CHICKEN GLOBIN mRNA AND ITS PRECURSOR

A thesis submitted for

the Degree of

Doctor of Philosophy

in the

University of Adelaide

by

Robert John Crawford, B.Sc. (Hons.)

Department of Biochemistry

June 1977

TABLE OF CONTENTS

SUMMARY

1. The work described in this thesis concerns the characterization of globin mRNA from avian erythroid cells. In particular, the organization of sequences within chicken globin mRNA was determined, and the possibility that polysomal globin mRNA is derived from a larger molecular weight precursor investigated.

2. Several globin complementary DNAs (cDNAs) of different lengths were prepared, characterized and used to study the organization of sequences within polysomal chicken globin mRNA. Formamide polyacrylamide gel electrophoresis showed that chicken globin mRNA contains about 650 nucleotides. Since only 435 of these code for globin, a further 215 are not translated, and their function and position are not known. This work has produced the following conclusions:

(a) 45 - 50 of these untranslated nucleotides are present as poly(A) at the 3' terminus.

(b) The 3' untranslated region of chicken globin mRNA is at least 90 nucleotides in length.

This minimal estimate is based on data derived from hybridization of defined lengths of chicken globin cDNA to rabbit globin mRNA. The percentage of chicken globin mRNA sequences which hybridize to rabbit globin mRNA is directly proportional to the length of cDNA in each case. This relationship holds for lengths of cDNA from 115 to 620 nucleotides. The low percentage homology for short cDNA molecules is not due to their being short *per se*. In homologous mRNA-excess hybridizations (chicken cDNA/chicken mRNA) all preparations were completely protected from S1 nuclease digestion.

(c) The pattern of nucleotide divergence in the 3' untranslated region of chicken and rabbit globin mRNAs is different to that of the coding regions of these molecules. The evidence suggests that nucleotide sequences have diverged to a greater extent in the 3' untranslated region than in the coding region.

The combined data is used to formulate a regional map of chicken globin mRNA.

3. Experiments in which chicken globin cDNA was re-annealed to complete gemone chicken DNA were carried out. These showed that there are a few copies only (2 - 5) of the globin gene in the chicken genome. Hence high levels of globin protein synthesis in chicken erythroid cells are not achieved by globin gene reiteration.

4. A method that enables gene specific priming of complementary DNA synthesis was developed. Reverse transcriptase can use "short" polymerase I cDNA to prime further cDNA synthesis along the mRNA template. The product of extended cDNA synthesis is identical in length and hybridization properties to oligo(dT) primed "full length" transcriptase cDNA.

5. With the view to studying the primary transcription product of chicken globin genes, a method was developed for the isolation of total RNA from avian erythroblasts. Loss of RNA, and RNA degradation during the isolation procedure were shown to be insignificant. In particular, the 40S - 45S precursor to 18S and 28S ribosomal RNAs was detected within isolated erythroblast RNA. Furthermore, control experiments involving keratin mRNA showed that mRNA sequences present at the level of about one to ten molecules per cell could be recovered in the RNA preparation. The isolated erythroblast RNA preparation was considered to be representative of erythroblast RNA present *in vivo*.

6. Globin cDNA was used to probe erythroblast RNA for globin RNA sequences larger than 10S polysomal globin mRNA. Two types of hybridization were carried out. RNA-excess hybridizations, using cDNA of high specific radioactivity showed that globin RNA sequences could not be detected in the size range 20S to 50S.. However, a vast excess of polysomal globin mRNA masked the detection of possible higher molecular weight precursors in the 10S to 20S region. In a second series of experiments, cDNA-excess hybridization was used. Erythroblasts were pulse-labelled with ^{3}H-uridine and fractioned RNA was hybridized to excess unlabelled globin cDNA. In these experiments, newly made globin RNA sequences can be detected exclusively. An RNA species that contains globin sequences was located in the 14S region of sucrose gradients. The half-life of this RNA species is about 4 minutes. It has the properties of a higher molecular weight precursor to polysomal globin mRNA.

STATEMENT

This thesis contains no material which has been accepted for the award of any other Degree or Diploma of any University, and to the best of my knowledge and belief contains no material previously published or written by any other person, except when due reference is made in the text.

R. J. CRAWFORD

June 1977

ACKNOWLEDGEMENTS

I should like to express my thanks to Dr. J. R. E. Wells for his stimulating and friendly supervision during the course of my work, and to Dr. Dave Kemp for his valuable practical advice during the initial stages of this work. I am grateful to Drs. Roger Harlow and Andy Scott, and Messrs. Paul Kreig, Allan Robins and other members of the laboratory who contributed enthusiastic discussion and encouragement, and to Mrs Jan Dinan for competent assistance.

Thanks are also due to Professor W. H. Elliott for allowing me to work in the Department of Biochemistry in a stimulating and pleasant environment, and to Bronwyn Young who typed this thesis.

The generous cooperation of Dr. M. A. Chirigos and Dr. J. W. Beard in providing avian myeloblastosis virus, and transcriptase I is gratefully acknowledged; also Dr. W. J. Rutter, Dr. D. L. Williams and Dr. S. Longacre for cooperation in supplying partially purified reverse transcriptase (from which transcriptase II was prepared).

I also wish to thank my wife Dianne for her companionship and for providing many home comforts during these three years.

During this work I was financially supported by a Commonwealth Postgraduate Research Award, and my wife (not necessarily in that order).

ABBREVIATIONS

cDNA	complementary DNA
$Cr_o t$	product of initial RNA concentration (moles nucleotide, litre^{-1}) and time (sec)
$C_o t$	product of initial DNA concentration (moles nucleotide, litre^{-1}) and time (sec)
dNTP	deoxynucleotide triphosphate
oligo(dT)	oligo (thymidylic acid)
oligo(U)	oligo (uridylic acid)
poly(A)	poly (adenylic acid)
A	optical absorbance (subscript denoting wavelength at which measured)
BSA	bovine serum albumin
mRNA	messenger RNA
rRNA	ribosomal RNA
tRNA	transfer RNA
transcriptase	avian myeloblastosis virus RNA dependent DNA polymerase (EC 2.7.7.-)

polymerase I	*E.coli* DNA polymerase I (EC 2.7.7.7)
S1 nuclease	(EC 3.1.4.21)

CHAPTER 1.

INTRODUCTION AND LITERATURE SURVEY

CHAPTER 1. INTRODUCTION AND LITERATURE SURVEY

1.1 INTRODUCTION

This thesis is particularly concerned with eukaryote cell differentation at the molecular level - the process that enables cells with identical genomes to synthesize different proteins. The model system used for this study is the adult avian erythroid cell series. Within this system, a major differentation step occurs when pre-erythroid stem cells are committed to the erythroid series. The synthesis of globin proteins within these cells is one of the principal phenotypic developments of this differentation step.

This work studies aspects of globin gene expression within adult avian erythroid cells, and is principally concerned with characterising the products of transcription of globin genes. Such a study is relevant in determining factors influencing globin gene expression. It is likely that some of the untranslated regions within polysomal globin mRNA are responsible for regulating globin protein synthesis, either at the transcriptional or post-transcriptional level. Accordingly, the nucleotide length of untranslated regions at both the 5' and 3' end of polysomal globin mRNA were estimated. An extensive search for the primary transcription products of adult avian globin genes was also carried out.

The remainder of this chapter is devoted to reviewing the literature on eukaryote differentiation. Evidence concerning the origin of eukaryote mRNA is discussed in detail.

The review is not exhaustive, but it does attempt to cover the important findings relevant to the work described in this thesis. It is complete to January 1977.

1.2 LITERATURE SURVEY

1.2.1 Differentiation in Eukaryotes

Eukaryote differentiation starts shortly after fertilization. Progeny cells of a single fertilized ovum (Blastomeres) are organised into tissues that perform many different specialized functions.

Although many cellular changes occur during differentiation, this thesis is principally concerned with events occurring at the molecular level. This review is confined to surveying molecular aspects of differentiation only.

Within differentiated cells, only a small proportion of the proteins that can be coded for by the genome, are synthesized. Hence much of the genetic information available to the cell is not expressed. One possible way that genetic information might be repressed as differentiation proceeds, is by the progressive deletion of regions of DNA coding for proteins whose further syntheses are no longer required for cellular function. However, several different approaches have shown that this proposed mechanism of control does not operate. The most spectacular demonstration was provided by Gurdon and Woodland (1970) who showed that a fully differentiated Xenopus laevis cell had retained all the genetic information needed for the development of another complete adult. Experiments involving phytohaemagglutinin stimulation of differentiated lymphocytes *in vitro* (Pogo *et al.* 1960), and the activation of chicken erythrocyte nuclei upon fusion with Hela cells (Harris 1970) have also showed that gene sequences are not physically lost during differentiation. Rather, with rare exceptions, the genome within different cell types remains

constant, and differentiation occurs when different regions of the genome are expressed.

Prokaryote gene expression appears to be mainly controlled at the level of transcription. However the control of eukaryote gene expression is more complex. Eukaryote transcription, unlike that in prokaryotes, is generally not coordinate with translation, so that control mechanisms operating post-transcriptionally can assume importance. There are four principal levels at which eukaryote genes can be regulated.

1) Control of transcription of DNA into RNA (Gurdon, 1968).

2) Control at the level of RNA transfer from the site of transcription in the nucleus to the site of translation in the cytoplasm (Scherrer and Marcaud, 1968; Warocquier and Scherrer, 1969).

3) Control at the level of translation of mRNA into protein (Harris, 1970; Ilan and Ilan, 1971; Maxwell *et al.*, 1971).

4) Post translational controls may operate at the level of active protein assembly from precursor subunits.

Control of eukaryote gene transcription is the main concern of this thesis, and is therefore the level of control discussed in most detail in this review.

1.2.2 Transcriptional Control

The selective transcription of separate genes within cells containing identical genomes is the essence of transcriptional control. Hybridization data, showing that different RNA sequences are synthesized in different tissues (Smith, Church, and McCarthy, 1969; Grouse, Chilton,

and McCarthy, 1972; Axel *et al.*, 1976) provides ample evidence that it is an important level of control in eukaryote gene expression, although a proportion of mRNA sequences are common to cells performing different specialized functions (Axel *et al.*, 1976). These mRNA sequences are probably needed for essential cellular functions common to different tissues.

How eukaryote transcriptional control operates at the molecular level is not yet well understood, although extra-nuclear events must be involved. Globin mRNA synthesis (Ramirez *et al.*, 1975) and ovalbumin mRNA synthesis (Harris *et al.*, 1973) are both influenced by specific hormonal stimuli. Similarly cytoplasmic factors appear to exert some form of positive transcriptional control, since transcription within chicken erythrocyte nuclei is re-activated after fusion with Hela cell cytoplasm (Harris, 1970).

The ultimate response to transcriptional control factors external to the nucleus, must occur at the chromatin level. How the information is mediated is not clear, but it must result in the recognition of specific transcriptional initiation and termination signals along the chromatin. It has been proposed that histone proteins (Stedman and Stedman, 1950) non-histone proteins (Gilmour and Paul, 1973; 1975; Stein *et al.*, 1975) and chromosomal RNA (Huang and Bonner, 1965) might be involved in the recognition of specific regions of DNA, although later evidence suggests that histones do not play a specific role in the control of gene transcription (DeLange *et al.*, 1969). Similarly chromosomal RNA is probably not a control element (Artman and Roth, 1971; Tolstoshev and Wells, 1974). However it is possible that RNA sequences are involved

in maintaining the gross morphology of chromatin (Benyajati and Worcel, 1976).

Genetic studies were initially used to characterise prokaryote transcriptional control mechanisms (Jacob and Monod, 1961), eventually leading to a clear understanding of the molecular basis for prokaryote control at the site of DNA. This approach is generally not possible with eukaryotes, principally because of their diploid nature and much greater genome complexity (Britten and Kohne, 1968). Nevertheless, it is useful to discuss whether there is evidence for similarities in prokaryote and eukaryote transcriptional control.

1) RNA Polymerase

Specificity of transcription is mediated in part by RNA polymerase in both prokaryotes and eukaryotes, although there are important differences between these enzymes from prokaryote and eukaryote sources. Within E.coli, RNA polymerase specificity involves the interaction of subunits with the core enzyme. In particular the σ factor permits the recognition of some specific start signals along the template DNA (Burgess *et al.*, 1969) whereas the ψ factor which possibly recognizes different initiation sequences in a similar manner, is required for the transcription of E.coli ribosomal RNA genes (Travers *et al.*, 1970). Entirely different forms of prokaryote polymerases may also occur (Chamberlain et al., 1970).

Within eukaryotes, there are at least three major forms of RNA polymerase in the cell. One is programmed to synthesize ribosomal

RNA, another HnRNA, while a third is involved in 5S and tRNA synthesis (Roeder and Rutter, 1969; Wobus *et al.*, 1971; Chambon *et al.*, 1972). Hence it is evident that specific transcription of eukaryote genes is controlled to some extent at the RNA polymerase level.

RNA polymerase interaction with other protein components of chromatin appears to be an essential feature of control at the RNA polymerase level. Xenopus RNA polymerase III selectively directs the in vitro transcription of Xenopus oocyte 5S genes from chromatin only, and not naked DNA (Parker and Roeder, 1977).

2) Regulatory molecule interaction with DNA sequences

Finer regulation of gene transcription in prokaryotes is mediated by factors other than RNA polymerase, such as repressor and activator proteins, and operator sequences (Jacob and Monod, 1961). The role of each of these elements has been worked out in detail for the lac operon. mRNA synthesis for this operon is prevented in the absence of inducer, by the interaction of repressor protein with the operator (Gilbert and Muller-Hill, 1966) - a specific sequence of 21 base-pairs immediately preceding the structural genes.

The cyclic AMP binding protein positively controls transcription of the lac operon, as well as other operons associated with catabolite repression, by influencing the binding of RNA polymerase to the promotor sites (de Crombrugghe *et al.*, 1971; Eron *et al.*, 1971; Nissley *et al.*, 1971).

Specific repressor proteins regulating phage gene expression

have also been detected (Ptashne, 1967; Chadwick *et al.*, 1970).

Hence there is ample evidence that prokaryote regulatory proteins can recognize unique nucleotide sequences. Chromatin reconstruction experiments, in which chromatin is dissociated into separate histone, non-histone and DNA components, reassembled and transcribed in vitro, suggest that eukaryote proteins may similarly recognize specific DNA sequences. These experiments, in which the products of transcription are examined by hybridization, suggest that it is the non-histone proteins that impart specificity to eukaryote gene transcription. Early studies could be criticized, since only repeated sequences were detected (Gilmour and Paul, 1970; Spelsberg *et al.*, 1971). However later experiments showed that specific transcription of unique structural genes occurred only when specific non-histone proteins were present on the chromatin (Gilmour and Paul, 1973; 1975; Stein *et al.*, 1975). Some criticism can still be levelled at these experiments, since bacterial RNA polymerase was used to transcribe the chromatin. No specificity can therefore be attributed to RNA polymerase in these systems.

The non-histone proteins are generally the least-well characterised component of chromatin, so that the isolation of specific regulatory proteins from eukaryotes may not yet be possible. Despite these problems, several groups claim to have detected tissue-specific non-histone proteins (Loeb and Cruezet, 1969; Platz *et al.*, 1970; Teng *et al.*, 1971) or changes in the pattern of non-histones occurring after hormonal induction of specific genes (O'Malley

et al., 1967; Spelsberg *et al.*, 1973). Such studies however, are inconclusive because it is uncertain whether such proteins are the cause or result of transcription, or even whether the putative non-histones are cytoplasmic contaminants, or true components of chromatin.

Specific RNA sequences are postulated to play a regulatory role in a model proposed for eukaryote transcriptional control (Britten and Davidson, 1969) - proteins were also suggested as possible candidates for this role (Davidson and Britten, 1973). To date there is no evidence that RNA plays a regulatory role in eukaryote gene transcription, although there are advantages to the cell that might be envisaged if RNA, rather than protein was the eukaryote regulatory unit. For example the complete control network might be contained in the nucleus, rather than depending on the transport of regulatory protein mRNAs, and regulatory proteins themselves across the nuclear membrane. Techniques are not yet available to test this hypothesis, although the availability of cloned gene fragments should greatly facilitate such studies.

3) Transcriptional control sequences with the genome

Transcriptional control sequences have been well characterised in prokaryotes, and are predicted to occur in eukaryotes. In particular, the nature of the lac operon operator of E.coli gives some insight into the specific interaction between operator and repressor protein.

The operator for this operon consists of 21 base-pairs - a sufficient number to avoid the chance of this sequence occurring randomly elsewhere in the genome, thereby ensuring that the interaction of operator and repressor protein is unique. 16 of these base-pairs are related by a 2 fold axis of symmetry, permitting two of the four lac repressor subunits to bind simultaneously to the operator.

The nucleotide sequence of the right hand operator of phage λ has also been determined (Maniatus, Jeffrey and Kleid, 1975). This sequence is also palindromic, leading to the suggestion that hair-pin loop formations might generally be involved in the specific recognition of DNA sequences by DNA binding proteins.

Recently, a different type of DNA sequence that regulates gene expression in bacterial plasmids was identified (for review see Cohen, 1976). These sequences, termed insertion sequences, are transposable, and are capable of regulating the expression of nearby genes as a consequence of their insertion into, or excision from prokaryotic genomes.

Detailed information about eukaryote genes control sequences is not available. The much greater genome size and genome complexity makes any attempted analysis of eukaryote DNA sequences difficult. As well as this, the occurrence of reiterated sequences within eukaryotes must be explained in any proposed mechanism for eukaryote transcriptional control - no such reiterated sequences have been detected in prokaryotes.

Unlike prokaryotes, eukaryote genes that are co-ordinately expressed (apart from histone genes, (Kedes *et al.*, 1975)) may not generally be clustered together in the genome. α and β globin genes, which are co-ordinately expressed in erythroid cells, are located on different chromosomes. Therefore one unit of transcription may include only one structural gene. Accordingly, eukaryote control mechanisms must permit the concurrent transcription of co-ordinated genes located at different sites in the genome.

Studies on the organization of eukaryote DNA sequences provide evidence that is consistent with repeated DNA sequences playing a regulatory role in gene transcription. Unique nucleotide sequences, corresponding to structural genes are interspersed with moderately repeated DNA sequences in the sea urchin genome (Davidson *et al.*, 1975) thus fulfilling one prediction of the Britten-Davidson model of gene regulation (Britten and Davidson, 1969; Davidson and Britten, 1973). The transcription of genes that are co-ordinately expressed might be controlled in part by interactions of specific molecules with these moderately repeated sequences.

Hair-pin loop structures have been detected in eukaryote DNA from widely differing sources (Schmid *et al.*, 1975; Cech and Hearst, 1975). However these inverted repeat sequences are generally 10 to 1000 times larger than those located in the control regions of prokaryote genes. Their role may therefore be more complex than to just act as a recognition site for a single regulatory molecule.

In a manner analagous to that of insertion sequences in bacterial

plasmids, eukaryote DNA sequences may also be transposed within the genome. Movement of immunoglobulin genes within the mouse genome has been detected in developing embryos (Hozumi and Tonegawa, 1976). It is possible that control elements of eukaryote genes may also be transposed within the genome. Hence the expression of one structural gene might be co-ordinated separately with the expression of other genes.

1.3 ORIGIN OF EUKARYOTE mRNA

It has been proposed that mRNAs in eukaryote cells are generally derived by post-transcriptional modification of larger RNA precursor molecules (Darnell, Jelinek and Molloy, 1973). An alternative view, held by Davidson and Britten (1973) stated that, as yet, there is no hard evidence that eukaryote mRNAs originate from higher molecular weight precursors.

This section of the literature review is concerned with the nature of primary transcripts of eukaryote structural genes, and considers these conflicting viewpoints.

1.3. Nuclear RNA

Three different classes of RNA have been identified within eukaryote nuclei (for recent view, see Perry, 1976). Firstly, precursor ribosomal RNA was detected within the nucleolus (Perry, 1962; Scherrer *et al.*, 1963), and later tRNA sequences were detected within a higher molecular weight precursor RNA species (Burdon and Clason, 1969).

A third class of RNA, termed heteregeneous nuclear RNA (HnRNA)

is also synthesized within eukaryote nuclei (Sibitani *et al.*, 1962; Georgiev and Mantieva, 1962). This class of nuclear RNA is characterised in part by its high molecular weight and heterogeneous size, varying from about 1000 nucleotides up to 30,000 nucleotides in length (for review see Lewin, 1975). Recent experiments using completely denaturing conditions have confirmed the molecular unity of HnRNA, and have shown that the high molecular weight is not generally due to the aggregation of smaller molecular weight RNA species (Derman *et al.*, 1976).

The ability to be rapidly labelled with radioactive RNA precursors, and a nucleotide composition that resembled that of both DNA and mRNA further characterised this class of nuclear RNA, and provided the initial circumstantial evidence that HnRNA might contain higher molecular weight precursors to eukaryote polysomal mRNA. Since these early experiments, several features of HnRNA have been directly compared with those of cytoplasmic mRNA. However, this approach has not proved a precursor relationship between HnRNA and mRNA - rather, it has provided additional circumstantial evidence only.

1.3.2 A comparison of HnRNA with mRNA sequences

1) Size of HnRNA and mRNA

The size distribution of HnRNA is generally about four to five times that of polysomal mRNA (Derman and Darnell, 1974) although in Dictostelium, it is on average only 20% larger (Firtel and Lodish, 1973), and in Balbiani rings, about the same size as mRNA (Egházi, 1976). Whatever the size difference, some degree of overlap is generally apparent.

It is possible that the larger HnRNA sequences are higher molecular weight precursors to polysomal mRNAs. The rapid turnover rate of HnRNA within the nucleus is consistent with this view. However, overlap in the size distributions of HnRNA and mRNA populations suggests that an alternative proposal, that some mRNAs are transcribed directly into polysomal length sequences, cannot be excluded.

2) Complexity analysis of HnRNA and mRNA

Complexity analysis of HnRNA and mRNA sequences has been carried out for several systems (Herman *et al.*, 1976; Ryffel and McCarthy, 1975; Galau *et al.*, 1975; Getz *et al.*, 1975) and has shown that HnRNA is generally five to ten times more complex than mRNA populations. On this basis, up to 10% to 20% of HnRNA sequences might be represented in mRNA. Estimates using other approaches however, indicate that only 2% of HnRNA nucleotides in mouse cells at least, enter polysomal mRNA (Bradhurst and McConkey, 1974). The larger proportion of HnRNA that is rapidly turned over within the nucleus (Soiero *et al.*, 1968) might well account for the remainder of HnRNA that definitely does not contain mRNA sequences.

HnRNA contains variable proportions of different repetitive sequences (Pagoulatos and Darnell, 1970; Spradling *et al.*, 1974). Since mRNA sequences appear tc be generally derived from unique regions of the genome, it seems likely that the rapidly turned-over fraction of HnRNA contains the majority of these repetitive sequences.

The function of the large proportion of HnRNA nucleotides not transported to the cytoplasm is unknown. It is possible that they contain protein coding sequences that are not expressed, or that they are involved in gene regulation. Recently it was proposed that areas of the genome containing particular structural genes to be expressed, were opened up by total transcription of this region (Crick, 1976). Specific transcriptional control sequences for the structural genes in this area of the genome would thus be exposed to regulatory molecules. Hence in this scheme, HnRNA molecules are equated with the total transcripts of this area of the genome, and are not necessarily seen as mRNA precursors. At this stage, there is no compelling evidence against this proposal.

3) Post-transcriptional modification of HnRNA and mRNA

a) Polyadenylation

With the exception of histone mRNAs (Adesnik *et al.*, 1972) poly (A) sequences are added post-transcriptionally to most eukaryote mRNAs (Edwards *et al.*, 1971; Lee *et al.*, 1971). The role of the 3' poly (A) sequence is uncertain, although it is likely that it confers some protection on mRNA sequences against degradation during translation (Huez *et al.*, 1974; Marbaix *et al.*, 1975).

Poly (A) sequences have also been found at the 3' end of some HnRNA molecules (Edwards *et al.*, 1971; Derman and Darnell, 1974). However kinetic analysis revealed that much

of the poly (A) within HnRNA decays rapidly within the nucleus, and is not transported to the cytoplasm (Perry *et al.*, 1974). Furthermore polyadenylation of mRNA sequences can occur within the cytoplasm (Derman and Darnell, 1974; Perry *et al* , 1974). Hence a direct relationship between the poly (A) within HnRNA, and that on mRNA, has not been demonstrated, although it is possible that some polyadenylated HnRNA sequences are transported to the cytoplasm, and are precursors to mRNA (Price *et al.*, 1974). This view is consistent with the proposal that mRNA sequences are located at the 3' end of some HnRNA molecules. However the recent discovery of 5' caps in both mRNA and HnRNA has led to an alternative proposal, that mRNA sequences are located at the 5' end of HnRNA (Perry *et al.*, 1975).

