THE ISOLATION AND CHARACTERIZATION OF ERYTHROID-EXPRESSED CLONES FROM A CHICKEN RETICULOCYTE cDNA LIBRARY

BY

Mark J. Federspiel

A DISSERTATION

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry

1987

ABSTRACT

THE ISOLATION AND CHARACTERIZATION OF
ERYTHROID-EXPRESSED CLONES FROM A
CHICKEN RETICULOCYTE cDNA LIBRARY

BY

Mark J. Federspiel

In an effort to expand the set of isolated erythroid-expressed genes, erythroid-specific cDNA clones from a chicken reticulocyte cDNA library were isolated and several identified. Further study of the expression of this set of genes might make possible the identification of new tissue-specific control elements and/or other important regulatory DNA sequences. A chicken reticulocyte cDNA library was constructed in λgt10 after optimizing the efficiency of each enzymatic step in the cDNA cloning protocol. The library was screened with a cDNA probe enriched for non-globin, reticulocyte-expressed messages. A hybridization-selection procedure was employed to subtract globin and erythroblast sequences from ^{32}P-labelled single strand reticulocyte cDNA. The isolated clones were characterized by restriction enzyme mapping, Northern and Southern blot analysis, and nucleotide sequencing. A carbonic anhydrase II clone, which is maximally expressed in red blood cells, was identified from restriction maps and comparison of its nucleotide sequence, to that of a previously isolated clone. Two clones, coding for ferritin heavy chain and ubiquitin, were identified by comparing putative amino acid sequences, deduced from the clones' nucleotide sequences, to a protein database. Ferritin heavy chain is an iron storage protein found in blood, liver, spleen and bone

marrow. Ubiquitin is a highly conserved 76 amino acid protein found in all tissues studied, and it has been implicated in a variety of cellular functions. A Charon 4A chicken genomic library was screened with the ubiquitin clone. The genomic organization of the two ubiquitin loci was deduced by analyzing the isolated Charon 4A clones by restriction enzyme analysis, ubiquitin fragment hybridization, and nucleotide sequencing. Our results are compared to recently reported studies of the chicken ubiquitin loci.

To Mom and Dad

ACKNOWLEDGEMENTS

I foremost thank Dr. Jerry Dodgson for his patient scientific advice and long financial support.

I thank Dr. Edward Fritsch for the scientific experiences of the Cold Spring Harbor Cloning Course and his lab.

Finally, I also acknowledge the scientific influences and support provided by Dr. Hsing Jien Kung and Dr. Wynne Lewis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The basic structure of a typical eucaryotic gene has been elucidated by analyzing genes with a wide variety of differential and developmental fates. Many of these genes have been isolated by taking advantage of the fact that they are transcribed to high levels during some stage of development (1). For example, the chicken globin genes were initially isolated using a cDNA probe from reticulocytes, a late cell type in the erythroid differentiation pathway in which globin sequences constitute 90% of the total cellular poly(A)$^+$ RNA (2-4). The study of various gene families has elucidated precise programs of specific expression of certain members of such families during tissue differentiation and/or animal development. Chicken globin gene expression begins specifically at the erythroblast stage during erythropoiesis, and different sets of globin genes, embryonic and adult, are expressed at different times in chicken development (5-7). DNA sequence elements which appear to regulate transcription of adjacent genes in *cis* have been identified through the analysis of the expression of mutant genes, either those identified *in vivo* or those constructed *in vitro* (8-13).

Although the expression of several eucaryotic genes has been examined in detail, the mechanisms by which a differentiating cell concomitantly regulates as many as a hundred or more genes remain obscure. Generally, only a few of the genes whose expression is

regulated in any given pathway of cellular differentiation have been isolated and studied in detail. The aim of this thesis research was the isolation and possible identification of genes which would expand the pool of available genes known to be expressed in chicken erythroid cells. Several such genes have been identified including erythroid-specific and "housekeeping" genes. Housekeeping genes code for proteins constitutively expressed in many or all tissues and include common metabolic enzymes and structural proteins. We primarily hoped to expand the set of isolated genes which are specifically expressed late in erythropoiesis as opposed to housekeeping genes. Further study of the expression of this set of genes might make possible the identification of new tissue-specific control elements and/or other important regulatory DNA sequences.

Chicken erythropoiesis offers many advantages as an experimental system in which to study the coordinate expression of tissue-specific genes: 1.) The stages of avian erythroid differentiation and development have been extensively studied. 2.) Erythroid cell lines generated by transformation with avian erythroblastosis virus (AEV) provide a system in which to conduct in vitro experiments. These cell lines also enable certain stage-specific cell populations to be isolated in quantities suitable for biochemical analysis. 3.) Chicken reticulocytes have 1/10 the mRNA sequence complexity of comparable mammalian cells (14). 4.) Many proteins found in red cells have been isolated and examined in some detail which may assist in the identification of the function of the clones we intend to isolate.

Normal avian erythropoiesis proceeds through a sequence of precursor cell types that are morphologically, biochemically,

antigenically and kinetically distinguishable (6, 15-24) (Figure 1). The earliest identifiable erythroid-specific precursor is the colony forming unit-marrow (CFU-M). When CFU-M cells were injected into the bone marrow of irradiated chickens, macroscopic erythrocytic clones (of cells) developed that exhibited the self-renewal properties characteristic of immature stem cells (17, 18). Two distinct classes of later erythroid progenitors were identified according to their respective outcomes when placed in an in vitro culture system dependent on anemic chicken serum. These are the burst forming unit-erythroid (BFU-E) and colony forming unit-erythroid (CFU-E) (19). Bursts are large aggregates of 3-20 clusters with each cluster containing 8-60 erythroid cells while colonies are a single compact cluster of 8-150 cells. Both cell types can also be distinguished by their differences in antigen expression and growth factor sensitivity. Im antigen expression (see Figure 1), a characteristic antigen of immature erythroid cells only becomes fully detectable by the CFU-E stage (18). Erythroid potentiating activity (EPA), produced in T cells, has been shown to stimulate the growth and cell division of early mouse erythroid precursor cells (25). Both BFU-E and CFU-E cells are stimulated by EPA. However, only CFU-E cells show an absolute requirement for erythropoietin, a growth factor produced in kidney cells, for continued erythroid differentiation to erythroblasts (21, 26). Erythroblasts are a late erythroid cell type characterized by the onset of hemoglobin production, and they are the last erythroid cell stage capable of cell division. Morphological differences are primarily used to distinguish the final erythroid stages - early, middle and late polychromatic erythrocytes, reticulocytes, and mature

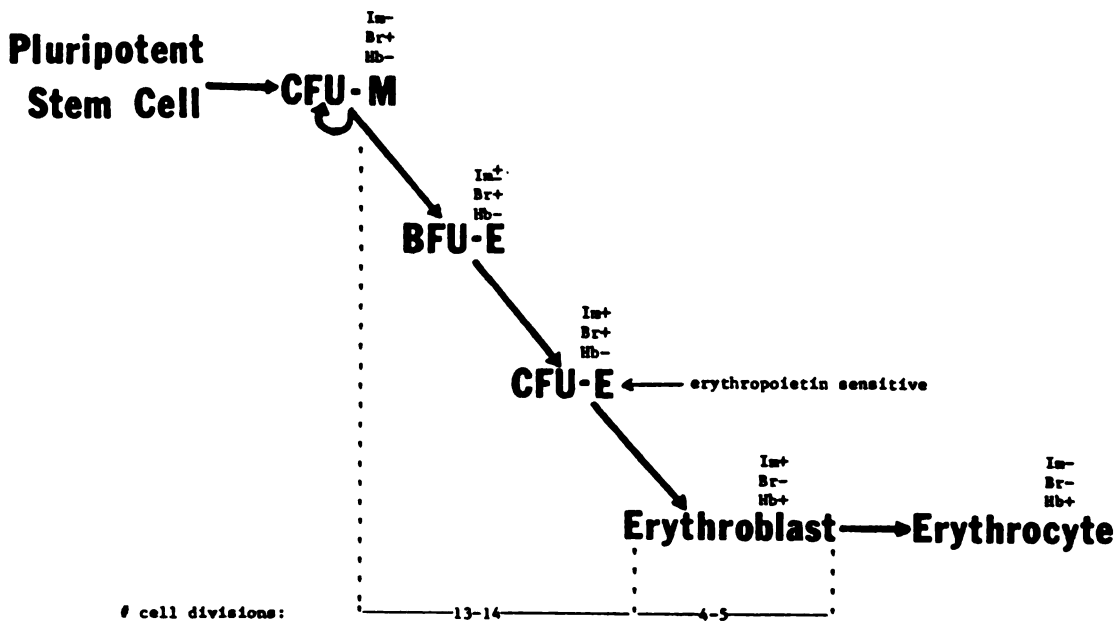

Figure 1. A schematic representation of the chicken erythroid differentiation pathway. The figure information is primarily from Gazzolo et al. (20). Abbreviations: colony forming unit-marrow (CFU-M), burst forming unit-erythroid (BFU-E), colony forming unit-erythroid, Im antigen (Im), brain antigen (Br), and hemoglobin (Hb).

erythrocytes. In these non-dividing stages, the overall cell size gradually decreases. The cell nucleus also compacts in part due to the replacement of histone H1 with the erythroid-specific histone H5 (27, 28). Avian erythroid cells retain their nuclei, but the mature erythrocyte has no detectable transcriptional or translational activities. Mathematical analysis of kinetic growth experiments estimates 17-19 cell generations are required for erythrocyte differentiation from the CFU-M stage, with the last 4-5 cell divisions occurring at the erythroblast stage (20).

Approximately three generations of differentiating erythroid cells arise during avian embryonic development. As development progresses, the site of erythroid cell production moves from the blood islands, to the yolk sac, and finally to the adult erythropoietic organ, the bone marrow (6, 29). At approximately 120 hr of embryo development, a characteristic switch from the production of primitive erythroid cells to definitive erythroid cells occurs. This switch can be observed morphologically and can also be observed by analyzing globin gene expression (5, 6) (Table 1). The embryonic globin genes π, ρ and ε, are only expressed in the primitive or first erythroid generation. The adult globin gene β, is only expressed in the second and successive generation of definitive erythroid cells. Two alpha genes, αA and αD, are expressed in all erythroid generations but their level of expression in primitive cells is only a fraction of that in definitive cells.

Several labs have developed continuous erythroid tissue culture cell lines from chicken bone marrow cells transformed with avian erythroblastosis virus (AEV) (19, 20). The target cells of AEV

Table 1. Chicken globin genes expressed in primitive and definitive erythroid cells[a].

PRIMITIVE (Embryonic ; 5 Days)				DEFINITIVE (Adult)			
α -type	%[b]	β -type	%[b]	α -type	%[b]	β -type	%[b]
π	70	ρ	70	α A	70	β	100
α A	20	ε	30	α D	30		
α D	10						

[a] from Brown and Ingram (5).

[b] percentage of total globin type.

Table 2. Distribution of erythroid cell types of various temperature-sensitive avian erythroblastosis (tsAEV) strains and cell lines grown at 36° and 42° [a,b].

	Percentage of Cell Types after Growth at 36°C or 42°C							
	36°C				42°C			
Cells	Ebl	ER	LR	Ery	Ebl	ER	LR	Ery
wt clone C2	100	0	0	0	99	0	0.5	0.5
ts34 clone 2	99	1	0	0	93	6	1	0
ts34 clone 4	97	3	0	0	19	23	34	24
ts34 clone 8	98	1.5	0	0	42	33	16	9
ts167 clone 1	100	0	0	0	25	13	37	25
ts167 clone 3	100	0	0	0	8	29	32	31
ts167 clone 6	99	1	0	0	1	3	31	65
ts167 clone 8	98	2	0	0	0	0	3	97
HD3 subline 32	99	1	0	0	32	38	22	8
HD3 subline 41	99	1	0	0	21	32	29	18

[a] from Beug et al. (23).

[b] 42° cells were grown in differentiation medium containg anemic chicken serum.

infection are the BFU-E, but the cells continue to differentiate to approximately the CFU-E stage (21) (Figure 2). The AEV transforming genes, v-erb-A and v-erb-B, immortalize the cells and arrest differentiation. AEV temperature-sensitive mutants (tsAEV) have also been isolated and, in cells transformed with these viruses, the differentiation block can be released at the non-permissive temperature (42°) in the presence of anemic chicken serum (23, 24). As seen in Table 2, tsAEV-infected erythroid cells are nearly 100% erythroblasts when grown at the permissive temperature (36°) (23). In this case, the term erythroblast is used to indicate all precursor cells prior to the hemoglobin-producing stage. Different tsAEV viruses show variable abilities in the extent to which the differentiation block is released at 42°. Differentiation following the temperature shift has an absolute requirement for anemic chicken serum, probably as a source of erythropoietin. Three highly enriched cell subpopulations, erythroblasts, reticulocytes and erythrocytes, can be isolated by fractionation of temperature-induced tsAEV cells by Percoll density gradient centrifugation (30). Adkins et al. compared the pattern of protein synthesis within these three subpopulations by two dimensional gel electrophoresis of ^{35}S-methionine labelled cell lysates (31). Twenty-eight major protein synthesis changes were observed as differentiation proceeded. Nine proteins increased in abundance in erythrocytes relative to erythroblasts, fifteen decreased, and four were found de novo in erythrocytes.

Cell populations from several precursor cell stages of chicken red cell differentiation can be isolated. The in vitro AEV-transformed erythroid cell system offers a relatively homogeneous erythroblast (all

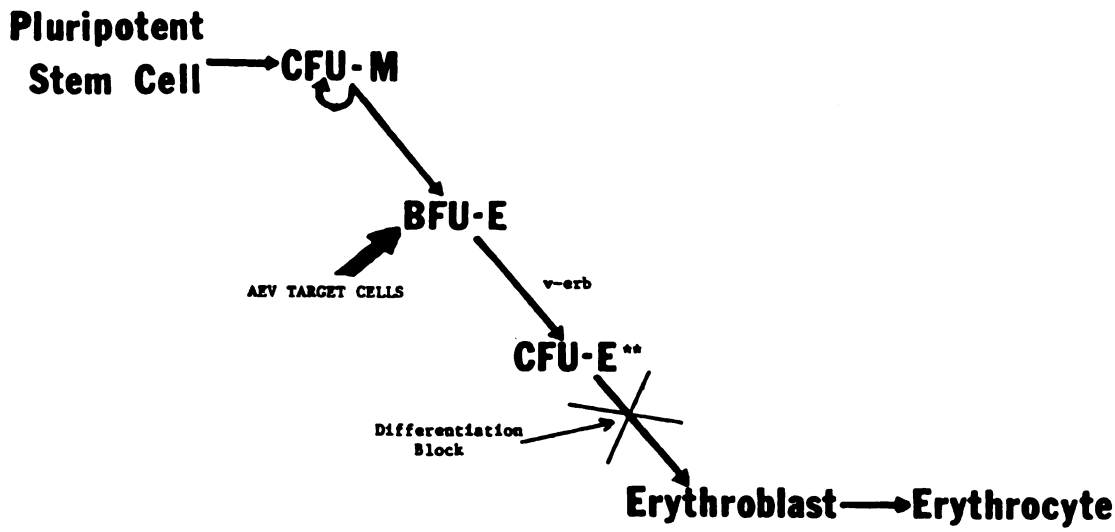

Figure 2. The effect of avian erythroblastosis virus (AEV) on chicken erythropoiesis. The target cells of AEV infection are the BFU-E which continue to differentiate to the CFU-E stage. The CFU-E cells are transformed (**) by the AEV transforming genes v-erb-A and v-erb-B.

or mostly CFU-E type) population at the permissive temperature. Non-homogeneous but highly enriched populations of erythroblasts, reticulocytes and erythrocytes, can also be fractionated from temperature-induced cells as mentioned above. A relatively homogeneous in vivo population of reticulocytes can be isolated from chickens with anemia induced by phenylhydrazine or bleeding (32). Reticulocytes are postmitotic and differ from most other eucaryotic cells studied in that their transcriptional repertoire is severely restricted. Lasky et al., in a kinetic analysis of the hybridization of reticulocyte poly(A)$^+$RNA to cDNA made to such RNA, estimated that globin mRNA accounts for 90% of the total mRNA pool in these cells (14). Analysis of the hybridization of enriched nonglobin cDNA to total reticulocyte mRNA indicated that approximately 100 nonglobin mRNA species comprise the remaining 10% with the concentration of each RNA species at about 1/200 the globin concentration (33). It is likely, however, that these studies led to a somewhat over-generalized picture of reticulocyte mRNA complexity due to the limitations of solution hybridization of bulk mRNA:cDNA populations.

Because of the ease of isolating erythroid cells and their relatively simple gene expression pattern, many erythroid cell proteins have been characterized. The globin genes were among the first genes cloned due to their abundant expression in reticulocytes. Since then, several genes expressed in erythroid cells, both erythroid-specific and housekeeping genes, have been cloned and characterized. Carbonic anhydrase II (34, 35), histone H5 (27, 28) erythroid antigens and other membrane proteins (18, 36-38), red cell cytoskeletal proteins (e.g., spectrin) (39, 40), and enzymes involved in heme synthesis (41, 42) and

other aspects of erythrocyte metabolism are among the known red cell proteins for which the corresponding genes have been isolated, at least in some species.

To achieve the goal of isolating other erythroid-specific genes, many of the advantages of chicken erythroid cells just discussed were utilized. The availability of relatively homogeneous cell populations of reticulocytes and erythroblasts allowed a reticulocyte-specific probe to be produced. This was done by employing a hybridization selection procedure refined by Davis et al. (43). Figure 3 shows a schematic of the overall procedure. Two gene sequence populations were subtracted from total reticulocyte cDNA: globin sequences and erythroblast sequences. This resulted in an enrichment of reticulocyte-specific, non-globin gene sequences. This enriched probe was then used to screen a previously-constructed reticulocyte cDNA library. The hybridizing clones were then characterized by recombinant DNA techniques.

In this thesis the research will be described in three parts. In Chapter I, the cDNA synthesis and library construction techniques and results will be discussed. Chapter II will describe the production of the erythroid-specific probe, the screening of the reticulocyte cDNA library, and the characterization of the isolated clones. The further characterization of one clone - the analysis of the genome organization of chicken ubiquitin will be described in Chapter III. Two publications are included in appendices. Appendix I contains the publication that describes the isolation and identification of the first chicken carbonic anhydrase II cDNA clone, pPE5-0.3 (34). This clone was isolated from the pBR322 reticulocyte cDNA library described

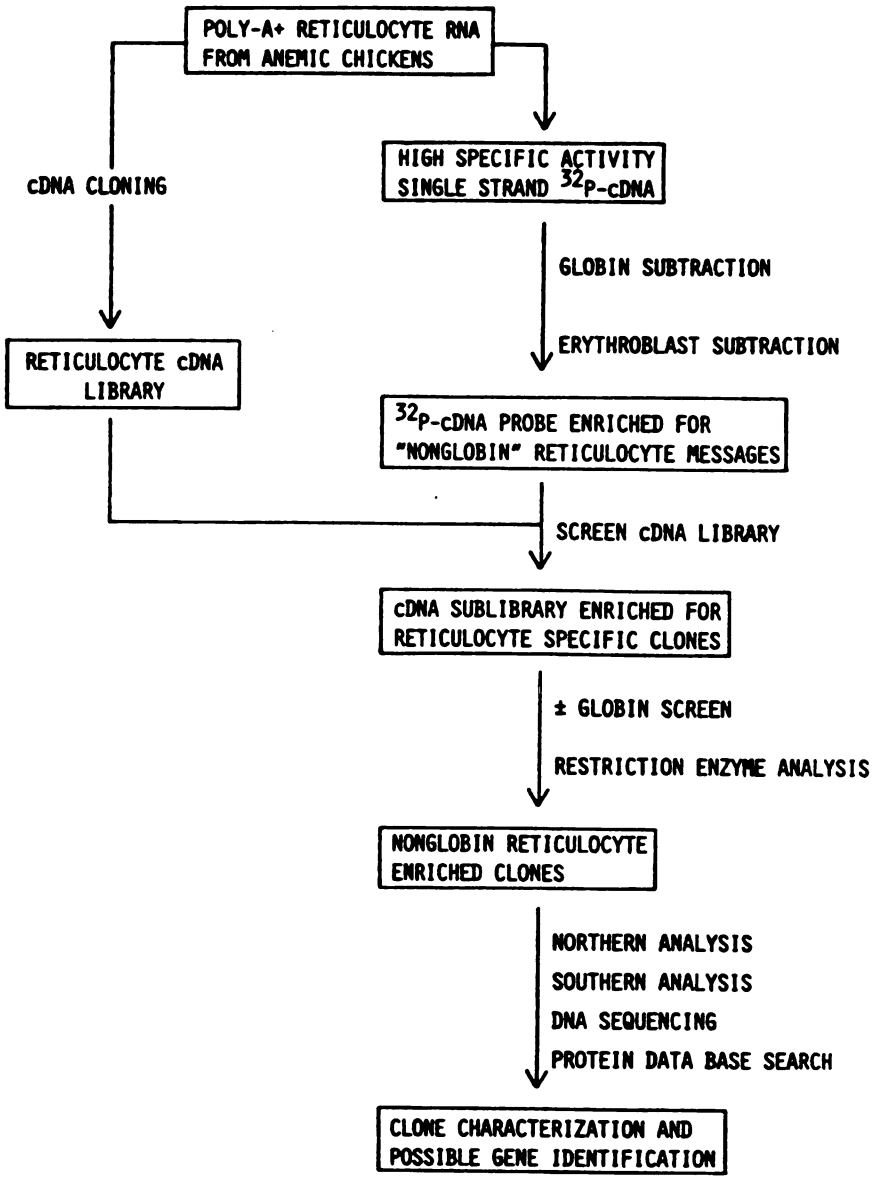

Figure 3. Overall scheme for the isolation of chicken erythroid-specific genes.

in Chapter I. Appendix II contains the publication describing the isolation of chicken erythroid δ-aminolevulinate synthase cDNA clones (42). Co-workers transferred the reticulocyte cDNA library constructed in λgt10 (described in Chapter I) into a λ fusion/expression vector λgt11. They then screened the resulting expression library to identify clones coding for δ-aminolevulinate synthase, the first enzyme of the heme biosynthesis pathway.

REFERENCES

1. Lewin, B. 1980. pp. 771-776. In Gene Expression 2. John Wiley and Sons, New York.

2. Engel, J.D. and J.B. Dodgson. 1978. J. Biol. Chem. 253:8239-8246.

3. Dodgson, J.B., J. Strommer and J.D. Engel. 1979. Cell. 17:879-887.

4. Engel, J.D. and J.B. Dodgson. 1980. Proc. Natl. Acad. Sci. U.S.A. 77:2596-2600.

5. Brown, J.L. and V.M. Ingram. 1974. J. Biol. Chem. 249:3960-3972.

6. Keane, R.W. and U.K. Abbott. 1980. Develop. Biol. 75:442-453.

7. Harrison, P.R. 1976. Nature. 262:353-356.

8. McKnight, S.L. and R. Kingsbury. 1982. Science. 217:316-324.

9. Davidson, E.H., H.T. Jacobs and R.J. Britten. 1983. Nature. 301:468-470.

10. Treisman, R., M.R. Green and T. Maniatis. 1983. Proc. Natl. Acad. Sci. U.S.A. 80:7428-7432.

11. Meyers, R.M., N. Lumelsky, L.S. Leuman and T. Maniatis. 1985. Nature. 313:495-498.

12. Kollias, G., N. Wrighton, J. Hurst and F. Grosveld. 1986. Cell. 46:89-94.

13. Choi, O.-R. and J.D. Engel. 1986. Nature. 323:731-734.

14. Lasky, L., N.D. Nozick and A.J. Tobin. 1978. Develop. Biol. 67:23-39.

15. Cline, M.J. and D.W. Golde. 1979. Nature. 277:177-181.

16. Harrison, P.R. 1976. Nature. 262:353-356.

17. Samarut, J., J.P. Jurdic and V. Nigon. 1979. J. Embryol. Exp. Morph. 50:1-20.

18. Samarut, J., J.P. Blanchet and V. Nigon. 1979. Develop. Biol. 72:155-166.

19. Samarut, J. and M. Bouabdelli. 1980. J. Cell. Physio. 105:553-563.

20. Gazzolo, L., J. Samarut, M. Bouabdelli and J.P. Blanchet. 1980. Cell. 22:683-691.

21. Samarut, J. and L. Gazzolo. 1982. Cell. 28:921-929.

22. Graf, T., N. Ade and H. Beug. 1978. Nature. 275:496-501.

23. Beug, H., S. Palmieri, C. Freudenstein, H. Zentgraf and T. Graf. 1982. Cell. 28:907-919.

24. Beug, H., G. Dorderlein, C. Freundenstein and T. Graf. 1982. J. Cell. Physio. Supplement I:195-207.

25. Gasson, J.C. et al. 1985. Nature. 315:768-771.

26. Marks, P.A., and R.A. Rifkind. 1978. Annu. Rev. Biochem. 47:417-448.

27. King, P.A., A.J. Robins, R. D'Andrea and J.R.E. Wells. 1983. Nucleic Acid Res. 11:619-627.

28. Renard, J. and A. Ruiz-Carrillo. 1986. J. Mol. Biol. 189:217-226.

29. Lucas, A.M. and C. Jamrog. 1961. In Atlas of Avian Hematology. Agriculture Monograph 25, U.S.D.A.

30. Graf, T., B. Royer-Pokora, E. Korzeniewski, S. Grieser and H. Beug. 1981. Hematology and Blood Transfusion - Modern Trends in Human Leukemia IV. 26:417-423.

31. Adkins, B., H. Beug and T. Graf. 1983. J. Cell. Biochem. Supplement 7B:21.

32. Longacre, S.S. and W.J. Rutter. 1977. J. Biol. Chem. 252:2742-2752.

33. Tobin, A.J., S.E. Selvig and L. Lasky. 1978. Develop. Biol. 67:11-22.

34. Yoshihara, C.M., M. Federspiel and J.B. Dodgson. 1983. Annals of New York Academy of Science - Biology and Chemistry of the Carbonic Anhydrases. 429:332-334.

35. Ferrell, R.E., S.K. Stroup, R.J. Tanis and R.E. Tashian. 1978. J. Biol. Chem. 258:4459-4463.

36. Kline, K., J.P. Allison and B.G. Sanders. 1982. Develop. Biol. 91:389-396.

37. Weise, M.J. and L-N. L. Chan. 1978. J. Biol. Chem. 253:1892-1897.

38. Nelson, C.H., J.P. Allison, K. Kline and B.G. Sanders. 1982. Cancer Res. 42:4625-4630.

39. Lazarides, E. and W.J. Nelson. 1982. Cell. 31:505-508.

40. Elgsaeter, A., B.T. Stokke, A. Mikkelsen and D. Branton. 1986. Science. 234:1217-1223.

41. Granick, S. and S.I. Beale. 1978. Advan. Embryol. 46:33-203.

42. Yamamoto, M., N.S. Yew, M. Federspiel, J.B. Dodgson, N. Hayashi and J.D. Engel. 1985. Proc. Natl. Acad. Sci. U.S.A. 82:3702-3706.

43. Davis, M.M., D.I. Cohen, E.A. Nielson, M. Steinmetz, W.E. Paul and L. Hood. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:2194-2198.

CHAPTER I

RETICULOCYTE cDNA LIBRARY CONSTRUCTIONS

In 1981, when this work was begun, cDNA cloning techniques varied widely in efficiency and cDNA cloning vectors were limited mainly to plasmids such as pBR322 (1-3). Only 10^3-10^4 cDNA clones per microgram cDNA were commonly obtained upon transformation into E. coli but often only nanogram quantities of clonable cDNA are available. Thus, the first priority in cDNA library construction was the establishment of reliable, high yield cDNA synthesis procedures. New cloning vectors (4, 5) and transformation procedures (6) have since been established employing different strategies that greatly increase the yield of cDNA clones. This chapter will describe the cDNA synthesis techniques employed in the construction of two reticulocyte cDNA libraries. The overall procedure developed for cDNA synthesis was a combination of new and published strategies (1-6). The methodology developed is basically similar to that of "Molecular Cloning. A Laboratory Manual" by Maniatis et al. (6).

