. Green, and T. Maniatis.** 1983. cis and trans activation of globin gene transcription in transient assays. Proc. Natl. Acad. Sci. USA 80:7428-7432.

40. **Trudel, M., J. Magram, L. Bruckner, and F. Costantini.** 1987. Upstream Gγ-globin and downstream β-globin sequences required for stage-specific expression in transgenic mice. Mol. Cell. Biol. 7:4024-4029.

41. **Weintraub, H., H. Beug, M. Groudine, and T. Graf.** 1982. Temperature-sensitive changes in the structure of globin chromatin in lines of red cell precursors transformed by tsAEV. Cell 28:931-940.

42. **Wright, S., A. Rosenthal, R. Flavell, and F. Grosfeld.** 1984. DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell 38:265-273.

43. **Yoshihara, C. M., J.-D. Lee, and J. B. Dodgson.** 1987. The chicken carbonic anhydrase II gene: evidence for a recent shift in intron position. Nucl. Acids Res. 15:753-770.

FIG. 1. Restriction map of clones used in transfections. Line A shows the relative positions of the αA- and the αD-globin genes. Filled boxes represent protein coding sequences and open boxes represent intron and 3' and 5' untranslated sequences. Arrows denote the direction of transcription. Line B shows the restriction map of the chromosomal DNA region containing the chicken adult α-globin genes. The EcoRI linker site arose only in a single λ recombinant during library construction but is shown since it is used in some subclones listed below. Line C and below designate regions of chromosomal DNA present in the various subclones shown. Triangles above the lines designate the position of inserted RSV LTR enhancers. Arrows above the LTR indicate the direction of RSV transcription in the provirus from which the LTR arose. R above the LTR indicates an LTR circularly permuted at its unique EcoRI site: such an LTR contains RSV sequences from -51 to +101 joined to -234 to -52 (15) in the direction of the arrow. S indicates an LTR permuted at the Sau3A site which contains sequences -109 to +101 joined to -234 to -110 in the direction of the arrow. M/A indicates that portion of the LTR from MspI to AccI (+46 to +101 joined to -234 to -109, in the direction of the arrow) which contains the enhancer element (15). Deletions of sequences are indicated by joining the undeleted flanking regions with lines angled above the level of the subcloned DNA. The synthetic XhoI linker used in two constructions is as indicated (at a deletion site in pBRα7-D/A and the equivalent replacement in pBRα7-X).

FIG. 2. Comparison of the expression levels of the αA-globin gene with and without an RSV LTR in stably transfected QT6 cells. The RSV LTRs are positioned in 3'- and 5'-regions flanking the gene in the same (S) and opposite (O) transcriptional orientations with respect to the gene: RNA samples prepared from pooled stable transfectants were assayed by S1 analysis with a 5' end-labeled αA-globin gene probe (-798 to +120, see MATERIALS AND METHODS). S1 nuclease levels used were 660 U/ml in lanes 1-6 and 1320 U/ml in lanes 7-11. (Equivalent results were obtained with either S1 level.) RNAs used were: lane 1, total anemic chicken reticulocyte RNA; and total RNA from cells transfected with: lane 2, pdBRα7-1.7L3'S; lane 3, pdBRα7-1.7L3'S but with only 1/10 of the sample loaded; and lane 4, pdBRα7-1.7. Lanes 5 and 6 are 10-fold longer exposures of lanes 3 and 4, respectively. Other RNAs used were: lane 7, total QT6 RNA; and RNA from cells transfected with: lane 8, pdBRα7-1.7L5'O; lane 9, pdBRα7-1.7L5'S; lane 10, pdBRα7-1.7L3'O; and lane 11, pdBRα7-1.7L3'S. 50 μg of RNA were used for all assays except reticulocyte RNA where 10 μg were used.

FIG. 3. Absence of αD-globin gene transcripts in transfected QT6 cells. RNA was prepared from stably-transfected QT6 cells and used for S1 analysis with an end labeled DNA fragment extending from -630 to +81 of the αD-globin gene (MATERIALS AND METHODS). The film was overexposed to detect any possible transcript in lanes 3-8. RNAs used were prepared from undifferentiated HD3 cells, lane 1; QT6 cells, lane 2; and QT6 cells transfected with: pHRα5-4.3, lane 3; pHRα5-4.3L3'S, lane 4; pHRα5-4.3L3'O, lane 5; pRStα5-1.8L5'S, lane 6; pRStα5-1.8L5'O, lane 7; pHV-5.6, lane 8; and pHV-5.6L3'S, lane 9. (RNA from cells transfected with pRStα5-1.8 lacking an LTR was not used in this experiment but also showed no measurable expression of the αD gene in other experiments.)

FIG. 4. Reduced αA-globin expression levels in QT6 cells transfected with DNAs containing both αA and αD genes. Transfected plasmid DNAs included pdBRα7-1.7 containing only the αA gene and flanking regions, and pHV-5.6 containing both the αA and αD genes, with and without the RSV LTR (L) in the same (S) and opposite (O) transcriptional orientations with respect to the gene(s), and at the 5'- and 3'- ends of the gene(s). In the case of the larger clone, an LTR was also placed in the intergenic region approximately midway between the two genes (M). RNA samples (50 μg) were hybridized with an antisense RNA probe transcribed from the first exon region of the αA gene, followed by RNase T1 and A digestions (MATERIALS AND METHODS). RNAs assayed were QT6 cell RNA, lane 1; anemic hen reticulocyte RNA, lane 2; and RNA from QT6 cells transfected with: pdBRα7-1.7, lane 3; pdBRα7-1.7L3'O, lane 4; pHRα5-2.9L3'O, lane 5; pHRα5-2.9L5'S, lane 6; pHV-5.6, lane 7; pHV-5.6L3'S, lane 8; pHV-5.6L3'O, lane 9; pHV-5.6LMS, lane 10; pHV-5.6LMO, lane 11; pHV-5.6L5'S, lane 12; and pHV-5.6L5'O, lane 13. Bands in lanes 2, 4, 5 and 6 were overexposed to see those present in lanes 7 to 13.