(b) 5' "capping", and methylation of internal residues.

7 methyl guanosine residues joined to RNA sequences by a 5' - 5' pyrophosphate linkage have been found at the 5' terminus of both HnRNA and mRNA (Adams and Corey, 1975; Perry *et al.*, 1975). Internal base methylations have also been detected in HnRNA and mRNA.

The significance of these similarities is not known. While the role of the methylated residues is unclear, the role of the 7 methyl guanosine may be to protect the RNA against 5' exonuclease attack (Furuichi *et al.*, 1977). It is uncertain whether the cap is necessary for translation. While it appears that several eukaryote mRNAs are not translated when

the cap is removed (Both *et al.*, 1975), some uncapped viral mRNAs still appear to be translated (Rose and Lodish, 1976).

Whatever the role of the cap, its presence at the 5' end of both mRNA and HnRNA provides circumstantial evidence for the proposal that some mRNA sequences are located at the 5' end of some HnRNA molecules. However, there is no evidence to refute a third proposal that mRNA sequences are internally situated within HnRNA molecules. The mRNA sequences could be derived by cleavage of both 5' and 3' HnRNA regions, and the subsequent addition of 3' poly (A) and 5' cap might then occur.

1.3.3 Identification of precursors to defined mRNAs

The study of complex mixtures of mRNAs has failed to prove the existence of mRNA sequences in higher molecular weight precursors. An alternative approach involves mRNAs of defined coding specificity that are produced in abundance in a particular cell type. In these experiments, DNA sequences complementary to a unique mRNA, are used to probe RNA fractions of known molecular weight.

Initially it was claimed that by using this approach, globin sequences could be detected within several higher molecular weight RNA size classes (Imaizumi *et al.*, 1973). However, these initial experiments still failed to demonstrate convincingly, a unique globin mRNA precursor species.

Other attempts were made to detect higher molecular weight precursors to both ovalbumin mRNA (McKnight and Schimke , 1974),

and silk fibroin mRNA (Lizardi, 1976). In each case, it was not possible to detect a precursor mRNA species that was different in size from polysomal mRNA. Although RNA degradation was not apparent during the RNA extraction for these experiments, the possibility that some RNA degradation had occurred, could not be specifically excluded. It is still difficult to be certain that ovalbumin and silk fibroin mRNAs are not derived from short-lived higher molecular weight precursors.

More recently, convincing evidence that some globin mRNAs at least originate from larger primary transcripts was provided (Curtis and Weissman, 1976; Ross, 1976). In both induced Friend cells and mouse foetal liver cells, globin mRNA sequences can be found in RNA molecules of 14S - 18S in size. The precise location of the polysomal globin mRNA sequence within this precursor, and the way in which the precursor is processed into polysomal-size mRNA has yet to be worked out.

CHAPTER 2.

MATERIALS AND METHODS

CHAPTER 2. MATERIALS AND METHODS

2.1 MATERIALS

General chemicals were of A.R. grade. E.coli DNA polymerase I was from Boehringer (Mannheim, Germany) and Proteinase K and Actinomycin D were from Merck (Darmstadt, Germany). Ribonuclease A (pancreatic), ribonuclease T1, bovine serum albumin (fraction V), phenylhydrazine (twice recrystallized) and unlabelled deoxynucleotide triphosphates were purchased from Sigma Chemical Co. (U.S.A.). All media was from Commonwealth Serum Laboratories (Melbourne, Australia). Oligo $(dT)_{12-18}$ was from P.L. Biochemicals (U.S.A.). Transcriptase I was provided by Dr. M. A. Chirigos and Dr. J. W. Beard, and partially purified transcriptase, from which transcriptase II was prepared was supplied by Dr. W. J. Rutter, Dr. D. L. Williams and Dr. S. Longacre.

Isotopes were obtained from the following sources:
α -^{32}P - dGTP and α -^{32}P - dCTP were a gift from Dr. R. H. Symons.
5 -^{3}H - dCTP (15 - 30 Ci/mmole) was from New England Nuclear (U.S.A.),
5 -^{3}H - dGTP (12 Ci/mmole) from the Radiochemical Centre, Amersham, England, and 5 - 6 -^{3}H uridine (30.7 Ci/mmole) from I.C.N., California, U.S.A.

2.2 METHODS

Unless otherwise stated, all operations were performed at 4°C.

2.2.1 Avian Erythroid Cells

1) Preparation of Cells

The procedures followed were essentially those of Williams (1971). Immature erythroid cells were obtained from White

Leghorn-Austalorp pullets (6 months old), which had been rendered anaemic by five daily injections of 2.5%. phenylhydrazine in 47.5% ($\frac{V}{V}$)ethanol pH 7.

a) Reticulocytes

Blood was obtained from anaemic chickens by heart puncture into NKM (0.15M NaCl, 0.005M KCl, 0.002M $MgCl_2$) containing 1 mgm/ml of heparin. The reticulocytes were washed several times by centrifugation (1000g, 5 minutes) and resuspension into NKM. The buffy, white-cell layer could be removed during this operation by aspiration.

b) Erythroblasts

Anaemic chickens were decapitated, and bone-marrow from the two long bones of each leg collected into ice-cold NKM. The cells from the bone-marrow were washed through cheese-cloth, and centrifuged on discontinuous BSA density gradients, consisting of BSA solutions of refractive indices 1.3800 (ρ = 1.068) and 1.3720 (ρ = 1.059). After centrifugation at 3,800g for 45 minutes at 4°C, erythroblasts were collected from the interphase of the two BSA solutions, and washed twice with NKM before use.

2) In vitro incubation of cells with ^{3}H-Uridine

For the *in vitro* incorporation of ^{3}H-uridine, washed erthroblasts were incubated in Eagles Basal Medium, containing 10% chicken serum (freshly prepared), and buffered with bicarbonate/CO_2 to Ph 7.0. ^{3}H-uridine, in the required amount, was dried within the incubation vessel prior to adding incubation medium. Cells were

suspended at 1% to 5% ($\frac{W}{V}$) in the medium, and slowly shaken at 39°C. The incorporation was stopped by centrifuging the cells (1000g, 3 mins) and washing with ice-cold NKM, before isolation of the labelled RNA.

2.2.2 Isolation and Analysis of RNA

All procedures involving RNA employed sterile glass-ware and solutions, and were carried out as rapidly as possible at 0°C unless otherwise stated.

1) Chicken Globin mRNA preparation

a) Polysome preparation

Washed reticulocytes were lysed by the addition of 2 volumes of ice-cold 2mM $MgCl_2$, and vortex mixing for 5 minutes. An equal volume of TKM (0.01M Tris-Cl pH 7.4, 0.2M KCl, 0.002M $MgCl_2$) was then added to restore tonicity, and the nuclei were removed by centrifugation (5000g, 10 mins). The supernatant was again centrifuged to remove any remaining nuclei, and the polysomes were collected by centrifuging (210,000g, 60 mins) through a pad of 50% ($\frac{W}{V}$) sucrose in half strength TKM, in a Beckman Ti 50 rotor. The polysome pellet was rinsed in ice-cold TK (0.01M Tris-Cl pH 7.4, 0.015M KCl).

b) Preparation of 10S RNA (Pemberton *et al.*, 1972)

Polysomes were suspended in TK buffer by aspiration with a pasteur pipette, and a half volume of 1.0M Tris-Cl pH 9.0, followed by $\frac{1}{10}$ volume of 0.3M EDTA pH 7.6 added. This dissociated the polysomes into ribonuclear protein (RNP) particles, which were

fractionated by centrifugation through linear 10% - 40% ($\frac{W}{V}$) sucrose gradients in TK (160,000g, 16 hours) in a Beckman SW 41 rotor. The 20S RNPs were collected using an ISCO fractionator, and ethanol precipitated by the addition of $\frac{1}{10}$ volume of 1M NaCl, two volumes of RNA'ase-free ethanol and storage at -15°C for at least 4 hours. The RNP's were collected by centrifugation (20,000g, 20 minutes) and dried in vacuo. They were then dissolved in 10mM Tris Cl pH 9.0, 0.1% SDS containing 250 µg/ml of Protease K, incubated at 37°C for 3 to 5 minutes to degrade any protein, and fractionated once more on sucrose gradients in NET (0.01M NaCl, 0.01M Tris-Cl pH 7.4, 0.001M EDTA) buffer. The 10S RNA was collected and ethanol precipitated.

2) *Isolation of Total RNA from Erythroblasts*

0.5ml to 1.0ml of packed erythroblasts (about 5 x 10^9 cells) previously washed twice in ice-cold NKM, were resuspended in 100 µl to 200 µl of NKM. The cell lysis solution consisted of phenol: chloroform (1:1) equilibriated with 10% ($\frac{V}{V}$) SET 9 (2% SDS, 0.01M EDTA, 0.02M Tris-Cl pH 9.0) and heparin 1mgm/ml, and the erythroblast suspension was added drop-wise to 10 mls of this solution as it was vortex mixed.

On completion of lysis, 10 mls of SET 9 were added. The mixture was shaken vigorously for 1 minute, and centrifuged at 20,000g for 3 minutes to separate the phases. The top phase was separated by pasteur pippette and each phase re-extracted. The aqueous phases, containing erythroblast nucleic acids, were pooled and ethanol precipitated.

The ethanol precipitate from phenol extraction was dissolved in 0.5 ml SET 7.4 (2% SDS, 0.01M EDTA, 0.02M Tris-Cl pH 7.4) and centrifuged on 5% - 20% linear sucrose gradients in ½ strength SET 7.4 over a 1 ml pad of 50% sucrose. Centrifugation was at 160,000g for 4 hours in a Beckman SW41 rotor, at 17°C. After centrifugation, 0.8 ml was carefully pippetted off the top of the gradient, and the remainder of the gradient containing nucleic acids larger than 4S to 5S in size, was ethanol precipitated.

The nucleic acid precipitate containing erythroblast RNA and DNA, was dissolved in 0.75 ml - 1.0 ml of 0.001M EDTA, 0.01M Tris-Cl pH 7.4, and 3 volumes of 4.0M sodium acetate, 0.005M EDTA pH 6.0 were added (Kern, 1975). After standing at 0°C for at least 5 hours, precipitated RNA was pelleted by centrifugation (20,000g, 15 mins). The supernatant, containing most of the DNA was separated by gentle pippetting. The RNA was washed once in 2.0 mls of 3.0M sodium acetate 0.004M EDTA pH 6.0, pelleted again by centrifugation, dissolved in 0.001M, EDTA, 0.01M Tris-Cl pH 7.4, and ethanol precipitated.

The volume of 3.0M sodium acetate solution from which the RNA is precipitated is a critical feature of this procedure, in preventing significant RNA losses. The volumes mentioned here are those used for nucleic acids from 0.5 ml to 1.0 ml packed erythoblasts. If RNA was to be isolated from a smaller number of erythroblasts without significant losses occurring, then the

volume of sodium acetate solution was reduced proportionally.

The purified RNA was fractionated on 10% - 40% sucrose gradients in NET. Prior to loading, erythroblast RNA solutions were heated at 65°C for 10 minutes to ensure disaggregation of the RNA (Haines *et al.*, 1974).

3) *Sizing of nucleic acids on formamide gels* (Pinder, Staynov, and Gratzer, 1974).

Nucleic acids were electrophoresed on polyacrylamide gels containing formamide to estimate their size. The gels were 4% or 10% acrylamide, 0.02M diethyl barbituric acid pH 9.0 (adjusted with 1.0M NaOH) in formamide. These were overlaid with formamide buffered to pH 9.0 with diethyl barbituric acid, and the sample layered under this in buffered formamide containing 5% sucrose and a trace of bromophenol blue. The electrode buffer was 0.02M NaCl. Gels were electrophoresed at a constant current of 3 to 4 mA per gel, and then scanned at 280 nm, or stained in toluidine blue 0.05% in 50mM sodium acetate pH 5.5, and destained in water. For the size determination of labelled nucleic acids, the gels were sliced into 1mm slices, and dissolved in NCS scintillant (Ward, Wilson and Gillian, 1970). Radio-activity was determined in a Packard scintillation spectrometer.

2.2.3 cDNA Synthesis and Purification

Siliconized glassware was used in all experiments involving DNA.

Two hybridization techniques were used to detect specific RNA sequences. Small amounts of cDNA labelled to high levels of specific radioactivity with ^{32}P - dCTP, or ^{3}H - dCTP and/or

^{3}H - d GTP were used in RNA-excess hybridizations. In cDNA-excess hybridizations, relatively large amounts of cDNA were required to detect particular RNA sequences labelled in cell culture. For these hybridizations cDNA was prepared on a larger scale and labelled with low levels of specific radioactivity (^{32}P - dCTP 0.6 Ci/mole) to detect synthesis.

The efficiency of transcription was determined on Whatman GF/A papers by precipitating and washing samples of the reaction systems in cold 5% trichloroacetic acid, 0.5% sodium pyrophosphate.

1) Synthesis of cDNA used for RNA excess hybridization

a) Globin transcriptase cDNA

Chicken globin cDNA approaching a nucleotide length equivalent to that of globin mRNA, was synthesized using an avian myeloblastosis virus RNA-dependent DNA polymerase (transcriptase I), which was the kind gift of Dr. M. A. Chirigos and Dr. J. W. Beard. Intermediate length (300-400 nucleotides) cDNA was synthesized using an avian myeloblastosis virus RNA-dependent DNA polymerase (transcriptase II) isolated by the procedure of Kacian and Spiegelman (1974). Using saturating amounts of each enzyme, globin cDNA was synthesized in a 50 µl or 100 µl reaction volume containing 120 µM of labelled deoxynucleotide triphosphates, 0.66 mM unlabelled deoxynucleotide triphosphates, 100 mM KCl, 8 mM $MgCl_2$, 8 mM dithiothreitol 50 mM Tris-Cl pH 8.3, 100 µg/ml Actinomycin D, 20 µg/ml $(dT)_{12-18}$ as a primer, 0.5 µg globin mRNA and 2-20 units of enzyme.

Reaction mixtures were incubated at 37°C for 60 minutes.

b) Globin Polymerase I cDNA

Globin cDNA of 100 - 200 nucleotides was synthesized using E.coli DNA Polymerase I by the method of Bernard (personal communication), and similar to that described by Proudfoot and Brownlee (1974a). A 50 µl reaction volume contained 18 µM ^{3}H-dCTP, 50 µM dTTP and dATP, 100 µM dGTP, 50 mM KCl, 1 mM $MnCl_2$, 67 mM glycine buffer pH 9.2, 50 µg/ml Actinomycin D, 20 µg/ml $(dT)_{12-18}$, 0.5 - 1 µg globin mRNA, and 2 units enzyme. Incubations were carried out at 37°C for 60 minutes.

c) 18S and 28S Ribosomal cDNA

$(dT)_{12-18}$ could not be used to prime the synthesis of cDNA to non-polyadenylated RNA species. Hence random oligonucleotide sequences of salmon sperm DNA, prepared as described by Taylor *et al.*, (1976) were used to prime the synthesis of cDNAs to chicken 18S and 28S ribosomal RNAs. A 50 µl reaction volume contained 100 mM KCl, 3 mM $MgCl_2$, 8 mM dithiothreitol, 0.67 mM each of dATP, dTTP and dCTP , 120 µM of labelled dCTP, 50 mM Tris-Cl pH 8.3, 100 µg/ml Actinomycin D, 1 - 2 µg RNA, 100 µg primer DNA, and 5 units of transcriptase I. Incubations were carried out as described for the synthesis of other cDNAs.

2) Transcriptase Globin cDNA used in cDNA-excess Hybridizations

Large scale cDNA synthesis was similar to that described for the synthesis of high specific radioactivity transcriptase cDNA, except that cDNA was labelled with ^{32}P - dCTP of specific activity 0.6 Ci/mol at a concentration of 300 µM. The reaction volume of 500 µl contained 50 - 80 µg mRNA, 25 µg/ml dT_{12-18}, and saturating amounts of

transcriptase I. Incubations were at 37°C for 60 minutes.

3) Purification of cDNA

cDNA synthesis was stopped by the addition of SDS and NaOH to a final concentration of 0.1% and 0.3M respectively. After alkaline hydrolysis for 60 minutes at 37°C, the reaction mixture was neutralized and cDNA isolated by Sephadex G-50 chromatography in 0.2 M NaCl, 0.001 M EDTA, 0.01 M Tris-Cl pH 7.4. cDNA was collected by ethanol precipitation, using E.coli tRNA as carrier.

2.2.4 Nucleic Acid Hybridization

1) RNA - cDNA Hybridizations

Hybridizations were carried out in 50 µl - 100 µl of hybridization buffer (0.18 M NaCl, 0.01 M Tris-Cl, 0.001 M EDTA, 0.05% SDS pH 7.4), overlaid with paraffin oil. RNA-excess hybridization contained varying amounts of RNA, and 2000 - 10000 c.p.m. of cDNA. Globin cDNA-excess reactions contained 0.5 µg - 1.0 µg globin cDNA, and labelled RNA. Reaction mixtures containing the labelled RNA sample, without globin cDNA were also prepared to determine the background c.p.m.

Samples were heated briefly at 100°C and incubated at 60°C. Reactions were stopped by cooling rapidly to 0°C.

2) Assay of RNA - cDNA Hybrids

a) RNA-excess hybridizations

Hybridizations involving labelled cDNA were assayed using S1 nuclease (prepared and assayed as described by Scott, 1975). 40 µl - 80 µl of the hybridization mixture was diluted in 300 µl

of 0.03 M sodium acetate, 0.05 M NaCl, 0.001 M $ZnSO_4$, 5% glycerol pH 4.6), containing 12 µg of sonicated denatured carrier DNA. Two 150 µl samples were incubated with and without 4 units of S1 nuclease for 30 minutes at 45°C. 75 µg of carrier BSA was then added, followed by 1 ml of 10% TCA to precipitate undegraded DNA. After 15 minutes at 0°C, the precipitate was collected onto GF/A filters, washed with TCA, and ether, and dried. Radioactivity was determined in a scintillation spectrometer.

b) Globin cDNA-excess hybridizations

Labelled RNA sequences hybridized to globin cDNA were detected using an RNA'ase assay, similar to that described by Ross (1976). 40 µl - 80 µl of the hybridization mixture was diluted in 500 µl of 0.2 M NaCl, 0.004 M EDTA, 0.01 M Tris-Cl pH 7.4, containing 30 µg E.coli tRNA, and 5 µg DNA. 4 µg RNA'ase A and 100 units T_1 RNA'ase were added, and the solution was incubated at 37°C for 40 minutes. 1 ml of 20% TCA was added to precipitate undegraded RNA, followed by 50 µg of carrier BSA and 200 µg of 2' 3' UMP. After 15 minutes at 0°C, the precipitate was collected onto GF/A filters, and washed and counted as before.

3) DNA - DNA Re-annealing (Kemp, 1975)

Re-annealing was carried out in 10 ml screw-topped tubes containing 10 mg/ml sonicated chicken DNA and about 50,000 c.p.m./ml of cDNA or total ^3H - labelled DNA in hybridization buffer. The DNA was denatured by boiling for 5 minutes, and incubated at 60°C to the *Cot* value desired. To assay the reaction, 50 µl aliquots were

diluted into 2 ml of high salt assay buffer [0.3 M NaCl, 0.03 M sodium acetate, 0.001 M $ZnSO_4$, 5% glycerol pH 4.6], and frozen until assayed. When all *Cot* values were sampled, 1 ml aliquots were incubated with or without 150 units of S1 nuclease for 30 minutes at 37°C. All samples were made up to the same DNA concentration to standardize the effects of quenching, and TCA precipitated as before.

2.2.5 ^3H - Labelling of Total Chicken DNA (Schachat and Hogness, 1973)

Total chicken genome DNA was labelled in 25 µl reaction volumes at 10°C - 12°C for 90 minutes. The reaction mixture contained 0.05 M Na-PO_4 (pH 7.5), 0.01 M $MgCl_2$, 0.001 M EDTA, 0.1 mM each of unlabelled dATP, dGTP and dTTP, 25µCi of ^3H - dCTP (final concentration of 0.06 mM), 5 µg of chicken DNA, 5 ng DNA'ase I, and 5 µl of E.coli DNA polymerase I. Reactions were stopped by boiling for 90 seconds, and the labelled DNA separated from other ingredients by Sephadex G-50 chromatography (Section 2.2.3).

CHAPTER 3.

ISOLATION AND CHARACTERISATION OF AVIAN GLOBIN mRNA

CHAPTER 3. ISOLATION AND CHARACTERISATION OF AVIAN GLOBIN mRNA

3.1 INTRODUCTION

The ability to isolate eukaryote mRNAs has been vital to investigations into eukaryote gene organisation. It has enabled the mRNAs themselves to be readily characterised, and has led to some understanding of the manner in which sequences are arranged within a structural gene. Similarly, it has allowed basic questions concerning the arrangement of structural genes within the genome (e.g. gene reiteration frequency) to be answered.

This chapter briefly discusses relevant aspects of chicken globin mRNA isolation, and also describes the initial characterisation of this mRNA.

3.2 mRNA ISOLATION

Globin mRNA is the major species within reticulocytes and over 90% of protein synthesis within these cells is directed towards globin synthesis (Williams, 1970). Furthermore, all globin mRNAs so far isolated are about 2 - 2.2 x 10^5 Daltons in size, (Lewin, 1975b) and sediment at 9S to 10S during sucrose gradient centrifugation. On this basis, a population of 9S - 10S RNA molecules isolated from reticulocyte polysomes, is likely to contain globin mRNA as the major mRNA species.

The method described in Section 2.2 is essentially that for duck globin mRNA isolation (Pemberton *et al.*, 1972). Chicken reticuloyte polysomes were dissociated with EDTA, and centrifuged on 10% - 40% sucrose

gradients. The 20S shoulder, (Fig. 3.1) containing globin mRNA as a ribonuclear protein complex, was collected from these gradients. Protease K digestion of the 20S RNPs, yielded a discrete 10S RNA species, as shown in Fig. 3.2. The protease K digestion step was carried out, to ensure the destruction of any residual RNA'ase activity.

In our hands, a critical feature of this procedure is that the reticulocytes must be taken from the circulation on the 6th day of the phenylhydrazine injection course (i.e. after 5 daily injections). Cells collected after 6 or 7 injections failed to yield the characteristic 10S RNA species seen in Fig. 3.2. Reduced levels of 18S and 28S ribosomal RNAs, and markedly increased yields of smaller RNA species suggested that RNA isolated from these cells was partially degraded. This is probably due to higher RNA'ase levels within cells collected after 6 or 7 injections.

3.3 GLOBIN mRNA IDENTIFICATION AND CHARACTERISATION

3.3.1 Translation of 10S RNA

The translation experiments described here, were carried out by Dr. A. C. Scott (1975). They showed that globin is the sole product when reticulocyte 10S RNA is translated under optimal conditions in a wheat embryo system. The globin product, allowing for the slightly different molar concentrations of leucine in the different chains, consists of 59% α globin, and 41% β globin (Scott, 1975). The conclusion from this is that the 10S RNA species contains biologically active globin mRNA.

3.3.2 Formamide Gel Analysis of Globin mRNA

The mobility of nucleic acids during formamide gel electrophoresis

Figure 3.1

Preparation of RNP Particles from Sucrose Gradients

Polysomes were isolated from avian reticulocytes as described in Section 2.2.2, and dissociated by EDTA treatment. Centrifugation was carried out at 4°C for 16 hr. in a Beckman SW41 rotor, on 10% - 40% sucrose gradients, at 37000 r.p.m. The 20S RNP shoulder, containing 10S RNA was collected and ethanol precipitated. Centrifugation was from right to left.