The overall cDNA synthesis scheme is outlined in Figure 1. Briefly, the first cDNA strand is synthesized by reverse transcriptase which initiates from an oligo (dT) primer hybridized to the polyadenylated portion of mRNA. Upon removal of the RNA template, a 3'-hairpin loop is created by the secondary structure of the cDNA strand which is sufficient to prime DNA Polymerase I - Klenow fragment-

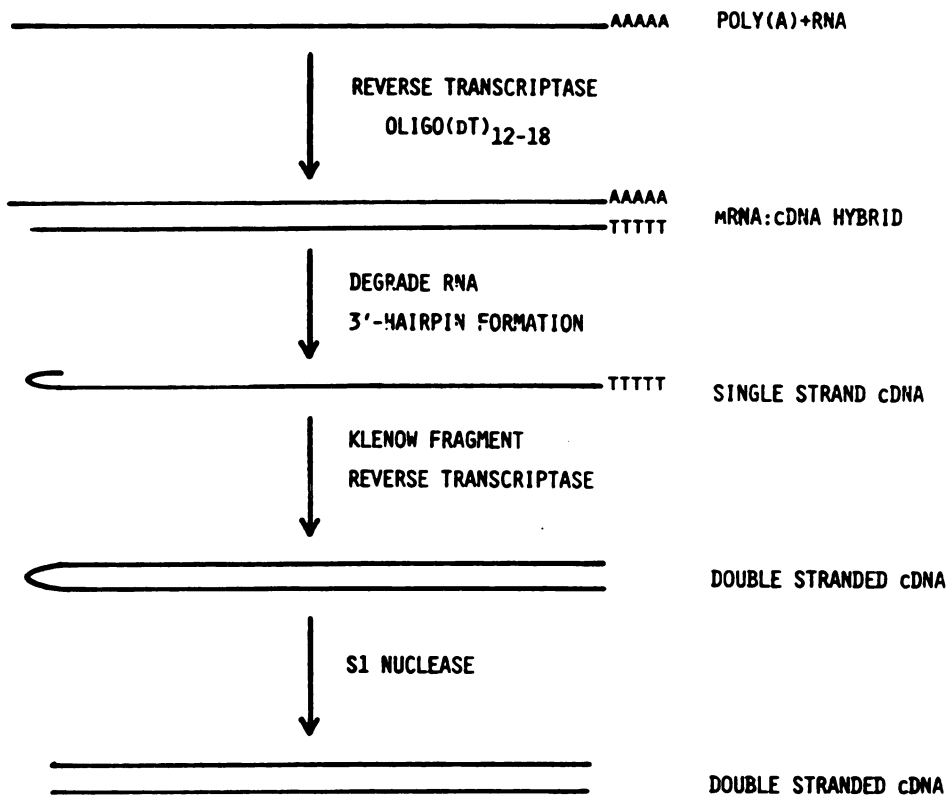

Figure 1. A schematic of the overall cDNA synthesis protocol.

catalyzed second strand synthesis. To maximize the length of the cDNA second strand, reverse transcriptase is also employed in this step. Finally, S1, a single strand-specific nuclease, digests the hairpin loop leaving double stranded cDNA ready for cloning. The experimental details of these steps, which are described in Methods, are the final result of many quantitative and qualitative experiments. Effort was concentrated on testing aspects of several published methods (1-3) to establish routine conditions and procedures for each step, titrating all enzymes involved and improving overall product recovery (described below).

The first reticulocyte cDNA library was inserted into pBR322 using a homopolymer tailing procedure (6-8). Terminal deoxynucleotide transferase synthesizes a single stand deoxynucleotide tail from 3'-OH-ends. Synthesized cDNA was tailed with deoxycytidine (dC) by terminal transferase and annealed to deoxyguanine (dG) - tailed vector. The vector was PstI-digested pBR322. Cutting at the PstI site interrupts the ampicillin (amp) resistance gene but leaves the tetracycline (tet) resistance gene intact. Annealed cDNA:vector clones were transformed into E. coli. The phenotype of transformants in which cDNA was successfully inserted into plasmid DNA should be tetracycline resistant (tet^R) and ampicillin sensitive (amp^S). Problems occurred with screening large numbers of plasmid transformants as well as with clone viability in storage, so the decision was made to construct another cDNA library in a lambda vector. The pBR322 library was used to screen for the chicken carbonic anhydrase II (CAII) gene using a homologous mouse CAII probe (9). This also allowed us to test the

reliability of the overall cDNA synthesis procedure and how well the RNA pool was represented.

The second reticulocyte cDNA library was constructed in gt10 using a linker addition procedure (6). Synthetic EcoRI linkers were ligated onto synthesized cDNA creating flanking EcoRI sites. The cDNA could then be ligated into the single EcoRI site of λgt10 which is within the λ repressor (C_I) gene. The two pathways of lambda growth, lytic and lysogenic, are controlled by an intricate balance of host and lambda gene products and are employed here as a selection for cDNA transformants (5, 6). Central to this control is the C_I gene product which is a repressor of lambda early transcription and consequently blocks late gene expression. High repressor concentrations plus the int gene product lead to the insertion of the lambda DNA into the genome of the bacterial host - the lysogenic pathway. Low repressor concentrations release the inhibition on the maintenance promotor, P_m, allowing lambda to enter the lytic cycle which results in a many fold replication of lambda DNA and eventual lysis of the host cell.

The product of a lysogenic infection is a cloudy plaque because those bacteria in which lambda is integrated into the host genome will be resistant to further infection and produce the turbidity within the cleared area resulting from the lytic cycle infection. A purely lytic infection, though, produces a clear plaque morphology. The small percentage of lytically growing lambda that give rise to the cloudy plaque can be further reduced by infecting an E. coli hfl (high frequency lysogenization) mutant (5). Deletion of the host hflA gene enhances the stability of several lambda factors resulting in a stimulation of C_I synthesis and a high frequency of lysogenization. So

on an hfl host, a C_I^+ phage, e.g., intact λgt10, will give a faint, almost invisible plaque, but a C_I^- phage will still give a clear plaque since no λ repressor is made. The insertion of cDNA into the EcoRI site of λgt10 interrupts the C_I gene resulting in a lytic infection with a clear plaque morphology. By amplifying the primary cDNA-vector ligations on an E. coli hflA mutant, the resulting high titer phage stock will be 98-99% C_I^-. The C_I^- phage must have resulted from insertion of cDNA (or linker DNA) into the EcoRI site or be naturally occurring C_I^- mutants.

EXPERIMENTAL PROCEDURES

Materials

Phenylhydrazine was purchased from Aldrich. Heparin, HEPES, and proteinase K were purchased from Sigma. Oligo (dT)-cellulose type 775 and oligo $(dT)_{12-18}$ were purchased from Collaborative Research. RNasin, terminal deoxynucleotide transferase, and deoxynucleotides were purchased from P.L. Biochemicals. Reverse transcriptase was purchased from Life Sciences. EcoRI - 10 mer synthetic linkers, DNA polymerase I - Klenow fragment, S1 nuclease, T4 DNA ligase, T4 polynucleotide kinase, and all restriction enzymes were purchased from either Bethesda Research Labs, New England Biolabs, or Boehringer Mannheim. [α-^{32}P]dCTP, [^{3}H]dCTP, and PstI-digested, dG-tailed pBR322 were purchased from New England Nuclear.

Reticulocyte RNA Isolation

The method employed was a modification of the procedure of Longacre and Rutter (10). Hens were injected with 2.5% (w/v) phenylhydrazine, pH 7, on six consecutive days according to the following schedule: 0.7, 0.7, 0.4, 0.4, 0.5, and 0.6 ml. On the seventh day, the chickens were bled by heart puncture and the blood added to four volumes NKM solution (140 mM NaCl, 5 mM KCl, 2 mM $MgCl_2$, and 0.1% heparin). The cells were recovered by centrifugation at 1500 x g removing the supernatant and buffy coat. The cells were then

washed three times with cold NKM and recovered as before. Two volumes of lysis buffer (2 mM $MgCl_2$, 2 mM dithiothreitol, 10 mm Tris-HCl, pH 7.5, and 10 mM iodoacetate) were used to resuspend the final cell pellet and the mixture was incubated at 0° for 30 min. The nuclei and cell debris were then removed by centrifugation at 16,000 x g, 4°. The supernatant was made to 1.0% SDS, 10 mm EDTA, and 100 µg/ml proteinase K and incubated at 37° for 60 min. One tenth volume of 3 M NaOAc, pH 5.0 was added, mixed, and the solution extracted with one half volume phenol (equilibrated with 0.1 M Tris-HCl pH 7.0) and one half volume chloroform. The phases were separated by centrifugation at 1500 x g, 25°. The aqueous phase was extracted with one volume chloroform and separated as above. The final aqueous phase was ethanol precipitated with 2.2 volumes ethanol at -20° for 18 hr. The RNA was recovered by centrifugation at 12,000 x g for 20 min at 0°; the pellet dried under vacuum, and resuspended in ddH_2O. The RNA recovery was determined from A_{260} measurement (40 µg/O.D.). All steps were done with ribonuclease free technique (6).

Poly(A)$^+$RNA Separation

All steps were carried out under ribonuclease free conditions. Oligo (dT)-cellulose type 775 was pretreated with 0.1 N NaOH at 25° for 20 min. The cellulose was then washed with water until the pH reached neutral, packed into a suitable column, and washed with 1 volume high salt buffer (500 mM NaCl, 10 mM Tris-HCl pH 7.5, 0.05% SDS, 1 mM EDTA). Total RNA in elution buffer (10 mM Tris-HCl pH 7.5, 0.05% SDS, 1mM EDTA), was heated at 55° for 5 min, cooled and made 500 mM in NaCl. The solution was added to the cellulose column, eluted, and the column

washed with high salt buffer until the eluate A_{260} was near 0. The column was then washed with several volumes of low salt buffer (150 mM NaCl, 10 mM Tris-HCl pH 7.5, 0.05% SDS, 1 mM EDTA). The poly(A)$^+$RNA was then eluted with elution buffer, ethanol precipitated as before, and resuspended in ddH_2O.

cDNA Synthesis

First Strand Synthesis. The reaction mixture (100 µl) contained the following: 100 mM Tris-HCl pH 8.3, 140 mM KCl, 10 mM $MgCL_2$, 100 µg/ml oligo$(dT)_{12-18}$, 30 mM β-mercaptoethanol, 1.1 mM each of dGTP, dCTP, dTTP and dATP, 60-100 Ci [α-^{32}P]dCTP, 50 units RNasin, 98 units reverse transcriptase, and 5-10 µg poly(A)$^+$RNA. All first strand solutions, equipment, and techniques were ribonuclease free. All components, except the RNA, were combined and incubated at 25° for 10 min. Synthesis was initiated with the addition of the RNA, incubated at 42° for 2 hr, and terminated with EDTA to 25 mM. An aliquot of the cDNA synthesized, with [α-^{32}P]dCTP incorporated, was quantitated by TCA precipitation (6). The reaction was then extracted with one half volume phenol, one half volume chloroform; and the aqueous phase passed over a one ml Sephadex G-50 column (6). The excluded volume was made 0.16 N in NaOH, incubated at 68° for 10 min, neutralized, ethanol precipitated, and resuspended in 10 mM Tris-HCl pH 7.5.

Second Strand Synthesis. The reaction mixture contained the following: 10 mM HEPES pH 6.9, 80 mM KCl, 3 mM $MgCl_2$, 2.3 mM dithiothreitol, 1 mM each of dGTP, dCTP, dTTP, and dATP, and 40-50 units of DNA polymerase I - Klenow fragment per microgram of first strand cDNA synthesized. The reaction was incubated at 15° for 15-20

hr and terminated with EDTA to 25 mM. The mixture was phenol:choloroform extracted, ethanol precipitated and resuspended in the following: 100 mM Tris pH 8.3, 140 MM KCl, 10 mM $MgCl_2$, 30 mM β-mercaptoethanol, and 1.1 mM each of dGTP, dCTP, dTTP, and dATP. The Klenow synthesized second strand cDNA was lengthened by incubating the mixture at 42° for 2 hr with 90 units of reverse transcriptase. The terminated reaction was extracted, ethanol precipitated, and the double stranded cDNA resuspended in S1 nuclease buffer (200 mM NaCl, 50 mM NaOAc pH 4.5, and 1 mM $ZnSO_4$).

S1 Nuclease Digestion. A set of small scale pilot reactions were carried out to determine the optimal concentration of S1 required to digest only the hairpin loop connecting the two cDNA strands. To 25,000 cpm aliquots of cDNA; 10, 5, 2.5 and 1.25 units of S1 nuclease were added. Incubations proceeded at 37° for 30 min and were terminated with EDTA to 25 mM. The digestion products were analyzed on alkaline agarose gels (6), which were 1.4% agarose and run with 30 mM NaOH and 2 mM EDTA as electrophoresis buffer. The full scale S1 digestion was an exact scale up of the optimal S1 pilot digestion conditions. The completed digest was then phenol:chloroform extracted, ethanol precipitated, and the S1-digested double stranded cDNA resuspended in ddH_2O.

pBR322 cDNA Library Construction

Oligodeoxycytidine Addition. The reaction mixture contained 1.4 mM β-mercaptoethanol, 7.5 mM dCTP, 400 mM potassium cacodylate, pH 6.9, 1 mM $CoCl_2$, and 500 units/ml terminal deoxynucleotide transferase. The reaction was initiated by the addition of the double stranded cDNA,

incubated at 37°, and terminated with EDTA to 25 mM. The optimal incubation time to add 20-25 deoxycytidine residues varied with each enzyme lot and was determined by a pilot time course reaction using [^{3}H]dCTP. Following termination, the reaction was phenol:chloroform extracted and passed over a one ml Sephadex G-50 column previously equilibrated with annealing buffer (100 mM NaCl, 10 mM Tris-CHl pH 7.5, 1 mM EDTA).

Annealing and Transformation Reactions. PstI - digested, oligodeoxyguanine-tailed pBR322 was combined with the deoxycytidine-tailed cDNA in a 8:1 molar ratio and diluted with annealing buffer to a final DNA concentration of 0.8 μg/ml. The mixture was incubated at 68° for 10 min, then 42° for 2 hr, cooled to 20°, and transformed into E. coli strain HB101 by a modification of Hanahan's χ1776 procedure (11). The completed transformation reaction was plated on LB plates with tetracycline (15 μg/ml).

Screening the cDNA Library. Background ampicillin resistant colonies were screened out by duplicate plating on tetracycline and ampicillin (50 μg/ml) plates. Ampicillin sensitive and tetracycline resistant colonies were screened for carbonic anhydrase II (CAII) by a modification of the method of Grunstein and Hogness (12). In short, colonies were grown on two replica nitrocellulose filters, lysed with alkali, and neutralized. The released DNA was bound to the filters by baking at 80° for 2 hr under vacuum, and probed by hybridization with a ^{32}P-nick translated labelled CAII probe. A 300 bp PstI fragment from the coding region of a mouse CAII cDNA clone obtained from Peter Curtis, was isolated by the method of Girvitz (13) and labelled with ^{32}P by nick translation (6). Hybridization was carried out with 30%

formamide at 42° for 30 hr (6). The filters were then washed first with 2 X SSC (0.3 M NaCl, 0.03 M sodium citrate pH 7.0) at 25° then washed with 0.2 X SSC at 55°. The filters were exposed to Kodak X-Omat R film.

λgt10 cDNA Library Construction

EcoRI Linker Addition. The ends of the synthesized cDNA were repaired by DNA polymerase I - Klenow fragment (in second strand buffer) to create blunt ends prior to linker addition (6). Synthetic EcoRI - 10 mer linkers were phosphorylated by T4 polynucleotide kinase in the following reaction mixture: 1 μg linkers, 66 mM Tris-HCl, pH 7.6, 1 mM ATP, 1 mM spermidine, 10 mM $MgCl_2$, 15 mM dithiothreitol, 2 μg/ml BSA, and 2 units T4 polynucleotide kinase. The mixture was incubated at 37° for 1 hr, terminated by heating at 68° for 10 min, and then added to the repaired cDNA. Six units of T4 DNA ligase were added and the mixture incubated at 15° for 18 hr. The linkers were then extensively digested with EcoRI and the digest electrophoresed on a 1% agarose gel run in TAE (40 mM Tris-OH, 20 mM acetic acid, 2 mM EDTA). Two size fractions of linker added cDNA, 400 bp-800 bp and 800 bp-3 kb, were isolated by the method of Girvitz (13).

Vector:cDNA Ligation and In Vitro Packaging. A set of trial ligations was used to determine the optimal EcoRI digested λgt10 (vector):cDNA ratio giving the highest plaque forming units (pfu) per microgram of vector yield. The completed ligations were *in vitro* packaged by the method of Hohn (14), and grown on two *E. coli* strains: C600 and C600 *hfl* (5). The full scale ligation, with the remainder of the cDNA, was carried out under the determined optimal conditions and

in vitro packaged. A separate library was constructed with each of the isolated cDNA size fractions.

Amplification. 380,000 pfu from each library were plated separately, on 18 150 mm LB plates with C600 hfl and grown for 10 hr at 37°. The plates were harvested by scraping the nearly confluent plaques into a tube, and adding an equal volume of SM (100 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 mM $MgSO_4$, 0.01% gelatin) and 0.01% chloroform. After mixing extensively the agar was pelleted by centrifugation at 4000 x g for 10 min at 4°. This λgt10-cDNA stock was titered and stored at 4°.

RESULTS AND DISCUSSION

Isolation of Reticulocyte Poly(A)$^+$ Cytoplasmic RNA

Reticulocytes were isolated from the circulating blood of chickens with induced anemia. The phenylhydrazine treatment schedule of Longacre and Rutter (10) induces the highest anemia with the lowest mortality. The total cytoplasmic RNA yield by this method averaged 60 mg RNA per 100 ml of blood. The RNA yield was quantitated by A_{260} measurement using 25 A_{260} O.D. units equalling 1 mg RNA as a conversion. Approximately 2% of the total cytoplasmic RNA was separated by oligo (dT)-cellulose chromatography as poly(A)$^+$RNA which agrees with the estimate of Lasky et al. (15) for reticulocytes. Agarose gel analysis of the selected poly(A)$^+$RNA showed the predicted 9S globin RNA band along with very minor contamination with 18S ribosomal RNA and transfer RNA (results not shown). Agarose gel analysis of ^{32}P-labelled first strand cDNA synthesized from the poly(A)$^+$RNA showed a smear ranging in size from 200 bases to 3 kb with a 600 bases band corresponding to globin (results not shown).

cDNA Synthesis

The overall cDNA synthesis scheme is outlined in Figure 1. Through incorporation of ^{32}P into the first strand cDNA, the first and second strand synthesis products as well as the S1 nuclease test digestions could be analyzed by autoradiography of alkaline agarose

gels. Under the denaturing conditions, the double stranded molecules denature but the first and second strand cDNA strands are connected by the 3'-hairpin loop. Thus the size of the double stranded cDNA population should be approximately twice that of the first strand. After S1 nuclease digestion, however, the two strands are no longer connected, and therefore the size of the labelled cDNA should revert to approximately that of the first strand cDNA. Also, by incorporating ^{32}P into the first strand cDNA, the yield of cDNA synthesized, the efficiency of each enzymatic reaction, and the overall recoveries could be quantitated by TCA precipitation.

In reviewing the established cDNA synthesis protocol, several critical areas determined the overall success. Reverse transciptase quality and reaction conditions were shown to be critical for efficient, full length first strand cDNA synthesis. A high molar excess of enzyme to template (20 fold), the optimal pH of 8.3, the optimal monovalent cation concentration (140 mM KCl) for strand elongation, the required divalent cation (10 mM Mg^{++}), and a high concentration of all four deoxynucleotides (1 mM) were shown to dramatically increase first stand synthesis (results not shown). RNA quality was another critical factor. To prevent RNA degradation, ribonuclease free techniques as well as a potent ribonuclease inhibitor, RNasin, were used. Under these conditions approximately 20% of the initial mRNA was copied to generate first strand cDNA. Theoretically, all of the mRNA could template first strand cDNA synthesis by reverse transcriptase. In reality, only 10-30% conversion is typically as reported in the literature (1-6).

An absolute requirement for efficient second strand cDNA synthesis was the complete degradation of the RNA template that primed the first strand. Published degradation conditions were either so harsh that the synthesized cDNA and RNA were both degraded, or inadequate, leaving a large proportion of the RNA intact. Incomplete RNA hydrolysis significantly inhibited enzyme initiation and second strand elongation. The conditions developed for treating the completed first strand reaction, 0.16N NaOH at 68° for 10 min, optimally degraded the RNA yet left the cDNA strand intact.

Two enzymes were used to synthesize the second cDNA strand primed off of the 3'-hairpin loop formed naturally by the first strand cDNA. Initially, the second strand was synthesized by DNA polymerase I - Klenow fragment. This enzyme's optimal reaction conditions included a pH optimum of 6.9, a 15° incubation temperature to minimize snapback DNA, and a long incubation time (18 hr) to allow the enzyme time to initiate off of the unstable hairpin loop. An additional enzymatic synthesis step was added to further lengthen the second strand and thereby increase the percentage of full length double-stranded cDNA. It has been observed that there are several positions along a DNA single strand, perhaps relating to its intrinsic secondary structure, that result in a pause or stop in replication (6). It is thought that each enzyme has its own specific types of strong-stops, that are not necessarily recognized by other enzymes. In this procedure, by continuing second strand synthesis with reverse transcriptase, cDNA second strand length increased an average of 10% as estimated from an alkaline agarose gel. The average product recovery, after the total

second strand synthesis, was 33% of the ^{32}P incorporated into the first strand cDNA synthesized.

The optimal S1 nuclease conditions must be determined for each preparation of cDNA due to S1's ability to digest even double stranded DNA at high enzyme concentrations. It was also imperative that the full scale S1 reaction, with the remaining double stranded cDNA, be scaled up exactly to the conditions of the appropriate S1 pilot reaction. Approximately 75% of the double stranded cDNA molecules were resistant to S1 nuclease digestion. A plateau of 80% resistance has been reported in the literature (2). The recovery of the completed double stranded, S1-digested cDNA averaged 25% of the original first strand synthesized (Table 1).

A crucial factor in the overall success of the cDNA synthesis involved methods of handling the various reactants and reaction mixtures. The frequent extractions, precipitations, buffer changes, and other manipulations of the cDNA initially resulted in substantial product loss. Considerable time, therefore, was spent in minimizing the number of transitional clean-up steps to maximize product recovery without sacrificing enzyme efficiency. Silanized tubes and micropipets substantially increased product yields by reducing adhesion of cDNA and RNA to surfaces. Another valuable technique was the use of Sephadex G-50 (1 ml) spin columns (6) to remove reactants and change buffer systems in a minimum volume. The most severe loss of product occurred during second strand synthesis despite all improvements. The large reaction volumes needed and two synthesis reactions resulted in an average 63% loss. Table 1 summarizes the enzymatic yields and recovery

Table 1. A summary of the cDNA synthesis protocol efficiences [a].

Reaction Step	Reaction Recovery (%)[b]	DNA Recovery (ug) SS[c]	DS[d]	Overall Recovery (%)[e]
1.) First Strand	---	2.1	---	100
-Processing[f]	89	1.9	---	89
2.) Second Strand				
-Klenow Fragment of DNA Polymerase I	75	1.4	2.8	66
-Reverse Transcriptase	50	0.7	1.4	33
3.) S1 Nuclease	85	0.6	1.2	28

[a] data is from one experiment.

[b] efficiency of each step (recovered product divided by initial reactant cocentration).

[c] only the first cDNA strand was ^{32}P-labelled (SS-single strand).

[d] double strand (DS) DNA concentration was estimated by doubling the SS concentration.

[e] recovery relative to the first strand.

[f] includes RNA degradation, extraction, Sephadex G-50 exclusion, and ethanol precipitation.

efficiencies of a representative cDNA preparation used to construct the λgt10 reticulocyte library.

cDNA Library Constructions

One of the main advantages of studying chicken reticulocytes is the relatively small number of expressed genes. Lasky et al. (15) estimate only about 100 mRNA species comprise the nonglobin 10% of the mRNA of reticulocytes. The number of clones required for a cDNA library with a given probability that every such mRNA would be represented can be estimated by the following equation:

$$N = \frac{\ln (1-P)}{\ln (1-n)}$$

where N = the number of clones required, P = probability desired (usually 0.99), and n = fractional proportion of the total mRNA population that a single type of low abundance mRNA represents (6). To construct a cDNA library with a 99% probability that each of the 100 nonglobin mRNA species are represented (P = 0.99, n = 0.001) only 4600 clones are required. The isolation of a sequence present at even a ten fold lower abundance would only increase the number of clones needed to 46,000, a number easily within the efficiency of the described cDNA cloning procedures.

The first reticulocyte cDNA library was constructed in a plasmid vector, pBR322, by the homopolymer tailing method (Figure 2). By adding a deoxycytidine single strand "tail" to the synthesized cDNA, and a deoxyguanine "tail" to PstI-digested pBR322 with terminal deoxynucleotide transferase, the two DNA pieces could be annealed. The regenerated plasmid was transformed into E. coli strain HB101. The

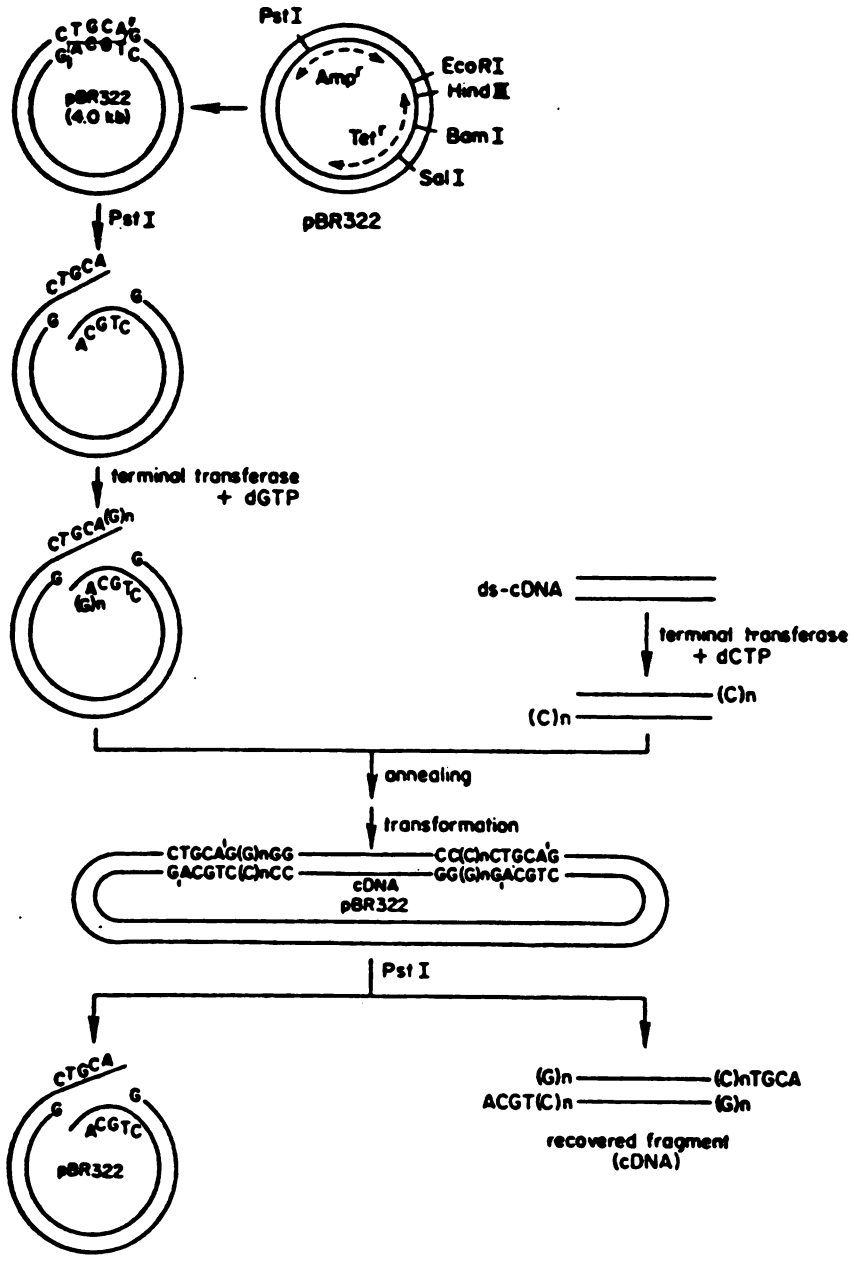

Figure 2. A schematic of cDNA library construction by homopolymer tailing. This figure is taken, in part, from Maniatis et al. (6). See text for details.

terminal transferase reaction, however, was exceedingly sensitive to inhibitors. Tris, EDTA, and ethanol in even trace amounts are all potent inhibitors and repeated extractions and precipitations must be done to remove these common reactants. The terminal transferase reaction employed a large excess of enzyme to ensure that the number of residues added was essentially independent of cDNA concentration. For each lot of enzyme, a pilot reaction, using [^{3}H]dCTP, was done to establish a residue addition time course standard curve. The reaction time was determined that added 20-25 deoxycytidine residues to the cDNA, which in previously published experiments (16) was shown to be optimal for dC-dG annealing as measured by the yield of transformants. The transformation procedure, a modified Hanahan X 1776 procedure (11), routinely produced 10^7-10^8 colonies per microgram of intact plasmid. The annealed cDNA:vector mixture, in a 1:8 molar ration, produced 6.7 x 10^5 transformants per microgram vector - an expected 100 fold decrease compared to intact plasmid. Ligation and transformation of the vector alone gave 5 x 10^2 colonies per microgram vector.