FIG. 5. S1 nuclease analysis of αA-globin gene transcript levels in cells stably co-transfected with αA-globin and αD-globin DNAs. The plasmid pdBRα7-1.7L3'S was transfected in the presence and absence of αD-globin gene DNAs; pHRα5-4.3, pHRα5-4.3L3'S, pRStα5-1.8L3'S, and a chicken H3.2 histone gene plasmid containing a 3'-LTR, pH3-L3'S. RNA samples (50 μg) were assayed for αA transcript levels as in Fig. 2 with 660 U/ml (even-numbered lanes) or 1320 U/ml (odd-numbered lanes) of S1 nuclease. RNAs were from QT6 cells with: pdBRα7-1.7L3'S, transfected singly, lanes 1 and 2 and pdBRα7-1.7L3'S cotransfected with pHRα5-4.3, lanes 3 and 4; with pHRα5-4.3L3'S, lanes 5 and 6; with pRStα5-1.8, lanes 7 and 8; with pRStα5-1.8L3'S, lanes 9 and 10; and with pH3-L3'S, lanes 11 and 12.

FIG. 6. S1 analysis of hybrid globin gene expression in stably transfected QT6 cells. QT6 cells were stably transfected with the clones pdBRα7-1.7; pdBRα7-X, an αA-globin plasmid with the 8 bp sequence at bp -15 to -7 replaced by an XhoI linker; and pdBRα7-D/A, a hybrid plasmid clone consisting of the αD-globin promoter and the αA-globin gene body, joined at an XhoI linker at bp -15 to -7. S1 (1320 U/ml) analysis was as described in Fig. 2. RNA was extracted from anemic hen reticulocytes, lane 1; QT6 cells, lane 2; or from QT6 cells transfected with pdBRα7-1.7, lane 3; pdBRα7-X, lane 4; and pdBRα7-D/A, lane 5. 50 μg of each RNA was used except for total reticulocyte RNA of which 10 μg were used.

FIG. 7. αA-globin gene expression in QT6 cells transfected with deleted derivatives of pHV-5.6. A. RNase protection analysis as in Fig. 4 of RNA (50 μg) isolated from undifferentiated HD3 cells, lane 1; QT6 cells, lane 2; and QT6 cells stably transfected with: pHV-5.6, lane 3; pHV-5.6L3'S, lane 4; pHV-5.6dB1, lane 5; or pHV-5.6dB2, lane 6. B. Same as A except exposed for 5 times longer. pHV-5.6dB1 deletes two BamHI fragments (807 bp and 80 bp) 5' to and within the αD gene; pHV-5.6dB2 deletes a BamHI fragment of 1.56 kb from intron 2 of αD to -814 relative to the αA gene (see Fig. 1).

Figure 2

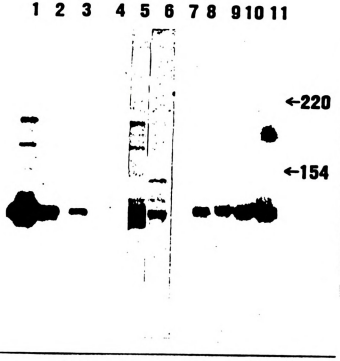


Fig. 3

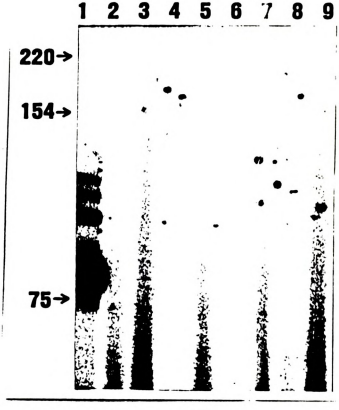


Fig 4

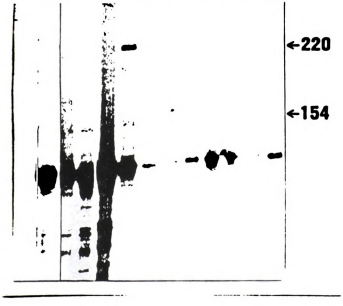


Fig 5

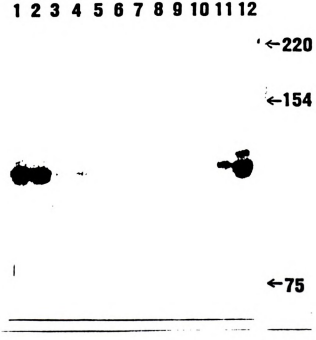


Fig 6

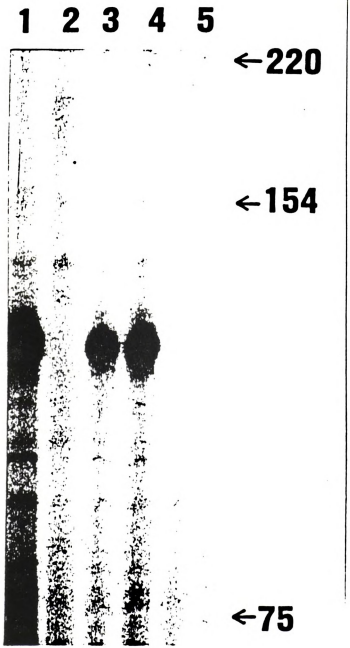


Fig 7