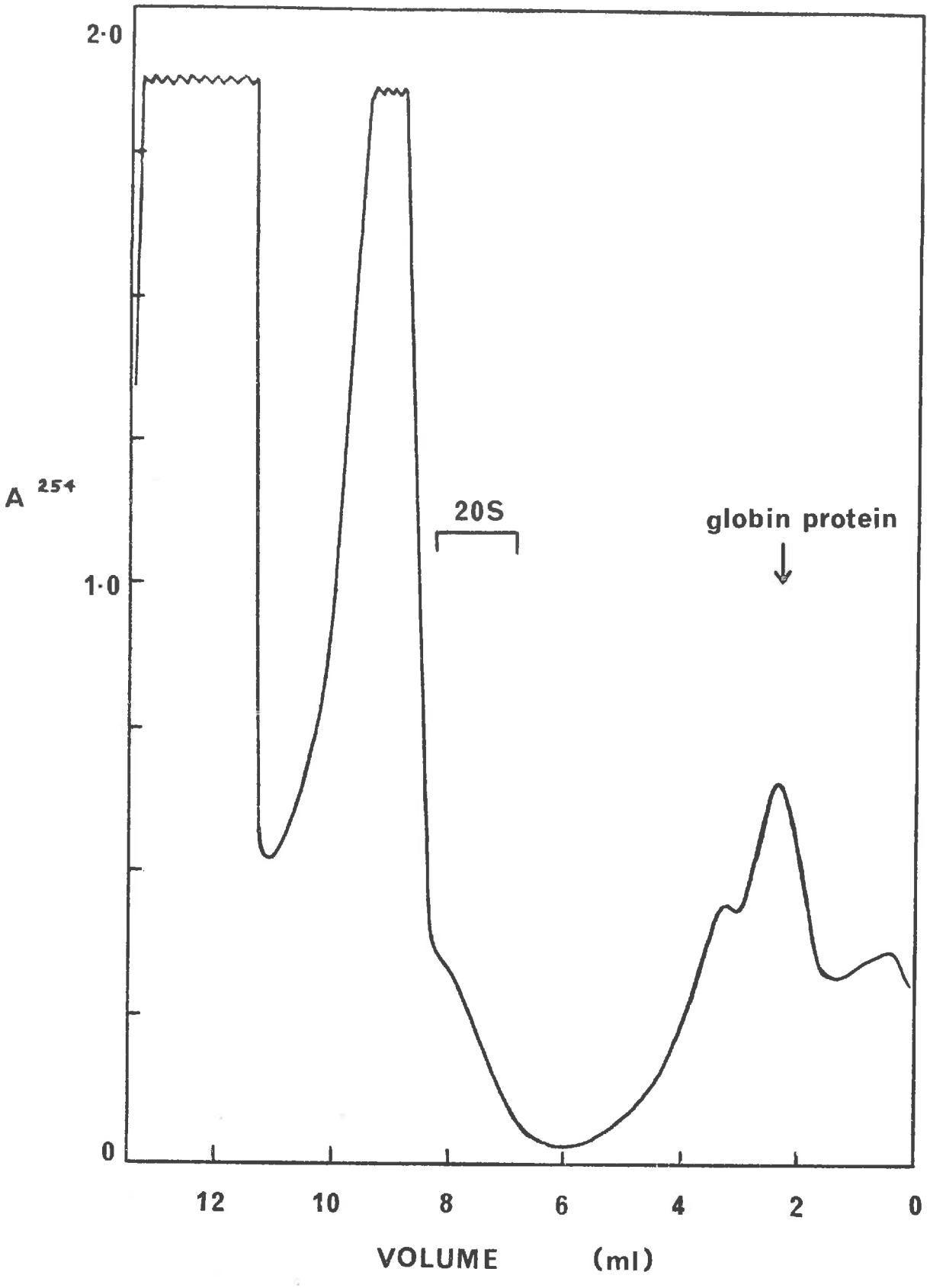
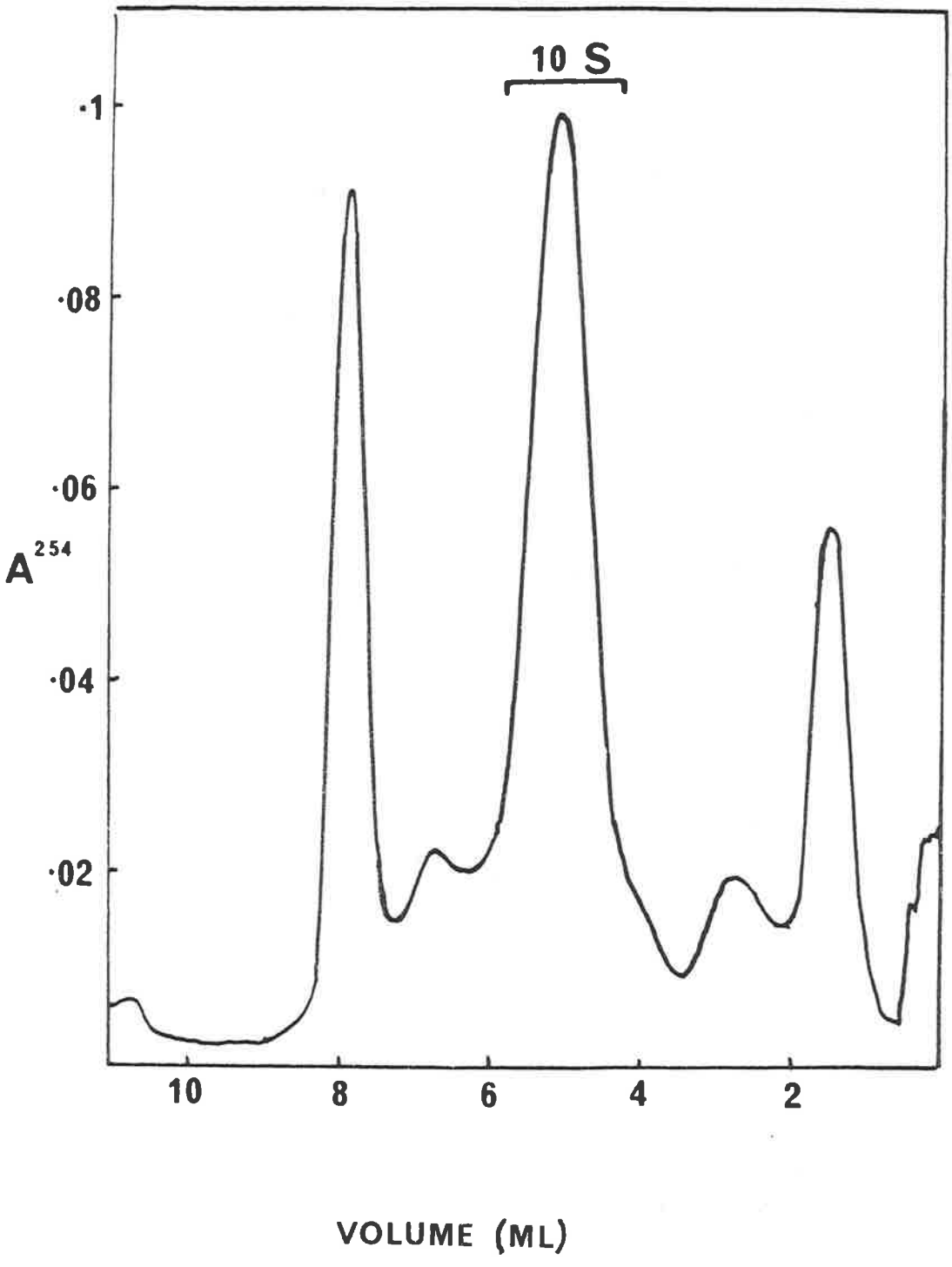

2·0
1·0
0
A 254
20S
globin protein
12
10
8
6
4
2
0
VOLUME (ml)

Figure 3.2

Isolation of 10S RNA from 20S RNPs

Ethanol precipitated 20S RNPs were taken up in 0.01M Tris pH 9.0, 0.001M EDTA, 0.1% SDS, and heated at 65°C for 10 minutes prior to centrifugation on 10% - 40% sucrose gradients. RNA sedimenting at 10S was collected as globin mRNA, and ethanol precipitated. Centrifugation was from right to left.


10 S
·1
·08
·06
A254
·04
·02
10
8
6
4
2
VOLUME (ML)

is independent of secondary structure, and is solely dependent on molecular weight (Pinder *et al.*, 1974). For this reason, formamide gel electrophoresis is an accurate method for measuring the size of chicken globin mRNA. Because it removes secondary structures within nucleic acid molecules, it is also useful in revealing hidden breaks within the mRNA.

Reticulocyte 10S RNA was electrophoresed under completely denaturing conditions in a 4% polyacrylamide gel system in formamide (Pinder *et al.*, 1974). As shown in Fig. 3.3, the mRNA was resolved into two closely migrating species, which by analogy with rabbit globin mRNA, are presumably α and β globin mRNAs (Hamlyn and Gould 1975). The failure to detect any other major RNA species (particularly smaller RNA species) on the gel, suggests that the 10S RNA is predominately intact chicken globin mRNA.

Relative to ribosomal RNA markers, the average molecular weight of the globin mRNA is 2.2×10^5 (equivalent to 650 nucleotides). This is in good agreement with previous estimates for the nucleotide length of globin mRNA (Lewin, 1975b).

3.3.3 Length of Poly(A)

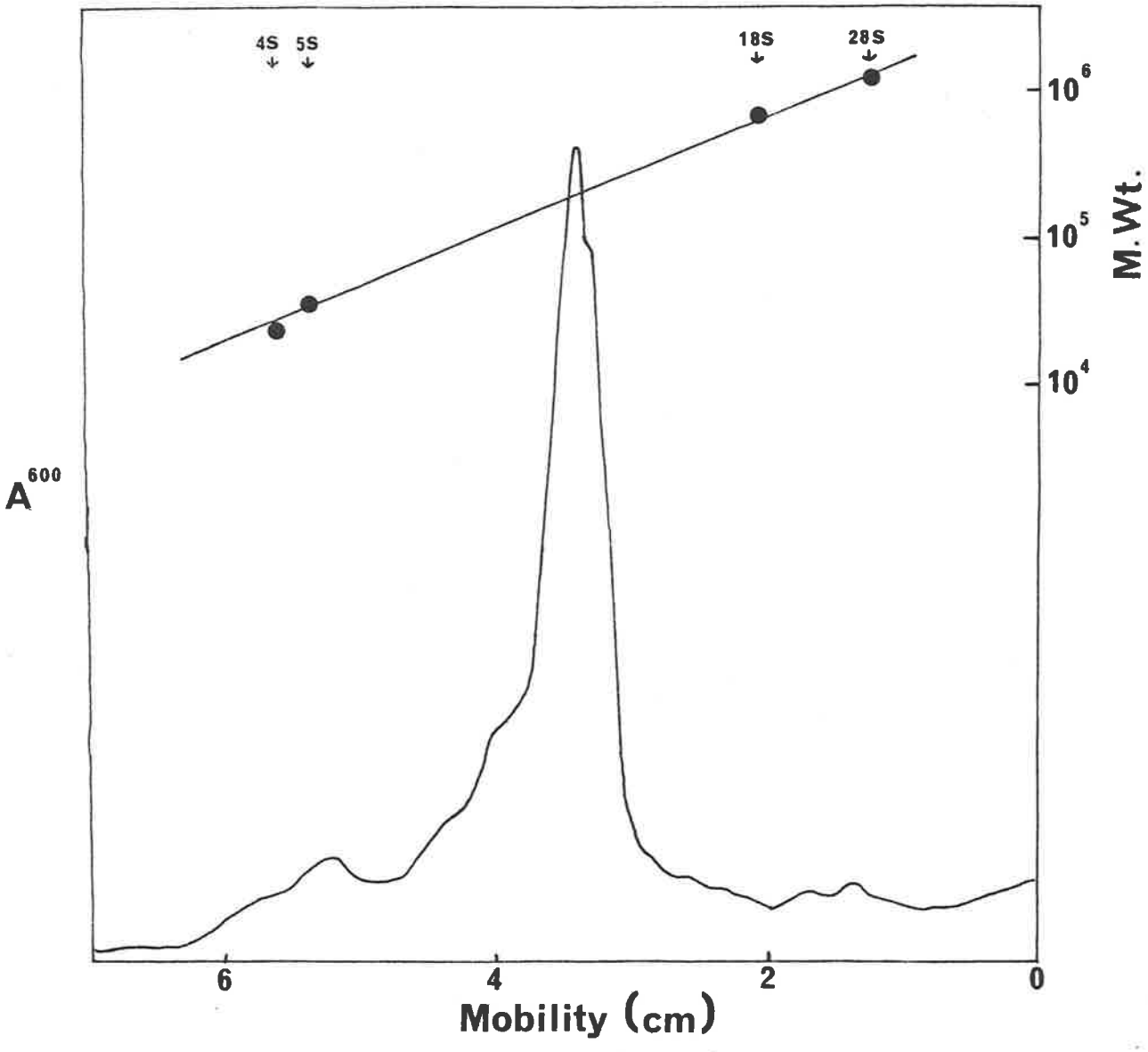
Amino acid sequence analysis has showed that α globin consists of 141 amino acid residues, and that β globin contains 146, (Croft, 1973). On this basis, approximately 435 nucleotides of globin mRNA are directly required for coding. Since globin mRNA contains 650 nucleotides, there are about 215 nucleotides whose function is unaccounted for.

All globin mRNAs characterised so far have been shown to contain

Figure 3.3

Formamide Gel Analysis of Purified Chicken Globin mRNA

10S chicken globin mRNA (5 µg/gel) was electrophoresed on 4% polyacrylamide gels in formamide, as described by Pinder *et al.*, (1974), using 28S, 18S, 5S and 4S RNA as markers. The gels were stained for 10 minutes in 0.05% toluidine blue (pH 5.5), destained overnight in water, and scanned at 600 nm using a Gilford gel scanner.


4S
5S
18S
28S
10^6
10^5
10^4
M.Wt.
A^{600}
6
4
2
0
Mobility (cm)

a 3' poly(A) region (Lewis, 1975b) so that it was likely that at least some of the extra nucleotides within chicken globin mRNA, were present as poly(A).

Experiments to determine the length of poly(A) within globin mRNA were carried out by Dr. A. C. Scott (1975). Globin mRNA was digested with RNA'ase as described by Morrison *et al.*, (1973) under conditions where 18S RNA was totally degraded, and poly(A) sequences remained intact. The products of this digestion were electrophoresed under denaturing conditions on 10% polyacrylamide gels in formamide. Relative to 5S RNA, 4S RNA and oligo$(dT)_{15}$ markers, the length of the poly(A) region was estimated to be 45 - 50 residues. This is in good agreement with the estimate of poly(A) sequence length in mouse globin mRNA (Morrison *et al.*,1973).

3.4 CONCLUSIONS

One of the significant findings of the work described in this chapter, was that chicken globin mRNA could be readily isolated from reticulocytes by the procedure described for duck globin mRNA purification (Pemberton *et al.*, 1972). The translation experiments showed that globin mRNA was the only detectable mRNA in the preparation. Further evidence provided by formamide gel analysis, indicated that the 10S RNA contains no detectable levels of RNA, other than globin mRNA. Hence, by these criteria, the 10S RNA isolated as described, was essentially purified chicken globin mRNA.

Chicken globin mRNA was shown to consist of 650 nucleotides (i.e. molecular weight of 2.2×10^5), and to contain 45 - 50 residues as poly(A).

CHAPTER 4.

PREPARATION AND CHARACTERISATION OF cDNA

The contents of this chapter have been published

Gene Specific Priming of Complementary DNA Synthesis,

Crawford, R. J. and Wells, J. R. E. (1976) *Mol. Biol. Reports, 3,* 167.

CHAPTER 4. PREPARATION AND CHARACTERISATION OF COMPLEMENTARY DNA

4.1 INTRODUCTION

Along with the development of techniques for eukaryote mRNA isolation, the discovery of enzymes possessing reverse transcriptase activity has been of major significance in eukaryote gene studies. Accordingly, complementary DNA copies of chicken globin mRNA were prepared, with the view to investigating aspects of globin gene organisation within the chicken genome. cDNA probes to chicken 18S and 28S ribosomal RNAs were also synthesized.

Both avian myeloblastosis virus RNA dependent DNA polymerase (Verma *et al.*, 1972; Kacian *et al.*, 1972) and E.coli DNA polymerase I (Proudfoot and Brownlee, 1974a, 1974b) have been used to direct the synthesis of specific cDNAs from RNA templates. In each case cDNA synthesis is absolutely dependent on the presence of a suitable primer within the synthesizing reaction mix.

4.2 CHICKEN GLOBIN cDNA SYNTHESIS

cDNA synthesis from specific eukaryote mRNA templates has generally depended on a poly(A) tract being present at the 3' end of the mRNA. This has enabled oligo (dT), which presumably base-pairs with the 3' poly(A), to act as the primer for cDNA synthesis (Verma *et al.*, 1972). As stated in Section 2.2.3, two different preparations of AMV reverse transcriptase (termed transcriptase I, and transcriptase II) and E.coli DNA polymerase I were used to synthesize chicken globin cDNA. In each case, oligo (dT) was used to prime cDNA synthesis, so that the chicken globin cDNAs each possessed a 5' oligo (dT) tail.

As shown in Table 4.1, the amount of cDNA synthesized from 1 µg of chicken globin mRNA template depended on the enzyme used. Transcriptase I was the most efficient enzyme, synthesizing 0.7 µg of cDNA (assuming 25% G), whereas the polymerase I system synthesized approximately 0.1 µg. The difference in apparent efficiencies of the enzymes could be accounted for, at least in part, by differences in the length of the cDNA products.

4.3 SIZE OF CHICKEN GLOBIN cDNAs

Both sucrose gradient centrifugation, and polyacrylamide gel electrophoresis in formamide were used to assess the size of each chicken globin cDNA.

When centrifuged on 10% - 40% sucrose gradients, the major fraction of transcriptase I cDNA sedimented as a single sharp 9S peak, whereas the products from transcriptase II and polymerase I were smaller, and more heterodisperse. Transcriptase II cDNA sedimented at about 5S to 7S, and polymerase I cDNA at about 3S to 5S (Fig. 4.1). Hence results from sucrose gradient centrifugation show that each cDNA type sediments with a different characteristic S value, and that the distribution of size classes within transcriptase II and polymerase I cDNAs is broader than in transcriptase I cDNA.

The size of each cDNA type was more accurately measured by polyacrylamide gel electrophoresis in formamide. The assumptions were made, that the 4 deoxynucleotides are represented equally within the cDNAs, and that the mobility of DNA is equivalent to that of RNA in this gel system.

Relative to ribosomal RNA markers, transcriptase I cDNA is 600 - 620 nucleotides in length (Fig. 5.3). This represents the complete length of the

Table 4.1 PROPERTIES OF CHICKEN GLOBIN cDNAs

Enzyme	µg cDNA synthesized from 1 µg mRNA	Average nucleotide length	$Cr_o t$ ½ value of chick cDNA-chick mRNA hybridization	Average nucleotide length of 5' oligo(dT) tail
Reverse Transcriptase I	0.7	620	7×10^{-4}	18 - 20
Reverse Transcriptase II	0.2	300 - 400	1.5×10^{-3}	18 - 20
E. coli DNA Polymerase I	0.1	100 - 200	5×10^{-3}	18 - 20

cDNAs were synthesized with the various enzyme preparations as described in Methods. The length of the cDNA and its oligo(dT) tail were determined from formamide gel analyses as shown in Figures 5.3 and 4.5 respectively and $Cr_o t$ ½ values from data in Figure 4.2.

Figure 4.1

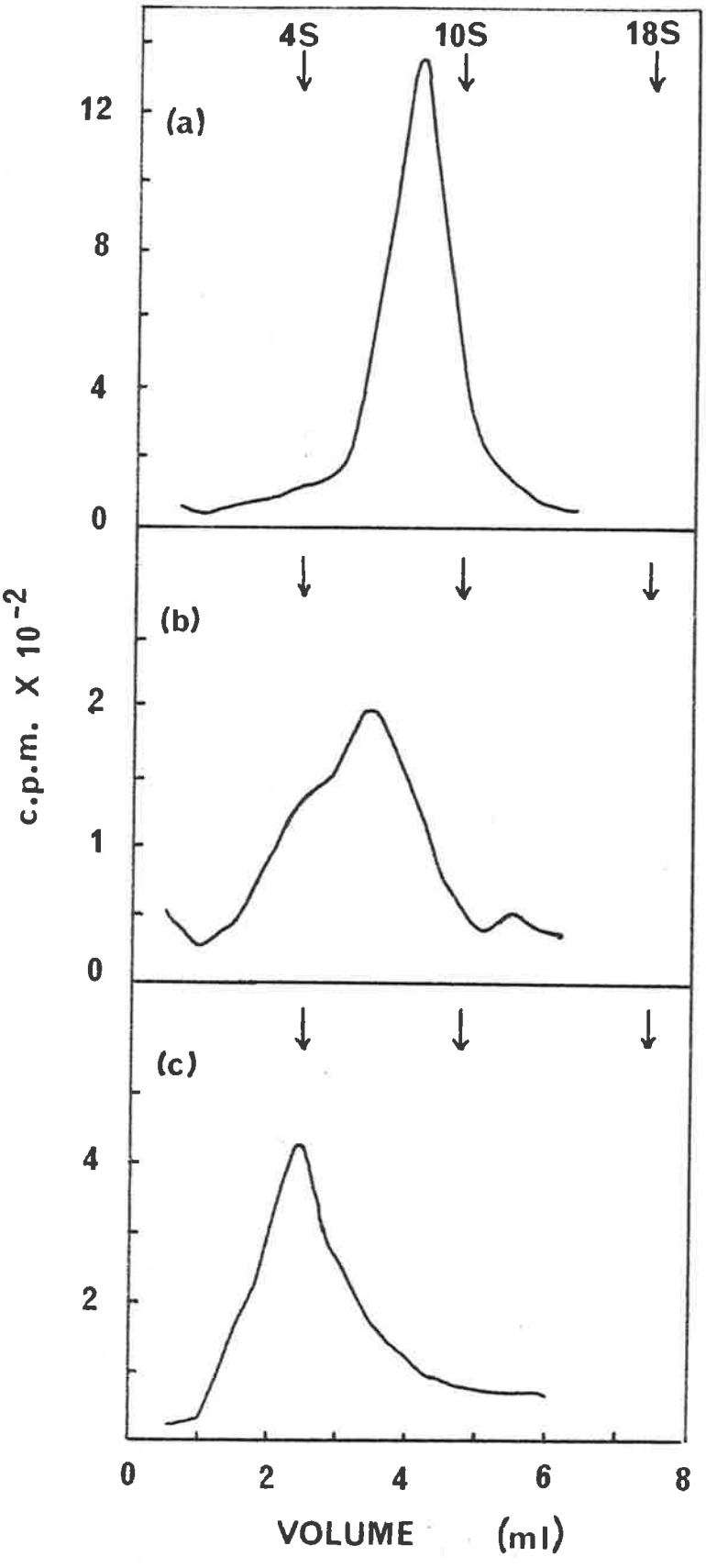
Sucrose Gradient Centrifugation of Chicken Globin cDNAs

Transcriptase I, transcriptase II and polymerase I chicken globin cDNAs, labelled with ^{3}H-dCTP (see Section 2.2.3) were centrifuged on 10% - 40% sucrose gradients for 16 hr. at 37000 r.p.m. 0.5 ml fractions were collected, and 20 µl samples counted for radioactivity. Centrifugation was in the direction left to right.

(a) transcriptase I cDNA

(b) transcriptase II cDNA

(c) polymerase I cDNA


4S
10S
18S
(a)
(b)
(c)
12
8
4
0
2
1
0
4
2
0
2
4
6
8
c.p.m. X 10^{-2}
VOLUME (ml)

RNA template when allowance is made for the poly(A) region of the mRNA, and the oligo (dT) region of cDNA (see Section 4.6.2). In comparison, polymerase I produced cDNA molecules 100 - 200 nucleotides in length, whereas transcriptase II synthesized intermediate length cDNA of 300 - 400 nucleotides.

As initially suggested, sequence data (Proudfoot and Brownlee, 1974b) and oligo (dT) primer dependence (Verma *et al.*, 1972) indicate that the synthesis of each cDNA starts at the 3' end of globin mRNA. It therefore follows that different length cDNAs result when synthesis stopped at different points along the mRNA sequence, and that chicken globin polymerase I cDNA is a copy of a small region at the 3' end of the mRNA.

4.4 HYBRIDIZATION PROPERTIES OF CHICKEN GLOBIN cDNAs AND CHICKEN GLOBIN mRNA

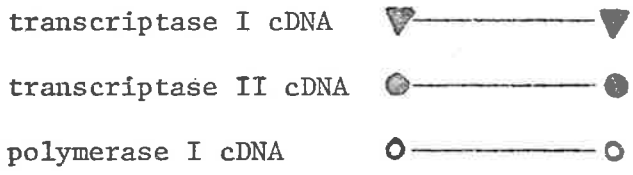
In mRNA excess reactions, chicken globin cDNAs were hybridized separately to chicken globin mRNA. Different $Cr_o t$ values, over the range 10^{-5} to 10^{-1}, were achieved by varying the mRNA concentration, and the extent of hybridization for each $Cr_o t$ value was measured using S1 nuclease (Kemp, 1975). The $Cr_o t$ ½ value (i.e. the $Cr_o t$ value at which half the tracer cDNA is hybridized with mRNA) is a comparative measure of the rate of hybridization.