10,692 transformants were screened for background tet^R amp^S colonies yielding 1296 (12%) plasmids with no inserted sequences. 49% of the 9396 tet^R amp^S transformants hybridized to a globin probe consisting of the three adult globin clone fragments ^{32}P-labelled by nick translation. Approximately 70% of the resulting 4800 "non-globin" clones were lost due to the laborious manual screening procedures used. During the long storage periods needed to perform the various screening procedures used, many of the individual colonies (stored on plates) that made up the original library lost viability. This problem might have been obviated by storing mini-glycerol stocks of the isolated

individual colonies. The remaining 1300 viable clones were screened for chicken cDNAs homologous to the mouse carbonic anhydrase II gene (CAII) to assess the representation of CAII cDNAs in the library. The probe was a nick-translated, ^{32}P-labelled 300 bp fragment from the coding portion of a mouse CAII cDNA clone provided by Peter Curtis (Wistar Institute). Three positive autoradiographic signals, as well as nine other clones, were examined. The average insert size was 300 bp ranging from 75-500 bp. Nucleotide sequences were determined for two of the CAII-positive clones, with insert sizes of 370 bp and 300 bp. Open reading frames were identified and possible amino acid sequences deduced. The amino acid sequences corresponding to the inserts of both cDNA clones were 60% homologous with the known CA II sequences (Appendix I). Thus, we identified two CAII clones in approximately 780 non-globin reticulocyte cDNA clones screened. (From the results above, we expect 40% of the 1300 clones to be globin clones.) This is in close agreement with the estimate that CAII mRNA comprises approximately 0.1% of the reticulocyte poly(A)$^+$ cellular RNA (see above). Due to the small insert sizes and to viability problems in maintaining large numbers of colonies, the pBR322 library was abandoned. At this time the phage cDNA cloning systems λgt10 and λgt11 became available, and it was felt that this would be a more suitable vector for this project.

The second reticulocyte cDNA library was constructed in λgt10 using a linker addition protocol (Figure 3). The gt10 vector offers the advantages of easily screening large numbers of clones, high yield cDNA cloning, and convenient storage at 4° with little loss in viability. Plasmid clone libraries can be stored for long time periods

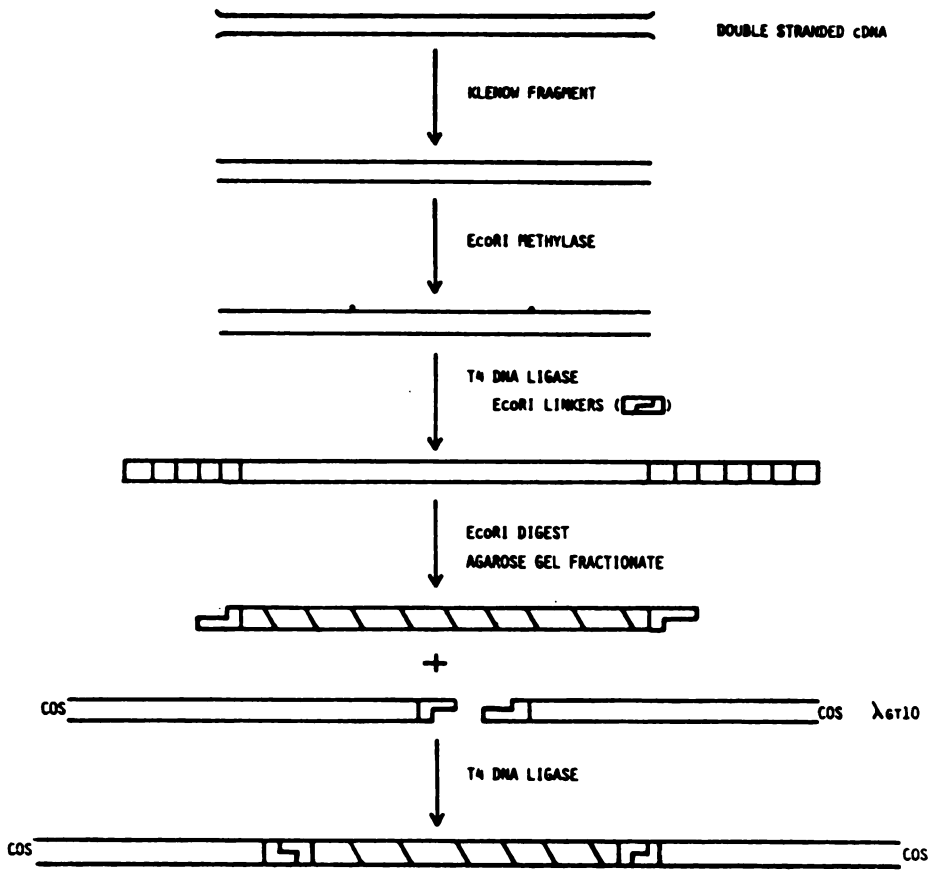

Figure 3. A schematic of cDNA library construction by linker addition. See text for details. Methylated EcoRI sites are denoted by (*). Abbreviation: cohesive ends (COS).

in bulk (e.g., as a pooled glycerol stock). However, we were afraid that differences in growth rates of different clones would significantly alter the representation of the plasmid library, so we attempted to store it as individual colonies on plates. As described above, this led to seriously reduced clone viability.

As shown with the CAII screening of the pBR322 library, non-globin genes were represented but with small insert sizes. The cDNA synthesis procedure was altered slightly to help increase the size of the synthesized cDNA ready for cloning. The extent of the S1 digestion was reduced to digest less of the double stranded cDNA.

When using a linker addition protocol, one usually methylates the restriction sites of the synthesized cDNA corresponding to the enzyme to be used to cut the linkers. In our case, treating the cDNA with EcoRI methylase would prevent any cDNA EcoRI site from being digested when the linkers were cut to expose the EcoRI ends. Due to problems with the methylation reaction, adequate methylation was never achieved. Because full length cDNA clones were not required for the isolation protocol described in Chapter II, we did not feel that the small number of cDNA clones containing EcoRI sites would cause problems. After EcoRI linkers were added and extensively digested, the cDNA was size fractionated on agarose gels, isolating the 400-800 bp and 800 bp - 3 kb cDNA size ranges. Each size range was inserted into λgt10, creating two separate cDNA libraries.

The two critical areas that determined the success of cloning into λgt10 both deal with clone selection. In λgt10, the cDNA inserts are cloned into the single EcoRI site in the repressor (C_I) gene of λgt10. Thus phage with inserts are C_I^- and grow lytically producing clear

plaques. Lambda without inserts grow lysogenically, producing cloudy plaques or virtually no plaques at all, if grown on an E. coli hfl (high frequency lysogen) mutant. The first critical area is the isolation of a λgt10 stock with a very low spontaneous C_I^- background, i.e., less than 0.25% (17). Upon in vitro packaging, the vector DNA stock should give a titer of 10^8 pfu/μg DNA. Infectivity of the DNA should be reduced at least three logs after EcoRI digestion prior to cloning, and upon ligation of vector DNA clone, one should recover 5-30% of its original infectivity with little or no increase in the C_I^- background (Table 2). The second critical area is the complete removal of all digested EcoRI linkers. A synthetic linker can interrupt the C_I gene coding sequence producing clear plaques. Thus it is imperative to size fractionate an extensively digested cDNA-EcoRI linker ligation to purify the cDNA away from the excess linkers.

After testing several phage isolates, a λgt10 vector stock that fulfilled all the requirements listed above was found. However, upon test ligations of this vector DNA with cDNA the C_I^- percentage could only be increased two fold over the no cDNA insert control. Normally one would like a 5-10 fold increase (Table 2). Consequently, only one half of the C_I^- pfu of the constructed libraries are expected to have cDNA inserts. This was judged not to be a major problem since large numbers of cDNA transformants were available compared to the relatively small number of clones needed for a "complete" library. The two libraries constructed were a 400 bp - 800 bp library yielding 2.8×10^6 independent C_I^- pfu and a 800 bp - 3 kb library yielding 1.7×10^6 C_I^- pfu. 380,000 C_I^- pfu from each library were amplified resulting in amplified bulk library stocks for storage of 180 ml at 5.6×10^9 C_I^-

Table 2. λgt10 vector characteristics and ligation efficiencies.

	Expected λgt10 DNA Characteristics	Actual λgt10 Vector Characteristics: total infectivity (per ug vector)	% of original infectivity	C_I-% [c]
Total infectivity	10^8 pfu/ug vector	6 X 10^7 pfu	100	0.2
Infectivity after EcoRI digestion	Greater than 3 logs less infectivity	1 X 10^3 pfu	0.0017	---
No insert EcoRI λgt10 ligation	5-30% of original infectivity	5.1 X 10^6 pfu	8.5	0.5
EcoRI λgt10 with insert ligation	5-30% of original infectivity	1.0 X 10^7 pfu [a] 1.0 X 10^7 pfu [b]	16.7 16.7	1.2 1.6

[a] inserts were the 400 - 800bp fraction of reticulocyte cDNA.

[b] inserts were the 800bp - 3kb fraction of reticulocyte cDNA.

[c] spontaneous background C_I- plaques should be less than 0.25% in the vector stock.

pfu/ml (400-800 bp) and 190 ml at 5.3 x 10^9 C_I^- pfu/ml (800 bp - 3 kb). The size fractions were chosen to clone most of the globin messages (550-600 bp) in the library with smaller inserts in order to aid in screening for non-globin clones (see Chapter II).

Both reticulocyte cDNA libraries were transferred into the λfusion/expression vector λgt11 by another laboratory as described in Appendix II. They then screened the gt11 cDNA library with a heterologous antibody for δ-aminolevulinate (ALA) synthase. Several cDNA clones encoding ALA synthase were isolated and characterized. ALA synthase is the first enzyme in the heme biosynthesis pathway. These results, along with those to be described in Chapter II, demonstrate that a variety of non-globin, reticulocyte messages are represented within these cDNA libraries.

REFERENCES

1. Efstratiadis, A. and L. Villa-Komaroff. 1979. pp. 15. In Genetic Engineering. Vol I. J.K. Stelow and A. Holla, eds., Plenum Press, New York.

2. Wickens, M.P., G.N. Buell and R.T. Schimke. 1978. J. Biol. Chem. 253:2483-2495.

3. Kurtz, D.T. and C.F. Nicodemus. 1981. Gene. 13:145-152.

4. Okayama, H. and P. Berg. 1982. Mol. Cell. Biol. 2:161-170.

5. Huynh, T.V., R.A. Young and R.W. Davis. 1985. pp. 49-78. In DNA Cloning: A Practical Approach. Vol I. D.M. Glover, ed., IRL Press, Washington, D.C.

6. Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. In Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

7. Villa-Komaroff, L., A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S.P. Naker, W.L. Chick and W. Gilbert. 1978. Proc. Natl. Acad. Sci. U.S.A. 75:3727-3731.

8. Peacock, S.L., C.M. McIver and J.J. Monohan. 1981. Biochem. Biophys. Acta. 655:243.

9. Yoshihara, C.M., M. Federspiel and J.B. Dodgson. 1983. Annals of the New York Academy of Science - Biology and Chemistry of the Carbonic Anhydrases. 429:332-334.

10. Longacre, S.S. and W.J. Rutter. 1977. J. Biol. Chem. 252:2742-2752.

11. Hanahan, D. 1985. pp. 109-136. In DNA Cloning: A Practical Approach. Vol I. D.M. Glover, ed., IRL Press, Washington, D.C.

12. Grunstein, M. and D. Hogness. 1975. Proc. Natl. Acad. Sci. U.S.A. 72:3961-3964.

13. Girvitz, S.C., S. Bacchetti, A.J. Rainbow and F.L. Graham. 1980. Anal. Biochem. 106:492.

14. Hohn, B. and K. Murray. 1977. Proc. Natl. Acad. Sci. U.S.A. 74:3259.

15. Lasky, L., N.D. Nozick and A.J. Tobin. 1978. Develop. Biol. 67:23-39.

16. Deng, G. and R. Wu. 1981. Nucleic Acids Res. 9:4173-4188.

17. St. John, T. personal communication.

CHAPTER II

ISOLATION AND CHARACTERIZATION OF ERYTHROID EXPRESSED cDNA CLONES

The specific goal of this research was the isolation and characterization of genes expressed during erythroid differentiation as described above. To do this we used two homogeneous erythroid precursor populations that can be isolated in large quantities. An *in vivo* population consisting almost entirely of reticulocytes, a late erythroid precursor which is still active in transcription, can be isolated from the blood of anemic chickens (1). Another homogeneous erythroid precursor population can be isolated from an *in vitro* cell line derived from bone marrow cells transformed with avian erythroblastosis virus (AEV) (2). Erythroid differentiation is arrested by AEV at approximately the CFU-E stage. For simplicity, this population will be called erythroblasts. Temperature sensitive AEV mutants (tsAEV) can be induced to terminally differentiate by growth at the permissive temperature (42°) in medium supplemented with anemic chicken serum (2). Three highly enriched erythroid cell populations, erythroblasts, reticulocytes and erythrocytes, can be fractionated by Percoll density centrifugation from induced tsAEV cells (3). In a study comparing protein synthesis in these three populations, twenty-eight major protein synthesis changes were observed by two dimensional gel electrophoresis of ^{35}S-methionine labelled cell lysates

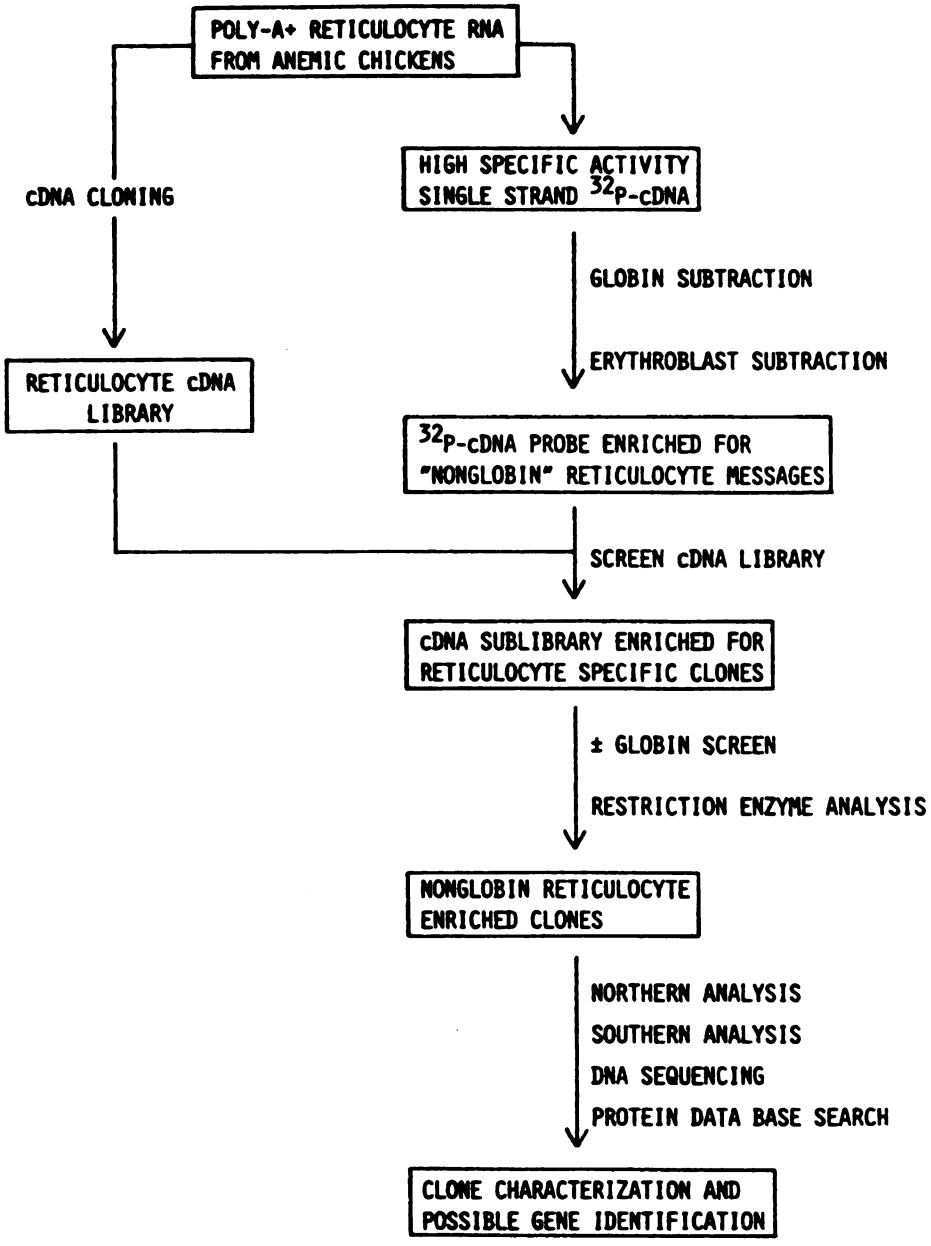

Figure 1. Overall scheme for the isolation of chicken erythroid-specific genes.

(4). It was therefore hoped that these visible differences in the proteins that these cells were making would be correlated with significant differences in the poly(A)$^+$RNA present in the reticulocytes versus that of the erythroblast cells.

Of course, the most obvious change in erythroblast versus reticulocyte RNA is the presence of very high levels of each of three adult globin mRNAs in the latter. Since these three genes have already been studied in some detail, we attempted to minimize their representation in our cloning protocol. The three adult globin genes were subcloned into Promega Biotec's SP6 transcription vectors (5). The SP6 vectors have a specific promotor site, recognized by the bacterial SP6 polymerase, upstream of a polylinker cloning site. Run-off RNAs are transcribed from the promotor in vitro, through the cloned DNA, to a site of restriction enzyme digestion at the 3' end of the region which one desires to copy. Coding or non-coding strand RNA can be transcribed depending on the direction of the insert relative to the vector.

Figure 1 outlines the experimental approach for isolating erythroid-expressed genes employing the refined hybridization-subtraction procedure of Davis et al. (6). To enrich for non-globin reticulocyte gene sequences, two RNA populations, globin and erythroblast, were subtracted from the cDNA made to the reticulocyte poly(A)$^+$RNA. High specific activity, ^{32}P-labelled first strand cDNA was transcribed from poly(A)$^+$ reticulocyte RNA isolated from anemic chickens. Globin cDNAs were subtracted by the hybridization of SP6-transcribed globin RNA to the reticulocyte ^{32}P-cDNA by the subtraction procedure outlined in Figure 2. Hydroxylapatite

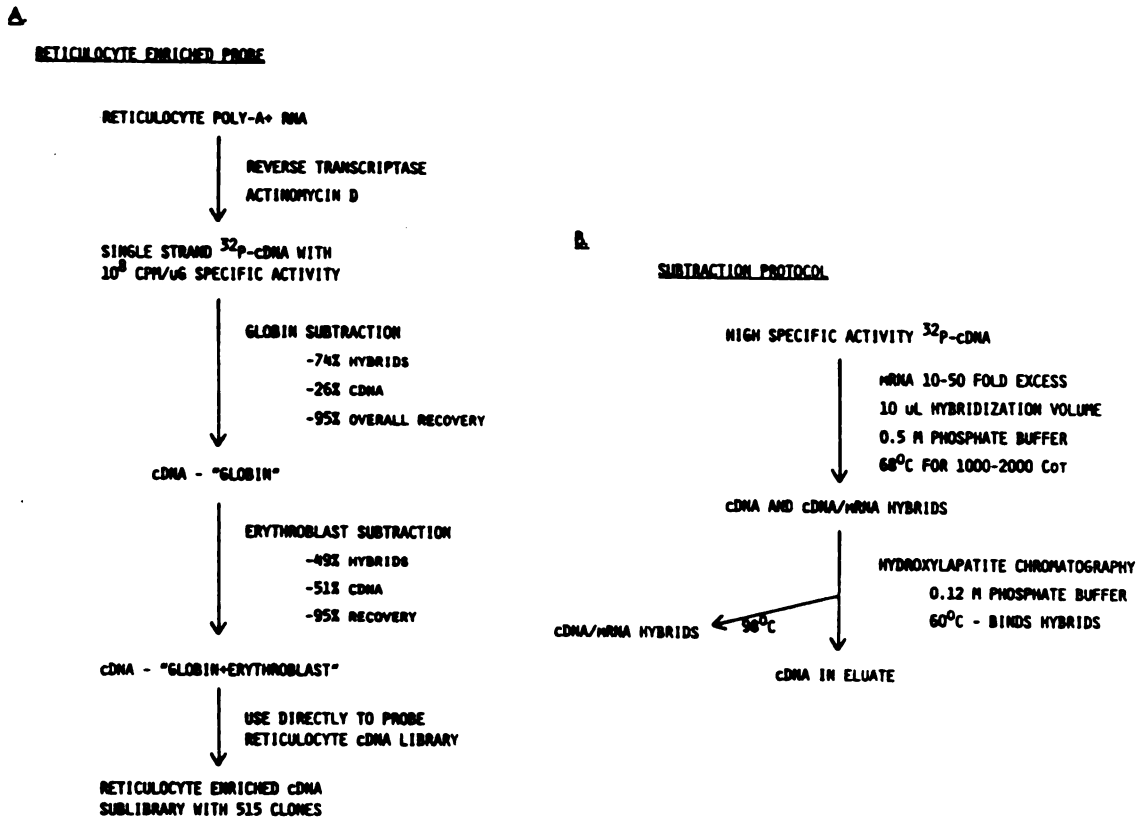

Figure 2. A schematic of the preparation of a reticulocyte-enriched cDNA probe. The overall procedure, including results, of producing an enriched reticulocyte cDNA population is shown in (A). A detailed summary of the subtraction procedure is outlined in (B).

chromatography separated the cDNA:SP6-RNA hybrids (double stranded molecules) from the unhybridized cDNA (single strand molecules). Hydroxylapatite, under the appropriate conditions, specifically binds double stranded nucleic acids. With an increase in temperature or phosphate concentration, the double stranded molecules can also be eluted. Sequences which were common to both erythroblasts and reticulocytes were subtracted by the hybridization of the single-stranded cDNA off the first column to erythroblast poly(A)$^+$RNA isolated from the AEV-erythroid cell line LSCC HD3 (7). The resulting reticulocyte-enriched cDNA probe was used to screen the reticulocyte λgt10 cDNA library (Chapter I) leading to the isolation of a reticulocyte-enriched cDNA sublibrary.

This chapter describes the production of the reticulocyte-enriched cDNA probe, the screening of the reticulocyte λgt10 cDNA library, and the characterization of some members of the enriched sublibrary. Since the globin subtraction was not 100% efficient, the members of the cDNA sublibrary initially chosen for study were found to contain contaminating globin sequences. Further characterization led to the identification of eight non-globin clones which were analyzed in some detail. The size and tissue specificity of the mRNA homologous to each clone were determined by analysis of Northern blots. Fine-structure restriction enzyme maps were generated, and the nucleotide sequence of the cloned cDNA inserts were determined. The functions of several clones were identified. A clone specifically expressed in reticulocytes, clone 37, was identified as coding for carbonic anhydrase II (CAII) by comparing its restriction map and partial nucleotide sequence to that of a chicken CAII cDNA clone which I

previously had isolated (Appendix I). Two other clones, 104 and 200, were identified by the comparison of their corresponding amino acid sequences, as deduced from an open reading frame in the nucleotide sequence, to the Protein Sequence Database using the program FASTP (8). Clone 104 was identified as coding for ferritin heavy chain, an iron storage protein found mainly in the blood, liver, spleen, and bone marrow. Clone 200 was identified as ubiquitin, a small 76 amino acid protein found in all tissues. Ubiquitin has several known and putative functional roles and is the subject of Chapter III. The other potential erythroid specific clones have not been identified.

EXPERIMENTAL PROCEDURES

Materials

The SP6 transcription system including the vectors pSP64 and pSP65, and the SP6 polymerase were purchased from Promega Biotec. Hydroxylapatite was purchased from Bio-Rad. Collodial tubes were purchased from Schleicher and Schuel. The LSCC HD3 cell line was a gift from H.J. Kung (Case Western Reserve). Dulbecco's Modified Eagle's Medium (4500 mg/ml glucose), chicken serum, and fetal bovine serum were purchased from Gibco. HEPES and proteinase K were purchased from Sigma. All restriction enzymes were purchased from either New England Biolabs, Boehringer Mannheim, or Bethesda Research Labs.

RNA Isolation

Reticulocyte poly(A)$^+$RNA was isolated from chickens with induced anemia as described in Chapter I.

Erythroblast poly(A)$^+$RNA was isolated from the AEV-transformed erythroid cell line LSCC HD3. This continuous cell line was grown on Dulbecco's Modified Eagle Medium (4500 mg/ml glucose) with 8% fetal bovine serum, 2% chicken serum, and 10 mM HEPES pH 7.3. A 500 ml (10^7 cells/ml) culture, grown in a spinner flask, was centrifuged at 1500 x g for 10 min. The cells were washed in phosphate buffered saline, recentrifuged, and the cells resuspended in two volumes 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 incubated at 37° for 1 hr. The mixture was sonicated using a large probe in 15 sec bursts at 4° until the solution lost its viscosity. 200 μg/ml of fresh proteinase K was added to the solution followed by further incubation for 1 hr at 37°. The poly(A)$^+$RNA was isolated by two rounds of oligo(dT)-cellulose chromatography as described in Chapter I.

RNA was synthesized from the three adult globin genes subcloned into SP6 vector plasmids. The transcription reaction mixture contained the following: 40 mM Tris-HCl, pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine, 500 μM each of ATP, UTP, GTP and CTP, 10 mM dithiothreitol, 1000 units/ml RNasin, 40 μg/ml SP6 linear plasmid DNA, 50 μCi [α-^{32}P]UTP, 5-10 units/μg DNA SP6 polymerase. The reaction was carried out at 40° for 2 hr and then digested with 30 μg/ml ribonuclease-free DNase I at 37° for 10 min. The solution was made 3.5 M in NH_4OAc, and tRNA was added to 10 μg/ml followed by phenol:chloroform extraction and ethanol precipitation. The RNA pellet was resuspended in 2.5 M NH_4OAc, ethanol precipitated, and resuspended in water. All steps were carried out using ribunuclease free technique. The RNA synthesized was quantitated by TCA precipitation and the transcript size evaluated by agarose gel electrophoresis.