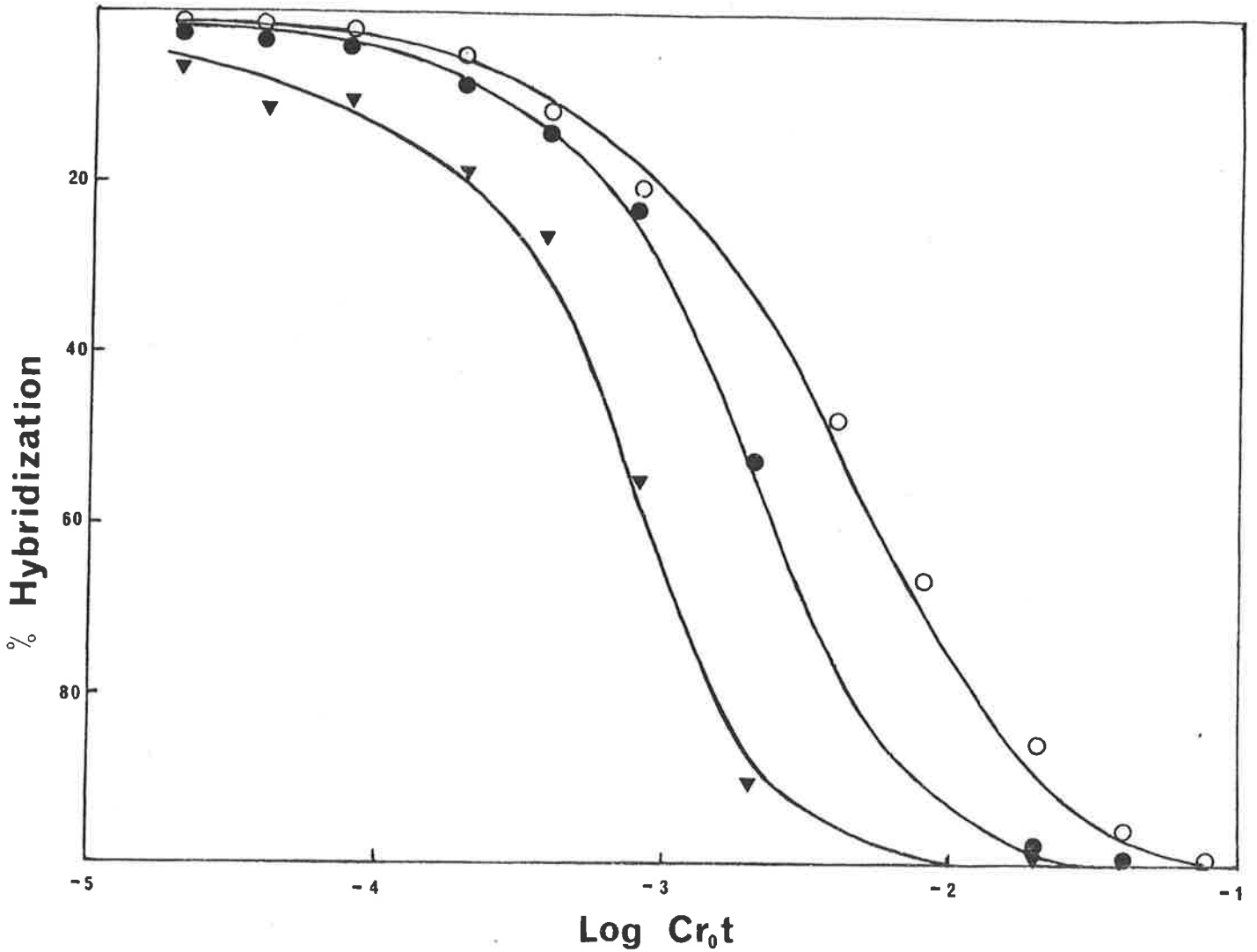
The $Cr_o t$ ½ values obtained (Fig. 4.2) reflect the length differences between the cDNAs. As previously explained, polymerase I cDNA is a copy of a small region at the 3' end of chicken globin mRNA. Consequently, at a given concentration of mRNA, this cDNA hybridizes back to globin mRNA at a slower rate than both the transcriptase cDNAs, since in these mRNA-excess hybridizations, the real Cr_o value is not represented by the concentration of

Figure 4.2

Chicken Globin mRNA/cDNA Hybridization

Various dilutions of chicken globin mRNA were each mixed with 2000 c.p.m. of different chicken globin cDNA preparations (Table 4.1) in a final volume of 50 µl. Hybridizations were at 60°C for 4 hr in siliconized microtubes in 0.18 M NaCl hybridization buffer overlaid with paraffin oil. The percentage hybridization was determined by S1 nuclease assay (see Section 2.2.4).




% Hybridization
20
40
60
80
-5
-4
-3
-2
-1
Log Cr0t

total mRNA nucleotides, but by the concentration of mRNA nucleotides taking part in the reaction. In other words, only the 100 to 200 mRNA nucleotides immediately adjacent to the 3' poly(A) region are involved in the polymerase I cDNA - mRNA hybridization reaction. Hence one would predict that the hybridization rate in these experiments to be approximately proportional to the nucleotide length of the cDNA. As shown in Fig 4.2 and Table 4.1, the experimental $Cr_o t$ ½ values demonstrate this approximate relationship. For example, the $Cr_o t$ ½ value for transcriptase I cDNA (600 - 620 nucleotides) is 7×10^{-4}, compared with 5×10^{-3} for the polymerase I cDNA (100 - 200 nucleotides in length).

These hybridization experiments also show that both the transcriptase cDNAs, and polymerase I cDNAs can be protected completely from S1 nuclease by globin mRNA, indicating that each cDNA is a faithful transcript of the template mRNA. This serves as one important control for experiments discussed in Chapter 5, in which chicken globin cDNA molecules are hybridized to rabbit globin mRNA.

4.5 POLYMERASE I cDNA PRIMING OF FURTHER cDNA SYNTHESIS

It seemed possible that since the short polymerase I cDNA was complementary to sequences at the 3' end of globin mRNA (Proudfoot and Brownlee, 1974a, 1974b), then it might prime further cDNA synthesis in a transcriptase cDNA synthesizing system, towards the 5' end of the template. To test for polymerase I cDNA priming, ^{3}H-dC + ^{3}H-dG labelled polymerase I globin cDNA was prepared as before, and separated from residual oligo (dT) primer by two cycles of sucrose gradient centrifugation. The purified polymerase I cDNA would not prime dT incorporation on a poly(A) template, so that no detectable levels of free oligo (dT) primer remained in the

polymerase I cDNA preparation.

Extended cDNA synthesis involved a preliminary hybridization with globin mRNA, followed by cDNA synthesis using transcriptase I enzyme and unlabelled deoxynucleotide triphosphates, as described in Methods. Polymerase I cDNA - primed - cDNA synthesis was assayed by detecting an S value shift in the labelled polymerase I cDNA primer, from 3S - 5S to 9S, as shown in Fig. 4.3.

In this experiment it is clear that polymerase I cDNA has altered in S value from 3S - 5S to 9S, indicating that polymerase I cDNA has primed further cDNA synthesis with considerable efficiency. Although there was some variability in the extent of priming, in some experiments over 90% of the radioactivity in polymerase I cDNA was converted to 9S cDNA.

Additional evidence for the successful extension of polymerase I cDNA to a size equivalent to a full length of the messenger template, was obtained by hybridization of the extended cDNA product with globin mPNA. As previously discussed, transcriptase I cDNA hybridizes to chicken globin mRNA with a $Cr_o t$ ½ value of 7×10^{-4}, whereas unextended polymerase I cDNA does so with a $Cr_o t$ ½ of 5×10^{-3}. If the cDNA synthesized by extending labelled polymerase I cDNA with unlabelled deoxynucleotides, represents a full length copy of the mRNA, then it should hybridize to globin mRNA with a $Cr_o t$ ½ of 7×10^{-4}.

Accordingly 9S extended cDNA, labelled only in the polymerase I cDNA region was selected from the sucrose gradient shown in Fig. 4.3, and hybridized to chicken globin mRNA. As shown in Fig. 4.4, the $Cr_o t$ ½ value for this hybridization was 7×10^{-4}, and the curve was indistinguishable from that of oligo (dT)-primed transcriptase I cDNA.

These results indicate that transcriptase I cDNA, primed specifically

Figure 4.3

Sucrose Gradient Analysis of Polymerase I cDNA Primed Chicken Globin cDNA

Chicken globin cDNAs were centrifuged separately in 5% - 20% linear sucrose gradients at 4°C for 16 hr at 37000 r.p.m. 0.4 ml fractions were collected, and 20 µl samples counted for radioactivity. The positions of 4S tRNA and 10S globin mRNA are shown. Fractions as indicated were pooled and used for the hybridizations shown in Fig. 4.4.

(a) polymerase I cDNA

(b) oligo(dT) primed transcriptase I cDNA

(c) cDNA synthesized by extending labelled polymerase I cDNA as primer with unlabelled deoxynucleotides, catalysed by transcriptase I.

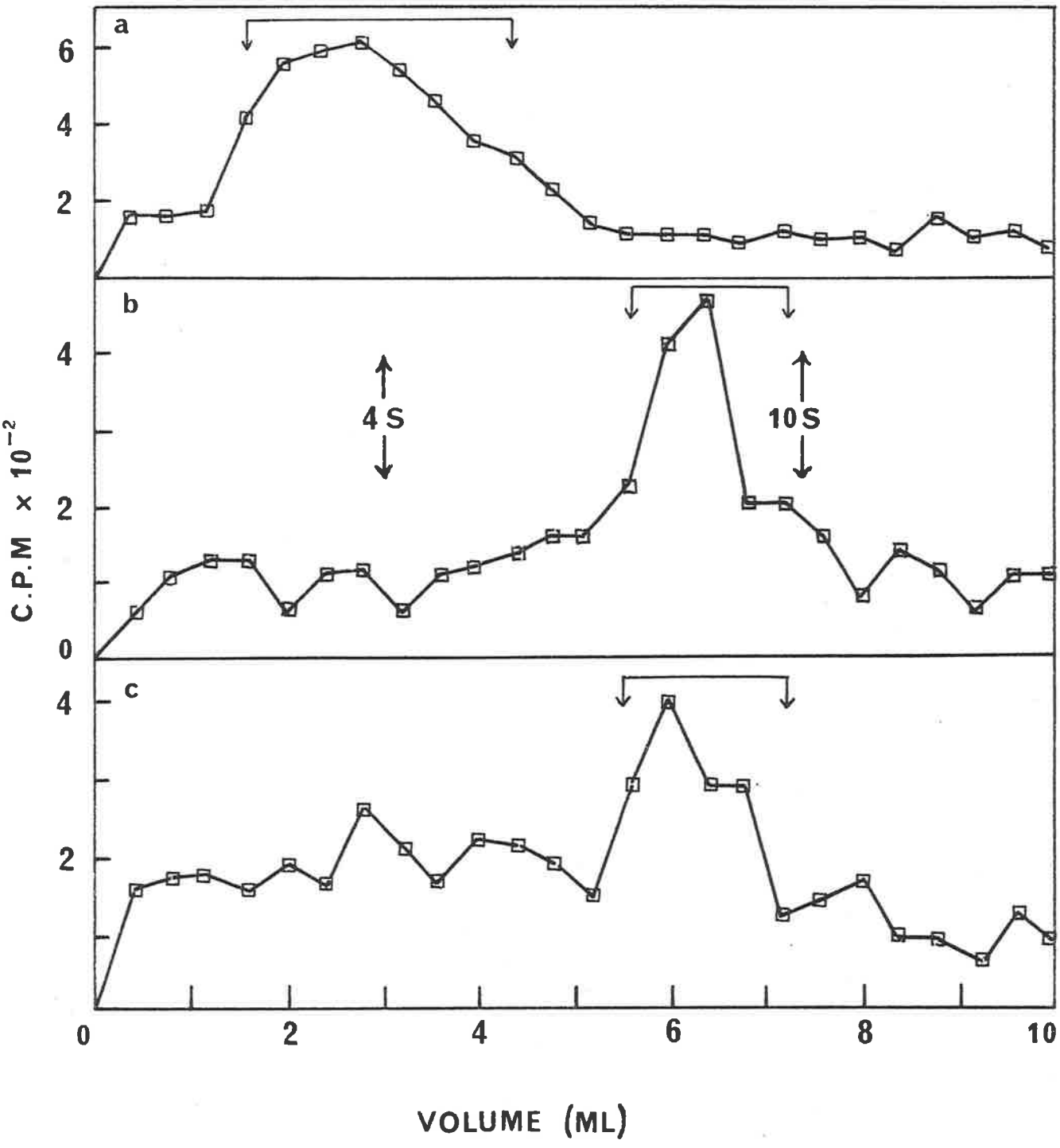

a
b
c
4 S
10 S
C.P.M x 10^{-2}
VOLUME (ML)
0
2
4
6
8
10

Figure 4.4

Chicken Globin mRNA/cDNA Hybridization

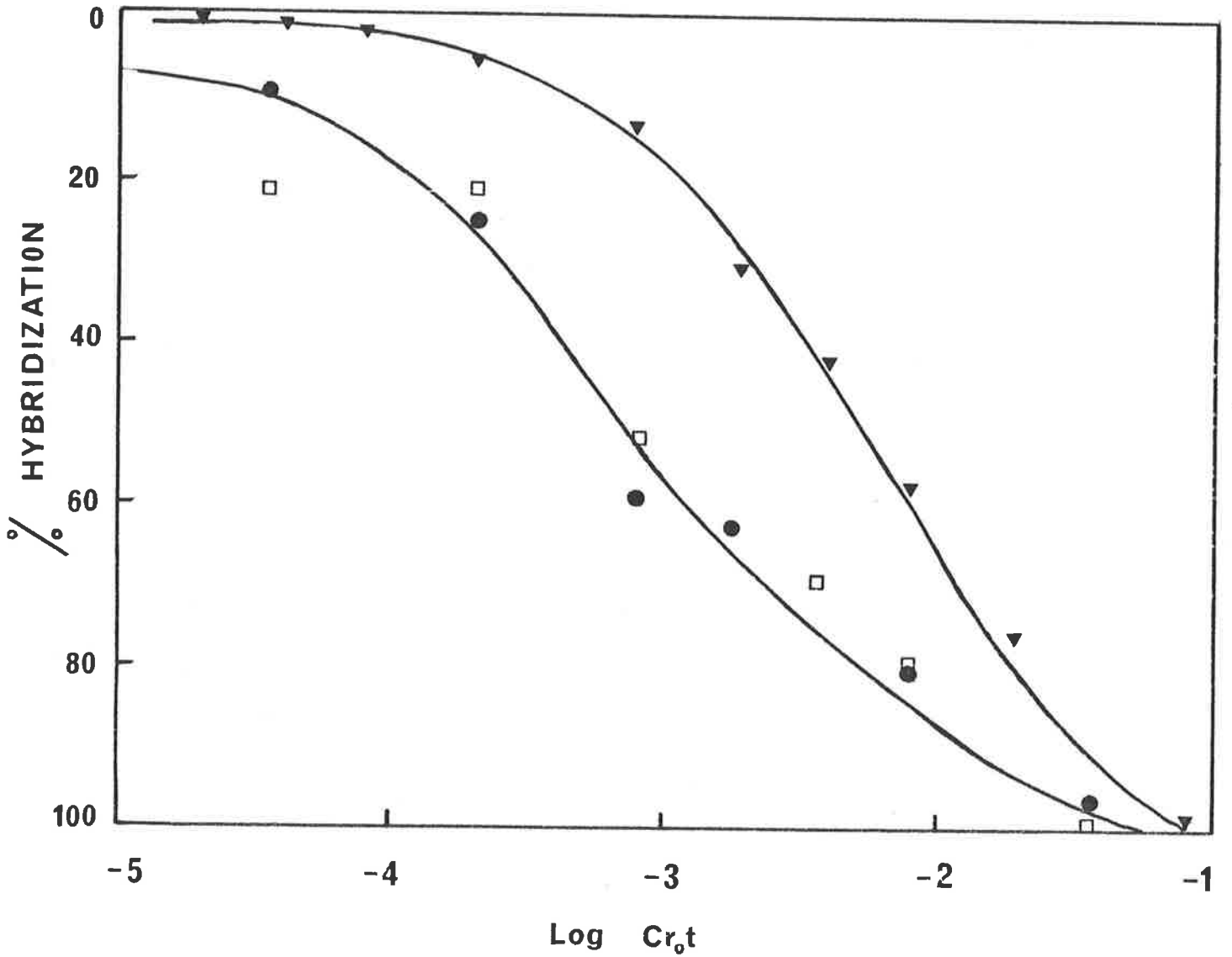
Chicken globin cDNAs, labelled with ^{3}H-dCTP were hybridized separately to various dilutions of chicken globin mRNA in a final volume of 50 µl as described in Fig. 4.2. The percentage hybridization was determined by S1 nuclease assay.

Transcriptase I cDNA oligo(dT) primed ▫———▫

transcriptase I cDNA synthesized by extending labelled

polymerase I cDNA with unlabelled deoxynucleotides ●———●

polymerase I cDNA ▼———▼


0
20
40
60
80
100
% HYBRIDIZATION
-5
-4
-3
-2
-1
Log Cr_0t

by polymerase I cDNA, is identical in both size and hybridization properties to oligo (dT) primed transcriptase I cDNA.

Hence it is apparent that cDNA synthesized by polymerase I will prime further cDNA synthesis in a transcriptase I cDNA synthesizing system.

4.6 CHARACTERISATION OF 5' END OF CHICKEN GLOBIN cDNAs

4.6.1 Sequence adjacent to Oligo (dT)

Since each cDNA synthesis was dependent on the presence of oligo (dT) it seemed likely that synthesis was primed by oligo (dT) base-pairing with the poly(A) at the 3' end of mRNA, and that each cDNA contained a 5' - oligo (dT) tail.

Two sets of transcription experiments were carried out with the aim of determining which deoxynucleotides were incorporated into the cDNA immediately adjacent to oligo (dT). In each case, the level of deoxynucleotide incorporation was measured by acid precipitation, as described in Methods. The results are summarized in Table 4.2.

In the first set (first six entries of Table 4.2), all incubations contained dTTP in addition to oligo(dT), mRNA and either one or two of the other dNTPs. Only one dNTP was labelled in each incubation. In all experiments in which dGTP was omitted, negligible incorporation of other dNTPs was seen (Table 4.2). With ^{3}H-dGTP as the only added dNTP in this system 0.7 p.mol. of dGTP per p.mol. of mRNA was incorporated by transcriptase I, whereas 0.3 p.mol. per p.mol. of mRNA was incorporated by polymerase I. When the relative lengths of these two cDNA preparations are considered, the level

Table 4.2 DETERMINATION OF DEOXYNUCLEOTIDES ADJACENT TO OLIGO (dT) IN CHICKEN GLOBIN cDNA

dNTPs present in incubation	Reverse transcriptase I p.mol. ^{3}H dNTP incorporated per p.mol. mRNA	E. coli DNA polymerase I p.mol. ^{3}H dNTP incorporated per p.mol. mRNA
T + ^{3}H-G	0.7	0.2
T + C + ^{3}H-G	0.6	0.3
T + A + ^{3}H-G	0.5	-
T + G + ^{3}H-C	0.5	0.2
T + A + ^{3}H-C	0.05	-
T + ^{3}H-C	0.03	0.03
^{3}H-G	0.14	-
^{3}H-C + G	0.07	-
^{3}H-C	0	-
No mRNA	0	0
Normal cDNA synthesis with all dNTPs present	0.7 p.mol. cDNA per p.mol. mRNA	0.2 - 0.3 p.mol. cDNA per p.mol. mRNA

of dGTP incorporation in these experiments is approximately equivalent to the level of initiation on complete cDNA synthesis. Hence we conclude that deoxyguanosine is incorporated into chicken globin cDNA immediately adjacent to oligo(dT).

In the second series of incorporation experiments, dTTP was omitted from reaction mixtures (Table 4.2, entries 7 to 10). With dGTP as the only dNTP present, incorporation by transcriptase I was lower than in the presence of dTTP. This is expected if oligo(dT) is not absolutely aligned with the 3' poly(A) tract in mRNA. Nevertheless the incorporation of dGTP was substantial, (0.14 p.moles/p.mole mRNA) and independent of dATP, dTTP or dCTP. However, dCTP incorporation was negligible when present as the only dNTP, but became significant in the presence of (unlabelled) dGTP. This indicates that deoxycytosine is the second nucleotide to be incorporated into chicken globin cDNA. This information can be used to deduce that the third dNTP to be incorporated in cDNA synthesis is not deoxyguanosine or deoxycytosine by referring back to the first set of experiments in Table 4.2. Here it is seen that deoxyguanosine incorporation is not increased in the presence of dCTP, neither is deoxycytosine incorporation increased in the presence of dGTP. Thus the 5' deoxynucleotide sequence in chicken globin cDNA molecules is 5' oligo(dT) G.C. (A or T).

4.6.2 Length of Oligo(dT) Tail

To utilize the heterologous hybridization data described in Chapter 5, it was necessary to estimate the length of the oligo(dT) tail on the cDNA preparations. α-^{32}P-labelled dGTP (a gift of

Dr. R. H. Symons) was used in incubations similar to those described in Table 4.2, and the incorporated ^{32}P product was isolated as material excluded from Sephadex G-50. The size of the ^{32}P containing molecule was measured on 10% polyacrylamide gels in formamide, using appropriate markers. As Fig. 4.5 shows, when both transcriptase I and polymerase I were used, α-^{32}P-labelled dGTP was incorporated into an oligo(dT) molecule that was 18 - 20 nucleotides in average length. The presence of either dCTP or dATP in the incubation did not detectably alter the size of this molecule.

We conclude that chicken globin cDNAs contain a 5' oligo(dT) tail 18 - 20 residues in length. This estimate is in agreement with the estimate for the length of the 5' oligo(dT) tail in light chain immunoglobulin cDNA (Rabbits and Milstein, 1975).

4.7 REITERATION FREQUENCY OF GENES CODING FOR GLOBIN WITHIN THE CHICKEN GENOME

Globin synthesis accounts for 90% of the protein synthesis taking place within avian erythroid cells (Williams, 1970). Globin gene reiteration is one possible way by which these cells might achieve such high levels of globin synthesis. This hypothesis predicts that there are many globin genes within the genome, and concurrent transcription of these genes provides the cell with sufficient globin mRNA to maintain high levels of globin protein synthesis. To test this hypothesis, the reiteration frequency of globin genes within the chicken erythrocyte genome was determined.

cDNA hybridizes only to complementary sequences, with a high degree of specificity. It can also be prepared so that it contains high levels of

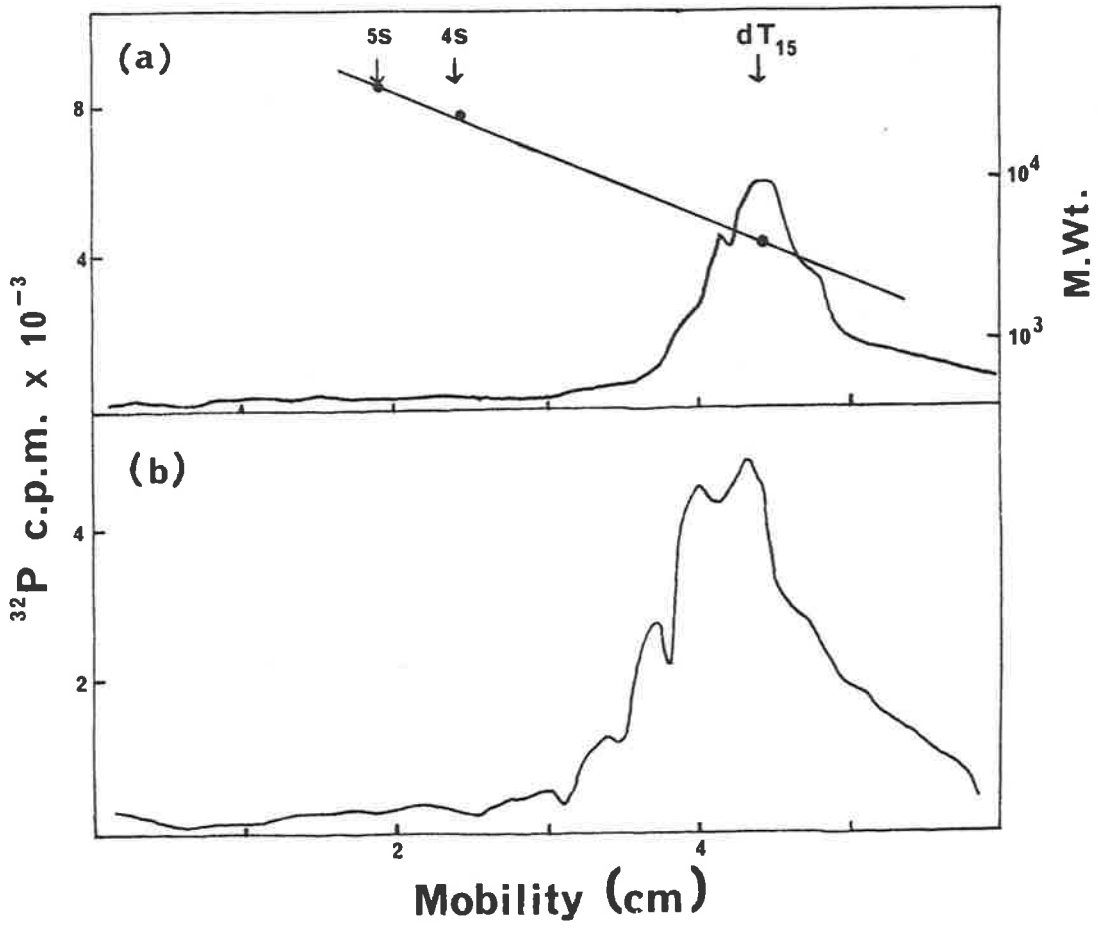
Figure 4.5

Size Determination of Oligo(dT) Tail on Chicken Globin cDNA

$(dT)_{12-18}$ was incubated with dTTP and ^{32}P-dGTP as the only dNTPs in cDNA synthesizing systems for reverse transcriptase and polymerase I as described in Section 2.2.3. ^{32}P-containing fragments excluded from Sephadex G-50 (column 1.2 x 20 cm) were electrophoresed on 10% polyacrylamide gels in formamide. Gels were cut into 1 mm slices counted for ^{32}P radioactivity.

(a) transcriptase I oligo(dT) tail

(b) polymerase I oligo(dT) tail


(a)
5S
4S
dT15
8
4
10^4
10^3
M.Wt.
(b)
4
2
2
4
32P c.p.m. x 10^-3
Mobility (cm)

specific radioactivity. These two particular characteristics make cDNA an extremely sensitive probe for detecting particular gene sequences within the genome.

The rate at which cDNA anneals with complementary sequences is directly dependent on the concentration of those sequences (Britten and Kohne, 1968). Hence by comparing the rate of re-annealing of cDNA and complete genome DNA, with a kinetic standard, the reiteration frequency of a particular gene sequence within the genome can be estimated. This generally was the approach used to determine the reiteration frequency of globin gene sequences within the chicken erythrocyte genome.

To estimate the rate of re-annealing of unique sequences within the chicken erythrocyte genome, total erythrocyte DNA was labelled to high levels of specific radioactivity (See section 2.2.5), and sheared to an approximate size of 500 nucleotides. The labelled, sheared DNA was then re-annealed to a vast excess of unlabelled erythrocyte DNA, giving the re-association curve shown in Fig. 4.6. Unique sequences, constituting the major portion of the DNA, reannealed with a $C_o t$ ½ of 1.7×10^3. There is an indication here also, that some sequences reannealed in a broad transition at lower $C_o t$ values, suggesting that about 20% of the DNA sequences are repeated within the genome. This is in agreement with another estimate of the proportion of reiterated sequences in the chicken genome (Rosen *et al.*, 1973).

In comparison, chicken globin cDNA reannealed in a sharp transition with a $C_o t$ ½ of 1.2×10^3. Hence the rate of reannealing of globin gene sequences is up to twice as fast as that for unique sequences, indicating that there are only a few copies (2 - 5) of the globin gene in

Figure 4.6

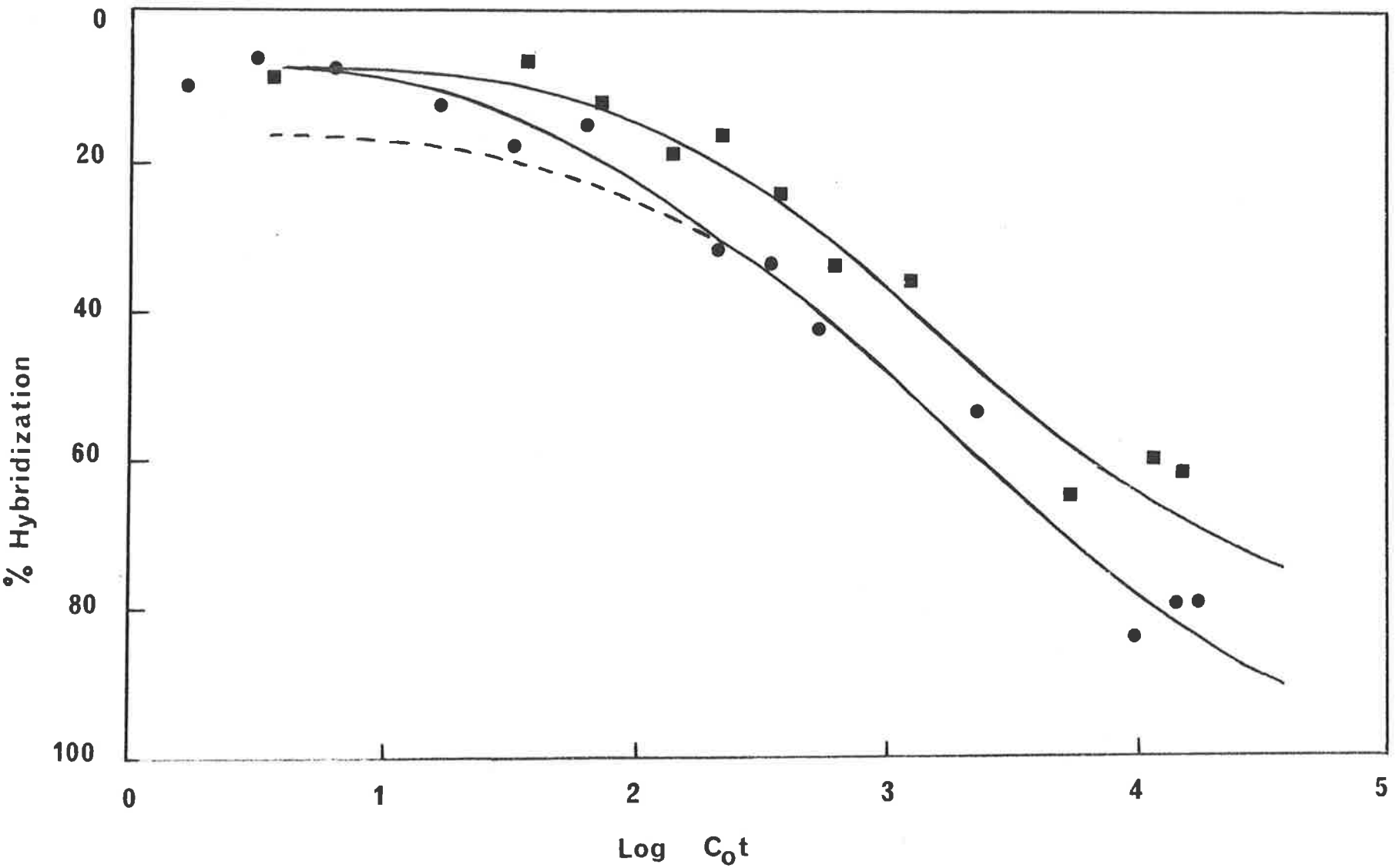
Reassociation of Chicken Globin cDNA with Excess Chicken Erythrocyte DNA

Reannealing was carried out as described in Section 2.2.4, and the percentage hyrbidization determined by S1 nuclease assay in high salt assay buffer.

^{3}H-chicken DNA annealed to chicken DNA ●————●

^{3}H-chicken globin cDNA annealed to chicken DNA ■————■

---------- denotes a possible course for the unique sequence reannealing curve.


0
20
40
60
80
100
% Hybridization
0
1
2
3
4
5
Log $C_o t$

the chicken genome.

This $C_o t$ analysis has therefore showed that high levels of globin protein synthesis are not achieved in chicken erythroid cells, by globin gene reiteration.

4.8 SYNTHESIS OF cDNA TO CHICKEN RIBOSOMAL 18S AND 28S RNA

As previously stated, the synthesis of cDNA is primer dependent. Oligo(dT) is unsuitable as a primer for ribosomal cDNA synthesis, since chicken 18S and 28S ribosomal RNAs do not contain significant poly(A) regions (R. Richards, personal communication). However, Taylor *et al.*, (1976) have shown that cDNA synthesis can generally be primed by random sequences of oligo deoxynucleotides. Presumably by chance there are sequences within the randomly cleaved DNA that are complementary to small sequences within the chicken ribosomal RNAs.

Chicken ribosomal cDNA was prepared as described in Section 2.2.3. Using transcriptase I, about 0.25 µg of ribosomal cDNA was transcribed from 1 µg of input ribosomal RNA.

The chicken ribosomal cDNAs were not characterised. However, experiments described in Chapter 6 indicate that each cDNA can successfully be used as a probe for 18S and 28S ribosomal RNA sequences.

4.9 DISCUSSION

The work described in this chapter has principally been concerned with the characterisation of chicken globin cDNAs. In particular it has showed that cDNAs of three different lengths can be synthesized by using three separate enzyme preparations. Two of the cDNAs were prepared using

two separate preparations of AMV reverse transcriptase. The reason that transcriptase II generally synthesized shorter length cDNA than transcriptase I is not known. Efstradiatis *et al.*, (1976) have demonstrated that below a thresh-hold deoxynucleotide concentration of about 120 μM, significant levels of shorter length globin cDNAs are synthesized. However this is not the reason for the length difference of transcriptase I and transcriptase II cDNAs, since the deoxynucleotide concentrations were identical for both sets of incubations, and was above 120 μM.

Neither is it due to different amounts of enzyme, since saturating amounts of enzyme were used in each case. Rather, the two different size classes of cDNA appear to be due to a fortuitous difference in the properties of the two enzyme preparations.

The reverse transcriptase activity of polymerase I was first described in 1974 (Proudfoot and Brownlee, 1974), and later used in experiments to sequence the 3' end of rabbit globin mRNA (Proudfoot, 1976). Proudfoot (1976) showed that the 5' exonuclease activity of "uncleaved" polymerase I competes to some extent with its polymerase properties. However, removal of the 5' exonuclease activity with subtilisin reduced the degradation of cDNA apparent when "uncleaved" enzyme was used. "Uncleaved" enzyme was used here to synthesize polymerase I cDNA, and good yields were still obtained. In these preparations, the reaction may have been stopped while the equilibrium of the reaction was still in favour of synthesis.

The properties of the cDNAs were compared in hybridization experiments involving chicken globin mRNA. The respective hybridization rates can be explained in terms of the length of each cDNA. These experiments illustrate clearly that it is important when determining the

complexity of a mRNA preparation by comparison with a kinetic standard, that the length of the cDNAs used in each hybridization is considered. For example, adult rabbit globin contains one α globin and one β globin chain, so that the total molecular weight of non identical sequences (i.e. complexity) of purified rabbit globin mRNA is 4×10^5 (twice the molecular weight of rabbit globin mRNA). The $Cr_o t$ ½ value for rabbit globin mRNA - rabbit globin transcriptase I cDNA hybridization is 5×10^{-4} (Kemp, 1975), and is a measure of this complexity. It can be used as a kinetic standard to determine the complexity of other nucleic acid preparations.

Since the $Cr_o t$ ½ value for chicken globin mRNA - chicken globin transcriptase I cDNA hybridization is 7×10^{-4}, then the complexity of chicken globin mRNA is 5.6×10^5. Hence there are equivalent to 2.8 non identical globin mRNA species within the preparation. This is in good agreement with the composition of adult chicken haemoglobin, which contains one β globin, and two related but non-identical α globins (Moss and Thompson, 1969).

One of the significant findings of the work described in this chapter, is that polymerase I cDNA will prime further cDNA synthesis using transcriptase I. Generally the priming ability of polymerase I cDNA approaches that of oligo(dT), and in some experiments 90% of the radioactivity in polymerase I cDNA was recovered in 9S cDNA.

The priming ability of polymerase I cDNA is useful for several reasons. Firstly it provides additional proof that polymerase I cDNA represents a short complementary copy of a small region of nucleotides immediately adjacent to the 3' poly(A). However it might be of primary

importance in potentially enabling gene specific priming of cDNA synthesis. For example, using globin polymerase I cDNA, it should be possible to specifically synthesize globin cDNA from an RNA preparation containing many other RNA species. Using this approach it may be possible to detect, and synthesize cDNA to possible higher molecular weight globin precursor mRNA. This technique should be sufficiently sensitive to detect a globin mRNA precursor that is only 10% larger than polysomal globin mRNA.

The 5' terminal region of chicken globin cDNA was characterized in detail. The cDNAs each contain a 5' oligo(dT) tail, about 18 - 20 residues in length.

These experiments also show that, by complementarity, the 3' sequence of chicken globin mRNA is poly(A) C.G. (A or T).... This sequence is identical to the corresponding region of rabbit α and β globin mRNAs, and by analogy with rabbit globin mRNA forms a small section of the 3' untranslated region (Proudfoot and Brownlee, 1974b). Its function is unknown, but is probably well defined. Although poly(A) polymerase from Hela cells does not seem to recognise a specific 3' sequence *in vitro* (Brown and Sugimoto, 1973), it is possible that this small 3' sequence on globin mRNAs is part of a recognition site for such an enzyme *in vivo*.

Evidence is provided here that the globin gene is not repeated to any significant extent, within the chicken erythroid genome. Hence high levels of globin synthesis in these cells is not a result of gene reiteration or amplification. This is consistent with other estimates for the number of globin gene repeats in other eukaryote genomes (Lewin, 1975b).

CHAPTER 5.

ORGANISATION OF SEQUENCES IN AVIAN GLOBIN mRNA

The contents of this chapter have been published

Organization of Sequences in Avian Globin mRNA,

Crawford, R. J., Scott, A. C. and Wells, J. R. E.

(1977) *Eur. J. Biochem.*, *72*, 291.

CHAPTER 5. ORGANIZATION OF SEQUENCES IN AVIAN GLOBIN mRNA

5.1 INTRODUCTION

Molecular weight determination of all specific eukaryote mRNAs so far isolated indicate that they contain many more nucleotides than required for coding (Lewin, 1975b). Poly(A) sequences are added to the 3' end of most eukaryote mRNAs after transcription, and similarly 7 - methyl guanosine is added post-transcriptionally to the 5' end. However, these sequences together do not account for all the untranslated nucleotides.

In particular chicken globin mRNA has approximately 215 extra nucleotides (see Section 3.3.2) : 45 - 50 of these can be attributed to the 3' poly(A) sequence, so that the remaining 165 - 170 untranslated nucleotides, are probably part of the unit transcribed from the structural gene. This chapter is particularly concerned with the location of untranslated sequences within chicken globin mRNA molecules. It describes a novel approach, involving hybridization of different length chicken globin cDNAs to rabbit globin mRNA. The results can be used to give a minimal estimate of the length of the untranslated region at the 3' end of chicken globin mRNA.

5.2 RATIONALE FOR THE DETERMINATION OF SEQUENCE ORGANISATION IN CHICKEN GLOBIN mRNA

For convenience, globin mRNA and globin cDNA are referred to as single entities. However, it is appreciated that this means three types of sequences for chicken globin (α_1, α_2, and β) (Moss and Thompson, 1969), and two for rabbit (α and β) respectively. These results are therefore the average for α and β chains.

This investigation was prompted by three pieces of information. First, globin mRNA sequence data available at the time of the investigation indicated that there is a substantial region of at least 52 untranslated nucleotides adjacent to the 3' poly(A) tract of rabbit globin mRNA (Proudfoot and Brownlee, 1974b). [More recently additional sequence data has been described, and the complete length of the 3' untranslated regions of rabbit α and β globin mRNAs determined (Proudfoot, 1976).] Secondly, amino acid sequence homology between rabbit and chicken globins suggested the possibility of significant nucleotide homology in the coding regions of these molecules, and further that the degree of homology might be linear in this region. Thirdly, as described in Section 4.3, defined lengths of chicken globin cDNA could be prepared from chicken globin mRNA.

Based on this information, this work tests the hypothesis that an untranslated region of nucleotides adjacent to the 3' poly(A) tract of chicken and rabbit globin mRNA had a different pattern of homology to that of the translated regions of the respective globin mRNAs. Experimentally this hypothesis was tested by hybridizing relatively short molecules of chicken cDNA (containing a high percentage of transcripts from non-coding nucleotides) and in a second set of experiments hybridizing long molecules of chicken cDNA (higher percentage of coding transcripts) to an excess of rabbit globin mRNA.

5.3 CHARACTERISATION OF RABBIT GLOBIN mRNA HYBRIDIZATION TO CHICKEN GLOBIN cDNA

Preliminary characterisation of chicken globin cDNA hybridization with rabbit globin mRNA was necessary to establish the $Cr_o t$ value at which

hybridizations reach completion. These preliminary heterologous hybridizations were carried out in a manner similar to that previously described for chicken mRNA excess experiments, (see Section 2.2.4), and showed that no further heterologous hybridization occurred beyond a $Cr_o t$ value of 1. The percentage hybridization at completion gives a minimum estimate for the degree of homology between rabbit and chicken globin mRNA sequences. However, it does not indicate the absolute homology between these mRNAs, since formamide gel analysis of the S1 nuclease products showed that sequences of less than 10 - 15 nucleotides, do not have sufficient stability to maintain their double-stranded form under these S1 assay conditions. Hence hybrid sequences of less than 10 to 15 nucleotides are digested by S1 nuclease, and do not therefore contribute positively to this estimate of homology between rabbit and chicken globin mRNAs.

Furthermore it is important to remember that the deoxynucleotide used to label the chicken globin cDNA, is deoxycytidine. Therefore double-stranded regions of the rabbit mRNA - chicken cDNA hybrid must contain deoxcytidine within the particular region of cDNA, if it is to be scored as as area of complementarity by this assay. Areas of complementarity not containing dC within the immediate cDNA region, are not detected as regions of complementarity. In retrospect, it may have been better to have included labelled dATP and dGTP in the reaction mixes for cDNA synthesis where the cDNA was to be used for heterologous hybridizations.

5.4 HETEROLOGOUS HYBRIDIZATIONS BETWEEN DIFFERENT LENGTH CHICKEN cDNAs AND RABBIT GLOBIN mRNA

Amino acid sequence data for rabbit and chicken globin chains can

be used to demonstrate that there is a high degree of overall homology between the respective globin chains (67% for β chains) (Braunitzer *et al.*, 1966, Matsuda *et al.*, 1971) and 71% for rabbit α and chicken α_1 (Flamm *et al.*, 1971) and chicken α_2 (Matsuda *et al.*, 1973) globin chains, and that the differences in homology are distributed in a relatively linear fashion along the chains (Fig. 5.1). This suggests that there may similarly be extensive nucleotide sequence homology between the translated regions of rabbit and chicken globin mRNAs, and that the differences in homology may also be distributed linearly throughout the translated region. If rabbit and chicken mRNAs did contain extensive regions where nucleotide differences were linearly distributed, then as discussed before, hybridization experiments involving rabbit globin mRNA and several chicken globin cDNAs of defined length, should detect such a distribution.

To test this hypothesis, defined lengths of chicken globin cDNA were isolated. Transcriptase I cDNA was used essentially as prepared, since as previously discussed, it represents a complete copy of the mRNA template. Transcriptase II and polymerase I cDNAs were further fractionated into size classes by separately collecting the heavy, intermediate and light regions from sucrose gradients (Fig. 5.2). The nucleotide length of each size class was determined by formamide gel analysis (Fig. 5.3). The respective nucleotide length of each cDNA size class, without the 5' - oligo(dT) tail is listed in Table 5.1.

The defined length cDNA preparations were then hybridized to rabbit globin mRNA, and the reactions taken to a $Cr_o t$ value of 1 to ensure completion (see Section 5.3). The extent of hybridization in each case was determined by S1 nuclease assays. The results of two separate experiments, in which

Figure 5.1

Distribution of Amino Acid Differences in

Rabbit and Chicken Globins

Rabbit α and chicken α_2 comparison (Flamm *et al.*, 1971; and Matsuda *et al.*, 1973) ●——————●

Rabbit β and chicken β comparison (Braunitzer *et al.*,1966; and Matsuda *et al.*, 1971) ■——————■

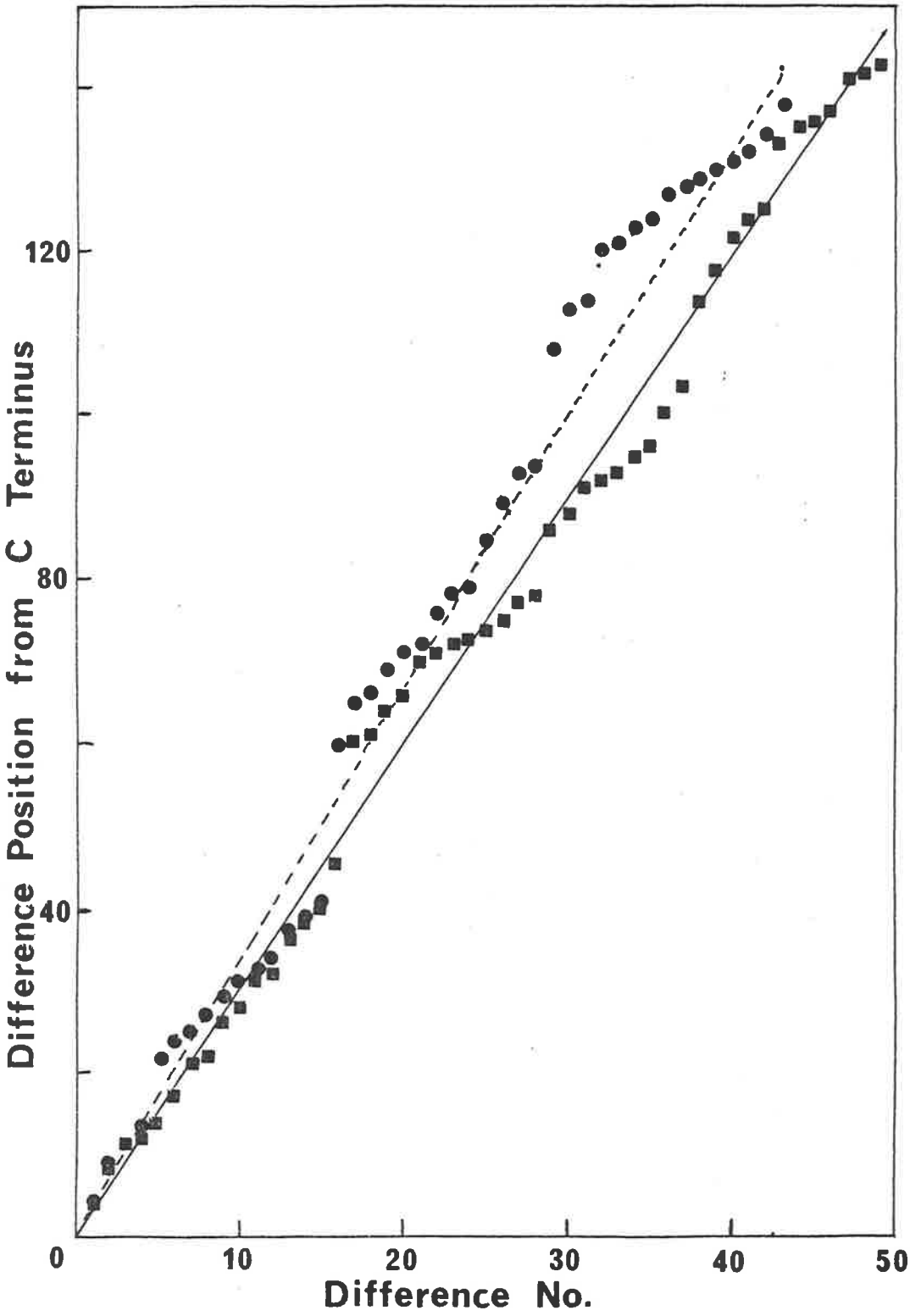

Difference Position from C Terminus
120
80
40
0
10
20
30
40
50
Difference No.