Reticulocyte-Enriched Probe

cDNA Synthesis. The first strand cDNA reaction mixture contained the following: 100 mM Tris-HCl, pH 8.7, 140 mM KCl, 10 mM $MgCl_2$, 100 μg/ml oligo(dT)$_{12-18}$, 30 mM β-mercaptoethanol, 1.1. mM each of dGTP, dATP and dTTP, 36 μM dCTP, 500 units/ml RNasin, 1 mCi [α-^{32}P]dCTP, 98 units reverse transcriptase, 100 μg/ml actinomycin D,

and 10 μg poly(A)$^+$ reticulocyte RNA. The reaction was carried out at 42° for 2 hr and terminated with EDTA addition to 25 mM. Ribonuclease free technique was used until completion of the cDNA synthesis. The solution was then made 0.16 N in NaOH, and incubated at 68° for 10 min. Following neutralization with HCl and phenol:chloroform extraction, and acqueous phase was passed over a 1 ml Sephadex G-50 column. The excluded cDNA was made 200 μg/ml in tRNA and ethanol precipitated.

Globin Subtraction. A 10-20 fold sequence excess of the three SP6 globin synthesized RNAs was combined with the high specific activity first-strand cDNA. These nucleic acids were then precipitated and resuspended in 5-10 μl TE. The RNA and cDNA were hybridized in 0.5 M sodium phosphate buffer, pH 6.8, 0.5% SDS and 1 mM EDTA (total volume less than 20 μl) in a sealed glass micropipet. The solution was heated at 90° for 2 min, then incubated at 68° for 18 hr. The hybridized solution was diluted with 1 ml HAP buffer (0.12 M sodium phosphate buffer pH 6.8, 0.1% SDS) and loaded onto a water-jacketed, 1 ml hydroxylapatite column previously equilibrated with 0.12 M phosphated buffer, 0.1% SDS at 60°. The column was washed with HAP buffer at 60°, and 1 ml fractions were collected. One microliter aliquots of each fraction were counted in a scintillation counter. The peak fractions of single strand cDNA were combined and dialyzed in a collodial tube against 50 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA for 6 hr at 25°. The double-stranded molecules were eluted by washing the column with HAP buffer at 98° for purposes of quantitation. The dialyzed single strand cDNA molecules were extracted with sec-butanol to reduce the total volume to less than 0.5 ml, followed by phenol:chloroform extraction and ethanol precipitation.

Erythroblast Subtraction. The single stranded cDNA fraction from the globin subtraction was combined with a 10-50 fold sequence excess of poly(A)$^+$ erythroblast RNA isolated from LSCC HD3 cells (see above) and precipitated with ethanol. The hybridization reaction and hydroxylapatite chromatography were performed as before. The combined single-strand cDNA peak fractions constituted the reticulocyte-enriched probe.

Screening the λgt10 Reticulocyte cDNA Library. 40,000 pfu from each size fraction λgt10 reticulocyte library (see Chapter 1) were plated with C600 hfl onto five 150 mm LBM plates (8,000 pfu/plate). The plates were incubated at 37° for 7 hr and stored at 4°. Duplicate nitrocellulose filter replicas were made of each plate. These were denatured, neutralized, washed and baked at 80° for 2 hr as described by Benton and Davis (9). The baked filters were then prewashed in 50 mM Tris-HCl, pH 8.0, 1 M NaCl, 1 mM EDTA, at 42° for 2 hr in a rotating water bath. The filters were prehybridized in 50% formamide, 0.5% Ficoll, 0.5% polyvinylpyrrolidine, 0.5% bovine serum albumin, 5 X SSC (1 X SSC = 150 mM NaCl, 15 mM Sodium Citrate, pH 7.0), 0.1% SDS, 1 mM EDTA, 10 mM Tris-HCl, 100 μg/ml denatured salmon sperm DNA, 10 μg/ml yeast tRNA, 10 μg/ml poly(rC) and 10 μg/ml poly(rA), at 42° for 18 hr with gentle shaking. Hybridization was initiated by adding the reticulocyte-enriched probe directly to the prehybridization mixture. Incubation continued at 42° for 60 hr with gentle shaking. The filters were then washed first with 2 X SSC at 25° and then with 0.2 X SSC at 68°. They were exposed to Kodak X-Omat film. All plaques showing hybridization to both duplicate filters were picked and

together they constitute the two reticulocyte-enriched cDNA sublibraries, one sublibrary for each original size fraction library screened.

Clone Characterization

An initial screen for contaminating globin clones was done by spotting 1 µl of each cDNA plaque lysate stock in a grid pattern on an E. coli lawn, making filter replicas, and hybridizing the replicas to a nick-translated, ^{32}P-labelled globin probe (containing the three adult globin sequences) and a ^{32}P-reticulocyte cDNA probe. Small DNA stocks of non-globin, reticulocyte-positive clones were made by the plate lysate method of Fritsch (10). Clone DNA was digested with EcoRI, electorphoresed on 1.2% agarose gels, Southern blotted and rescreened with the probes described above. The inserted cDNA of clones which continued to hybridize to the reticulocyte probe but not to the globin probe was isolated by the method of Girvitz (11). These DNA fragments were subcloned into the EcoRI site of the plasmid pUC8 (12) grown in the E. coli strain JM107.

Nick-translated, ^{32}P-labelled subclone DNA was hybridized to Northern blots of poly(A)$^+$ reticulocyte, erythroblast and pancreas RNAs (2-12 µg of each) electrophoresed in formaldehyde gels (10). 1.2% agarose formaldehyde gels were run in 20 mM MOPS, pH 7.0, 1 mM EDTA, 5 mM sodium acetate, and 2.2 M formaldehyde. RNA samples were ethanol precipitated, resuspended in sample buffer (20 mM MOPS, pH 7.0, 1 mM EDTA, 5 mM sodium acetate, 50% formamide, 2.2 M formaldehyde), and heated at 60° for 10 min before loading. After electrophoresis the gel was soaked in a large volume of 20 X SSC at 25° for 30 min. The RNA

was transferred to nitrocellulose with 20 X SSC and the filter was baked at 80° for 2 hr under vacuum. Hybridizations were carried out as described above at 42° for 15-20 hr.

Restriction enzyme maps were generated by single and double enzyme digestions (10). Appropriate DNA fragments were treated with calf alkaline phosphatase and labelled with T4 polynucleotide kinase (Chapter 1). The ^{32}P-labelled DNA fragment was sequenced by the method of Maxam and Gilbert (13) as modified by Smith and Calvo (14). Possible amino acid sequences, as deduced from nucleotide sequence open reading frames, were compared to the Protein Sequence Database by the computer program FASTP (8). The Protein Sequence Database is part of the Protein Identification Resource release of August 6, 1985 by The National Biomedical Research Foundation, Georgetown University Medical Center. The database contains 738,997 residues in 3309 sequences.

RESULTS AND DISCUSSION

Reticulocyte-Enriched Probe

Figure 2A outlines the procedure employed in producing a reticulocyte-enriched probe. For use as a probe, several modifications in the first strand cDNA synthesis procedure described in Chapter I were added. High specific activity cDNA ($>10^8/\mu g$) needed to be synthesized, rather than the cDNA with a nominal ^{32}P-incorporation used to quantitate reactions. Reverse transcriptase can tolerate dCTP concentrations as low as 36 μM (dATP, dTTP and dGTP at 1.0 mM) and still transcribe full length cDNA copies with reasonable yields. This condition routinely yielded a specific activity of 2-3 X 10^8 cpm/μg and copied 10-15% of the poly(A)$^+$RNA to cDNA. Another modification was the addition of actinomycin D to reduce the 3'-hairpin loop that can spontaneously form during the synthesis of the cDNA probe (6). The hairpin loop has enough double-stranded character to increase the "single-stranded" cDNA's affinity for hydroxylapatite (see below).

The three adult globin gene sequences and the erythroid housekeeping message sequences were subtracted from the ^{32}P-labelled, reticulocyte cDNA by two rounds of hybridization-subtraction. The key features of the subtraction protocol are shown in Figure 2B. A large sequence excess (10-50 fold) of the RNA was hybridized to labelled cDNA in a small volume (10-20 μl) to drive the hybridization to completion in an overnight incubation. Most RNA:cDNA hybridization reactions near

completion at a Cot of 500-2000. The speed of the procedure was important due to the rapid breakdown of the high specific activity cDNA. The unhybridized cDNA was separated from the cDNA:RNA hybrids by hydroxylapatite chromatography. Double stranded nucleic acids bind to hydroxylapatite at 60° in 0.12 M phosphate buffer, while single stranded molecules elute. The peak single strand fractions were pooled, dialysed to remove the phosphate buffer, and precipitated. The double strand hybrid molecules were eluted at 98° in 0.12 M phosphate buffer to determine the recoveries. The overall recovery off the hydroxylapatite column (comparing the recovered double-strand hybrid CPM plus the precipitated single-strand cDNA cpm to the total cpm loaded) was 95%.

Globin Subtraction. Globin gene sequences were subtracted from the reticulocyte cDNA by using RNA transcribed from the coding strand of subcloned globin chromosomal DNAs in the SP6 transciption system. The two adult alpha globin SP6 subclones were a gift from J.D. Engel (Northwestern University). The A subclone, pSP αA(+), contains the 1.7 kb BamHI/EcoRI fragment from pBRα7-1.7 (15) in pSP64 (Figure 3). The αD subclone, pSP αD(+), contains the 1.1 kb Pst/EcoRI fragment from πSVαD (16) in pSp64 (Figure 3). The adult beta gene (β), on a 2.5 kb EcoRI/SacI fragment of pβ1BR15 (17), was subcloned into pSP65 and named pSP β(+) (Figure 3). Upon run-off transcription by SP6 polymerase, the SP6 subclones yielded 1-2 μg of RNA per μg of linear plasmid template. In the subtraction, 1.65 μg of ^{32}P-labelled reticulocyte cDNA was hybridized to 16 μg of αA, 8 μg of αD and 15 μg of βRNA, and the products were separated on hydroxylapatite. Subtraction of 74% of the cDNA was accomplished using the SP6 globin RNA.

Figure 3. Restriction enzyme maps of the SP6-globin clones. The SP6 vector plasmids, pSP64 and pSP65, have an ampicillan resistance gene, an origin of replication and an SP6 promotor site just 5' of a M13 polylinker cloning region. The SP6 promotor/polylinker regions of both vectors are shown at the top of the figure. The three adult globin genes (αA, αD, β) were cloned into the appropriate SP6 vector to result in the coding strand being transcribed by SP6 polymerase. Transcription initiates at ↱, reads through the cloned DNA to a 3'-site of digestion (marked by *).

Erythroblast Subtraction. The subtraction protocol was repeated with RNA from an erythroid precursor cell line to eliminate gene messages common to both early and late erythroid cells. The unhybridized cDNA (0.43 μg) from the globin subtraction was hybridized to 20 μg of poly(A)$^+$ erythroblast RNA from the AEV-erythroid cell line, LSCC HD3. The erythroblast RNA effected subtraction of 49% of the cDNA. The two subtractions together removed 87% of the reticulocyte cDNA sequences.

The overall enrichment for non-globin, reticulocyte-specific messages was not as effective as expected. A significant number of contaminating globin clones must be present in the isolated sublibraries since the globin selection only removed 74% of the reticulocyte cDNA rather than the 90% estimated by Lasky et al. (18). Nevertheless the enriched probe was used for several reasons. The foremost reason was the warning of Mark Davis (Stanford University) that more than two subtraction rounds could lead to severe losses in product recovery. However, two factors helped reduce the problem of contamination with globin sequences. First, HD3 erythroblasts transcribe small amounts of RNA from all three adult globin genes as seen on Northerns (data not shown). The concentration of globin mRNA in HD3 cells has not been quantitated, but its presence should further deplete globin sequences in the erythroblast subtraction round. Second, two size-selected λgt10 cDNA libraries were constructed in the hope of reducing the number of globin clones in the larger 800 bp - 3 kb, library. Thus, we hoped the final probe would at least be adequate to screen the library with large inserts.

Screening the λgt10 Reticulocyte cDNA Libraries. The remaining non-hybridized cDNA (13%; 200 ng) was used as a probe to screen 40,000 reticulocyte λgt10 cDNA clones from each of the two separate libraries constructed with different size fractions of cDNA (Chapter I). Approximately 50% of the λgt10 cDNA libraries have cDNA inserts. All positive autoradiographic signals were isolated yielding 295 clones from the 400-800 bp library and 220 clones from the 800 - 3 kb library, 1.5% and 1.1% of the clones with inserts, respectively. This is 5-10 fold fewer clones than we expected since the probe contained 13% of the reticulocyte cDNA sequences. This may relate to difficulties in identifying a true positive plaque. Whether a plaque shows a positive signal above the background depends on several factors including plaque size, size of the cDNA insert, efficiency of DNA transfer, and the efficiency of the hybridization. The cDNA probe was hybridized for 3 X Cot $\frac{1}{2}$ (10), so most positive plaques should have hybridized to near completion. However sequences which were significantly more rare than the 0.1% level estimated for non-globin messages by Lasky et al. (18) may not have hybridized to completion. Furthermore, since 200 ng of cDNA were used in the hybridization, a sequence which, for example, still represented only 0.1% of the enriched cDNA would account for 200 pg of the probe. Therefore, these rare cDNA sequences might not be able to saturate the phage DNA on the filters, and this would further lower their signals, even if hybridization did go to completion. For these reasons, it is likely that only moderately rare sequences (0.1% of reticulocyte cDNA before enrichment) gave positive signals in this screen. These problems could be ameliorated by improved enrichment protocols (see below) and/or using much larger amounts of cDNA.

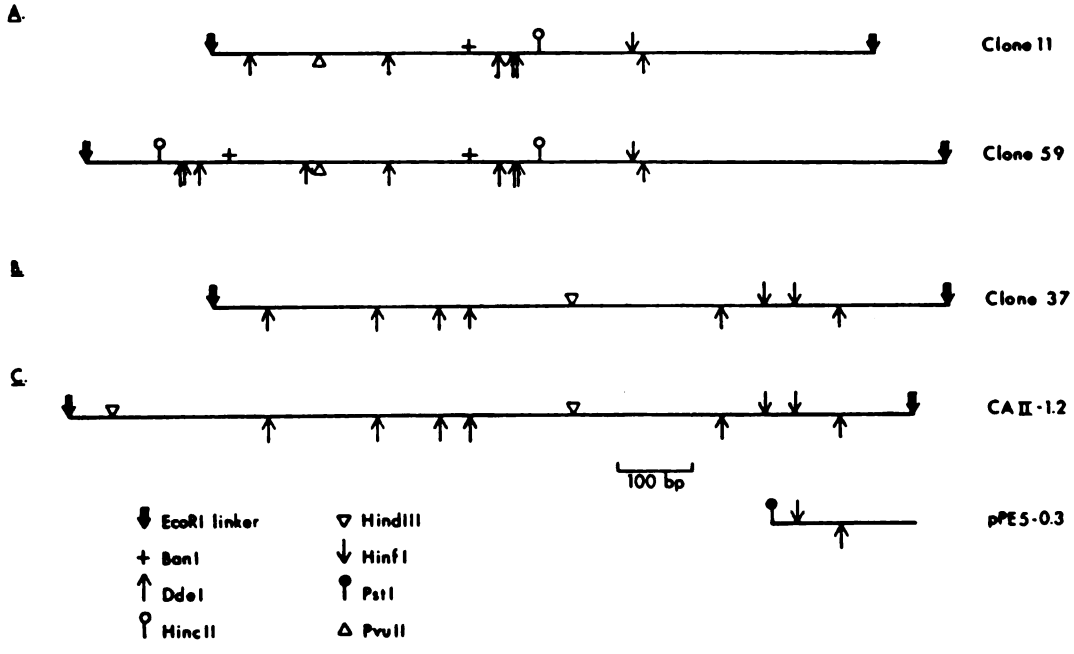

Figure 4. Restriction enzyme maps of chicken ∝D-globin and carbonic anhydrases II cDNA clones. Two ∝D globin cDNA clones, clones 11 and 59 (A), and two carbonic anhydrase II (CAII) cDNA clones, clone 37 (B) and clone CAII-1.2 (C), were isolated from a λgt10 reticulocyte cDNA library (Chapter I). The CAII clone, pPE5-0.3 (C) was isolated from a pBR322 reticulocyte cDNA library (Appendix 1).

Figure 5. Northern blot analysis of clones 11 and 59 (A), 37 (B), CAII-1.2 (C), and 22 (1350 bp) (D). Clones 11 and 59 code for αD-globin, clones 37 and CAII-1.2 code for carbonic anhydrase II, and clone 22 (1350 bp) is unidentified. In (A-C), 2 μg of poly(A)$^+$RNA from reticulocytes (R), AEV-transformed erythroblasts (E), and pancreas (P), were run on 1.2% agarose formaldehyde gels and Northern blots were prepared. The filters were probed with clone DNA ^{32}P-labelled by nick translation. In (D), 12 μg of poly(A)$^+$RNA from reticulocytes and erythroblasts were screened as above.

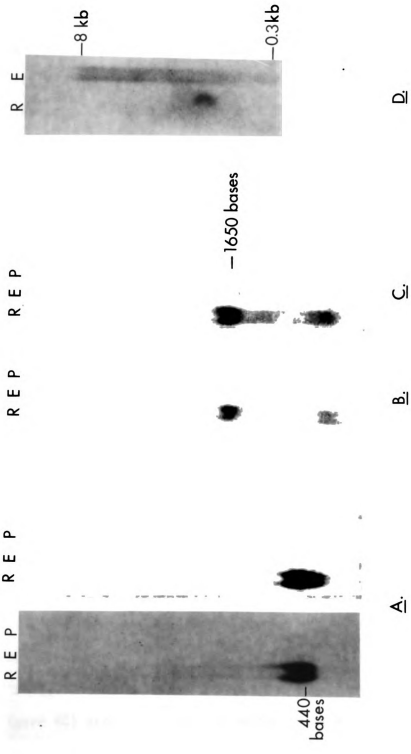

Figure 5.

Clone Characterization

Initially, duplicate Southern blots of insert DNA of 60 clones were probed with ^{32}P-labelled globin sequences and reticulocyte cDNA. Twelve reticulocyte-positive, globin-negative clones were chosen for further characterization. Ten of the twelve clones cross-hybridized to each other. Restriction mapping showed that all their sequences were contained on two partially overlapping clones, 11 (800 bp insert) and 59 (1200 bp insert) shown in Figure 4A. Northern analysis showed an approximate 440 base reticulocyte-specific RNA hybridized to these two clones (Figure 5A). Both clones were identified by DNA sequencing as containing αD globin sequences although both had unknown sequences flanking each end of the αD cDNA sequence. Clone 11 contained a virtually complete αD cDNA sequence but rather than ending in the expected poly(A) sequence run, the αD sequence ends in a run of T residues (Figure 6). This suggests an unusual inversion or recombination event took place during the preparation of the double strand cDNA in this clone. Clone 59 contains a complete αD cDNA sequence part of which was duplicated and reversed (Figure 7). Similar artifactual rearrangements of DNA sequences during cDNA cloning procedures have been noted in the literature (19).

One of the two non-globin clones in this set of twelve, clone 37, hybridized to a 1650 base reticulocyte-specific RNA (Figure 5B). Clone 37's restriction map is shown in Figure 4B. Similar restriction maps (Figure 4C) and identical hybridization to a 1650 base reticulocyte-specific RNA (Figure 5C) were observed with carbonic anhydrase II (CAII) cDNA clones previously isolated from the pBR322 reticulocyte cDNA library (Chapter I, Appendix I) and the λgt10 cDNA

Figure 6. The nucleotide sequence of clone 11. Clone 11 is a cDNA clone coding in part for αD-globin. The initial ATG of the αD sequence (ATG) and the truncated polyadenylation signal (AATTTT) are marked. The schematic (under the nucleotide sequence) comparing clone 11 to the αD mRNA sequence illustrates an unusual cDNA cloning anomoly. The αD coding strand (▭;▬ non-coding strand) ends in the middle of the polyadenylation signal and an expected run of (A)s is found in the opposite strand. The 20 bp of the αD mRNA coding sequence between the truncated polyadenylation signal and the poly(A) run have been lost. The αD region is also flanked by unidentified DNA sequences (▨).

```
1                   20                  40                  60                  80                  100
ACCACCTACCCCCCGACCAAGACCTACTTCCCCCACTTCGTTCTGTCACACGGCTCCGCTCAGATCAAGGGGCACGGCAAGAAGGTAGTGGCTGCCTTGA
TGGTGGATGGGGGGCTGGTTCTGGATGAAGGGGGTGAAGCAAGACAGTGTGCCGAGGCGAGTCTAGTTCCCCGTGCCGTTCTTCCATCACCGACGGAACT

                    120                 140                 160                 180                 200
TCGAGGCTGCCAACCACATTGCCACCAGCCACCAGCCCGCCCCACCAGCTGCCACCATGCTGACTGCCGAGGACAAGAAGCTCATCCAGCAGGCATGGGA
AGCTCCGACGGTTGGTGTAACGGTGGTCGGTGGTCGGGCGGGGTGGTCGACGGTGGTACGACTGACGGCTCCTGTTCTTCGAGTAGGTCGTCCGTACCCT

                    220                 240                 260                 280                 300
GAAGGCCGCTTCCCACTCGGAGGAGTTTGGAGCTGAGGCTCTGACTAGGATGTTCACCACCTATCCCCAGACCAAGACCTACTTCCCCCACTTCGACCTT
CTTCCGGCGAAGGGTGAGCCTCCTCAAACCTCGACTCCGAGACTGATCCTACAAGTGGTGGATAGGGGTCTGGTTCTGGATGAAGGGGGTGAAGCTGGAA

                    320                 340                 360                 380                 400
TCGCATGGCTCTGACCAGGTCCGTGGCCATGGCAAGAAGGTGTTGGGTGCCCTAGGCAACGCCGTGAAGAACGTGGACAACCTCAGCCATGCCATGGCTG
AGCGTACCGAGACTGGTCCAGGCACCGGTACCGTTCTTCCACAACCCACGGGATCCGTTGCGGCACTTCTTGCACCTGTTGGAGTCGGTACGGTACCGAC

                    420                 440                 460                 480                 500
AGCTGAGCAACCTGCATGCCTACAACCTGCGTGTTGACCCCGTCAATTTCAAGCTGTTGTCGCAGTGCATCCAGGTGGTGCTGGCTGTACACATGGGCAA
TCGACTCGTTGGACGTACGGATGTTGGACGCACAACTGGGGCAGTTAAAGTTCGACAACAGCGTCACGTAGGTCCACCACGACCGACATGTGTACCCGTT

                    520                 540                 560                 580                 600
AGACTACACCCCTGAAGTGCATGCTGCCTTCGACAAGTTCCTGTCTGCCGAGTCTGCTGTGCTGGCTGAGAAGTACAGATAAGCCACCATCTACAACTTC
TCTGATGTGGGGACTTCACGTACGACGGAAGCTGTTCAAGGACAGACGGCTCAGACGACACGACCGACTCTTCATGTCTATTCGGTGGTAGATGTTGAAG

                    620                 640                 660                 680                 700
AAGTCTTCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TTCAGAAGTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

                    720                 740                 760                 780                 800
TTTTTTTTTTTTTTTTTTTTTGAGGAAAGCAACAAACAATCCTGCATGTTTTATTGATATTGAAGTACAGTCCCATTAAAATTGGCTAAATGTCACTTAA
AAAAAAAAAAAAAAAAAAAACTCCTTTCGTTGTTTGTTAGGACGTACAAAATAACTATAACTTCATGTCAGGGTAATTTTAACCGATTTACAGTGAATT

                    820                 840                 860
AAAAAAAAAACAACACTGCAATTCAATACACAAAACCCTTGAATAACAGTGTGATCAGTCAGGA
TTTTTTTTTTTGTTGTGACGTTAAGTTATGTGTTTTTGGGAACTTATTGTCACACTAGTCAGTCCT
```

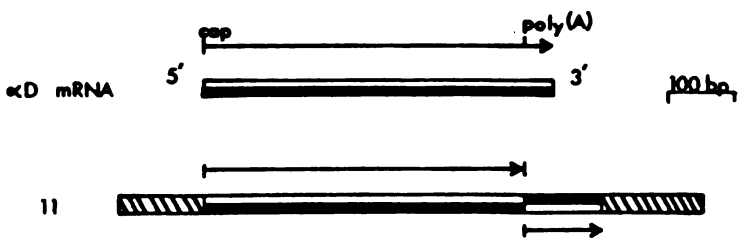

Figure 6.