Figure 5.2

Further Fractionation of Chicken Globin cDNAs into Size-Classes by Sucrose Gradient Centrifugation

Transcriptase I, transcriptase II and polymerase I chicken globin cDNAs were centrifuged on sucrose gradients as described for Fig. 4.1. Different size-class fractions were pooled as shown, and the nucleotide length of these fractions was measured by formamide gel electrophoresis as shown in Fig. 5.3.

Centrifugation was in the direction left to right.

(a) transcriptase I cDNA

(b) transcriptase II cDNA

(c) polymerase I cDNA

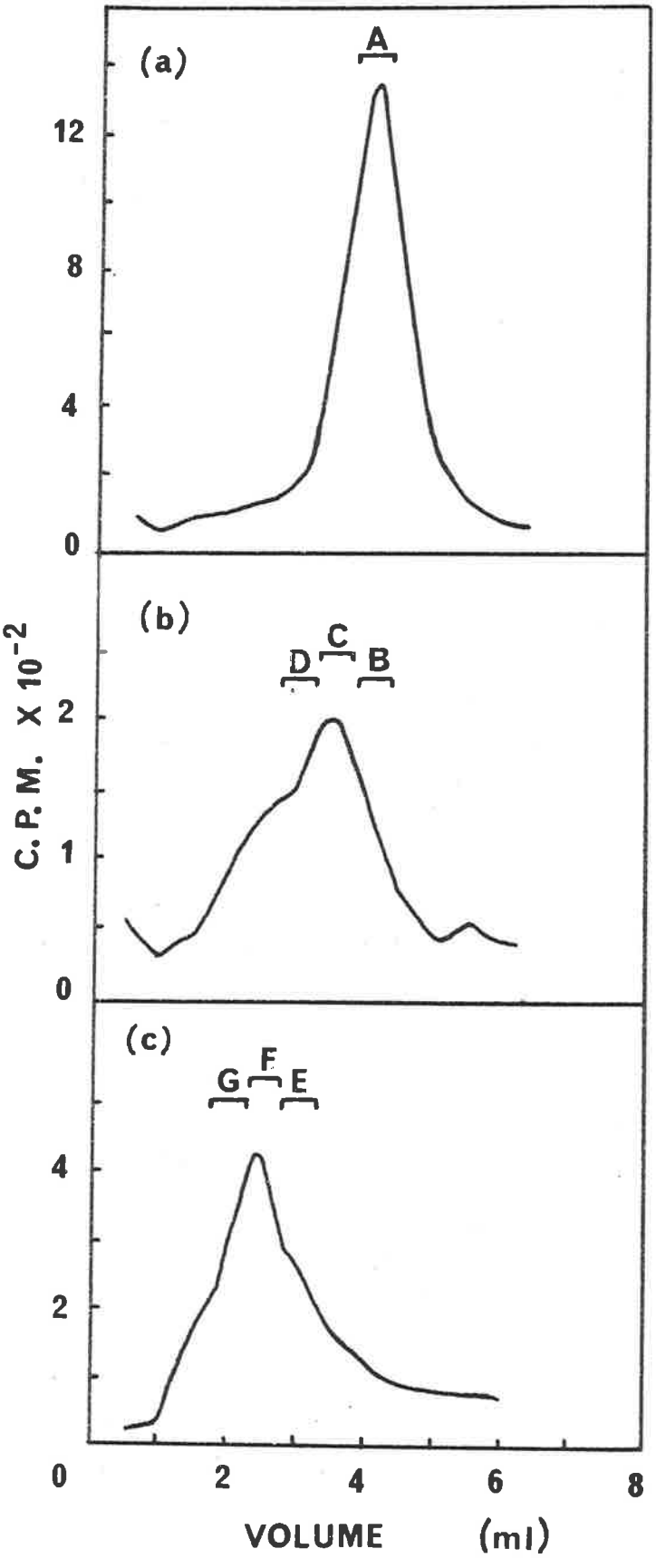

(a)
A
12
8
4
0
(b)
D
C
B
2
1
0
C.P.M. X 10^{-2}
(c)
G
F
E
4
2
0
2
4
6
8
VOLUME (ml)

Figure 5.3

Size Determination of Chicken Globin cDNA on Formamide Gels

Separate size-class cDNAs were isolated as shown in Fig. 5.2, and electrophoresed in 4% polyacrylamide formamide gels. The mobility of the marker RNAs was determined by staining in 0.5% toluidine blue. The gels were then cut into 1 mm slices, counted, and the average nucleotide length of each cDNA calculated.

(A) transcriptase I

(B) and (C) transcriptase II

(E), (F), (G) polymerase I

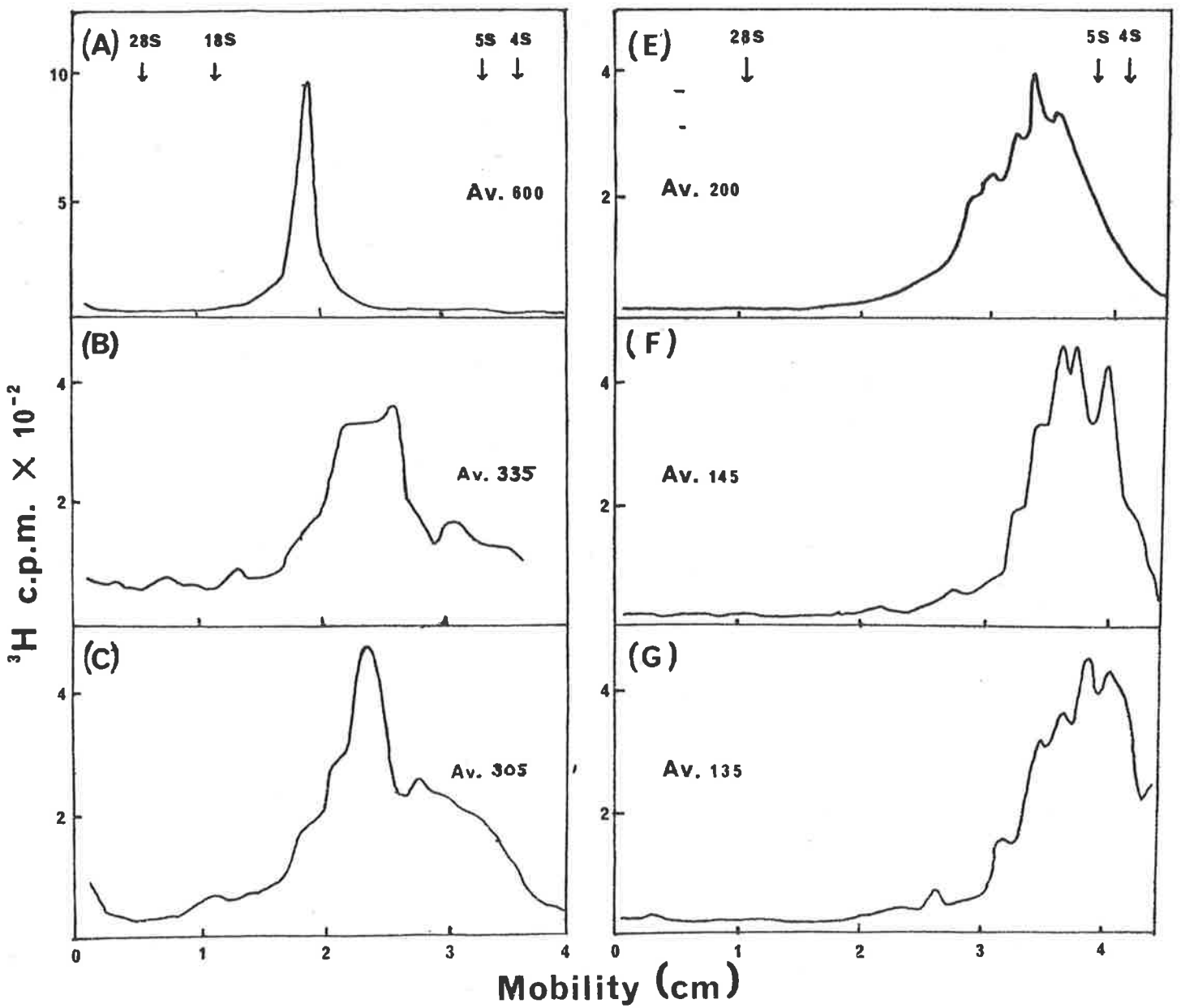

(A)
28S
18S
5S 4S
Av. 600
(B)
Av. 335
(C)
Av. 305
(E)
28S
5S 4S
Av. 200
(F)
Av. 145
(G)
Av. 135
^{3}H c.p.m. X 10^{-2}
Mobility (cm)

Table 5.1 CHICKEN GLOBIN cDNA - RABBIT GLOBIN mRNA HYBRIDIZATION

cDNA fraction from sucrose gradient centrifugation (Fig.5.2)	Average Nucleotide length without 5' oligo(dT) tail	% cDNA hybridization to rabbit globin mRNA Expt. 1.	Expt. 2.
A	600	38.6	37.1
B	315	30.3	-
C	285	-	26.7
E	185	26.5	24.8
F	125	11.2	14.8
G	115	-	12

Chicken globin cDNA preparations of defined sizes were hybridized to rabbit globin mRNA to $Cr_o t$ 1 and % hybridization determined by S1 nuclease assays.

different length chicken globin cDNAs were hybridized to separate rabbit globin mRNA preparations, are shown in Table 5.1. With the largest chicken globin cDNA, about 40% of the total cDNA was protected against S1 nuclease digestion by rabbit globin mRNA. This estimate that 40% of the nucleotide sequences in rabbit and chicken globin mRNA are homologous, is probably a minimal one (see Section 5.3). However, Leder *et al.*, (1973) obtained a similar result in hybridizations between duck and rabbit globin sequences assayed with S1 nuclease.

The data for shorter chicken globin cDNA hybridizations to rabbit globin mRNA is also shown in Table 5.1, and from this data it is clear that the extent of hybridization increases with increasing lengths of chicken globin cDNAs.

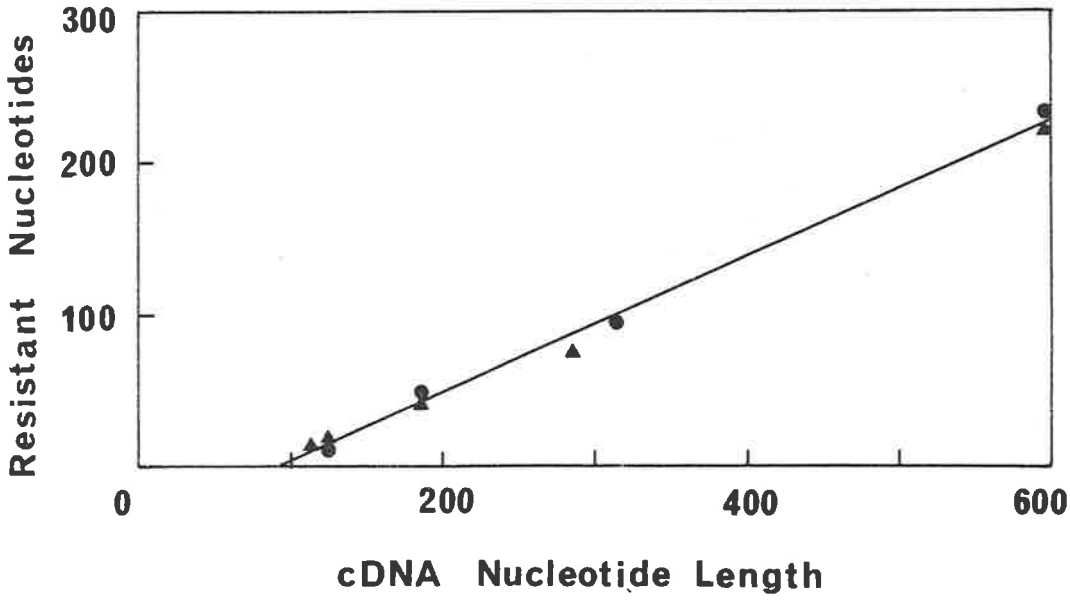
When the number of cDNA nucleotides resistant to S1 nuclease digestion is plotted against cDNA nucleotide length, as in Fig. 5.4, a straight line relationship is evident. Since each length of chicken globin cDNA completely hybridized to chicken globin mRNA with a characteristic $Cr_o t$ ½ value (Fig. 4.2), the data in Fig. 5.4 demonstrates conclusively that there is a major region within chicken globin cDNA that is significantly complementary to rabbit globin mRNA, and that within this region there is a linear distribution of the complementary sequences.

When this line of linear complementarity (Fig. 5.4) is extrapolated to a point of zero homology, an intercept value equivalent to 90 chicken globin cDNA nucleotides is obtained. On this basis, we suggest that chicken globin cDNA can be divided into two distinct regions: one in which there is substantial complementarity to rabbit globin mRNA coding sequences, and one corresponding to the 3' end of rabbit globin mRNA where there is very little

Figure 5.4

Rabbit Globin mRNA/Chicken Globin cDNA Hybridization

Rabbit globin mRNA (15 µg/ml) was hybridized to chicken globin cDNAs in a 50 µl reaction volume in 0.6 M NaCl hybridization buffer. The extent of hybridization at $Cr_o t$ 1 was determined by S1 nuclease assay. Under these conditions, chicken globin mRNA completely protected the cDNAs against S1 nuclease digestion. The cDNA nucleotide length in this plot refers to the length without the 5' oligo(dT) tail. The line was computer-fitted to the experimental data. The different symbols represent data from two separate experiments.


Resistant Nucleotides
300
200
100
0
200
400
600
cDNA Nucleotide Length

apparent complementarity (Fig. 5.5).

The limits of this assay are such that only hybrid sequences of 10 - 15 nucleotides in length, or more, that contain dC, contribute to this estimate of complementarity. Hence region A (Fig. 5.5) extending 90 nucleotides in from the 5' oligo(dT) sequence, does not contain labelled sequences of 10 - 15 nucleotides (or more) which are complementary to the corresponding region of rabbit globin mRNA. In contrast, region B extending from the 90 nucleotide point towards the 3' end of the cDNA, contains significant nucleotide lengths labelled with dC that are complementary to rabbit globin mRNA. In addition, the complementary sequences within region B, are linearly distributed.

5.5 LOCATION OF SEQUENCES IN CHICKEN GLOBIN mRNA

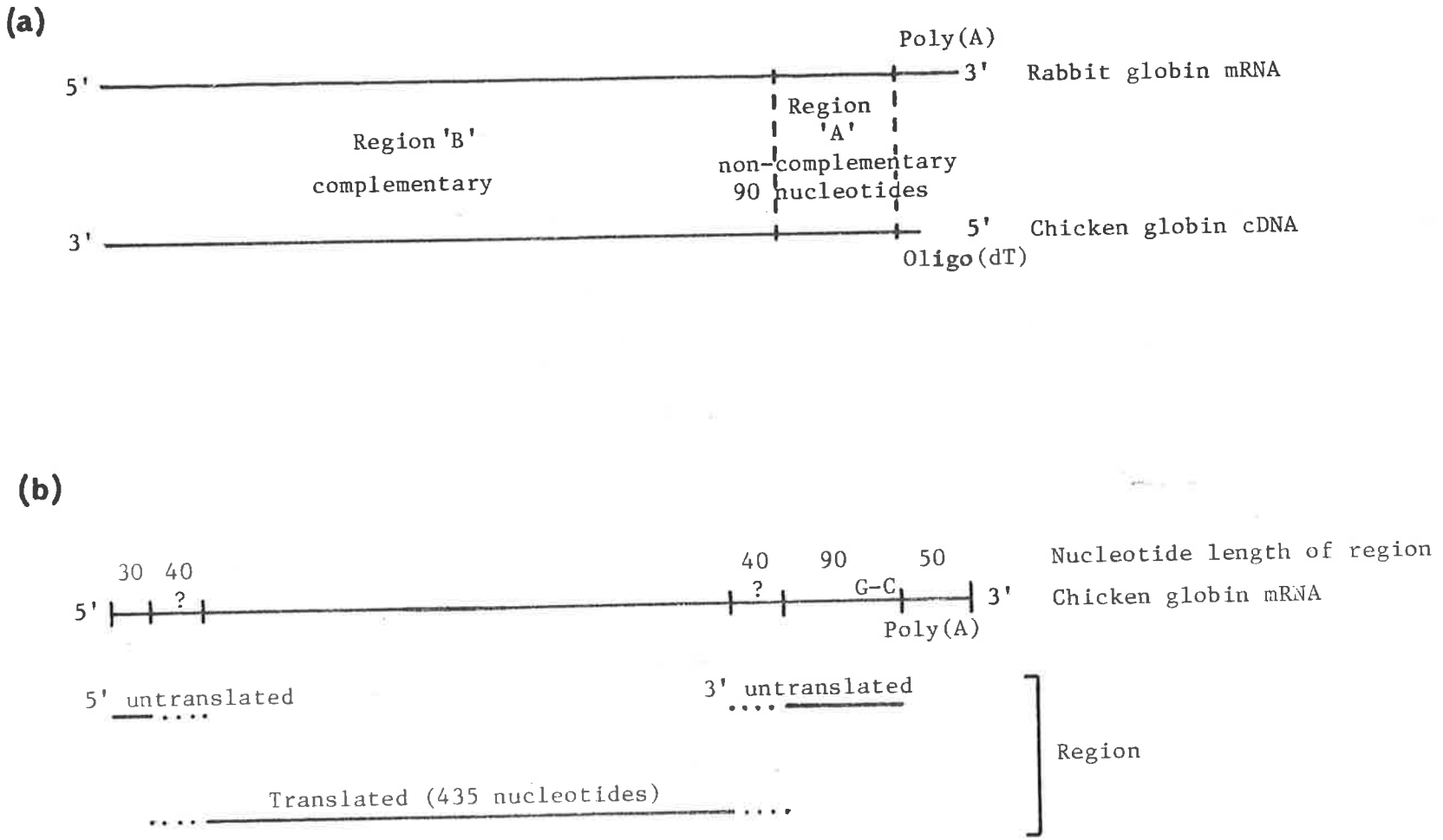
The average length of the translated region of globin mRNA is 435 nucleotides. By correlating the region of linear homology within rabbit and chicken globin mRNAs with the linear amino acid homology of the globin proteins, it seems likely that the globin coding sequences lie within the region of linear nucleotide homology of mRNA. There is additional evidence in support of this proposal. In attempting to locate the 435 translated nucleotides within globin mRNA, it is significant that 8 amino acids at the C-termini of rabbit and chicken α globins contain only one amino acid difference, and that this difference can be accounted for by a single nucleotide change. Similarly, there is one amino acid difference, again accountable for by a single nucleotide change, in the 7 amino acid sequence at the C-termini of the β globins. This further suggests that the sequence coding for the C-termini, present towards the 3' end of the mRNA, are

Figure 5.5

Globin Nucleotide Sequence Organization

a) Major areas of complementarity (B) and non-complementarity (A) between rabbit globin mRNA and chicken cDNA are shown.

b) Regional map of chicken globin mRNA based on data given in this paper. A minimum of 90 nucleotides for the 3' untranslated region adjacent to the poly(A) tract has been estimated with the possibility that this region may extend up to a further 40 nucleotides (Fig. 5.4 and text). Allowing for the translated region, this leaves a region of 30 - 70 untranslated nucleotides at the 5' end of the mRNA. ····· represents the region of uncertainty.


(a)
Poly(A)
5'
3'
Rabbit globin mRNA
Region 'A'
Region 'B'
non-complementary
complementary
90 nucleotides
3'
5'
Chicken globin cDNA
Oligo(dT)
(b)
30 40
40 90 50
Nucleotide length of region
?
?
G-C
5'
3'
Chicken globin mRNA
Poly(A)
5' untranslated
3' untranslated
Region
Translated (435 nucleotides)

within the region of mRNA estimated to contain linear homology, rather than in the 90 nucleotide sequence immediately adjacent to the 3' poly(A) attachment site.

It is likely that a stretch of untranslated nucleotides between the coding sequence and the poly(A) sequence is a general characteristic of eukaryote mRNAs (Proudfoot and Brownlee, 1974b; Proudfoot, 1976; Cheng *et al.*, 1976). It was apparent at the time of this investigation that for rabbit β globin mRNA, this sequence was at least 52 nucleotides in length (Proudfoot and Brownlee, 1974b). On the basis of the evidence described here, we propose that the 3' untranslated region of chicken globin mRNA is at least 90 nucleotides in length (Fig. 5.5).

This estimate is a minimal one since it is based on an extrapolation. However data contributing to this estimate is based on all lengths of cDNA used. Furthermore, there is no detectable deviation from the linear relationship between the length of chicken cDNA used, and the number of nucleotides within this cDNA protected by rabbit globin mRNA. This is true even when cDNA preparations of 115, and 125 nucleotides were used (Fig. 5.4), so that it is unlikely that the length of this putative untranslated region is more than 130 nucleotides.

Accordingly, a regional map for chicken globin mRNA is proposed (Fig. 5.5). In this map, sequences are located within an average chicken globin mRNA molecule, based on the average of α_1, α_2, and β chicken globin mRNAs.

5.6 DISCUSSION

Using heterologous hybridizations, results in this chapter describe

how chicken globin cDNA can be divided into two distinct regions on the basis of its apparent complementarity with rabbit globin mRNA. As stated in Section 5.3, the limits of the assay used, are firstly that only hybrids greater than 10 to 15 nucleotides are scored as hybrid sequences. Secondly, only hybrids containing dC are detected. The work described in this chapter concerns the organization of sequences within chicken globin mRNA, and accordingly a regional map of chicken globin mRNA has been proposed. This map has been derived from 4 sets of data. Firstly, it has been shown experimentally that the overall length of the mRNA is 650 nucleotides, and secondly that the 3' poly(A) tract is 45 - 50 nucleotides long. Thirdly, heterologous hybridizations have shown that a large homologous region, and a 90 nucleotide non-homologous sequence can be located within rabbit and chicken globin mRNAs. Globin protein sequence data has suggested, by correlation, that chicken globin mRNA translated sequence is present within the region that is homologous to rabbit globin mRNA, and that the 90 nucleotide non-homologous sequence adjacent to the 3' poly(A) is untranslated. 90 nucleotides is a minimum estimate of the size of the 3' untranslated region, although as previously discussed, it is unlikely to be more than 130 nucleotides.

Finally, an average figure for the coding region of this globin mRNA of 435 nucleotides means that the 5' ends of these molecules contain a non-coding region of 40 - 70 nucleotides.

Since this investigation was completed, Proudfoot (1976) has determined the lengths of the 3' non-coding regions of rabbit globin mRNAs. Their lengths are 89 ± 2 nucleotides for α globin mRNA, and 95 ± 2 nucleotides for β globin mRNA. This data is in close agreement with the

90 nucleotide estimate, described here, for the length of the 3' non-coding region of chicken globin mRNA.

Conclusions concerning the relative evolutionary divergence of the coding and 3' non-coding regions can also be tentatively made from the experiments discussed in this chapter. Since no nucleotide homology was detected in the 90 nucleotide sequence immediately adjacent to the 3' poly(A), the results superficially indicate that the nucleotide sequences within the 3' non-coding regions of rabbit and chicken globin mRNAs have diverged to a greater extent than those within the translated regions. However, remembering the limits of the assay used, alternative interpretations are not excluded. Taking two extreme examples, the data could indicate firstly that there are no homologous G residues contained within the 3' non-coding regions of the globin mRNAs. Hybrid sequences between the 3' untranslated region of rabbit globin mRNA and the corresponding region of chicken globin cDNA would therefore be unlabelled, and not detected. Alternatively there may be only a single nucleotide difference every 7 - 8 residues within the 3' untranslated regions of rabbit and chicken globin mRNAs. Hence the hybrids formed in these regions, between rabbit globin mRNA and chicken globin cDNA might be totally digested by S1 nuclease and remain undetected by this assay, despite possible complementarity.

It is therefore more accurate to state from these results that the pattern of nucleotide sequence divergence within the 3' non-coding regions of chicken and rabbit globin mRNAs is different to that of the coding regions. Sequencing data has provided more detailed information on the evolutionary divergence of the 3' non-coding region of eukaryote mRNAs (Proudfoot and Brownlee, 1976; Cheng *et al.*, 1976) and this is referred to in the final

discussion (Chapter 8).