Figure 7. The nucleotide sequence of clone 59. Clone 59 is a cDNA clone coding in part for αD-globin. The initial ATG of the αD sequence (ATG) and the polyadenylation site (AATAAA) are marked. The schematic (under the nucleotide sequence) comparing clone 59 to the αD mRNA sequence, illustrates another cDNA cloning anomoly. A complete D coding region (▭ coding strand, ▬ non-coding strand) is contained in clone 59. A 300 bp portion of the coding region (▭▭▭) is repeated and found at the 5' end of the clone in the reverse orientation. An unidentified DNA sequence (▭▭) follows a run of (A)s at the 3' end of the clone.

```
1                                                  50                                                 100
GGCAGCCATGCACTTCAGGGGTGTAGTCTTTGCCCATGTGTACAGCCAGCACCACTTGGATGCACTGCGACAACAGCTTGAAATTGACGGGGTCAACACG
CCGTCGGTACGTGAAGTCCCCACATCAGAAACGGGTACACATGTCGGTCGTGGTGAACCTACGTGACGCTGTTGTCGAACTTTAACTGCCCCAGTTGTGC

                                                   150                                                200
CAGGTTGTAGGCATGCAGGTTGCTCAGCTCAGCCATGGCATGGCTGAGGTTGTCCACGTTCTTCACGGCGTTGCTAGGGCACCCAACACCTTCTTGCCAT
GTCCAACATCCGTACGTCCAACGAGTCGAGTCGGTACCGTACCGACTCCAACAGGTGCAAGAAGTGCCGCAACGATCCCGTGGGTTGTGGAAGAACGGTA

                                                   250                                                300
GGCCACGGACCTGGTCAGAGCCAGGCGAAAGGTCGAAGTGGGGGAAGTAGGTCTTGGTCTGGGGATAGGTGGTGATCATCCTAGTCAGAGCCTAGCCTCC
CCGGTGCCTGGACCAGTCTCGGTCCGCTTTCCAGCTTCACCCCCTTCATCCAGAACCAGACCCCTATCCACCACTAGTAGGATCAGTCTCGGATCGGAGG

                                                   350                                                400
AGCCACCAGCCCGCCCCACCGCTTGCCACGATGCTGACTGCCGAGGACAAGAAGCTCATCCAGCAGGCATGGGAGAAGGCCGCTTCCCACTAGGAGGAGT
TCGGTGGTCGGGCGGGGTGGCGAACGGTGCTACGACTGACGGCTCCTGTTCTTCGAGTAGGTCGTCCGTACCCTCTTCCGGCGAAGGGTGATCCTCCTCA

                                                   450                                                500
TTTGGAGCTGAGGCTCTGACTAGGATGTTCACCACCTATCCCCAGACCAAGACCTACTTCCCCCACTTCGACCTTTCGCCTGGCTCTGACCAGGTCCGTG
AAACCTCGACTCCGAGACTGATCCTACAAGTGGTGGATAGGGGTCTGGTTCTGGATGAAGGGGGTGAAGCTGGAAAGCGGACCGAGACTGGTCCAGGCAC

                                                   550                                                600
GCCATGGCAAGAAGGTGTTGGGTGCCCTAGGCGGCGCCGTGAAGAACGTGGACAACCTCAGCCAGGCCATGGCTGAGCTGAGCAACCTGCATGCCTACAA
CGGTACCGTTCTTCCACAACCCACGGGATCCGCCGCGGCACTTCTTGCACCTGTTGGAGTCGGTCCGGTACCGACTCGACTCGTTGGACGTACGGATGTT

                                                   650                                                700
CCTGCGTGTTGACCCCGTCAATTTCAAGCTGTTGTCGCAGTGCATCTAGGTGGTGCTGGCTGTACACATGGGCAAAGACTACACCCCTGAAGTGCATGCT
GGACGCACAACTGGGGCAGTTAAAGTTCGACAACAGCGTCACGTAGATCCACCACGACCGACATGTGTACCCGTTTCTGATGTGGGGACTTCACGTACGA

                                                   750                                                800
GCCTTCGACAAGTTCCTGTCTGCCGTGTCTGCTGTGCTGGCTGAGAAGTACAGATAAGCCACCATCTACAACTTCAAGTCTTCAATAAAGACACCATTGC
CGGAAGCTGTTCAAGGACAGACGGCACAGACGACACGACCGACTCTTCATGTCTATTCGGTGGTAGATGTTGAAGTTCAGAAGTTATTTCTGTGGTAACG

                                                   850                                                900
TGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATACCATTCATCAATG
ACGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATGGTAAGTAGTTAC

                                                   950                                                1000
TCCTTGATATCGTAGGGTTCACAAGCTTCACACCACGAACCAGCTCTAACCATGCTAATCAATGGATATTTCCCTTTGGGCTCAACGTGCCCAACAAATC
AGGAACTATAGCATCCCAAGTGTTCGAAGTGTGGTGCTTGGTCGAGATTGGTACGATTAGTTACCTATAAAGGGAAACCCGAGTTGCACGGGTTGTTTAG

                                                   1050                                               1100
TAACATTCGAATCAGAGGTGCCATTGAGCAGCCAGTCAACACTTACGCCACAGACTGACGCAAGTTCTGGTAAAAAGCGTGGTCGCTTAGTTTTACCGTT
ATTGTAAGCTTAGTCTCCACGGTAACTCGTCGGTCAGTTGTGAATGCGGTGTCTGACTGCGTTCAAGACCATTTTTCGCACCAGCGAATCAAAATGGCAA

                                                   1150
TTCGAGCTGCTCTATAGACTGCTGGGTAGTCCCCACCTTTTGAGCAAGTTCAGCCTAGTTAAGTCCAAGCTGA
AAGCTCGACGACATATCTGACGACCCATCAGGGGTGGAAAACTCGTTCAAGTCGGATCAATTCAGGTTCGACT
```

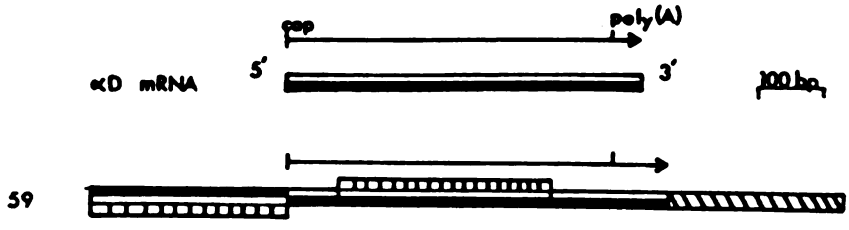

Figure 7.

library. Clone 37 was positively identified as coding for CAII by comparing a partial nucleotide sequence of the clone 37 insert to the known CAII sequence (J.D. Lee, personal communication; data not shown).

The final clone of the twelve, clone 22, contained two EcoRI fragments, 1350 bp and 700 bp. The 700 bp fragment hybridized to globin sequences. The 1350 bp fragment, on Northern blots, hybridized to a broad smear of erythroblast RNAs (300 bases - 8 kb) but the hybridization to the lower portion of this smear (300 bases - 1.8 kb) was only faint (Figure 5D). A repeated sequence in the 1350 bp fragment of clone 22 may be the source of this hybridization. Clone 22 was not studied further.

The screening protocol used above was modified in an attempt to more effectively identify the contaminating globin clones. All 220 clones of the sublibrary were screened for globin sequences by phage dot blots, using clone 11 as a probe for αD, and the SP6 plasmids of αA and β. A duplicate filter was also probed with ^{32}P-labelled reticulocyte cDNA. Because primary plaque lysates were used for the phage dot blots, some variation in hybridization intensity may result from differential growth on the plate from which filters were pulled. 140 clones (64%) showed strong hybridization to the globin probe. Purified DNA was isolated from sixty apparently non-globin clones. These DNAs were digested with EcoRI and electrophoresed on agarose gels. Southern blots were prepared from the gels and screened with the probes described above. Another fifteen of these clones hybridized to the globin probe totalling 155 of 220 or 71%.

To assess what proportion of the contaminating globin clones code for αD, phage dot blots of the original 60 clones characterized were

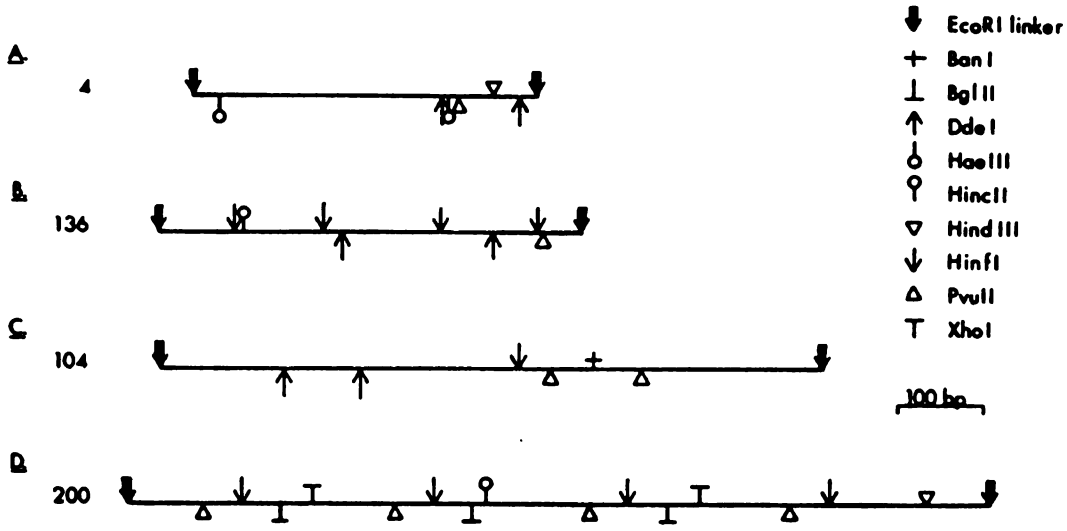

Figure 8. Restriction enzyme maps of clones 4 (A), 136 (B), 104 (C), 200 (D). Clones 4 and 136 are unidentified, clone 136 codes for ferritin heavy chain, and clone 200 codes for ubiquitin.

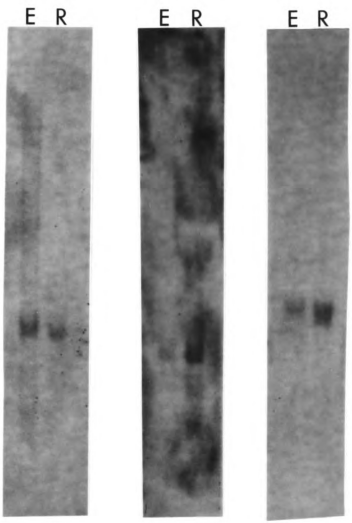

Figure 9. Northern blot analysis of clones 4 (A), 104 (B), and 200 (C). Clones 4 is unidentified, clone 104 codes for ferritin heavy chain, and clone 200 codes for ubiquitin. 5 μg of poly(A)$^+$RNA from erythroblasts (E) and reticulocytes (R) were run on 1.2% agarose formaldehyde gels and Northern blots were prepared. The filters were probed with clone DNA ^{32}P-labelled by nick translation. The high background in (B) was due to technical difficulties with the clone 104 probe.

probed with ^{32}P-labelled clone 11. Approximately 40 of 60 or 67% of the clones hybridized strongly to the αD sequence. Again, remember there can be some variation in hybridization intensity due to differential growth in the phage dot blot procedures. Since αD clones represent approximately 67% of the enriched sublibrary (94% of the globin clones), it is clear that the enrichment protocol used enriched for αD sequences as well as non-globin sequences. This may be due to a somewhat different optimal hybridization temperature for αD RNA:cDNA duplexes from that used in the subtraction protocol.

Six non-globin clones which showed positive hybridization to reticulocyte cDNA were chosen for further characterization. Initially, their cDNA inserts were subcloned into pUC8 plasmid DNA and the restriction maps of the resultant subclones were determined (Figure 8). ^{32}P-labelled plasmid clones were then used to probe Northern blots of reticulocyte and erythroblast poly(A)$^+$RNA (Figure 9), and Southern blots of HindIII-digested total chicken DNA (Figure 10). The nucleotide sequence of at least part of each cDNA insert was determined. Amino acid sequences were deduced from all open reading frames of the nucleotide sequence and these were then compared to the Protein Sequence Database in an attempt to identify the clones.

The identity of two clones, 4 and 136, were never positively established. The restriction map of clone 4 insert (400 bp) is shown in Figure 8A. Clone 4 hybridizes to a 1050 base RNA present in both reticulocytes and erythroblasts (Figure 9A). The 1050 base RNA may be slightly more abundant in erythroblasts. Clone 4 hybridized to two HindIII fragments of chromosomal DNA, 4.3 and 1.2 kb in size (Figure 10A). The nucleotide sequence of the entire clone 4 insert is shown in

```
1                   20                  40                  60                  80
GCAGAGGTTTACCGCAAGCATATTATGGGCCAAAACGTGGCAGATTGCATGCGTTACCTGATGGAGGAGGATGAAGATGC
CGTCTCCAAATGGCGTTCGTATAATACCCGGTTTTGCACCGTCTAACGTACGCAATGGACTACCTCCTCCTACTTCTACG

                    100                 120                 140                 160
TTACAAAAAACAGTTCTCCCAGTACATAAAGAACAATATAACGCCTGATGGGATGGAAGAAATGTATAAAAAAGCCCATG
AATGTTTTTTGTCAAGAGGGTCATGTATTTCTTGTTATATTGCGGACTACCCTACCTTCTTTACATATTTTTTCGGGTAC

                    180                 200                 220                 240
CTGCTATACGGGATAATCCAGTCCATGAAAAGAAACCCAAGAGAGAAGCCAAGAAAAAGAGAGTGAACAGTACCAAGATG
GACGATATGCCCTATTAGGTCAGGTACTTTTCTTTGGGTTCTCTCTTCGGTTCTTTTTCTCTCACTTGTCATGGTTCTAC

                    260                 280                 300                 320
TCCCTCGCCCAGAAGAAAGACCGTGTTGCTCAGAAGAAGGCTAGCTTCCTCAGGGCCCAGGAGCGTGCAGCTGACAGTTA
AGGGAGCGGGTCTTCTTTCTGGCACAACGAGTCTTCTTCCGATCGAAGGAGTCCCGGGTCCTCGCACGTCGACTGTCAAT

                    340                 360                 380                 398
AGACAACTTCCTCAATTTTCTGATGAAGCTTGGAAAAAGTGGCAATAAATTTATTGACTAAGAAAAAGAAAAAAAAAG
TCTGTTGAAGGAGTTAAAAGACTACTTCGAACCTTTTTCACCGTTATTTAAATAACTGATTCTTTTTCTTTTTTTTTC
```

Figure 11. Nucleotide sequence of clone 4 (Unidentified).

```
1                   20                  40                  60                  80
AAAAAAAAACAAACAGTGCTACTTCTAGACCACCACCAAATCGTACAGATAAAGAAAAAGGATATATGACTGATGATGGT
TTTTTTTTTGTTTGTCACGATGAAGATCTGGTGGTGGTTTAGCATGTCTATTTCTTTTTCCTATATACTGACTACTACCA

                    100                 120                 140                 160
CCGGGACTCGGAAGGGGAGTCGACCTTACAGAAACAGAGGGCATAGCAGACGTGGTCTAGGCTACGCTTCGGGAACTAAT
GGCCCTGAGCCTTCCCCTCAGCTGGAATGTCTTTGTCTCCCGTATCGTCTGCACCAGATCCGATGCGAAGCCCTTGATTA

                    180                 200                 220                 240
TCTGAAGCATCAAATGCTTCTGAAACGGAATCTGATCACAGAGATGAACTTAGCGATTGGTCATTAGCCCAAGCAGAAGA
AGACTTCGTAGTTTACGAAGACTTTGCCTTAGACTAGTGTCTCTACTTGAATCGCTAACCAGTAATCGGGTTCGTCTTCT

                    260                 280                 300                 320
GGAACGAGACAACTACCTTCGCAGGGGAGATGGGCGAAGACGTGGAGGTGGTGCAAGAGGGCAAGGAGGGAGGGGTCGTG
CCTTGCTCTGTTGATGGAAGCGTCCCCTCTACCCGCTTCTGCACCTCCACCACGTTCTCCCGTTCCTCCCTCCCCAGCAC

                    340                 360                 380                 400
GAGGATTCAAAGTTAATTCTTTTTGCAGGAAATGATGAGCAGTCACGAACAGATAACCGTCAGCGTAATCCAAGAGATGC
CTCCTAAGTTTCAATTAAGAAAAACGTCCTTTACTACTCGTCAGTGCTTGTCTATTGGCAGTCGCATTAGGTTCTCTACG

                    420                 440                 460                 480
AAAAGGGAGAACAGCAGAATGGATCTCTACTGGTAATTTTTATCTTCTTAGAACTATGCACTGTTTAAAATTTCATGAGA
TTTTCCCTCTTGTCGTCTTACCTAGAGATGACCATTAAAAATAGAAGAATCTTGATACGTGACAAATTTTAAAGTACTCT

                    500                 520                 540                 560
CATTATATACTAACAGTAGATTCTTCAGAACAGCAGCTGATACAGTAGCCTTGTGGTTCCTCACTTTCTAATTGCATTTACAA
GTAATATATGATTGTCATCTAAGAAGTCTTGTCGTCGACTATGTCATCGGAACACCAAGGAGTGAAAGATTAACGTAAATGTT
```

Figure 12. Nucleotide sequence of clone 136 (Unidentified).

Figure 11. Two amino acid sequences, deduced from two possible open reading frames of clone 4, were compared to the protein database. No significant correlations were found (data not shown).

The restriction map of clone 136 (580 bp insert) is shown in Figure 8B. Clone 136 did not detectably hybridize to either reticulocyte or erythroblast RNA (data not shown). Clone 136 did hybridize to a 4.0 kb HindIII fragment of chromosomal DNA (Figure 10B). The nucleotide sequence of clone 136 was determined (Figure 12). An amino acid sequence was deduced from one long open reading frame and compared to the protein database. No significant correlations were found (data not shown). Other amino acid sequences were generated from other small open reading frames of the clone 136 nucleotide sequence, but, again, no significant correlations were found (data not shown).

Three clones; 74, 104 and 112; gave identical restriction maps (Figure 8C). Clone 104 was characterized further and found to hybridize to a 950 base RNA in both reticulocytes and erythroblasts (Figure 9C). The abundance of this 950 base RNA is 2-10 fold higher in reticulocytes as compared to erythroblasts. Clone 104 hybridized to two HindIII fragments of chromosomal DNA, 7.0 and 4.5 kb in size (Figure 10C). The entire nucleotide sequence of the clone 104 insert was determined (Figure 13). One amino acid sequence, deduced from a long open reading frame in the nucleotide sequence, was compared to the protein database and showed a 91% homology to the human ferritin heavy chain protein sequence (Figure 14). A collaborator, J.D. Engel (Northwestern University), had concurrently identified a genomic ferritin heavy chain clone among a series of reticulocyte-positive clones isolated by J. Dodgson. This original ferritin heavy chain

```
                                1                                  10
                                MET ALA THR PRO PRO SER GLN VAL ARG GLN ASN TYR HIS GLN ASP CYS
GCCGACGCGCTCTCTCTTTCTCTCCCGCCAGCGCC ATG GCT ACG CCT CCT TCC CAG GTG CGC CAG AAC TAC CAC CAG GAC TGC

          20                                30                                  40
GLU ALA ALA ILE ASN ARG GLN ILE ASN LEU GLU LEU TYR ALA SER TYR VAL TYR LEU SER MET SER TYR TYR PHE
GAA GCC GCC ATC AAC CGG CAG ATC AAC CTG GAG CTG TAC GCC TCC TAC GTG TAC CTC AGC ATG TCC TAC TAT TTT

                           50                                 60
ASP ARG ASP ASP VAL ALA LEU LYS ASN PHE ALA LYS TYR PHE LEU HIS GLN SER HIS GLU GLU ARG GLU HIS ALA
GAC CGG GAT GAT GTG GCT CTG AAA AAC TTT GCC AAG TAC TTC CTG CAC CAG TCC CAC GAG GAG CGT GAA CAT GCT

          70                                80                                  90
GLU LYS LEU MET LYS LEU GLN ASN GLN ARG GLY GLY ARG ILE PHE LEU GLN ASP ILE LYS LYS PRO ASP ARG ASP
GAG AAG CTG ATG AAG CTG CAA AAC CAG AGG GGT GGA CGC ATC TTC TTG CAG GAC ATC AAG AAA CCG GAT CGT GAT

                          100                                110
ASP TRP GLU ASN GLY LEU THR ALA MET GLU CYS ALA LEU HIS LEU GLU LYS ASN VAL ASN GLN SER LEU LEU GLU
GAC TGG GAG AAT GGA CTG ACT GCA ATG GAG TGT GCC CTG CAC CTA GAG AAG AAT GTG AAC CAG TCG CTG TTA GAG

          120                               130                                 140
LEU HIS LYS LEU ALA THR GLU LYS ASN ASP PRO HIS LEU CYS ASP PHE ILE            THR LEU ASP GLU GLN
CTG CAC AAA TTG GCA ACT GAA AAG AAT GAC CCA CAC TTG TGT GAC TTC ATT ANN NNN NNT ACC TTG GAT GAG CAG

                          150                                160
VAL LYS ALA ILE LYS GLN LEU GLY ASP HIS VAL THR ASN LEU ARG LYS MET GLY ALA PRO LYS TYR GLY MET ALA
GTG AAA GCC ATC AAG CAG CTG GGT GAC CAT GTG ACC AAC CTG CGG AAG ATG GGG GCA CCC AAG TAT GGC ATG GCA

          170                               180
GLU TYR LEU PHE ASP LYS HIS THR LEU GLY GLU SER ASP SER
GAG TAC CTG TTT GAC AAG CAC ACC CTC GGG GAA AGT GAC AGC TGA AGGCTTGCCAAGAACCACCCTGTGGGTTTCAGGGGGACT


TCTACTTCACTGGTCCAGCAATGCATGCATCTTCAGCTATCTAGATAAACAGTTCTTTTACTCTGTACCAAATATCAGCCCTCTTTTCTCTTGTGTTTT
```

Figure 13. Nucleotide and amino acid sequences of clone 104. Clone 104 was identified as coding for ferritin heavy chain.

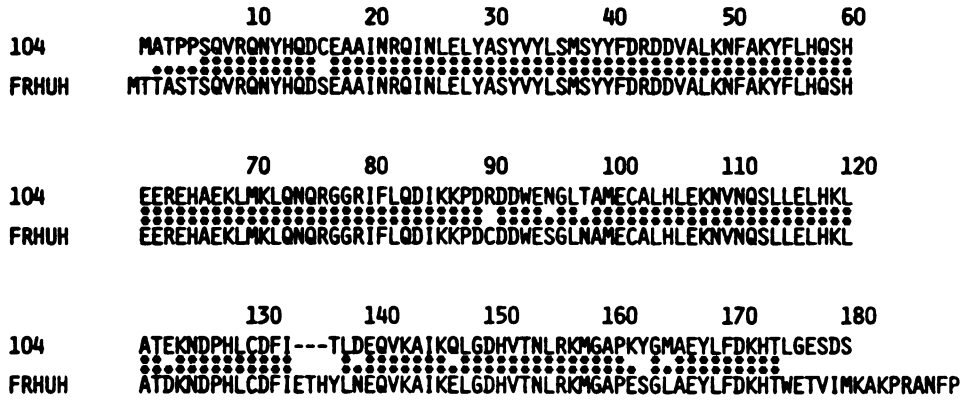

Figure 14. A comparison of the amino acid sequences of clone 104 and human ferritin heavy chain. The deduced amino acid sequence of clone 104 (Figure 13), when compared to the Protein Sequence Database by the program FASTP, was found to be 91% homologous to human ferritin heavy chain (FRHUH) ((*) homologous amino acid, (⁑) identical amino acid).

clone was used by the Engel lab to isolate further cDNA and genomic ferritin heavy chain clones. Clone 104 was sent to the Engel lab for any further characterization.

The restriction map of clone 200 is shown in Figure 8D. Clone 200 hybridized to a 1200 base RNA in both reticulocytes and erythroblasts (Figure 9D). The abundance of this 1200 base RNA is 2-10 fold higher in reticulocytes than erythroblasts. Clone 200 hybridized to a 5.5 kb HindIII fragment of chromosomal DNA (Figure 10D). An amino acid sequence of an open reading frame, deduced from a partial nucleotide sequence of clone 200, was compared to the protein database and positively identified as ubiquitin. The complete nucleotide sequence of clone 200 and the genomic organization of the chicken ubiquitin loci are described in Chapter III.

The enrichment of the reticulocyte probe was not as high as expected. The enriched sublibraries obviously contained far too many globin and HD3-expressed sequences for the subtraction steps to have been maximally effective. Nevertheless, some non-globin clones were isolated, several of which were more highly expressed in reticulocytes than in HD3 cells (Figure 10). The identity of some of these clones has been established. Some aspects of the subtraction procedures worked well. For example, the recovery of cDNA in both subtractions was near 95%, even though small amounts of cDNA were being handled. Other aspects of the procedure, however, need to be improved. A potentially serious problem may relate to the differences in optimal hybridization conditions for different mRNAs. This may have led to the high level of αD-globin clone contamination in the final sublibraries. Perhaps hybridizations could be started at a somewhat higher

temperature, and the temperature then lowered very slowly over a 10-15° range. Since recoveries were 95%, additional subtraction steps could probably be added to further enrich for specific sequences. Characterization of clones from the 800 bp - 3 kb cDNA library have also shown that the agarose gel size fractionation of the reticulocyte cDNA in the cloning procedure (Chapter I) was not adequate. The reasons for this remain unclear. Perhaps other size fractionation methods (gel exclusion chromatography, acrylamide gels) could be used. One must also be aware of potential cDNA cloning artifacts as seen with the two αD globin clones, 11 and 59. This is most important when trying to identify clones from deduced amino acid sequences, or from the size of the cDNA insert. It is concluded that the subtraction procedure described in this Chapter could be used to recover more reticulocyte-specific or reticulocyte-enriched cDNA clone sequences, but only with considerable improvements in the specific techniques employed.

REFERENCES

1. Longacre, S.S. and W.J. Rutter. 1977. J. Biol. Chem. 252:2742-2752.

2. Beug, H., S. Palmieri, C. Freudenstein, H. Zentgraf and T. Graf. 1982. Cell. 28:907-919.

3. Graf, T., B. Royer-Pokora, E. Korzenilwska, S. Grieser and H. Beug. 1981. Haematology and Blood Transfusion - Modern Trends in Human Leukemia IV. 26:417-423.

4. Adkins, B., H. Beug and T. Graf. 1983. J. Cell. Biochem. Supplement 7B:21.

5. Melton, D.A., P.A. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn and M.R. Green. 1984. Nucleic Acids Res. 12:7035-7056.

6. Davis, M.M., D.I. Cohen, E.A. Nielson, M. Steinmetz, W.E. Paul and L. Hood. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:2194-2198.

7. Beug, A., G. Doedeilein, C. Freudenstein and T. Graf. 1982. J. Cell. Physiol. Supplement I: 195-207.

8. Lipman, D.J. and W.R. Pearson. 1985. Science. 227:1435-1441.

9. Benton, W.D. and R.W. Davis. 1977. Science. 196:180.

10. Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. In Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory.

11. Girvitz, S.C., S. Bacchetti, A.J. Rainbow and F.L. Graham. 1980. Anal. Biochem. 106:482.

12. Vieira, J. and J. Messing. 1982. Gene 19:259-268.

13. Maxam, A.M. and W. Gilbert. 1977. Proc. Natl. Acad. Sci. U.S.A. 74:560-564.

14. Smith, D.R. and J.M. Calvo. 1980. Nucleic Acids Res. 8:2255-2274.

15. Dodgson, J.B. and J.D. Engel. 1983. J. Biol. Chem. 258:4623-4629.

16. Fischer, H.D., J.B. Dodgson, S. Hughes and J.D. Engel. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:2733-2737.

17. Dolan, M., J.B. Dodgson and J.D. Engel. 1983. J. Biol. Chem. 258:3983-3990.

18. Lasky, L., N.D. Nozick and A.J. Tobin. 1978. Develop. Biol. 67:23-39.

19. Volckaert, G., F. Tavernier, R. Derynak, R. Devas and W. Fiess. 1981 Gene. 15:215-223.

CHAPTER III

THE GENOMIC ORGANIZATION OF CHICKEN UBIQUITIN

Ubiquitin is a small protein, originally isolated from bovine thymus (1), consisting of 76 amino acids (Mr=8451). Ubiquitin has since been found in all eucaryotic cells investigated, and its amino acid sequence is identical in all species studied from insects to humans, making it one of the most conserved proteins known (2). Ubiquitin has been implicated in a variety of cellular functions in the nucleus, the cytoplasm, and at the cell surface (3). The molecular mechanisms of these proposed physiological functions of ubiquitin have not been thoroughly elucidated, however.

In the nucleus, ubiquitin is found conjugated to histones H2A and H2B (4, 5). The conjugate is formed through an isopeptide bond between the C-terminal glycine carboxyl group of ubiquitin and the ε-amino group of a lysine. Lysine 119 of histone H2A is conjugated with ubiquitin to form nuclear protein A24 (4). Experimental evidence suggests that A24 preferentially associates with actively transcribed genes (6). Histone ubiquitination varies with the cell cycle; during mitosis none of the histones contains ubiquitin (7). A specific lyase cleaves the isopeptide bond of the conjugate to recycle the ubiquitinated histones. Ubiquitin therefore, may have roles in the regulation of chromatin structure and in gene expression.

In the cytoplasm, ubiquitin is part of the ATP-dependent, non-lysosomal proteolytic pathway that is responsible for most of the rapid turnover of intra-cellular proteins under normal metabolic conditions (2). Presumably, ubiquitin-conjugated proteins are recognized as a substrate by certain intracellular proteases. A requirement for an intact cellular ubiquitin conjugation system has been supported by experiments with a mouse cell cycle mutant, ts85 (8). At the non-permissive temperature, ts85 cells have a defective ubiquitin-activating enzyme (E1), which is the first enzyme of the conjugating system. These cells fail to degrade abnormal intracellular proteins which are otherwise short-lived. Ubiquitin, therefore, appears to have a role in the regulation of protein turnover, including the turnover of specific regulatory proteins, through inducible protein degradation.

Ubiquitin has recently been found conjugated to cell surface proteins (9). A monoclonal antibody, MEL-14, specifically recognizes the mouse lymph node homing receptor on lymphocytes, inhibiting their binding to lymph node cells (10). A λ fusion/expression B-cell lymphoma cDNA library was screened with the MEL-14 antibody. Only ubiquitin coding cDNA clones were isolated (11). The antigenic determinant defined by MEL-14 resides in the C-terminal 13 amino acids of ubiquitin, but the MEL-14 antibody does not detect intact native ubiquitin or other cellular ubiquitinated proteins. Independent antibodies to ubiquitin appear to identify additional ubiquitinated cell surface proteins. Two possible roles have been proposed for ubiquitination of cell surface proteins. The first hypothesis proposes that ubiquitin is required for rapid intracellular protein degradation,

by the mechanism described above, after the receptor is internalized by the cell. The second hypothesis specifically suggests a role for ubiquitination in the receptor recognition process. MEL-14 specifically recognizes only ubiquitin and completely inhibits the binding of lymphocytes to the target cell receptors. It is therefore suggested that ubiquitin provides an essential part of the receptor binding site, and that ubiquitination could provide a large repertoire of receptor surfaces according to the numbers and locations of attached ubiquitin moieties.

In the hope of understanding and defining the multiple biological functions of ubiquitin, ubiquitin genes and mRNA's have been cloned and characterized from several species: human (12, 13), chicken (14, 15), Xenopus (16), Drosophila (17) and Saccharomyces (18). Since a similar gene organization pattern is seen in these species, only the human and chicken ubiquitin genes will be discussed.

Human cells contain three size classes of ubiquitin-specific RNA; 650, 1100, and 2500 bases (12). One cDNA clone, corresponding in size to the 2500 base RNA species, contains nine direct repeats of the ubiquitin amino acid sequence. The ubiquitin repeats are unusual in that they do not contain spacer sequences separating the repeats, and their coding regions are not interrupted by intervening sequences. This polyprotein organization was verified by cloning the corresponding chromosomal DNA. The ubiquitin monomer amino acid sequence is exactly the same in all nine repeats, but there exist many silent nucleotide sequence differences between the repeat units. Several cDNA clones were isolated that correspond to the 650 base RNA. These clones contain a single ubiquitin amino acid sequence plus an unidentified 80

amino acid carboxyl-terminal extension (13). Clones corresponding to the 1100 bp RNA have not been isolated but presumably would code for three or four ubiquitin repeats. Upon probing HindIII-digested human DNA with a ubiquitin specific probe, at least 12 hybridizing fragments of varying intensities are present. Since no known human ubiquitin gene contains a HindIII site, these data suggest at least 12 different chromosomal DNA fragments exist which contain one or more ubiquitin coding sequence each.

Both cDNA and genomic clones coding for ubiquitin have been isolated from chicken embryo fibroblasts (CEF) by Bond and Schlesinger (14, 15). Initially, clone 7, a 600 bp cDNA clone, was isolated from a cDNA library prepared from heat shocked CEF poly(A)$^+$RNA (14). Clone 7 was identified as ubiquitin by DNA sequencing. Clone 7 hybridized to a 1200 base RNA in CEF grown at the normal temperature, but upon heat shock, the 1200 base RNA concentration was induced five fold and a new 1700 base hybridizing RNA was seen. The 600 bp of clone 7 code for the last four (73-76) amino acids of one ubiquitin unit followed by two complete ubiquitin-coding repeats. Clone 7 also has a 120 bp 3'-untranslated region ending in a poly(A) run. As observed with other ubiquitin clones, the clone 7 coding region does not contain spacer sequences between repeats.

Recently, Bond and Schlesinger identified two ubiquitin loci (ubiquitin I and II) by the analysis of chicken genomic DNA clones which hybridized to clone 7 (15). The two genes differ in their organization, and, as yet, do not appear to be linked. One gene (ubiquitin I) contains four ubiquitin coding repeats, while the second gene (ubiquitin II) contains three repeats. Unique probes from the 5'

and 3' untranslated regions of ubiquitin I and II were used to determine which gene codes for the various mRNA species in CEF. Ubiquitin I was found to code for both the 1200 base and 1700 base mRNA's at the appropriate growth temperatures. The larger mRNA contains an unprocessed 674 base intron located in the 5'-untranslated region of ubiquitin I. This is the first ubiquitin gene reported to have an intron. RNA processing enzymes are inactivated by heat shock which could account for the existence of the 1700 base RNA only in heat-shocked cells. Ubiquitin II does not appear to code for an RNA species in normal or heat shocked CEF.

We have described the isolation and partial characterization of a ubiquitin cDNA clone, clone 200, from chicken reticulocytes in Chapter II. Comparison of the nucleotide sequence of clone 200 and the CEF ubiquitin cDNA clone 7, showed at least 17 nucleotide differences. Since the recent data of Bond and Schlesinger was not yet available, ubiquitin genomic clones were isolated in order to understand the organization of the chicken ubiquitin genes by their hybridization to clone 200. In this chapter, we describe the structure of these clones. Two different ubiquitin genes were identified. In comparing our data to those recently published by Bond and Schlesinger, certain differences in the results and/or interpretation thereof are noted.

EXPERIMENTAL PROCEDURES

Genomic Library Screening

500,000 bacteriophage from the Charon 4A chicken genomic library (19) were grown on 10 plates of E. coli 803 supF. Duplicate nitrocellulose filter replicas were screened by the method of Benton and Davis (20). The filters were hybridized with clone 200 insert DNA, ^{32}P-labelled by nick translation (specific activity of 1.3 x 10^8 cpm/μg) at 42° in a 50% formamide hybridization solution described in Chapter II. The filters then were washed with 2 X SSC at 25° followed by 0.2 X SSC at 65°. Eleven positive hybridizing plaques were isolated in the primary screening. Nine of the eleven plaques were plaque purified and plate lysate stocks of each were prepared (21). DNA was prepared from bacteriophage particles purified by equillibrium centrifugation in cesium chloride according to the modified method of Yamamoto et al. (21). The purified particles were disrupted with 20 mM EDTA, 0.5% SDS and 50 μg/ml proteinase K with incubation at 65° for 1 hr. The solution was extracted with phenol:chloroform, and the aqueous phase dialyzed against TE.

Other Methods

All other methods were carried out as described previously (Chapters I and II).

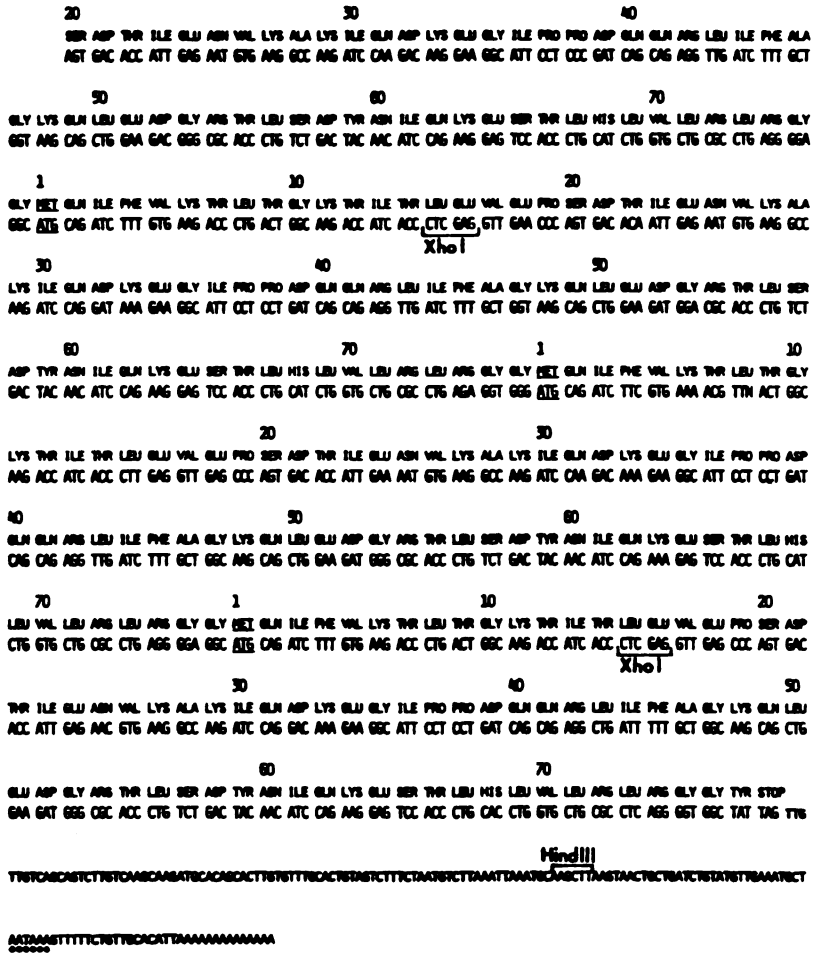

Figure 1. Nucleotide and amino acid sequences of clone 200. Clone 200 was identified as coding for ubiquitin. Clone 200 contains 3 3/4 ubiquitin coding regions, each numbered 1-76. The three ATG codons which begin each complete ubiquitin repeat and the polyadenylation signal (AATAAA) are marked.

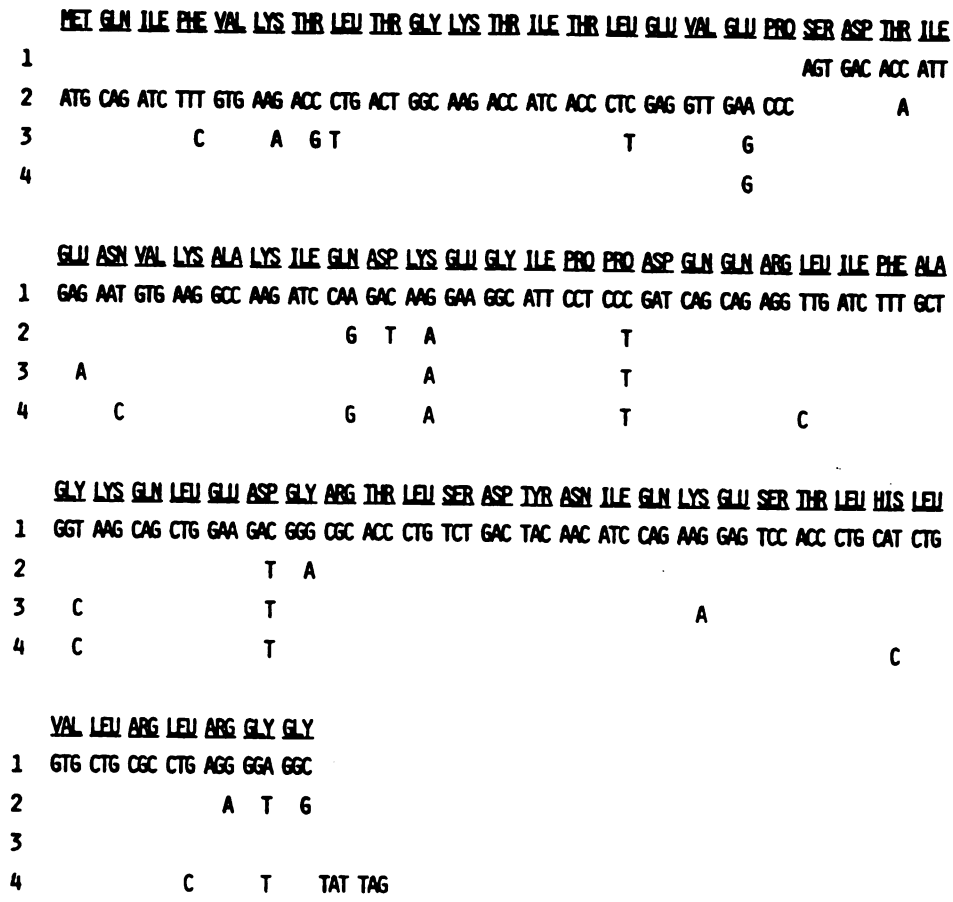

Figure 2. A comparison of the nucleotide sequences of clone 200's ubiquitin coding repeats. The first 228 nucleotides of clone 200 are listed as the prototype sequence of ubiquitin. The remaining ubiquitin repeat sequences are compared to the prototype sequence and only the sequence differences are given. Note, none of the nucleotide differences results in an amino acid change.

RESULTS AND DISCUSSION

The complete nucleotide sequence of the ubiquitin clone 200, isolated from the chicken reticulocyte λgt10 cDNA library (Chapter II), is shown in Figure 1. The nucleotide sequence encodes amino acids 20-76 of a ubiquitin coding region at its 5'-end, followed by three complete (76 amino acid) ubiquitin repeats. The final repeat is followed by an additional amino acid residue (TYR) preceding the stop codon. This non-ubiquitin C-terminal residue has been reported in ubiquitin genes of human (VAL), yeast (ASP) and chicken (TYR in ubiquitin I and ASN in ubiquitin II). Clone 200 has a 140 bp 3'-untranslated region that includes a putative polyadenylation site 20 bp 5' of a poly(A) run. The ubiquitin coding regions contain nucleotide sequence differences between repeats (33/854 or 4%) (Figure 2) but these differences do not result in amino acid changes. Characteristic restriction enzyme sites of clone 200 include two XhoI sites 456 bp apart in the ubiquitin coding repeats and the HindIII site in the 3'-untranslated region, 274 bp downstream of the 3'-XhoI site (Figure 2).

The nucleotide sequences of the two chicken ubiquitin cDNA clones, CEF clone 7 and reticulocyte clone 200, are compared in Figure 3. In the overlapping 600 bp, 17 nucleotide differences were found in the ubiquitin coding regions. The 3'-untranslated regions were identical for 70 bp and then diverged completely. Clone 7 does not contain a

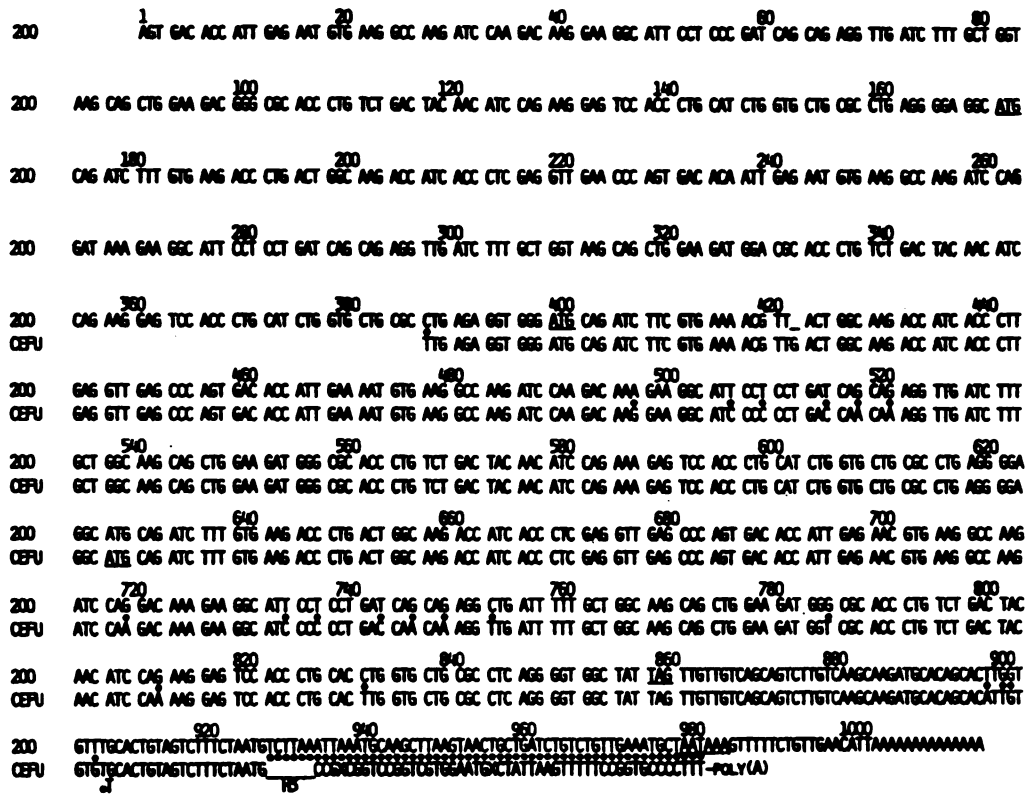

Figure 3. A comparison of the nucleotide sequences of two ubiquitin cDNA clones. The nucleotide sequences of clone 200, isolated from a reticulocyte cDNA library, and clone 7 (CEFU), isolated from a heat shocked chicken embryo fibroblast cDNA library and sequenced by Bond and Schlessinger (14), are compared over their overlapping 600 bp. Sequence differences are noted (*). The ATG's, the stop codon (TAG), and the polyadenylation signal (AATAAA) are marked.

polyadenylation signal but does end in a poly(A) run. Bond and Schlesinger have recently reported the nucleotide sequence of ubiquitin I and showed that clone 7 corresponds to a part of the ubiquitin I gene message (15). A comparison of the nucleotide sequences of clone 200 and the ubiquitin I genomic locus shows that their 3'-untranslated regions are virtually identical (1 bp difference) (Figure 4). There are seventeen nucleotide differences in the 854 bp overlap between clone 200 and the ubiquitin I ubiquitin coding regions; only one of these differences corresponds to a nucelotide difference noted between clones 200 an 7. One of the nucleotide differences between ubiquitin I and 200 eliminates a characteristic XhoI site, however, leaving only one XhoI site 274 bp 5' of the HindIII site. Since virtually all the nucleotide differences between clones 200 and 7 were not reproduced when comparing clone 200 and ubiquitin I, the sequence differences remaining may be attributable to DNA sequencing errors and/or allelic variations.

The two clones (clone 7 and ubiquitin I) sequenced by Bond and Schlesinger differ more from one another than either of them does from clone 200. Also, the 3'-untranslated sequence they report for clone 7 is clearly incorrect. Consequently, we presume that much of the variation involved is due to sequencing errors on their part. Errors in our data, allelic variation between the sources of chicken nucleic acids used, and sequence artifacts due to cloning may also contribute to these differences. Clone 200 most likely is a cDNA clone corresponding to the ubiquitin I gene message since Bond and Schlesinger reported only a 20% homology between the 3'-untranslated sequences of ubiquitin I and II. The large divergence between the

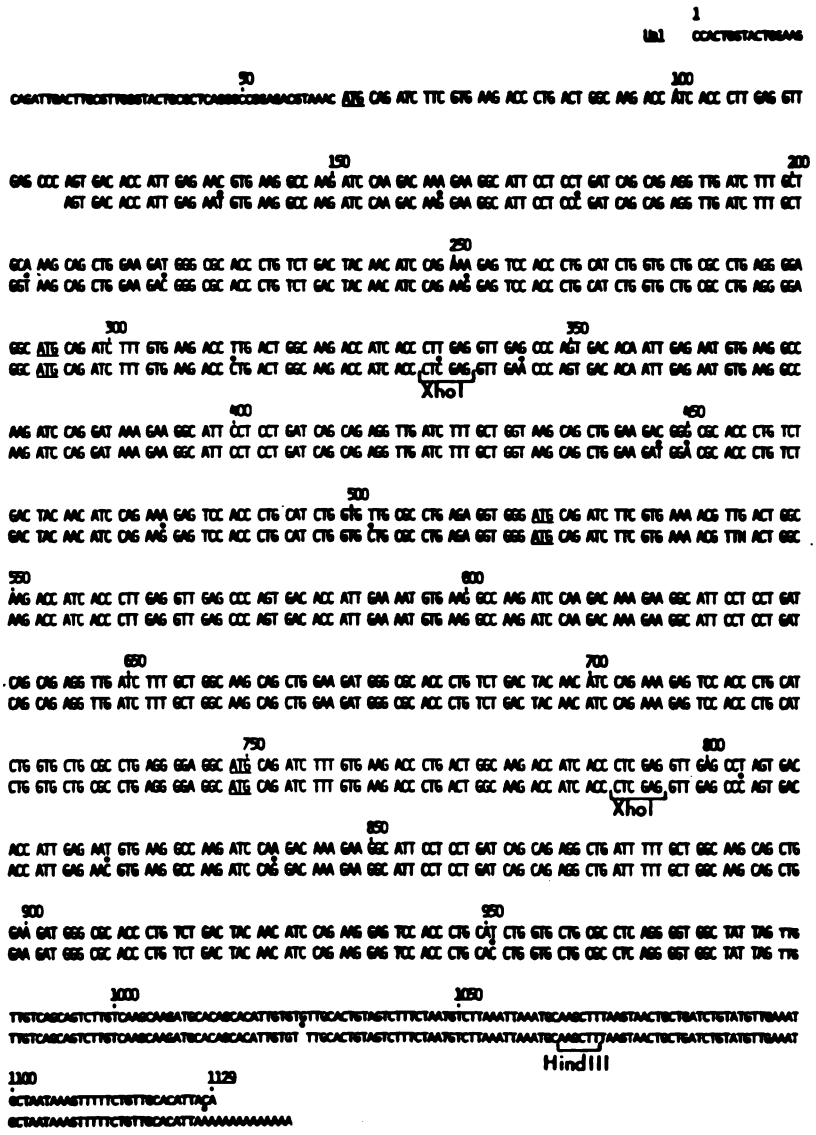

Figure 4. A comparison of the nucleotide sequences of clone 200 and the ubiquitin I genomic locus sequenced by Bond and Schlesinger (15). The sequence differences are noted (*).

3'-untranslated regions of clones 200 and 7 may be due to a cDNA cloning anomaly, or an error in reassembling the sequence information.

With the ubiquitin sequences of clone 200 as a probe, eleven genomic clones were isolated from screening approximately five genomic equivalents of the Dodgson et al. Charon 4A chicken genomic library (19). Nine of these clones were purified and DNA stocks prepared. Initial restriction enzyme mapping divided the nine clones into two major groups (1, 2, 4, 5, 6, 7, 9 and 3, 10). Clone 1, 2, 4, 7 and 9 gave identical restriction fragments, while clones 5 and 6 were very similar (data not shown). Clones 3 and 10 yielded completely different restriction fragments compared to the other seven clones (data not shown). Further restriction mapping was done on clones 2, 3, 5, 6 and 10. The restriction maps of these clones are shown in Figure 5 along with the maps of the genomic clones (UbI and UbII) reported by Bond and Schlesinger. A detailed comparison between our clones and the ubiquitin loci is hampered by the relatively few restriction sites mapped by Bond and Schlesinger. It should be noted that Bond and Schlesinger used the identical genomic library to ours, so we would expect that some of our ubiquitin clones would be identical to theirs.

The restriction maps of clone UbI and our clone 2 are identical with respect to the EcoRI, KpnI and SmaI sites (Figure 5A). This suggests, in fact, they are the same clone, since both clones were isolated from the same genomic library. Bond and Schlesinger also reported a detailed restriction map of the ubiquitin protein coding region (PCR) of UbI and nearby flanking regions. Their detailed UbI map was constructed by combining the restriction maps of two subclones. They propose that the ubiquitin I gene is transcribed from left to right

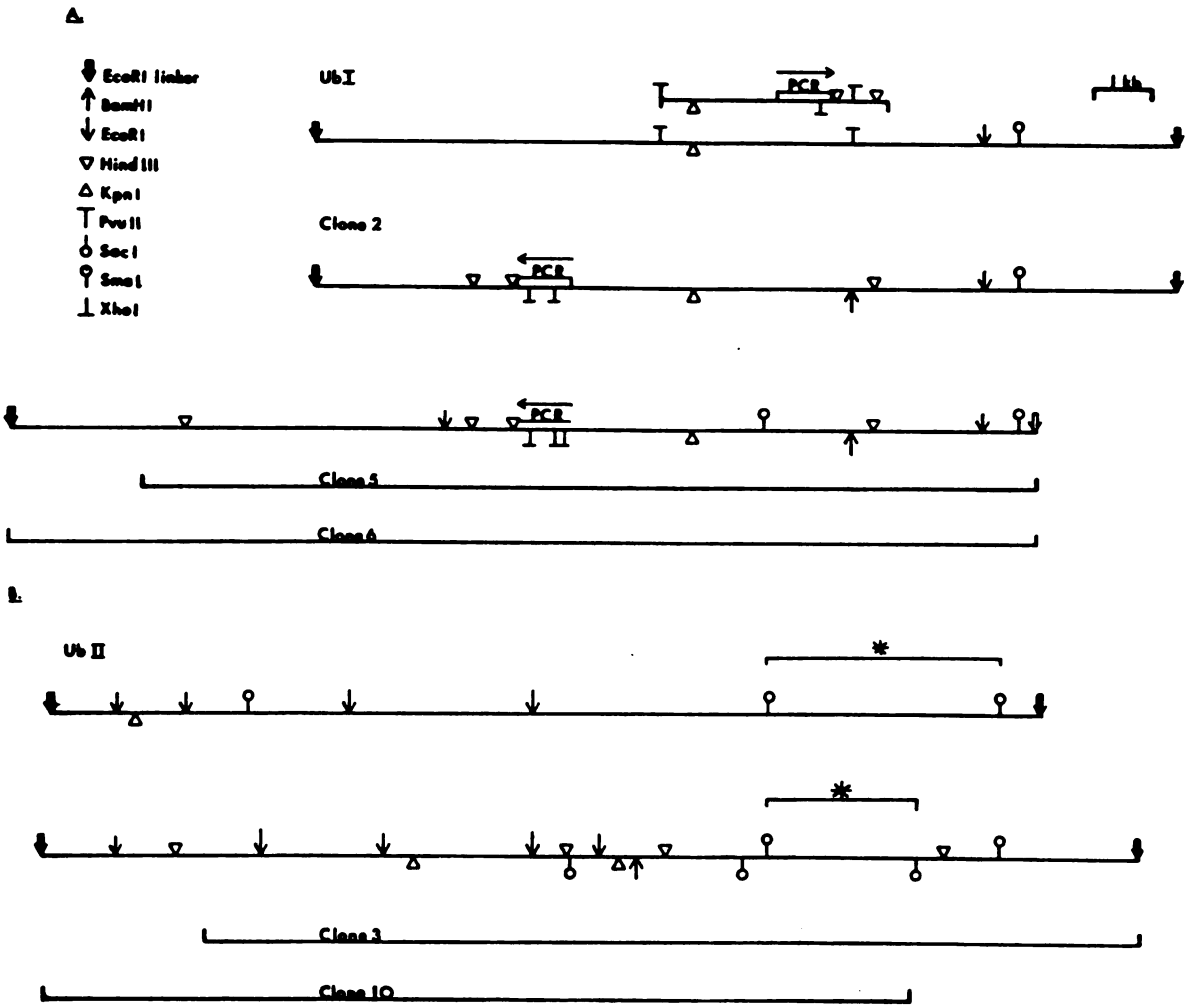

Figure 5. Restriction enzyme maps of chicken genomic clones containing ubiquitin sequences . Clones maps from two chicken genomic loci are shown, ubiquitin I (A) and ubiquitin II (B) (as defined by Bond and Schle singer, 15). In (A), the predicted ubiquitin protein coding regions (PCR) of the ubiquitin I locus clones are marked, with the arrow indicating the direction of transcription. The small restrition map above phage UbI, indicates restriction sites mapped from two overlapping subclones, but not directly compared to the UbI phage clone by the authors (15). In (B), the restriction fragments hybridizing to ubiquitin are marked (*).

with respect to the UbI phage restriction map (Figure 5A), and that it is located to the right of the KpnI site (on the 5.0 kb KpnI/EcoRI fragment). We conclude from our data that Bond and Schlesinger inverted the map of the subcloned ubiquitin I region around the KpnI site with respect to the UbI phage map. Consequently, we propose that the ubiquitin I gene is transcribed from right to left relative to Figure 5A, and that it is located to the left of the KpnI site on the 6.7 kb EcoRI/KpnI fragment. Two facts support our conclusions. First, the 0.7 kb HindIII fragment, which contains part of the 3'-untranslated region of the ubiquitin I gene, maps 2.2 kb to the left of the KpnI site. Second, the XhoI sites, which are found in the ubiquitin coding regions, also map to the left of the KpnI site and 274 bp to the right of the 0.7 kb HindIII fragment (Figure 5A). The locations of the HindIII and XhoI sites are shown in the nucleotide sequences of clones 200 and UbI (Figure 4). Clones 5 and 6 have very similar restriction maps to clones 2 and UbI (Figure 5A). Clones 5 and 6, however, differ in that they contain an additional EcoRI site and SmaI site that flank the PCR of clones 5 and 6, as well as a third XhoI site. As described below, these individual restriction site differences between clones 5 and 6 and clone 2 are likely due to allelic variation in the chicken genomic library.