In summary, the work described in this chapter has led to a proposal for the organization of sequences in chicken globin mRNA. In particular, it is proposed that there is a region of at least 90 untranslated nucleotides immediately adjacent to the poly(A). This is in close agreement with the length of the 3' untranslated region of rabbit globin mRNA (Proudfoot, 1976). In addition, it is likely that 40 - 70 nucleotides at the 5' end of chicken globin mRNA are untranslated. These nucleotides may form a ribosome binding site in a similar manner to the 5' untranslated region of bromosaic virus (Dasgupta *et al.*, 1975).

CHAPTER 6.

ISOLATION AND CHARACTERISATION OF PURIFIED ERYTHROBLAST RNA

CHAPTER 6. ISOLATION AND CHARACTERISATION OF PURIFIED ERYTHROBLAST RNA

6.1 INTRODUCTION

The ultimate aim of the remainder of the work described in this thesis was to characterise the primary transcription product of the globin gene. Hence a procedure for the isolation of total RNA from avian erythroid cells was developed. This chapter discusses the method chosen for the isolation, and evaluates its effectiveness.

It is particularly concerned with characterising the RNA isolation from chicken erythroid cells, and demonstrating that any presumptive precursor to polysomal globin mRNA is not lost or degraded during the isolation procedure.

6.2 RATIONALE FOR THE ISOLATION PROCEDURE

The isolation method is described in Section 2.2. It did not involve the prior isolation of erythroid cell nuclei because of the probability of RNA degradation, and is similar to the procedure previously described by McKnight and Schimke (1974), except for the following differences.

1) erythroid cells were lysed by the addition of ice-cold phenol : chloroform, with the view to destroying instantly the cellular RNA'ase activity. In this way it was anticipated that the molecular integrity of cellular RNA species might be maintained.

2) phenol : chloroform extraction was carried out at pH 9.0 to maximise the recovery of RNA, including all poly(A)-containing RNA, from the cells (Brawerman *et al.*, 1972).

The phenol : chloroform extract contained erythroid cell nucleic acids,

as well as other cellular material. The impurities probably included haemoglobin, since the ethanol precipitate pellet was pink in colour. Whatever their nature, these impurities interfered with initial attempts to separate the RNA from DNA.

6.2.1 Methods investigated to further purify the erythroblast phenol extract

1) *Repeated ethanol precipitation*

It was thought that residual phenol might be preventing the efficient separation of RNA from DNA. Accordingly, the phenol extracted material was repeatedly ethanol precipitated, to separate the nucleic acids from residual phenol. However, this technique did not noticeably improve the efficiency of RNA separation from DNA.

2) *Second phenol extraction*

It was possible that residual protein might be influencing RNA-DNA separation. However, when the nucleic-acid preparation was phenol : chloroform extracted for a second time, the extract retained some of its pink colour, and little improvement in the behaviour of the nucleic-acid mixture resulted. A second phenol extraction was therefore considered to be of little value.

3) *Protease K digestion*

The proteolytic activity of protease K, in the presence of 0.1% to 0.5% SDS (Gross-Bellard *et al.*, 1973) suggested that it might be useful in digesting the residual protein contamination in the erythroid cell nucleic-acid extract. Accordingly, digestions containing from 50 µg/ml to 250 µg/ml protease K were carried out. However, while the efficiency of RNA-DNA separation was markedly

improved, the ratio of 18S : 28S ribosomal RNAs in the final RNA preparation suggested that some RNA degradation had occurred during the digestion. Protease K was therefore not used further in the isolation of erythroid cell RNA.

4) *Sucrose gradient centrifugation*

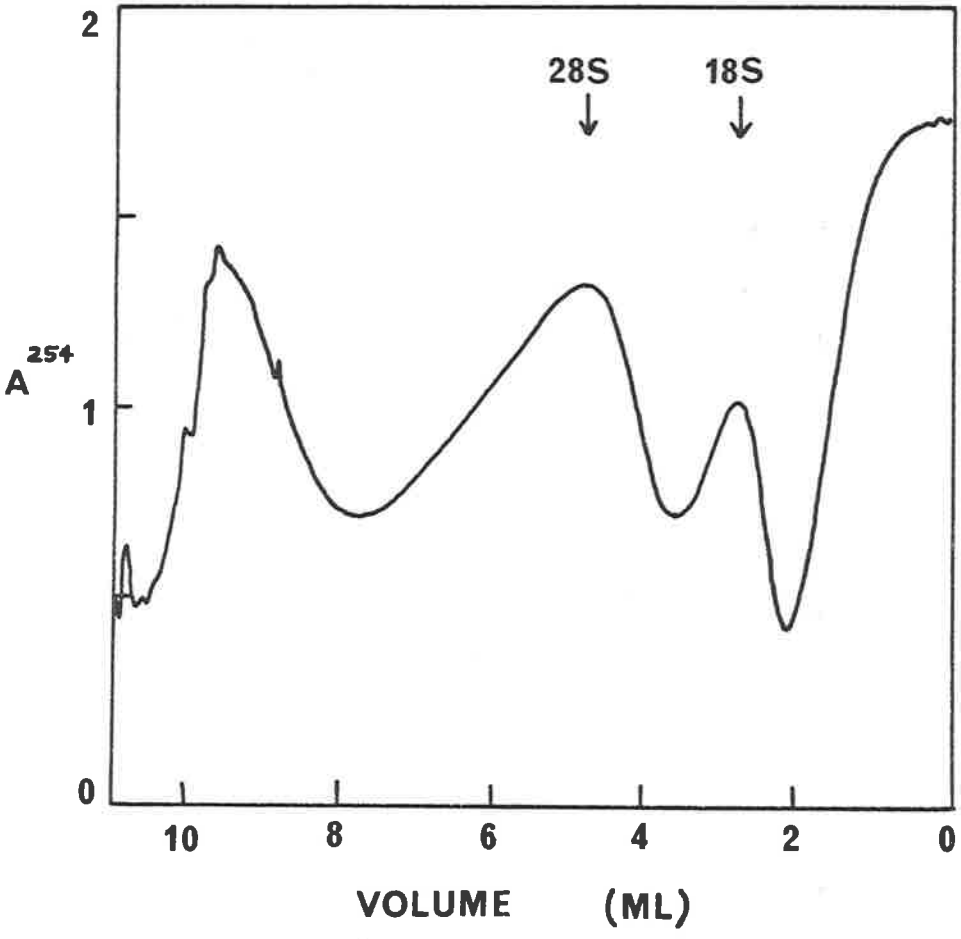
The protease K digestion experiments suggested that residual protein (probably haemoglobin) might be adversely influencing attempts to isolate erythroid cell RNA from DNA. It was known that during sucrose gradient centrifugation, the sedimentation rates of globin protein and globin mRNA sequences are about 4S, and 9S to 10S respectively (see Fig. 3.1). Hence the erythroid nucleic acid extracts were sedimented on sucrose gradients, in an attempt to separate the 4S components from the nucleic acid fraction containing the globin mRNA sequences. After this further purification step, the separation of RNA from DNA was markedly improved. Furthermore, the ratio of 18S : 28S ribosomal RNAs in the resultant RNA preparation, suggested that negligible RNA degradation had occurred. Accordingly, sucrose gradient centrifugation was adopted routinely for the further purification of erythroid cell phenol : chloroform extracts.

The A^{260} profile of such a gradient is shown in Fig. 6.1. The top 0.8 ml of the gradient, containing components up to 4S in size, was separated from the remainder of the gradient by pasteur pippette. A 1.0 ml pad of 50% sucrose at the bottom of the gradient ensured that none of the nucleic acid components were pelleted during centrifugation.

Figure 6.1

Sucrose Gradient Centrifugation of Phenol Extracted Nucleic Acids from Erythroblasts

Nucleic acids were phenol extracted from avian erythroblasts as described in Chapter 2, and ethanol precipitated. Material from 1 ml of packed erythroblasts was sedimented on 4 x 5% - 20% sucrose gradients in SET 7.4, at 37000 r.p.m. for 4 hr at 17°C in a Beckman SW41 rotor. The top 0.8 ml of each gradient was discarded, and the remainder of the nucleic acids (material larger than 4S) was ethanol precipitated.


2
1
0
A^{254}
28S
18S
10
8
6
4
2
0
VOLUME (ML)

6.2.2 Separation of RNA from DNA

Two principal methods have been described for separating RNA from DNA. One of these involves DNA'ase digestion (Soeiro and Darnell, 1969). However, it was felt that a method not involving enzymic digestion might be preferable, despite the availability of a technique for destroying any RNA'ase contaminating the DNA'ase (Zimmerman and Sandeen, 1966). Hence the alternative method, involving sodium acetate precipitation of RNA was used (Kern, 1975). The RNA precipitate was collected, and centrifuged on sucrose gradients. The sedimentation pattern of the resultant avian erythroblast RNA is shown in Fig. 6.2.

6.3 CHARACTERISATION OF THE RNA ISOLATION PROCEDURE

6.3.1 The ability to recover erythroblast RNA *in vitro*

Essentially two types of RNA loss can occur during the isolation of erythroblast RNA. Inherent inadequacies in any of the individual isolation steps, can potentially result in incomplete recovery of the RNA. Similarly, the specific loss of a particular RNA species can result in an *in vitro* RNA preparation that is not representative of that occurring *in vivo*. This section is particularly concerned with estimating the level of RNA recovery after each isolation step, and demonstrating that specific classes of RNA are not lost during the preparation procedure.

Nuclease degradation of RNA is another potential source of RNA loss, and experiments aimed at determining whether any marked RNA degradation had occurred during isolation, are also described in this chapter.

Figure 6.2

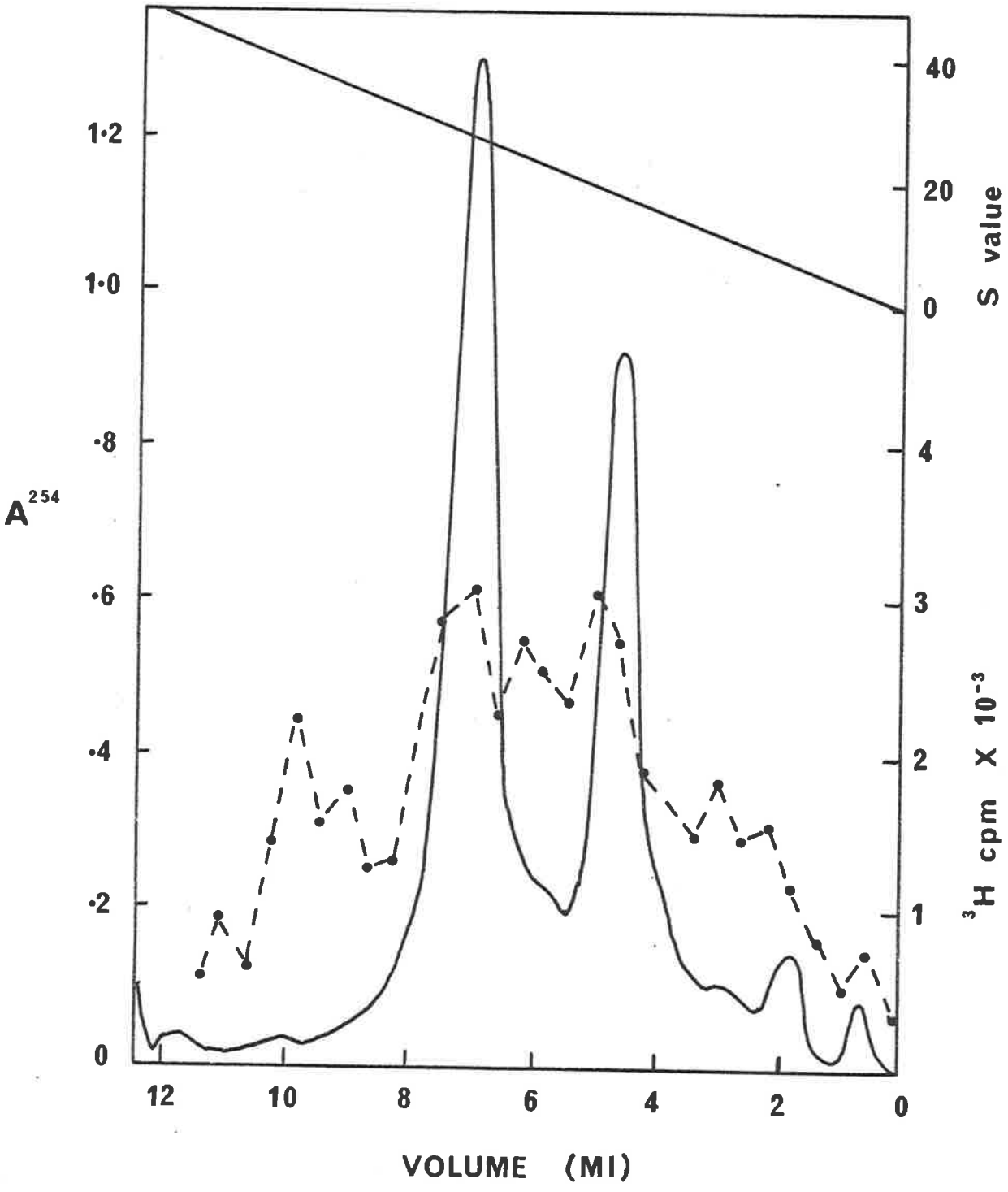
Sucrose Gradient Centrifugation of Isolated Chicken Erythroblast RNA

Erythroblasts were incubated with 100 μCi of ^{3}H-uridine for 50 minu and total RNA was isolated from one half of these cells. RNA from these ce was centrifuged on 10% - 40% sucrose gradients in NET, at 30,000 r.p.m. for 16 hr at 4°C in a Beckman SW41 rotor. 0.5 ml fractions were collected from the gradient, and aliquots counted for ^{3}H-radioactivity.

The remaining half of these cells was incubated for a further 90 mi after the addition of a 100 fold excess of unlabelled uridine, as described Fig. 6.5.

A^{260} profile ————————

^{3}H c.p.m. •----------•


1·2
1·0
·8
·6
·4
·2
0
A254
40
20
0
S value
4
3
2
1
3H cpm X 10-3
12
10
8
6
4
2
0
VOLUME (MI)

1) *Distribution of labelled RNA during isolation*

The efficiency of the RNA isolation steps can be measured most effectively, by labelling the erythroblast RNA *in vivo*, and determining the distribution of labelled RNA after each step. It is particularly important when attempting to locate globin gene primary transcripts, that labelled RNA (i.e. newly-made) is not specifically lost, since globin gene primary transcripts will be newly-made.

To test for the loss of newly-made RNA, erythroblasts were incubated *in vitro* with H^3-uridine, and total cellular RNA isolated by the method described. The percentage of labelled material recovered after each step of the procedure, directly indicated the proportion of newly-made RNA that was lost. The results of these experiments showed that negligible levels of incorporated uridine were lost in isolation steps preceding sodium acetate precipitation (Table 6.1).

However, RNA precipitation from 3.0 M sodium acetate was incomplete. About 15% of the labelled material could not be collected as a purified RNA precipitate. Sucrose gradient centrifugation showed that at least 40% of the non-precipitated, labelled RNA sedimented at 4S to 5S, and that the remainder sedimented in a manner similar to the labelled RNA within the purified RNA precipitate (data not shown).

This evidence is consistent with the proposal that the RNA collected as a precipitate, is equivalent to the RNA larger than 4S occurring within intact erythroblasts. The results of reconstruction experiments, using polysomal globin mRNA and keratin

Table 6.1

Recovery of ^{3}H-labelled Chicken Erythroblast RNA

During RNA Isolation Procedure

Isolation step	^{3}H-c.p.m. recovered ($\times 10^{-6}$)
incubation of erythroblasts	1.3
phenol extraction	1.6
sodium acetate precipitation	
precipitate	1.4
supernatant	0.16

0.25 - 0.5 ml packed erythroblasts were incubated for 15 minutes in presence of 1 mCi ^{3}H-uridine. Aliquots were TCA precipitated after each isolation step to determine the recovery of ^{3}H-labelled RNA.

mRNA further support this proposal (see below).

The inability of 4S and 5S RNA to be precipitated by 3.0 M sodium acetate has been previously documented (Kern, 1975).

2) *Distribution of globin RNA sequences during sodium acetate precipitation.*

Since the globin sequences in the erythroid RNA were of prime interest, it was important to establish that globin RNA was precipitated by sodium acetate. Hence in a control experiment, polysomal globin mRNA was mixed with DNA, and 18S and 28S ribosomal RNA. Sodium acetate precipitation was carried out, and the nucleic acids within the precipitate and supernatant hybridized separately to globin cDNA probe. The globin sequence content of both phases was calculated, using the $Cr_o t$ ½ value of 7×10^{-4} for purified globin mRNA - cDNA hybridization as a reference standard (see Section 4.4).

The results shown in Fig. 6.3 and Table 6.2, demonstrate that all the globin mRNA sequences were recovered either in the precipitated fraction or supernatant fraction. Hence there was no detectable loss of globin sequences during sodium acetate precipitation. Significantly, the $Cr_o t$ ½ values for the respective hybridizations show that the precipitate hybridization proceeded 30X faster than that of the supernatant. Hence 97% of the polysomal globin mRNA was precipitated by sodium acetate.

From these results it is apparent that globin mRNA sequences do not selectively remain in solution during sodium acetate precipitation. Accordingly, the globin RNA content of the purified

Table 6.2

Distribution of Globin RNA Sequences During Sodium Acetate Precipitation

	µg globin mRNA
prior to sodium acetate precipitation	1
sodium acetate precipitate	1
sodium acetate supernatant	0.03

1 µg of globin mRNA was precipitated in 3.0 M sodium acetate solution in the presence of 28S, 18S and 4S RNA, and 5 µg DNA. RNA in the supernatant and precipitate fractions was ethanol precipitated, and serially diluted to derive a $Cr_o t$ curve for the hybridization of each fraction with chicken globin cDNA. The concentration of RNA in each fraction was calculated by comparing the rate of hybridization of these fractions with the standard $Cr_o t\ ½$ value for chicken globin mRNA/cDNA hybridization (see Figs. 4.2 and 6.3).

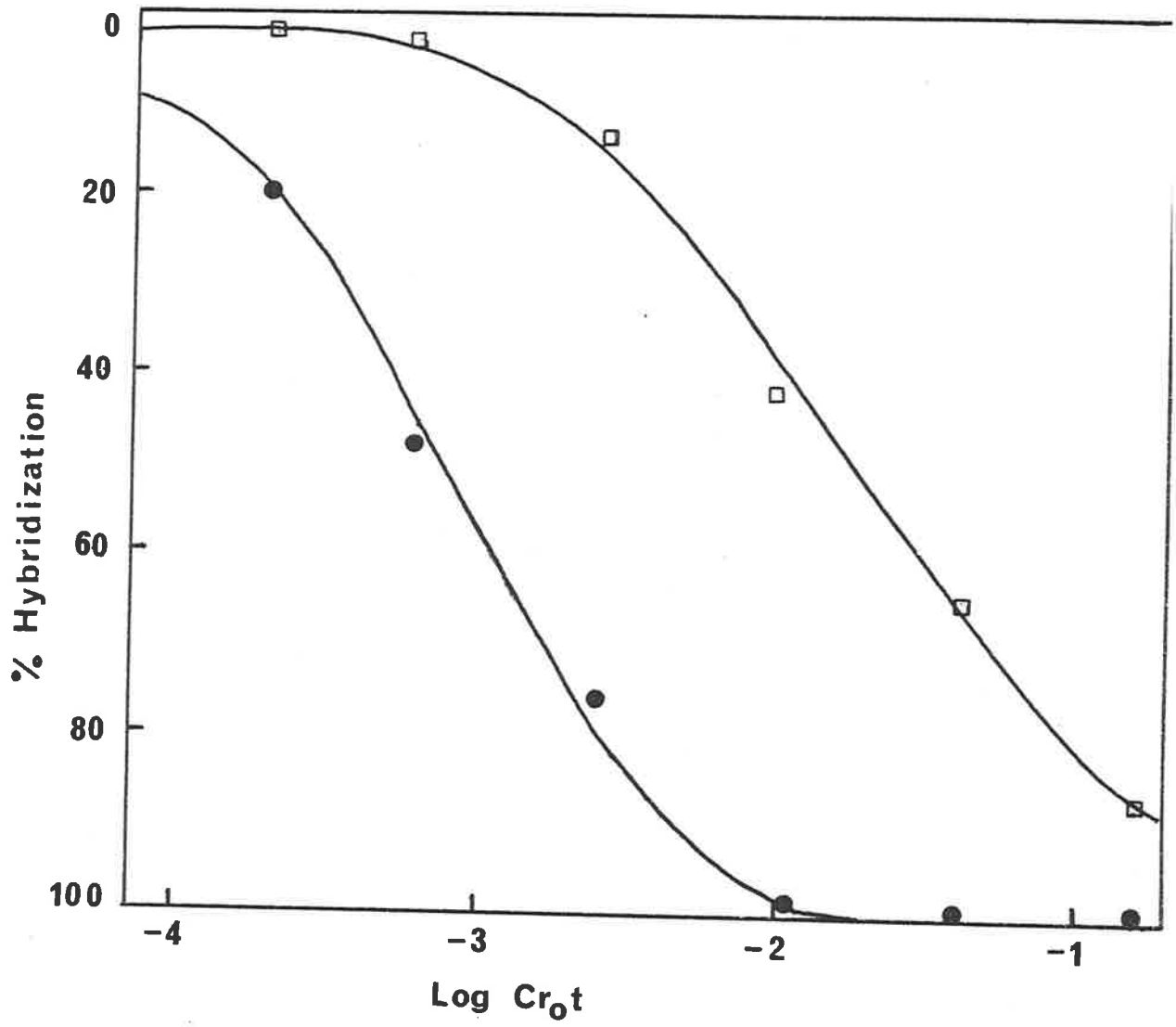
Figure 6.3

Distribution of Globin mRNA Sequences During Sodium Acetate Precipitation

1 μg of globin mRNA was precipitated in 3.0 M sodium acetate solution as described in Table 6.2. RNA in the supernatant and precipitate fractions was ethanol precipitated, and serially diluted samples were hybridized with chicken globin cDNA to derive a $Cr_o t$ curve for each fraction. The relative $Cr_o t$ ½ values indicated the difference in globin mRNA levels within the supernatant and precipitate fractions.

globin RNA within precipitated fraction ●——————●

globin RNA within supernatant fraction □——————□


0
20
40
60
80
100
% Hybridization
−4
−3
−2
−1
Log $Cr_o t$

erythroblast RNA should be equivalent to that occurring *in vivo*.

3) *The isolation of specific RNA sequences present at the level of one to ten molecules per cell.*

If the time taken for processing globin gene primary transcripts is shorter than that needed for transcription, then the number of molecules of this RNA species present at a given time within erythroblasts, might be as low as one per cell. It is probable that radio-activity studies similar to those described earlier in this chapter, would not detect the loss of specific RNA sequences present in these quantities. For this reason, an additional control experiment was carried out to determine whether RNA sequences initially present at the level of about one to ten molecules per cell could be recovered, and detected within the *in vitro* RNA preparation.