In Figure 6 the ubiquitin-hybridizing XhoI, XhoI/HindIII, and HindIII restriction fragments of clones 2, 5, and 6 are shown. The restriction digests were run on an agarose gel and a Southern blot was prepared. The filter was probed with ^{32}P-labelled ubiquitin sequences of clone 200. The characteristic 456 bp XhoI fragment and 274 bp XhoI/HindIII fragment hybridize to ubiquitin (clone 200) in clone 2.

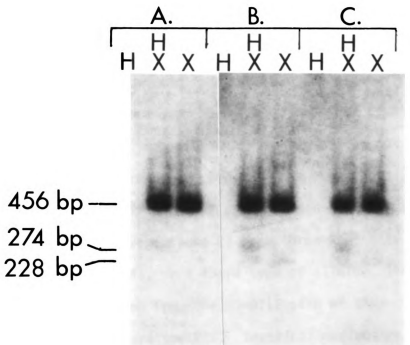

Figure 6. A comparison of the ubiquitin containing restriction fragments of clones 2 (A), 5 (B), and 6 (C). Clone DNA was digested with HindIII (H), HindIII + XhoI (HX) and XhoI (X), run on a 0.7% agarose gel, and a Southern blot was prepared. The filter was probed with the ubiquitin sequences from clone 200, ^{32}P-labelled by nick translation. Only the small restriction fragments found within the ubiquitin coding regions are shown.

These characteristic fragments also hybridize in clones 5 and 6 along with a new 228 bp XhoI fragment. Since the 228 bp XhoI fragment hybridizes to ubiquitin, this extra XhoI site is probably located within a ubiquitin repeat directly 5' to the 456 bp XhoI fragment. The ubiquitin repeats are 228 bp units. In clone 200, the two XhoI sites are located within the second and fourth repeats resulting in 456 bp fragment (2 x 228). Remember clone 200 does not extend completely through the first ubiquitin repeat. Consequently, we do not know if another XhoI site is in this first repeat in the cDNA clone. Therefore, we do not know if the cDNA clone originated from a two XhoI site ubiquitin I allele (e.g., clone 2) or a three XhoI site ubiquitin I allele (e.g., clone 5 and 6), or a third type of allele. The DNA of clones 5 and 6 was sequenced from the HindIII site of the 3'-untranslated region toward the PCR. Identical nucleotide sequences were obtained from clones 5 and 6. Their sequence corresponded exactly with the ubiquitin nucleotide sequences of the 3'-regions of clones 200 and UbI. We therefore conclude that our clone 2 and UbI of Bond and Schlesinger represent clones containing one allele of the ubiquitin I locus while clones 5 and 6 represent clones containing the other allele. The library used by both their group and ours was obtained from a single chicken (bred for heterozygosity for a restiction site polymorphism at the ovalbumin locus), and thus only two alleles of any given locus should be present therein.

Clones 3 and 10 are overlapping genomic clones whose restriction maps are similar to the UbII clone isolated by Bond and Schlesinger, at least near the ubiquitin hybridizing region (Figure 5B). A detailed comparison is again hampered by the fact that they report only a few

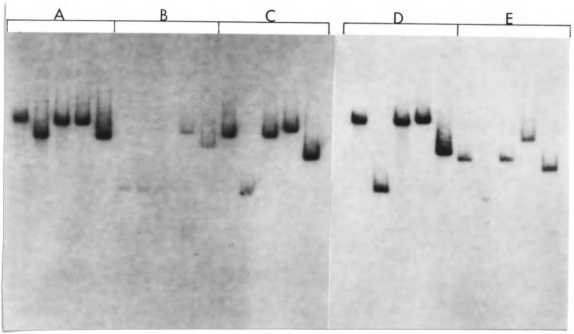

Figure 7. Differential hybridization of chicken genomic clones containing the two ubiquitin genomic loci. Clone DNA was digested, run on 0.7% agarose gels, and Southern blots were prepared. The filters were probed with ubiquitin sequences from clone 200 (homologous to the UbI locus, Figure 4). ^{32}P-labelled by nick translation. Clones 2 (A), 5 (C), and 6 (D) are homologous to the ubiquitin I locus while clones 3 (B) and 10 (E) are homologous to the ubiquitin II locus (see also Figure 5A and B).

flanking restriction sites. Both clone 3 and UbII contain a ubiquitin homologous region on a 4.3 kb SmaI fragment with a characteristic SacI site near the gene (Figure 5B). The ubiquitin-containing restriction fragments of clones 3 and 10 always hybridized weakly compared to the hybridization signals of clones 2, 5 and 6 (Figure 7). This difference was considerably greater than would be expected from the existence of four ubiquitin repeats in ubiquitin I and only three repeats in ubiquitin II. This may be attributable to a significant difference in the ubiquitin nucleotide sequence between ubiquitin I and ubiquitin II. Bond and Schlesinger reported that the last (3') ubiquitin protein coding repeats of clones UbI and UbII were 93% homologous in the nucleotide sequence. The restriction map of UbII has several differences with the maps of clones 3 and 10, especially in the order of EcoRI fragments. Clones 3 and 10 also have two KPNI sites and two SmaI sites rather than the one KPNI and three SmaI sites reported for UbII. Even with these mapping differences, we feel that clone 3 and 10 contain the same ubiquitin II locus as UbII as defined by Bond and Schlesinger, because of the similarity of restriction sites near the ubiquitin homologous region.

The mapping results of clones 2, 3, 5, 6, and 10 compare accurately with the results of a chicken genomic Southern probed with clone 200 (Figure 8). Chicken chromosomal DNA was digested with EcoRI, HindIII, and BamHI, singlely and in pairs, run on an agarose gel and a Southern blot was prepared. The filter was probed with ^{32}P-labelled ubiquitin sequences of clone 200. The different hybridization intensities of the ubiquitin I and II loci, noted above, are reflected in the results of all three single digest lanes. For example, a high

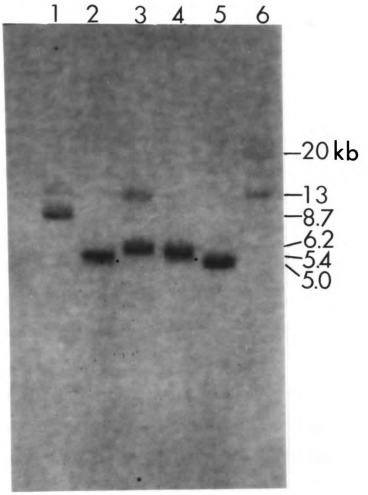

Figure 8. Southern blot analysis of the chicken genomic regions. Chicken chromosomal DNA (10 μg/lane) was digested with EcoRI (lane 1), EcoRI + HindIII (lane 2), EcoRI + BamHI (lane 3), HindIII (lane 4), HindIII + BamHI (lane 5), and BamHI (lane 6), run on a 0.8% agarose gel, and a Southern blot was prepared. The filter was probed with the ubiquitin sequences of clone 200, ^{32}P-labelled by nick translation. Faint hybridizing fragments are noted (•).

intensity 5.5 kb fragment and a weak intensity 5.0 kb fragment are produced in a HindIII digest. The ubiquitin-hybridizing regions of clones 2, 5, and 6 (high intensity, ubiquitin I) are contained on a 5.5 kb HindIII fragment while the ubiquitin regions of clones 3 and 10 (low intensity, ubiquitin II) are contained on a 5.0 kb HindIII fragment. A complete, one-to-one comparison of the clone mapping data with the genomic Southern cannot be made for all lanes for two reasons. First, at least one of the flanking sites of the larger DNA fragments is often not present within the cloned region of chicken chromosomal DNA. Second, the Southern blotting procedure does not always transfer different size DNA fragments equally. All the ubiquitin-hybridizing restriction fragments of known size correctly correspond to the appropriate chromosomal fragment on the genomic Southern (Figure 8).

Our data supports the results of Bond and Schlesinger, in that we found two different genomic loci in chickens containing sequences homologous to ubiquitin. We feel we have isolated clones from both alleles containing the ubiquitin I locus which are present in our library; clones 2 and UbI of Bond and Schlesinger contain one allele and clones 5 and 6 contain the other allele. Clones 3 and 10 correspond to the ubiquitin II locus, although Bond and Schlesinger did not report enough data on restriction sites flanking the ubiquitin II coding locus in clone UbII for a complete comparison of our clones to theirs to be made. The differences between their results and ours in the nucleotide sequence and clone map data can probably be resolved by a more detailed analysis of the ubiquitin loci and direct comparison of the clones isolated by the two labs.

REFERENCES

1. Goldstein, G., M. Scheid, U. Hammerling, E.A. Bayse, D.H. Schlesinger and G. Goldstein. 1975. Proc. Natl. Acad. Sci. U.S.A. 72:11-15.

2. Ciechanover, A., D. Finley and A. Varshavsky. 1984. J. Cell Biochem. 24:27-53.

3. Finley, D. and A. Varshavsky. 1985. Trends Biochem. Sci. 10:343-347.

4. Goldknopf, F.L. and H. Busch. 1977. Proc. Natl. Acad. Sci. U.S.A. 74:864-868.

5. West, M.H. and W.M. Bonner. 1980. Nucleic Acids Res. 8:4671-4680.

6. Levinger, L. and A. Varshavsky. 1982. Cell. 28:375-385.

7. Matsui, S.I., B.K. Sron and A. Snadberg. 1979. Proc. Natl. Acad. Sci. U.S.A. 76:6386-6390.

8. Ciechanover, A., D. Finley and A. Varshovsky. 1984. Cell. 37:57-66.

9. Gallatin, M., T.P. St. John, M. Siegelman, R. Reichert, E.C. Butcher and I.L. Weissman. 1986. Cell. 44:673-680.

10. Siegelman, M., M.W. Bond, W.M. Gallatin, T. St. John, H.T. Smith, V.A. Fried and I.L. Weissman. 1986. Science. 231:823-829.

11. St. John, T., W.M. Gallatin, M. Siegelman, H.T. Smith, V.A. Fried and I.L. Weissman. 1986. Science. 231:823-829.

12. Wiborg, O., M.S. Pedersen, A. Wind, L.E. Berglund, K.A. Marcher and J. Vunst. 1985. EMBO J. 4:755-759.

13. Lund, P.K., B.M. Moats-Staats, J.G. Simmons, E. Hoyt, A.J. D'Escole, F. Martin and J.J. VanWyte. 1985. J. Biol. Chem. 260:7609-7613.

14. Bond, U. and M.J. Schlesinger. 1985. Mol. Cell. Biol. 5:949-956.

15. Bond, U. and M.J. Schlesinger. 1986. Mol. Cell. Biol. 6:4602-4610.

16. Dworkin-Rastl, E., A. Shrutkowski and M.B. Dworkin. 1984. 6:4602-4610.

17. Izquierdo, M., C. Aribas, J. Galceran, J. Burke and V.M. Cabrera. 1984. Biochem. Biophys. Acta. 783:114-121.

18. Ozkaynak, E., D. Finley and A. Varshavsky. 1984. Nature. 312:663-666.

19. Dodgson, J.B., J. Strommer and J.D. Engel. 1979. Cell. 17:879-887.

20. Benton, W.D. and R.W. Davis. 1977. Science. 196:180-182.

21. Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. In Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

APPENDIX I

ISOLATION OF THE CHICKEN CARBONIC ANHYDRASE II GENE

Isolation of the Chicken Carbonic Anhydrase II Gene

CORINNE M. YOSHIHARA, MARK FEDERSPIEL, and JERRY B. DODGSON

Department of Microbiology and Public Health
Michigan State University
East Lansing, Michigan 48824-1101

The carbonic anhydrase (CA) gene family displays considerable variation in its expression pattern. In amniotes (birds, reptiles, and mammals) the three genetic loci that have been identified that encode isozymes CA I, CA II, and CA III, vary in their expression between classes of amniotes and in tissues within a class. The genes for CA I and CA II are both expressed in most mammalian red blood cells (RBC), for example, while in avians only the CA II gene is expressed in the RBC. In mammals, at least, the genes that encode CA I and CA II enzymes appear to be linked.[1-3]

In order to study the relationship between the structure and organization of CA genes and CA gene expression, it is necessary to first isolate the genes. We describe here the initial steps toward isolation of the chicken CA II gene.

A chicken RBC cDNA library was prepared in the plasmid pBR322 with poly-$(A)^+$ chicken anemic red cell cytoplasmic RNA by dG–dC tailing into the Pst I site.[4] Bacterial colonies that were $tet^R amp^S$ were transferred to nitrocellulose filters, lysed, and the liberated DNA fixed to the filters. The filters were first screened for those colonies containing recombinant plasmids with globin DNA inserts by hybridizing the filters with α and β globin-specific probes. The nonglobin colonies were selected and screened with ^{32}P-labeled mouse CA II cDNA.[5] The three colonies whose DNA gave a positive autoradiographic result were isolated. DNA sequence analysis demonstrated that one of the three clones isolated whose insert was approximately 300 bp was a *bona fide* chicken CA II cDNA clone.

The restriction map of the chicken CA II clone is shown in FIGURE 1 along with the strategy for sequence analysis. The amino acid sequence predicted from the nucleotide sequence shows extensive homology with the known amino acid sequences of human (65%), rabbit[6] (63%), and mouse[7] (60%) CA II (FIG. 2). The

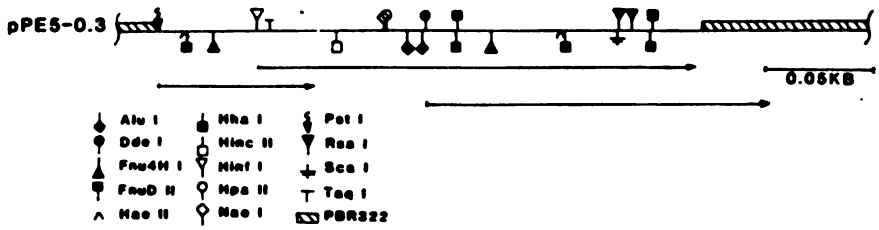

FIGURE 1. cDNA clone of carbonic anhydrase II. Restriction map of pPE5-0.3. Arrows indicate the direction and extent of the sequence determined by Maxam and Gilbert sequence analysis.

		GGG	AGY	CAC	AAC	GGA	CCC	GCG	CAC	TGG	YAC	GAG	YAC	TTC	CCC
					10						TYT		TYT		20
CHICKEN	CA II	Gly	Ser	His	Asn	Gly	Pro	Ala	His	Trp	His	Glu	His	Phe	Pro
HUMAN	CA II	—	Lys	—	—	—	—	Glu	—	—	His	Lys	Asp	—	—
RABBIT	CA II	—	Lys	—	—	—	—	Glu	—	—	His	Lys	Asp	—	—
MOUSE	CA II	Ser	Lys	—	—	—	—	Glu	Asn	—	His	Lys	Asp	—	—
		ATC	GCC	AAT	GGG	GAG	CGC	CAG	TCG	CCC	ATC	GCC	ATC	AGC	ACC
											30				
CHICKEN	CA II	Ile	Ala	Asn	Gly	Glu	Arg	Gln	Ser	Pro	Ile	Ala	Ile	Ser	Thr
HUMAN	CA II	—	—	Lys	—	—	—	—	—	—	Val	Asp	—	Asp	—
RABBIT	CA II	—	—	Asp	—	—	—	—	—	—	—	Asp	—	Asp	—
MOUSE	CA II	—	—	—	—	Asp	—	—	—	—	Val	Asp	—	Asp	—
		AAA	GCC	GCC	CGC	TAC	GAC	CCC	GCG	CYG	AAG	CCC	CTC	AGC	TTC
							40			Leu					
CHICKEN	CA II	Lys	Ala	Ala	Arg	Tyr	Asp	Pro	Ala	Pro	Lys	Pro	Leu	Ser	Phe
HUMAN	CA II	His	Thr	—	Lys	—	—	—	Ser	Leu	—	—	—	—	Val
RABBIT	CA II	Asp	—	—	Lys	His	—	—	Ser	Leu	—	—	—	Arg	Val
MOUSE	CA II	Ala	Thr	—	His	His	—	—	—	Leu	Gln	—	—	Leu	Ile
		AGC	TAC	GAT	GCC	GGC	ACG	GCC	AAA	GCC	ATC	GTC	AAC	AAC	GGG
			50										60		
CHICKEN	CA II	Ser	Tyr	Asp	Ala	Gly	Thr	Ala	Lys	Ala	Ile	Val	Asn	Asn	Gly
HUMAN	CA II	—	—	—	Gln	Ala	—	Ser	Leu	Arg	—	Leu	—	—	—
RABBIT	CA II	—	—	Glu	His	Pro	Ile	Ser	Arg	Arg	—	Ile	—	—	—
MOUSE	CA II	—	—	—	Lys	Ala	Ala	Ser	—	Ser	—	—	—	—	—
		CAC	TCC	TTC	AAC	GTG	GAG	TTC	GAC	GAC	TCC	TCC	GAC	AAG	TCA
									70						
CHICKEN	CA II	His	Ser	Phe	Asn	Val	Glu	Phe	Asp	Asp	Ser	Ser	Asp	Lys	Ser
HUMAN	CA II	—	Ala	—	—	—	—	—	—	—	—	Gln	—	—	Ala
RABBIT	CA II	—	—	—	—	—	—	—	—	—	—	His	—	—	—
MOUSE	CA II	—	—	—	—	—	—	—	—	—	—	Gln	—	Asn	Ala
		GTG	CYG	CAA	GGA	GGA	GCG	CTG	GAT	GGC	AGT				
			Leu		80										
CHICKEN	CA II	Val	Pro	Gln	Gly	Gly	Ala	Leu	Asp	Gly	Ser				
HUMAN	CA II	—	Leu	Lys	—	—	Pro	—	—	—	Thr				
RABBIT	CA II	—	Leu	Lys	Glu	—	Pro	—	Glu	—	Thr				
MOUSE	CA II	—	Leu	Lys	—	—	Pro	—	Ser	Asp	—				

FIGURE 2. Nucleotide sequence of pPE5-0.3. The predicted amino acid sequence of chicken carbonic anhydrase II is compared with homologous amino acid sequences of human, rabbit, and mouse carbonic anhydrase II. The amino acid sequence predicted by the nucleotide sequence is given below the coding regions along with its numbering. Y refers to C or T. Only those amino acids of the human, rabbit, and mouse CA II proteins that differ from the predicted chicken CA II sequence are shown. See reference 6 and citations therein for the human CA II sequence.

chicken cDNA clone contains sequence from the coding region at the 5'-end of the CA II mRNA from amino acids 7 to 86.

A comparison of amino acids 7 to 86 of chicken to the corresponding amino acids of mammalian CA I, II, III[8] indicates that chicken CA II is identical to 1 of the 12 invariant and unique residues of CA I; 1 of the 15 for CA III; and 5 of the 9 for CA II (residues 7, 26, 66, 68, and 75). Chicken CA II is identical with all of the 6 residues that are located in the active site regions of the CA isozymes (residues 28, 60, 63, 64, 66, and 68).

The chicken cDNA clone was used to isolate phage from a λ Charon 4A chicken genomic library.[9] The phage that hybridized to the cDNA clone presumably contain the chicken CA II gene. These recombinant clones are presently being characterized by restriction enzyme analysis. Future experiments will involve detailed restriction enzyme analysis, subcloning, and DNA sequence analysis (particularly the 5' flanking region) of the clones and will provide preliminary data for the study of the chicken CA gene family.

ACKNOWLEDGMENTS

We are grateful to Dr. Peter J. Curtis for providing the mouse carbonic anhydrase cDNA clone. We also thank Dr. Richard E. Tashian and Dr. David Hewett-Emmett for their advice and encouragement.

REFERENCES

1. CARTER, N.D. 1972. Carbonic anhydrase isozymes in Cavia porcellus, Carvia apera and their hybrids. Comp. Biochem. Physiol. 43B: 743–747.
2. DESIMONE, J., M. LINDE & R. E. TASHIAN. 1973. Evidence for linkage of carbonic anhydrase isozyme genes in the pig-tailed macaque, *Macaca nemestrina*. Nature (New Biol.) 242: 55–56.
3. EICHER, E. M., R. H. STERN, J. E. WOMACK, M. T. DAVISSON, T. H. RODERICK & S. C. REYNOLDS. 1976. Evolution of mammalian carbonic anhydrase loci by tandem duplication: Close linkage of Car-1 and Car-2 to the centromere region of chromosome 3 of the mouse. Biochem. Genet. 14: 651–660.
4. MANIATIS, T., E. F. FRITSCH & J. SAMBROOK. 1982. Molecular cloning manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.
5. CURTIS, P. J. 1983. Cloning of mouse carbonic anhydrase mRNA and its induction in mouse erythroleukemia cells. J. Biol. Chem. 258: 4459–4463.
6. FERRELL, R. E., S. K. STROUP, R. J. TANIS & R. E. TASHIAN. 1978. Amino acid sequence of rabbit carbonic anhydrase II. Biochim. Biophys. Acta 533: 1–11.
7. CURTIS, P. J., E. WITHERS, D. DEMUTH, R. WATT, P. J. VENTA & R. E. TASHIAN. 1984. The nucleotide sequence and derived amino acid sequence of cDNA coding for mouse carbonic anhydrase II. Gene. In press.
8. TASHIAN, R. E., D. HEWETT-EMMETT & M. GOODMAN. 1983. On the evolution and genetics of carbonic anhydrase I, II, and III. Isozymes: Current Topics in Biological and Medical Research 7: 79–100.
9. DODGSON, J., J. STROMMER & J. D. ENGEL. 1979. The organization of chicken globin genes. Cell 17: 879–887.

APPENDIX II

ISOLATION OF RECOMBINANT cDNAs ENCODING CHICKEN ERYTHROID δ-AMINOLEVULINATE SYNTHASE

Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3702–3706, June 1985
Developmental Biology

Isolation of recombinant cDNAs encoding chicken erythroid δ-aminolevulinate synthase

(antibody to δ-aminolevulinate synthase/red cell expression library)

MASAYUKI YAMAMOTO*, NELSON S. YEW*, MARK FEDERSPIEL†, JERRY B. DODGSON†, NORIO HAYASHI‡, AND JAMES DOUGLAS ENGEL*

*Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, IL 60201; †Departments of Microbiology and Biochemistry, Michigan State University, East Lansing, MI 48824; and ‡Department of Biochemistry, Tohoku University School of Medicine, Sendai, Japan

Communicated by David Shemin, January 17, 1985

ABSTRACT We report the isolation of cDNA clones encoding δ-aminolevulinate synthase (ALA synthase; EC 2.3.1.37), the first enzyme in the heme biosynthetic pathway in animal cells. The gene was isolated from a chicken erythroid cDNA library prepared in the bacteriophage λ fusion/expression vector gt11, using rabbit antibody raised against the relatively abundant chicken liver enzyme. The chicken liver and red cell ALA synthase isozymes share substantial crossreactivity to the antibody, thereby allowing isolation of the erythroid-specific gene by using the heterologous antibody in immune screening of the red cell cDNA library. Preliminary analysis documenting the tissue specificity of transcription indicates that the enzyme is encoded by a highly homologous set of messages, which appear to differ in size in various avian tissues. From analysis using strand-specific RNA probes, it appears that the different ALA synthase mRNAs detected may be transcribed from a family of genes that are closely related in nucleotide sequence and are each regulated in a developmentally specific manner.

δ-Aminolevulinate synthase (ALA synthase; EC 2.3.1.37), the first enzyme in the biosynthetic pathway leading to heme production, catalyzes the condensation of glycine and succinyl-CoA to form δ-aminolevulinic acid (1). The mRNA for the enzyme is encoded in the nucleus, from which it is transported to the cytoplasm, where it is translated into an enzymatically active precursor form. The ALA synthase preenzyme is processed into the mature mitochondrial enzyme (a proteolytic cleavage product of the preenzyme) as it traverses the mitochondrial membrane, where it finally occupies its normal cellular compartment (2).

Various studies have previously shown that both the ALA synthase gene and the enzyme are subject to a complex array of regulatory interactions. Thus, the enzyme form found in the liver is responsive to feedback regulation by the end product (heme) and can be induced by a variety of chemical effectors of hepatic porphyria, notably 3,5-dicarbethoxy-1,4-dihydrocollidine and allylisopropylacetamide. Furthermore, this induction appears to be regulated at both the transcriptional and translational levels (3–11). In the liver, heme also appears to directly interfere with the translocation of the preenzyme from the cytoplasmic to the mitochondrial compartments, causing abnormal increases in the cytoplasmic levels of the preenzyme (12, 13).

In contrast to these observations in hepatic tissue, the erythroid form of the enzyme is unresponsive to chemicals that normally induce porphyria, and this isozyme does not appear to accumulate in the cytoplasm of erythrocytes on treatment of animals with the same porphyrogenic agents (14–18). Instead, the biosynthesis of ALA synthase in erythroid cells appears to be strongly stimulated by chemical agents normally used for induction of anemia [e.g., phenylhydrazine (G. Kikuchi and M. Hasegawa, personal communication)]. Thus the enzyme activity appears to be regulated in a cell-specific manner. In keeping with the findings that there exist differentially regulated ALA synthase enzymes, it has been reported that the liver and erythroid counterparts of the enzyme also differ in size, both as preenzymes and as mature proteins (19).

We are interested in studying the regulatory mechanisms whereby ALA synthase becomes activated in chicken erythroid and liver cells in a developmentally specific manner. Furthermore, one report has claimed that ALA synthase is one of the earliest erythroid genes that is transcriptionally activated after dimethyl sulfoxide treatment of Friend erythroleukemia cells, thus suggesting the intriguing possibility that ALA synthase induction might serve as an early temporal marker for erythrocyte maturation (20). This served as an additional incentive to examine the regulation of this particular erythroid-specific gene in detail.

MATERIALS AND METHODS

Bacteriophage and Host Strains. Bacteriophage λ vectors gt10 and gt11 and lysogenic and lytic *Escherichia coli* host strains (Y1089, Y1090) were obtained from Tom St. John (Department of Pathology, Stanford University Medical School) (21, 22), and strain BSJ37 (23) was obtained from Ed Fritsch (Genetics Institute).

Erythroid cDNA Library Preparation and Screening. cDNA was prepared as described (ref. 24, pp. 230–238), and *Eco*RI linkers (Bethesda Research Laboratories) were ligated to the mixture of erythroid cDNAs. The linkers were then cleaved with *Eco*RI to reveal the cohesive ends, and the cDNAs were fractionated by gel electrophoresis into populations that were greater than and less than 800 base pairs (bp). The two pools were individually collected and separately ligated to λ gt10 DNA treated with *Eco*RI. The ligated "small" (i.e., <800-bp) and "large" (>800-bp) cDNA pools were then packaged *in vitro* (ref. 24, pp. 291–292) and plated on *hfl* host BSJ37. Since the packaged phage that contain recombinant cDNAs will not integrate in the *hfl* strain (the *Eco*RI inserts interrupt the λ cI gene, whose expression is essential for lysogeny) only recombinant bacteriophage efficiently produce plaques. The titers of the two libraries were 1.7×10^6 plaque-forming units (pfu) for the large inserts (*ca.* 10^4 pfu of cI^- revertants) and *ca.* 2.8×10^6 pfu for the small fragment library. The large insert library was amplified at 2×10^4 pfu/150-mm dish for 4 hr at 42°C on Y1090; the amplified phage were extracted from top agarose as previously described (25, 26). The phage

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

Abbreviations: ALA synthase; δ-aminolevulinate synthase; bp, base pair(s); pfu, plaque-forming units.

were then purified by cesium chloride equilibrium gradient centrifugation; DNA was isolated from the phage by addition of sodium dodecyl sulfate and proteinase K as described (ref. 24, p. 85).

The purified λ gt10 library was transferred into λ gt11 by treatment of the λ gt10 DNA cohesive ends with nuclease *Bal* 31 until *in vitro* packaging efficiency of the gt10 DNA had dropped by a factor of at least 10^4 (T. St. John, J. Rosen, and H. Gershenfeld; personal communication). The exonuclease-treated gt10 phage pool was then digested with *Eco*RI and added to ligated λ gt11 arms that had been digested with *Eco*RI and dephosphorylated by using column-purified calf intestine alkaline phosphatase (27). The phage were ligated at a roughly 2:1 weight ratio (λ gt10 library to λ gt11 arms) and packaged *in vitro*. Packaging efficiency was usually about 10^6 pfu/μg of total DNA added, resulting in 12–40% recombinants as judged by clear plaque formation on 5-chloro-4-bromo-3-indolyl β-D-galactoside (X-Gal) plates. We usually chose ligation ratios that gave about 20% recombinants, to avoid multiple cDNA inserts.

λ gt11 recombinants were plated for screening in top agarose at *ca.* 5×10^4 pfu per 150-mm Petri dish on host strain Y1090 (21). The plates were incubated for 4 hr at 42°C and then overlaid with nitrocellulose filters that had been soaked in 10 mM isopropyl β-D-thiogalactoside and dried. The overlays were then incubated for 4–12 hr more. The filters were then removed into a blocking solution (3% nonfat dry milk plus 0.1% Nonidet P-40 in Tris-buffered saline [20 mM Tris·HCl, pH 7.5/0.5 M NaCl/0.02% sodium azide (3% milk solution)] and placed on a rotator to agitate for 6–12 hr at 4°C in the presence of a sonic lysate of λ gt11-infected Y1090 cells (all subsequent treatment of the filters was also performed at 4°C). Anti-ALA synthase antibody was then added to the filters at a final concentration of 5 μg/ml, and the filters were allowed to bind to the antibody for an additional 6 hr. The filters were then removed from the antibody solution, washed four times for 30 min each in 3% milk solution, and then moved to a second dish, containing ^{125}I-labeled goat antibodies to rabbit IgG (5×10^5 cpm/ml). The filters were bound to the second antibody for 2–3 hr and again washed four times for 30 min in 3% milk solution. The filters were then dried and exposed for autoradiography at −70°C for 12–18 hr. Recombinant plaques were purified as previously described (26).

Antibodies. The antibody recognizing chicken liver ALA synthase was raised in rabbits; the specificity of this antibody has been previously characterized by demonstration that dilute antibody is able to specifically inhibit the enzymatic condensation reaction when presented with partially purified ALA synthase from either chicken liver or red blood cells (19). The antibody was purified by binding to, and elution from, staphylococcal protein A-Sepharose CL-4B (Pharmacia); we found this to be the minimal purification necessary to reduce background adequately for successful antibody screening. Prior to their use in screening, antibodies were treated by preabsorption with a λ gt11-infected Y1090 sonic lysate for a minimum of 12 hr at 4°C (22). Iodinated, affinity-purified, goat antibody to rabbit IgG was the generous gift of Susan K. Pierce (Northwestern University).

Other Methods. Positive hybrid selection and *in vitro* translation of the released mRNA was performed by modification (28) of the original method (29), adapted to the use of single-stranded RNA bound to filters. Rabbit reticulocyte lysate used for *in vitro* translation reactions was from Promega Biotec (Madison, WI) and was used according to the supplier's specifications. Subcloning in Sp6 vector pSp65 (30) was as previously described, as was the preparation of DNA template for the Sp6 *in vitro* transcription reactions; RNA blots were also performed as previously described (31). Isotopically labeled precursor amino acids and nucleotides were purchased from Amersham; labeled and unlabeled protein standards for gel electrophoresis were purchased from Amersham and Sigma, respectively. Protein blotting was performed as specified (32), with the exceptions that the filters were first blocked in 3% milk solution (above) and then washes after the first and second antibody binding reactions were in Tris-buffered saline/0.5% Triton X-100.

RESULTS

Characteristics of the Anti-ALA Synthase Antibody. To ensure that the antibody raised against the chicken liver form of ALA synthase would be successful for isolation of the erythroid ALA synthase gene in antibody screening, two experiments were undertaken to test the affinity of the antibody preparation to erythroid ALA synthase. In the first experiment, ALA synthase from *in vitro* translated anemic chicken red blood cell poly(A)$^+$ RNA was immunoprecipitated by using the antibody and protein A-Sepharose. The immunoprecipitated [^{35}S]methionine-labeled erythroid preenzyme was recovered in high yield from an *in vitro* translation mixture of total reticulocyte poly(A)$^+$ mRNA, as expected (Fig. 1; ref. 19). While it is clear that erythroid ALA synthase is the only protein recognized in immature erythroid cells, a clearly different molecular weight species of ALA synthase is immunoprecipitated from committed erythroid progenitor cells (Fig. 1, lane C). Whether this implies that these virally transformed cells exhibit ambiguous developmental properties in their pattern of gene expression or are developmentally "switched" in the type of ALA synthase isozyme expressed during erythroid maturation is currently unknown (see *Discussion*).

A second, more salient, experiment was performed to ask if ALA synthase would associate with the antibody after the enzyme was immobilized on nitrocellulose filter blots (32). The second experiment successfully demonstrated that the erythroid ALA synthase strongly associates with the antibody raised against the liver isozyme when the antigen is fixed to filters (Fig. 2).

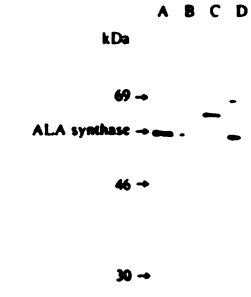

FIG. 1. Immunoprecipitation of ALA synthase from *in vitro* translation of chicken red cell RNA. Chicken red cell polysomes (0.5 A_{260} unit each) isolated from 11-day embryonic (lane A) or anemic adult (lane B) erythroid cells were translated in the presence of [^{35}S]methionine. Poly(A)$^+$ RNA from HD6 cells (tsAEV-transformed early erythroid lineage precursor cells; ref. 33) and from anemic adult chicken red blood cells (lanes C and D, respectively) was translated *in vitro*, also in the presence of radiolabeled methionine. All four samples were then immunoprecipitated from the translation reaction mixtures by using rabbit IgG antibodies to chicken liver ALA synthase (19) and protein A-Sepharose. These precipitates were then electrophoresed on 10% polyacrylamide/0.1% sodium dodecyl sulfate gels (34), fixed and bathed in EN^{3}HANCE (Amersham), dried, and exposed for fluorography. Sizes were derived by comparison of the electrophoretic mobility of ^{14}C-labeled protein standards run in a parallel lane of the same gel.

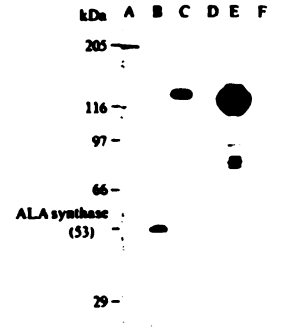

FIG. 2. Immunological detection of ALA synthase β-galactosidase fusion proteins in λ gt11 recombinants. Recombinant bacteriophage that were repeatedly positive on plaque purification were integrated into a lysogenic host strain (Y1089; ref. 21). The lysogenic cells were grown to OD_{600} = 0.3 at 32°C, at which time lytic replication was induced by shifting the temperature to 42°C for 15 min; after culturing for an additional 2 hr at 38°C, the cells were harvested and immediately lysed in sodium dodecyl sulfate sample buffer (50 mM Tris·HCl, pH 6.8/1.5% sodium dodecyl sulfate/50 mM dithiothreitol/4 M urea). Identical amounts of protein synthesized by induced λ lysogens [either treated (lanes C and E) or not treated (lanes D and F) by the addition of isopropyl thiogalactoside to 1 mM final concentration at the time of heat induction] were electrophoresed on 7.5% polyacrylamide/0.1% sodium dodecyl sulfate gels (34). The protein was then electrophoretically transferred to nitrocellulose as described (32). The nitrocellulose transfer was then blocked with 3% milk solution, and bound protein was allowed to react with IgG antibody to chicken liver ALA synthase; the filter was washed, exposed to affinity-purified goat ^{125}I-labeled antibody to rabbit IgG, washed, and exposed for autoradiography. Lane A is a parallel lane on the protein blot stained with india ink (35) to determine the position of commercial markers. Lane B contained total mitochondrial matrix proteins from anemic chicken red blood cells, bound to antibody at the same time as were the lysogenic bacterial proteins.

Isolation of ALA Synthase cDNA Recombinants. In our first successful screen of the λ gt11 expression library, we isolated 14 positive plaques (from *ca.* 2 × 10^6 total pfu screened; 28% recombinants) of which 6 were positive on repeated plaque purification. These 6 were grown as minipreparations for phage DNA isolation, and, while all 6 yielded isopropyl thiogalactoside-inducible fusion proteins larger than β-galactosidase, only 4 contained inserts that were released by digestion with *Eco*RI. The 2 of these that gave the strongest hybridization signal in protein blots were characterized further.

The two putative ALA synthase recombinants chosen for subsequent analysis are designated A4 and A14; the size of the inserts in these two recombinants are *ca.* 530 and 190 bp, respectively. [The fact that these recombinants contain cDNA sequences that are unexpectedly small (since the cDNAs were initially selected to contain inserts >800 bp) suggests that the size fractionation of the cDNA population was not accurate.] As an initial control, we first showed that the putative ALA synthase λ recombinants would produce the antigenically responsive determinant under β-galactosidase control, as expected of an ALA synthase hybrid fusion protein, and that the resultant fusion product was larger than native β-galactosidase. As shown in Fig. 2, the antibody binds proteins that are highly induced after treatment of recombinant A4 and A14 lysogens with isopropyl thiogalactoside (lanes C–F), and both fusion proteins are larger than the native β-galactosidase marker (116 kDa, lane A). Furthermore, as stated in the previous section, the 53-kDa erythroid ALA synthase enzyme (from whole red blood cell mitochondrial matrix) is the only protein recognized in a parallel lane on the same blot (Fig. 2, lane B).

Recombinant A4 Encodes Erythroid ALA Synthase. To prove that the isolated recombinants encode the red cell-specific ALA synthase protein, two complementary experiments were performed. First, if the two recombinants that gave strong antibody signals in protein blots (Fig. 2) both actually code for different or overlapping segments of the same structural gene, three criteria should be met for the corresponding reticulocyte mRNA encoding ALA synthase: recombinants A4 and A14 should hybridize to an RNA of the same size in RNA blotting analysis, hybridization to those red cell mRNAs should be strand specific, and finally, the mRNA that hybridizes to the strand-specific probes must be large enough to encode the erythroid ALA synthase preenzyme of 55-kDa (corresponding to a minimum mRNA size of 1500 nucleotides; ref. 19). The second experiment to verify the identification of these recombinants would be that ALA synthase mRNA should be selectable, in a strand-specific manner, from a population of total red blood cell poly(A)$^+$ RNAs, and that the *in vitro* translation product of that mRNA filter selection should be precipitable with the anti-ALA synthase antibody and should correspond in size to the precursor form of the enzyme found in red cells.

The results of RNA blot analysis are shown in Fig. 3, in which strand-specific radiolabeled probes have been hybridized to identical lanes of anemic adult chicken red blood cell poly(A)$^+$ RNAs. Strand-specific probes were created by subcloning the recombinant *Eco*RI inserts from the A4 and A14 phage in the vector pSp65 (30) in both orientations. The subclones were individually cleaved distal to the Sp6 promoter and recombinant segments of the ALA synthase gene (inserted at the *Eco*RI site of pSp65) and then transcribed by using Sp6 RNA polymerase and radiolabeled nucleotide precursors as previously described (31). The four individual

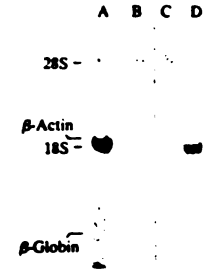

FIG. 3. Blot analysis of erythroid ALA synthase mRNA with strand-specific probes. One microgram of poly(A)$^+$ RNA isolated from the circulating red cells of anemic adult hens was electrophoresed on identical lanes of a vertical 1.2% formaldehyde agarose gels (24); the RNA was transferred to nitrocellulose, baked, and prehybridized as described (31). The inserts from recombinant bacteriophage A4 and A14 were excised with *Eco*RI and subcloned in pSp65 (30) in both fragment orientations. Both orientations of both parental recombinant cDNAs were transcribed in the presence of [α-^{32}P]GTP and Sp6 polymerase (31). These transcripts [called A14e (lane A), A14f (lane B), A4b (lane C), and A4l (lane D)] were then hybridized to individual lanes of red cell poly(A)$^+$ RNA at 55°C, washed in the presence of RNase A (31, 36) at 4 μg/ml, and exposed for autoradiography. Exposure times were 15 hr (lanes A and C) or 5 hr (lanes B and D). Identical lanes were hybridized to λGd1 [a chicken ribosomal gene recombinant (37)], pβ1BR15 [adult chicken β-globin genomic subclone (27)], and a β-actin cDNA clone (38) for internal size standardization (data not shown).

transcripts were then hybridized to separate lanes containing red blood cell mRNA; the blots were then treated with RNase (31) and exposed for autoradiography. As shown in Fig. 3, the two recombinants both recognize an mRNA of the same size (in only one of the two strand-specific probe orientations). The mRNA recognized by the two different recombinants is easily large enough (approximately 2000 nucleotides) to encode a protein of 55 kDa.

Additional proof that A4 encodes erythroid ALA synthase is provided by the hybrid selection/translation experiment shown in Fig. 4. Since we demonstrated above that the transcript produced by Sp6 subclone A41 was complementary to the putative ALA synthase mRNA, this implies that an unlabeled strand-specific transcript of that recombinant, when fixed to filters, should be able to preferentially select an mRNA whose product in an *in vitro* translation reaction should be immunoprecipitable with the anti-chicken liver ALA synthase antibody. Fig. 4 shows the results of such an experiment, in which the mRNA for the preenzyme was indeed selected from total red cell poly(A)$^+$ RNA with the synthetic Sp6 transcript from clone A41 but was not selected with the synthetic Sp6 RNA transcript from the complementary strand (prepared from subclone A4b). Thus, recombinant A4 does contain at least part of the erythroid ALA synthase gene.

Tissue Specificity of ALA Synthase Gene Expression. To determine whether or not we could distinguish between the various tissue-specific forms of ALA synthase at the mRNA level, further RNA blot analyses were performed. It was anticipated that, on the basis of these experiments, we could gain some insight as to whether or not the ALA synthase mRNAs transcribed in various chicken tissue and cell types were the same or different sizes [correlating with the different sizes of the proteins (19)] and that we might be able to make preliminary arguments regarding the possibility that the tissue-specific ALA synthase enzymes were transcribed from the same, or different gene(s). We again made use of the strand-specific Sp6 transcript complementary to erythroid ALA synthase mRNA. One might expect that if the red cell and liver (and perhaps other) forms of ALA synthase were encoded by different genes, they might have diverged sufficiently in nucleic acid sequence that a probe synthesized from one gene would have only partial homology to any or all of the heterologous gene(s). On the other hand, if various tissue-specific forms of ALA synthase mRNA were found to be highly homologous to a single conserved (coding sequence) probe, that would leave open the possibility that the tissue specificity of ALA synthase arises by differential cell-specific processing of transcripts produced from the same genetic locus.

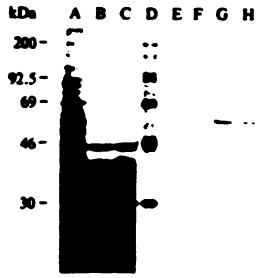

FIG. 4. Hybrid-release *in vitro* translation and immunoprecipitation with anti-ALA synthase antibody. Ten micrograms of red blood cell poly(A)$^+$ RNA (isolated from anemic adult hens) was hybridized to strand-specific Sp6 transcripts of recombinant subclone A4 fixed on nitrocellulose filters. The hybridized mRNA was released from the filters by brief boiling and then translated *in vitro* (28, 29). The *in vitro* translation reactions are shown after the following procedures: lane A, no filter selection (total red blood cell mRNA); lane B, hybridization of total red blood cell mRNA and release from filter-bound Sp6 transcript A4b; and lane C, hybridization and release from Sp6 transcript A41 (lane D shows the mobility of the ^{14}C-labeled protein markers on this 12.5% polyacrylamide/0.1% sodium dodecyl sulfate gel). Final lanes depict the immunoprecipitation reaction of total red cell [^{35}S]methionine-labeled protein after treatment of the translation reaction mixtures with either anti-ALA synthase antibody and protein A-Sepharose (lane H) or protein A-Sepharose alone (lane E). Lanes F and G show the results of immunoprecipitation of the translation products of lanes B and C, respectively, with anti-ALA synthase antibody and protein A-Sepharose.

The results of RNA blot analysis of various poly(A)$^+$ mRNAs isolated from cell lines and various chicken tissues are shown in Fig. 5. Fig. 5*A* shows the hybridization pattern of radiolabeled Sp6 A41 transcript to the mRNAs in a standard "high-stringency" wash solution (equivalent to 10 mM monovalent cation; 52°C) and Fig. 5*B* shows the same blot washed in a solution containing RNase A at 30°C. [We have previously established that the latter condition favors the release of all but specifically bound RNA probes in such blotting experiments (31)]. As is readily apparent from the data presented in Fig. 5, the RNA (probe)·RNA (mRNA) hybrids from different tissues are clearly distinguishable (*A*). Furthermore, the size of the mRNAs homologous to the erythroid-specific probe varies according to tissue and cell type. Thus, the largest cellular RNA that is detectable in

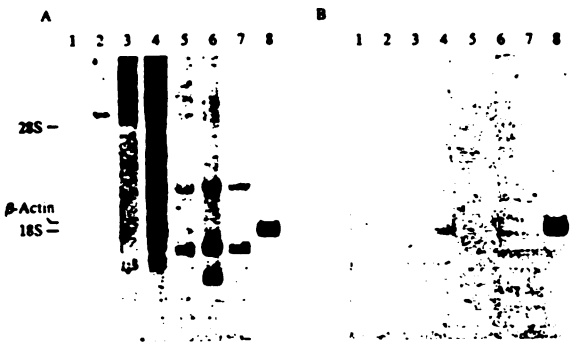

FIG. 5. Developmental specificity of ALA synthase mRNA synthesis. RNA was isolated from various chicken tissues and cell lines by banding of guanidinium thiocyanate-isolated RNA and poly(A)$^+$ RNA was isolated by two cycles of oligo(dT)-cellulose chromatography (31). These RNAs were electrophoresed, blotted, and hybridized to radiolabeled subclone A41 Sp6 transcript (see legend to Fig. 3). Chicken cellular RNA samples were isolated from the following sources: lane 1, MSB-1 cells (39); lane 2, 11-day chicken embryo fibroblasts; lane 3, 4.5-day whole chicken embryos; lane 4, 11-day embryonic brain; lane 5, 17-day embryonic liver; lane 6, 1-mo-old chick liver; lane 7, adult chicken liver; and lane 8, anemic adult hen red blood cells. The blot shown in *A* was washed in 7.5 mM NaCl/0.75 mM sodium citrate/0.1% sodium dodecyl sulfate at 52°C; exposure time was 5 hr. The blot shown in *B* is the same as that shown in *A*, but it had subsequently been washed in 300 mM NaCl buffer plus RNase A at 4 μg/ml (31, 36) after the exposure shown in *A* was taken; exposure time for *B* was 13 hr.

the liver (*ca.* 2800–3000 bases; Fig. 5*A*, lanes 5-7) is large enough to encode the liver-specific (73-kDa) preenzyme (19), while the other two homologous RNAs visualized in the same lanes are not. Similarly, in lanes containing chicken embryo fibroblast poly(A)$^+$ RNA (Fig. 5*A*, lane 2) or cultured MSB-1 cell (chicken lymphocytes transformed by Marek disease virus) poly(A)$^+$ RNA (Fig. 5*A*, lane 1; ref. 39), the homologous cellular RNAs are *ca.* 5500 and 5000 bases, respectively, and appear to be less abundant than ALA synthase mRNAs in either liver cells or anemic red cells.

Under normal high-stringency blot washing conditions, the liver ALA synthase mRNA appears to be highly homologous in nucleic acid sequence to the erythroid ALA synthase gene probe (Fig. 5*A*, lane 7). Only after treatment of the blot with RNase (Fig. 5*B*) can it be discerned that all these RNAs differ in primary nucleotide sequence except in the homologous tissue source. Thus, we conclude from these data that ALA synthase is encoded by a minimum of two (erythroid and liver) separate tissue-specific genes.

DISCUSSION

We report in this paper the isolation of a cDNA sequence corresponding to the erythroid-specific gene encoding the first enzyme contributing to heme biosynthesis in animal cells, ALA synthase. The mRNA encoding this particular isozyme is approximately 2000 nucleotides in length [including poly(A)] and appears, from this preliminary analysis, to be uniquely expressed in erythroid cells. While the presumptive mRNA encoding the liver ALA synthase is highly homologous to the erythroid gene, under very stringent washing conditions the liver ALA synthase mRNA can be distinguished from its erythroid counterpart on the basis of primary nucleotide sequence, thereby implying that the ALA synthase enzymes are encoded by a multigene family whose members are responsive to different developmental stimuli and regulation.

What is the likelihood that we have mistakenly identified a single ALA synthase locus as a family of genes encoding different tissue-specific enzymes, whereas only one gene actually exists and the nuclear RNA transcribed from that locus is differentially regulated, giving rise to multiple isozymes of ALA synthase? We see only one way that the presence of a single ALA synthase gene in chicken cells could be consistent with the results presented in Fig. 5. There is a high probability that the red cell enzyme shares only a subset of the (presumptive) multiple epitopes on the ALA synthase molecule found in the liver, against which the rabbit antiserum was raised. Furthermore, it is at least theoretically possible that while the major antigenic determinant is held in common and is highly conserved within the ALA synthase isozymes of both tissues, this predominant epitope is derived from two (or more) exons of the same transcription unit, one of which is used as part of the red cell ALA synthase mRNA and another as part of liver ALA synthase mRNA. This alternative explanation would also lead to the observed results presented in Fig. 5. However, we view this possibility as very unlikely, since differential use of equivalently functional coding sequence exons within a single transcription unit has not been previously observed. The more conservative conclusion, that the tissue-specific mRNAs arise from different genetic loci, seems likely to be correct. A final resolution to this, and several other questions naturally arising from data presented in this report [e.g., the origin of the large protein that is immunoprecipitable with the anti-ALA synthase antibody in erythroid progenitor cells (Fig. 1, lane C) and the origin of RNA species too small to encode liver ALA synthase in total liver RNA (Fig. 5, lanes 5, 6, and 7)] should be immediately forthcoming from analysis of the chromosomal gene(s).

We gratefully acknowledge the efforts of our laboratory colleagues (Hyeong-Reh Choi, Priscilla Wilkins Stevens, and Susan J. Stamler) for donation of RNA samples, and we further thank Tom St. John for fruitful discussions of antibody screening. We thank Drs. G. Kikuchi and N. Watanabe (Tohoku University School of Medicine) for giving us the anti-chicken liver ALA synthase antibody. This work was supported by grants from the National Institutes of Health (to J.B.D. and J.D.E.), by a postdoctoral fellowship from the Arthritis Foundation (to M.Y.), by a National Institutes of Health Research Career Development Award (to J.B.D.), and by an American Cancer Society Faculty Research Award (to J.D.E.).

1. Kikuchi, G., Kumar, A., Talmage, D. & Shemin, D. (1958) *J. Biol. Chem.* 233, 1214–1219.
2. Kikuchi, G. & Hayashi, N. (1981) *Mol. Cell. Biochem.* 37, 27–41.
3. Sassa, S. & Granick, S. (1970) *Proc. Natl. Acad. Sci. USA* 67, 517–522.
4. Tyrrell, D. L. J. & Marks, G. S. (1972) *Biochem. Pharmacol.* 21, 2077–2093.
5. Tomita, Y., Ohashi, A. & Kikuchi, G. (1974) *J. Biochem. (Tokyo)* 75, 1007–1015.
6. Yamamoto, M., Hayashi, N. & Kikuchi, G. (1983) *Biochem. Biophys. Res. Commun.* 115, 225–231.
7. Whiting, M. J. (1976) *Biochem. J.* 158, 391–400.
8. Srivastava, G., Brooker, J. D., May, B. K. & Elliott, W. H. (1980) *Biochem. J.* 188, 781–788.
9. Hayashi, N., Terasawa, M., Yamauchi, K. & Kikuchi, G. (1980) *J. Biochem. (Tokyo)* 88, 1537–1543.
10. Yamamoto, M., Hayashi, N. & Kikuchi, G. (1981) *Arch. Biochem. Biophys.* 209, 451–459.
11. Yamamoto, M., Hayashi, N. & Kikuchi, G. (1982) *Biochem. Biophys. Res. Commun.* 105, 895–900.
12. Hayashi, N., Kurashima, Y. & Kikuchi, G. (1972) *Arch. Biochem. Biophys.* 148, 10–21.
13. Yamauchi, K., Hayashi, N. & Kikuchi, G. (1980) *J. Biol. Chem.* 255, 1746–1751.
14. Wada, O., Sassa, S., Takaku, F., Yano, Y., Urata, G. & Nakao, K. (1967) *Biochim. Biophys. Acta* 145, 585–587.
15. Bottomly, S. (1968) *Biochim. Biophys. Acta* 159, 27–37.
16. Wood, J. S. & Dixon, R. L. (1972) *Biochem. Pharmacol.* 21, 1735–1744.
17. Wood, J. S. (1973) *Mol. Pharmacol.* 10, 389–397.
18. Bishop, D. F., Kitchen, H. & Wood, W. A. (1981) *Arch. Biochem. Biophys.* 206, 380–391.
19. Watanabe, N., Hayashi, N. & Kikuchi, G. (1983) *Biochem. Biophys. Res. Commun.* 113, 377–383.
20. Sassa, S. (1976) *J. Exp. Med.* 143, 305–315.
21. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1194–1198.
22. Young, R. A. & Davis, R. W. (1983) *Science* 222, 778–782.
23. Hoyt, M. A., Knight, D. M., Das, A., Miller, H. I. & Echols, H. (1982) *Cell* 31, 565–573.
24. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
25. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* 15, 687–701.
26. Dodgson, J. B., Strommer, J. & Engel, J. D. (1979) *Cell* 17, 879–887.
27. Dolan, M., Dodgson, J. B. & Engel, J. D. (1983) *J. Biol. Chem.* 258, 3983–3990.
28. Storti, R. V. & Szwast, A. E. (1982) *Dev. Biol.* 90, 272–283.
29. Ricciardi, R. P., Miller, J. S. & Roberts, B. E. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4927–4931.
30. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* 12, 7035–7056.
31. Brush, D., Dodgson, J. D., Choi, O.-R., Wilkins Stevens, P. & Engel, J. D. (1985) *Mol. Cell. Biol.*, in press.
32. Bittner, M., Kupferer, P. & Morris, C. F. (1980) *Anal. Biochem.* 102, 459–472.
33. Berg, H., Palmieri, S., Freudenstein, C., Zenegruf, H. & Graf, T. (1982) *Cell* 28, 907–919.
34. Dreyfuss, G., Adam, S. A. & Choi, Y. D. (1984) *Mol. Cell. Biol.* 4, 415–423.
35. Hancock, K. & Tsang, V. C. W. (1983) *Anal. Biochem.* 133, 157–162.
36. Zinn, K., DiMaio, D. & Maniatis, T. (1983) *Cell* 34, 865–879.
37. McClements, W. & Stalker, A. M. (1977) *Science* 196, 195–197.
38. Cleveland, D. W., Lopata, M. A., MacDonald, R. J., Cowan, N. J., Rutter, W. J. & Kirschner, M. W. (1980) *Cell* 20, 95–105.
39. Akiyama, Y. & Kato, S. (1974) *Biken J.* 17, 105–116.