Accordingly 10 ng of 12S polysomal keratin mRNA (a gift of B. Powell), an RNA sequence not expected to be normally present within avian erythroid cells, was added as a marker RNA to avian erythroblasts immediately after lysis with phenol : chloroform. Avian erythroblast RNA was then isolated and centrifuged on sucrose gradients. Fractions were collected, and keratin mRNA sequences located using labelled keratin cDNA probe. As shown in Fig. 6.4, keratin mRNA sequences were found in the 12S region of the gradient, indicating that the small levels of keratin mRNA sequences added immediately after lysis were not lost during isolation. In addition, the failure to detect significant levels of keratin mRNA in smaller regions of the gradient, suggested that

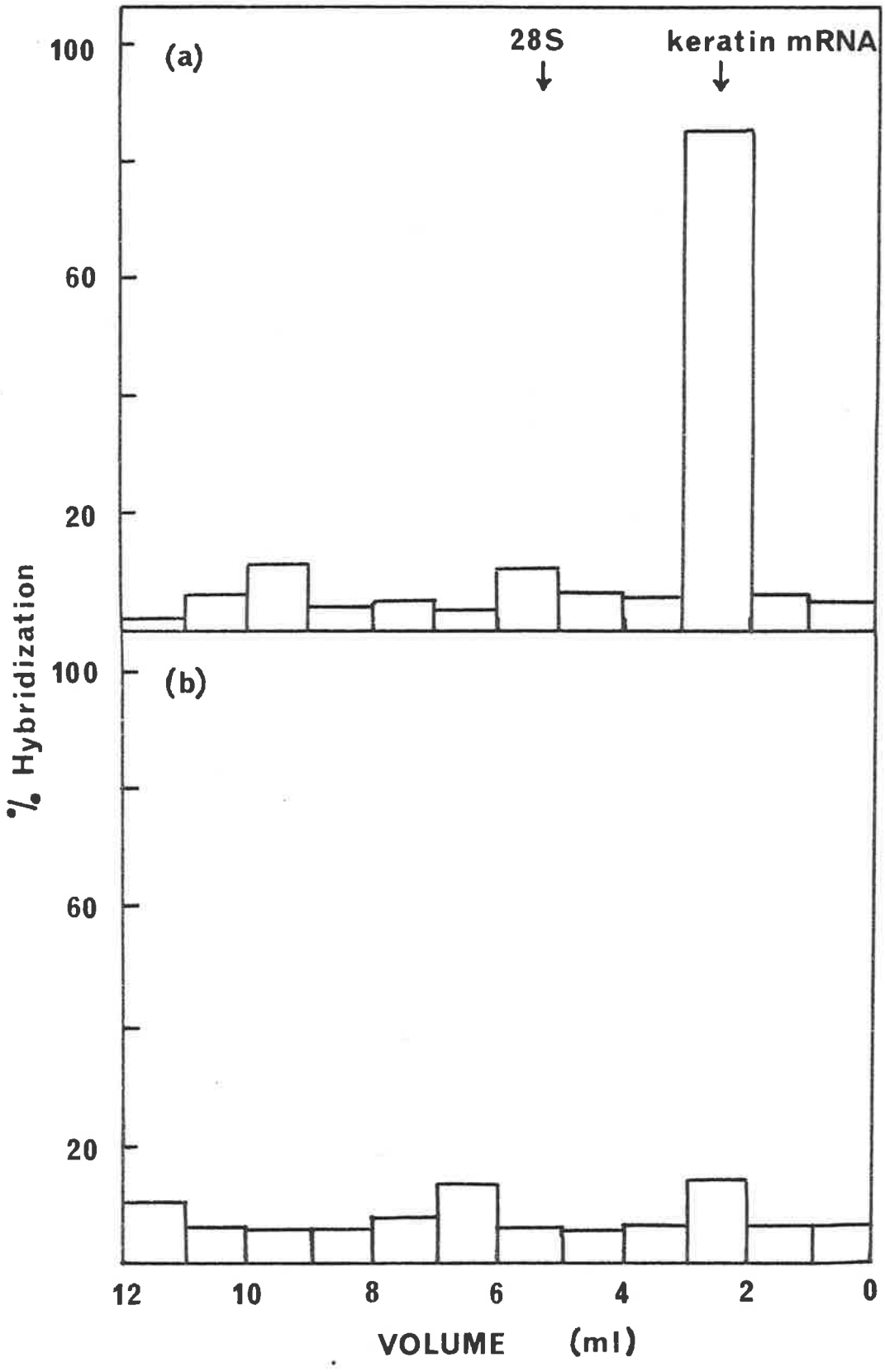
Figure 6.4

Isolation of RNA Sequences Present at the Level of One to Ten Molecules per cell, During Erythroblast RNA Preparation

Erythroblasts were lysed with ice-cold phenol : chloroform, and 5 - ng of chicken keratin mRNA (a gift of B. Powell) was mixed with the lysate prior to the addition of SET pH 7.4. Total RNA was purified, and centrifug on 10% - 40% sucrose gradients, at 30,000 r.p.m. for 16 hr at 4°C. 1 ml fractions were collected, and the RNA hybridized to high specific radioacti keratin cDNA to detect the keratin mRNA sequences.

a) 5 - 10 ng keratin mRNA added to erythroblast lysate

b) no keratin mRNA added.


(a)
28S
keratin mRNA
100
60
20
(b)
100
60
20
% Hybridization
12
10
8
6
4
2
0
VOLUME (ml)

no degradation of keratin mRNA sequences had occurred. It is therefore unlikely that similar levels of globin gene primary transcript would be either specifically lost or degraded during RNA isolation.

It is also significant that keratin RNA sequences were not located in regions larger than 12S, since it demonstrates that keratin mRNA sequences were not aggregated during sucrose gradient centrifugation. Hence globin RNA sequences located in regions larger than 10S are likely to be higher than polysomal mRNA in molecular weight, rather than 10S polysomal sequences in aggregated form.

4) *The extent of RNA degradation during erythroblast RNA isolation.*

As previously stated, nuclease degradation of RNA is an important potential source of erythroblast RNA loss. This section is aimed at determining if erythroblast RNA had been degraded during isolation.

a) Cytoplasmic RNA

The A^{260} profiles of erythroblast RNA sedimented in sucrose gradients, provides some evidence that negligible degradation of cytoplasmic RNA at least, had occurred. 18S and 28S ribosomal RNAs are the two major RNA species within intact erythroblasts in terms of quantity. Hence *in vitro* preparations of erythroblast RNA should also contain discrete 18S and 28S RNAs as major species. As shown in Fig. 6.2, both 18S and 28S ribosomal RNAs sediment in sucrose gradients as major, discrete peaks, suggesting that negligible ribosomal RNA degradation has occurred. The A^{260} profiles provide additional evidence that selective loss or degradation of one of the

ribosomal RNA species has not occurred during isolation. The length of 28S ribosomal RNA molecules is approximately twice that of 18S ribosomal RNA (Pinder *et al.*, 1974), and therefore a ratio of A^{260} values of about 2 : 1 is expected in the absence of any breakdown. As shown in Fig. 6.2, the distribution of A^{260} units in 18S and 28S RNA shows such a ratio within isolated erythroblast RNA. Hence the A^{260} profiles of RNA isolated from avian erythroblast suggest that cytoplasmic RNA at least, has not been degraded.

b) nuclear RNA - the detection of 40S - 45S ribosomal precursor RNA.

The A^{260} profile of erythroblast RNA in sucrose gradients, suggests that cytoplasmic RNA at least has not been degraded, but gives no indication on the fate of nuclear RNA during isolation. The primary transcription product of the globin gene is by definition, a nuclear RNA species, so that it was also important to demonstrate that nuclear RNA had remained intact.

Within higher eukaryotes, 18S and 28S ribosomal cistrons are transcribed into a single 40S - 45S precursor RNA species that contains both 18S and 28S ribosomal RNA sequences. The precursor is then processed within the nucleolus (an organelle within the nucleus) into mature 18S and 28S ribosomal RNA (Perry *et al.*, 1970). Pulse-chase experiments, in which ^{3}H-uridine was specifically chased into erythroblast 18S and 28S ribosomal RNA (Figs. 6.2 and 6.5) show tha ribosomal RNA is synthesized in avian erythroblasts. It therefore follows that isolated avian erythroblast RNA should contain a 40S - 45S precursor to the ribosomal RNAs, if the nuclear erythroblast

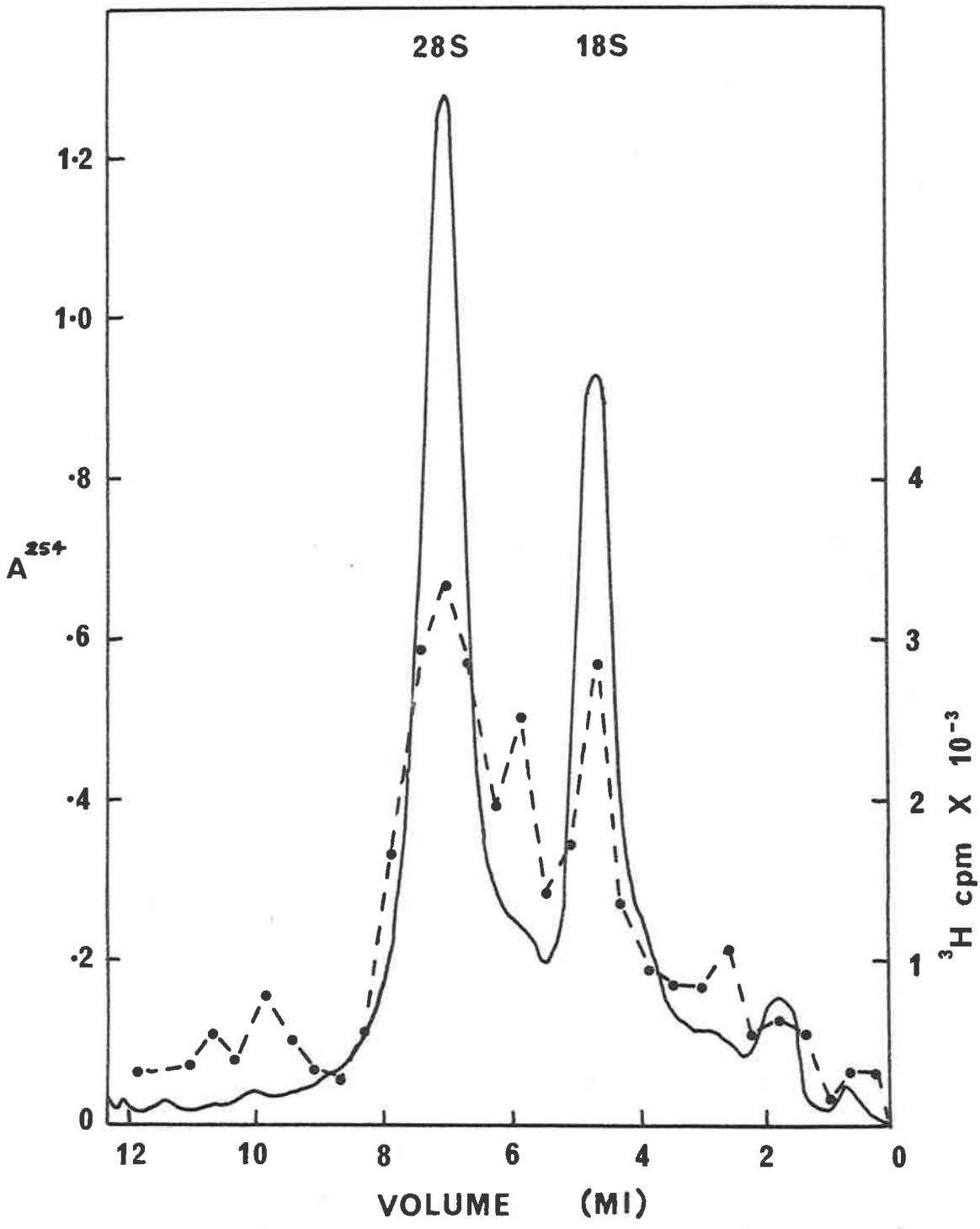
Figure 6.5

Incorporation of ^{3}H-uridine into Chicken 28S and 18S Ribosomal RNA

Erythroblasts were pulsed with 100 μCi ^{3}H-uridine as described in Fig. 6.2, and then incubated for a further 90 minutes after the addition of a 100 fold excess of unlabelled uridine. Total erythroblast RNA was isolated, and centrifuged at 30,000 r.p.m. for 16 hr at 4°C on 10% - 40% sucrose gradients. 0.5 ml fractions were collected, and aliquots counted for radioactivity.

A^{260} profile ————————

^{3}H c.p.m. •----------•


28S
18S
1·2
1·0
·8
·6
·4
·2
0
A254
4
3
2
1
3H cpm X 10-3
12
10
8
6
4
2
0
VOLUME (MI)

RNA has not been degraded.

The labelling profiles of erythroblast RNA initially indicated that intact ribosomal precursor RNA might be present in the preparation. When erythroblast RNA was labelled with ^{3}H-uridine prior to isolation, and the resultant RNA sedimented on sucrose gradients, a discrete 40S RNA species was detected (Fig. 6.2). However, more substantial proof was provided by hybridization experiments. ^{32}P- labelled DNA sequences, complementary to 18S and 28S RNA, were synthesized as described in Section 2.2, and hybridized separately to fractions of sucrose gradients in which the isolated 40S RNA species had been resedimented. As shown in Fig. 6.6, both 18S cDNA and 28S cDNA hybridized to the 40S RNA, showing conclusively that both 18S and 28S ribosomal RNA sequences are contained within the 40S RNA. Accordingly, it was concluded that this 40S RNA species contains precursor ribosomal RNA molecules, indicating that nuclear RNA could be isolated in an undegraded form from avian erythroblasts by the procedure described. Hence it is likely that the primary transcription products of globin genes are similarly undegraded during RNA isolation.

It is unlikely that the 40S RNA species is an aggregate of 18S and 28S ribosomal RNA, since heating steps were carried out prior to each centrifugation, to ensure that the RNA was disaggregated (Haines *et al.*, 1974). Furthermore, globin cDNA hybridization data (Chapter 7) provided no evidence that globin mRNA at least, was

Figure 6.6

Detection of 40S - 45S Precursor to Chicken 28S and 18S Ribosomal RNA

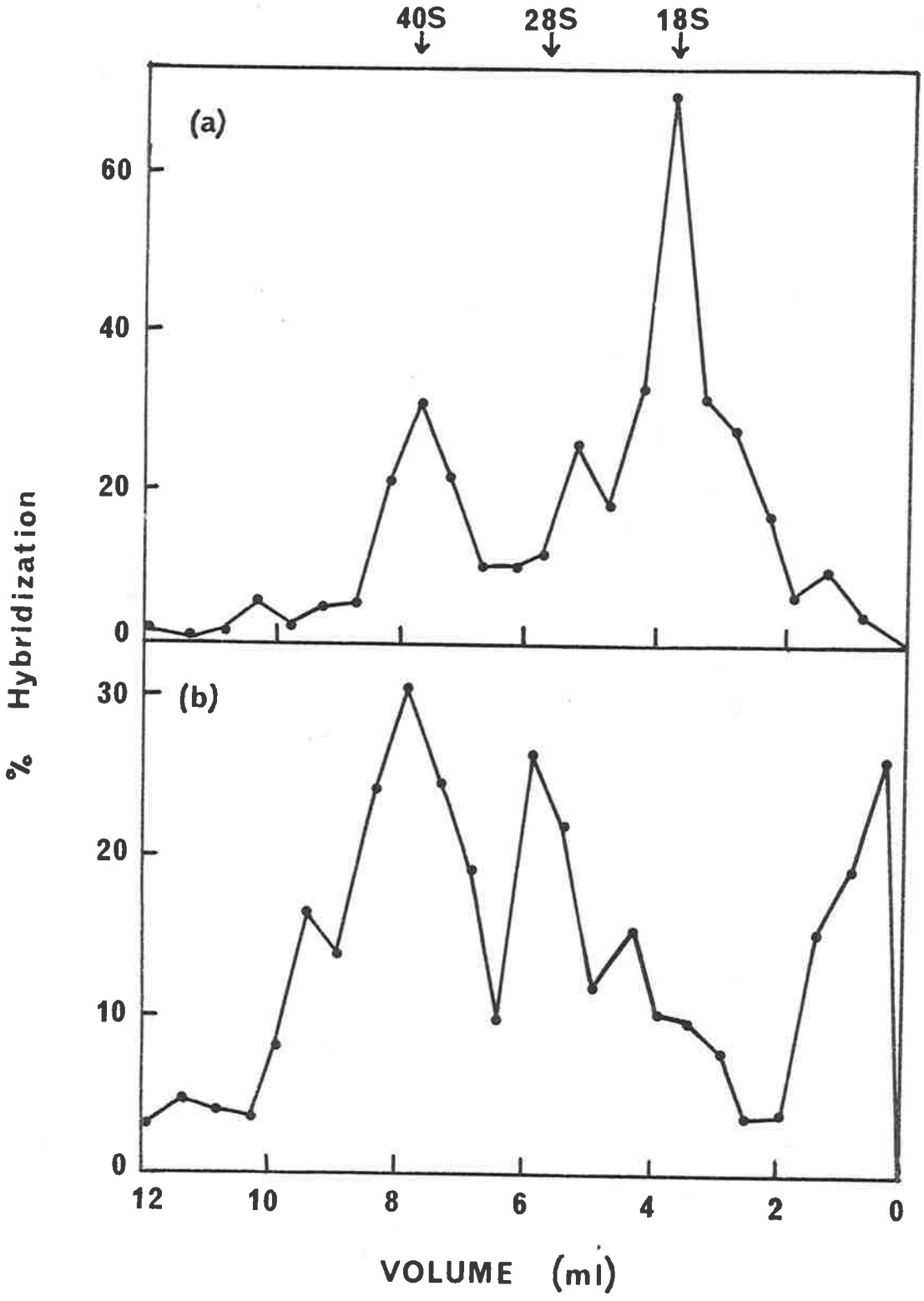
Labelled RNA sedimenting in the 40S - 45S region as shown in Fig. 6.2 was isolated, heated at 65°C for 10 minutes prior to each of the further cycles of sucrose gradient centrifugation. RNA hybridized to 18S cDNA was passaged through two further cycles of centrifugation, and the RNA hybridized to 28S cDNA passaged through three further cycles, to enrich the RNA for presumptive rRNA precursor. 0.5 ml fractions were collected, and the RNA in each fraction hybridized to 18S and 28S cDNA labelled with high specific radioactivity ^{32}P dCTP.

a) fractions hybridized with cDNA made on 18S rRNA template

b) fractions hybridized with cDNA made on 28S rRNA template

Footnote:

It is noted that the proportion of cDNA hybridized to mature rRNA species (i.e. 18S and 28S rRNA) compared with that to rRNA precursor, is higher in the gradient hybridized to 18S cDNA than in the gradient hybridiz to 28S cDNA. This is probably due to the fact that this "precursor" prepar ion was passaged through one less cycle of centrifugation, compared with th in which 28S sequences were detected. Similarly the 4S - 7S RNA that hybridized to 28S cDNA, probably represents breakdown products from 40S or 28S rRNA, generated by repeated heating and centrifugation of this RNA samp


40S
28S
18S
(a)
60
40
20
0
(b)
30
20
10
0
12
10
8
6
4
2
0
% Hybridization
VOLUME (ml)

aggregated during centrifugation.

6.4 DNA CONTENT OF ISOLATED ERYTHROBLAST RNA

It was predicted that the assay for the primary transcription product of the globin genes, might need to be sufficiently sensitive to detect about one molecule per erythroid cell. Hence it was important to ensure that globin sequences detected, were not present in DNA. For this reason, the level of DNA contamination in the RNA preparation was estimated.

The approach used involved determining the proportion of A^{260} units within the RNA preparation that were digested by RNA'ase. As shown in Table 6.3, all the A^{260} units were converted to an acid soluble form after RNA'ase digestion, indicating that within the limits of this assay, no DNA could be detected within the RNA preparation.

Hence the claim can reasonably be made, that any globin sequences detected, are present in RNA and not DNA.

6.5 DISCUSSION

The work described in this chapter was principally aimed at providing an avian erythroblast RNA preparation that contained the primary transcriptio product of globin genes. This is one essential step in determining whether adult avian polysomal globin mRNA originates from a higher molecular weight precursor. It is vital to this study, that neither selective losses of particular RNA species, nor RNA degradation have occurred during the isolation of RNA from erythroblasts. Hence the RNA isolation procedure was characterised in detail, and additional experiments were carried out to demonstrate that the resultant *in vitro* RNA preparation was equivalent to that occurring *in vivo*.

The procedure for RNA isolation was similar in many respects to that

Table 6.3

RNA'ase Digestion of Purified Chicken Erythroblast RNA

A^{260} units added to digestion mix	30
A^{260} units digested by RNA'ase	33

30 A^{260} units of erythroblast RNA were added to 100 µl of 0.25 mM EDTA, 0.15 M Tris-Cl pH 8.2, and incubated with 4 µg RNA'ase A and 100 units of T1 RNA'ase for 90 minutes at 37° C. Undigested material was precipitated by the addition of HCl to a final concentration of 0.3 N. Precipitated material was spun down, and the number of A^{260} units remaining in the supernatant directly indicated the amount of RNA in the erythroblast RNA preparation.

previously described by McKnight and Schimke (1974). In particular the extremely high nuclease levels within avian erythroblasts (unpublished observations) was an important factor in deciding on an approach for erythroblast RNA isolation. Accordingly, no attempt was made to isolate RNA from purified erythroblast nuclei, even though the globin gene primary transcripts would be localised within the nucleus. In addition, erythroblasts were lysed directly with phenol : chloroform, prior to the addition of aqueous phase, to minimise any nuclease degradation of the nucleic acids. Earlier experiments had shown that aqueous lysis of erythroblasts, resulted in extensive RNA degradation.

Steps were also taken to ensure that losses of specific classes of RNA did not occur. Hence phenol extraction was carried out at pH 9.0 to maximize the recovery of poly(A)-containing RNA (Brawerman *et al.*, 1972).

In our hands it was difficult to divide the resultant phenol extract into separate purified RNA and DNA components without a prior sucrose gradient centrifugation step. The exact nature of the contaminants that interfered with attempts to separate RNA from DNA is uncertain. Some of the experiments described suggested they might be denatured protein. Whatever their precise nature, RNA was readily separable from DNA after material less than 4S in size was removed from the nucleic acid extract.

Sodium acetate precipitation, in which DNA remains in solution, and RNA larger than 4S is selectively precipitated, was used to separate erythroblast RNA from DNA. Experiments using labelled RNA showed that there was negligible loss of erythroblast RNA in the isolation steps prior to sodium acetate precipitation. However, under the conditions of sodium acetate precipitation used here, about 10% of the erythroblast RNA larger than 4S, remained in solution, and was not collected in the purified RNA

pellet. (The 4S RNA selectively remaining in solution was not of interest since globin gene primary transcripts are at least 10S in size.) Although this amounted to a 10% loss of erythroblast RNA, control experiments using either polysomal globin mRNA or keratin mRNA, suggested that these losses probably did not involve the loss of specific RNA sequences of interest. Both globin mRNA sequences, and RNA sequences present at the level of one to ten molecules per cell, could be recovered in the isolated RNA. Hence globin gene primary transcripts, present within erythroblasts at about the level of one to ten molecules per cell, should be readily detectable within erythroblast RNA.

It is probable that the 10% of erythroblast RNA remaining in solution during sodium acetate precipitation, is indicative of a low level of RNA solubility in 3.0 M sodium acetate.

As well as the inherent inefficiencies of the isolation steps, nuclease degradation is potentially a major source of RNA loss during erythroblast RNA isolation. However, no significant degradation was detected within the erythroblast RNA preparation. Both 18S and 28S ribosomal RNA sequences sedimented as sharp, discrete peaks during sucrose gradient centrifugation, and the number of A^{260} units within the 28S RNA peak was about twice that in the 18S peak. Hence by these criteria, the 18S and 28S ribosomal RNAs within the *in vitro* RNA preparation are undegraded, and equivalent to that *in vivo*. This suggests that other cytoplasmic erythroblast RNA species are similarly undegraded.

However the primary transcription products of the globin genes are likely to be located within the nucleus of avian erythroblasts. It is therefore of prime significance that intact nuclear RNA molecules were

detected within isolated erythroblast RNA. In particular the 40S - 45S precursor to 18S and 28S ribosomal RNAs synthesized within the nucleus, was detected in the purified erythroblast RNA preparation, by hybridization. Hence ribosomal RNA precursor molecules at least, could be isolated in undegraded form from avian erythroblasts. On this basis it is likely that globin gene primary transcripts, also synthesized within avian erythroblast nuclei, are present in undegraded form within the RNA preparation.

In summary the experiments described in this chapter have demonstrated that total avian erythroblast RNA can be isolated without detectable RNA degradation or the selective loss of specific RNA sequences, other than 4S RNA. Accordingly this RNA isolation procedure should provide an RNA preparation containing the primary transcripts of avian globin genes.

The level of DNA contamination within this RNA preparation was estimated to be negligible. It therefore follows that globin sequences detected within *in vitro* erythroblast RNA, are present in RNA, and not DNA.

CHAPTER 7.

IDENTIFICATION OF A 14S PRECURSOR TO CHICKEN GLOBIN mRNA

CHAPTER 7. IDENTIFICATION OF A 14S PRECURSOR CHICKEN GLOBIN mRNA

7.1 INTRODUCTION

Hybridization experiments in different laboratories have led to varied claims concerning the primary transcription product of avian globin genes. In particular avian globin mRNA sequences have been detected in high molecular weight RNA fractions ranging in size from three times (Macnaughton *et al.*, 1974) up to greater than ten times the nucleotide length of polysomal globin mRNA (Imaizumi *et al.*, 1973; Spohr *et al.*, 1974; 1976).

However, these experiments generally did not determine whether the globin mRNA present in higher molecular weight RNA fractions is eventually cleaved into polysomal avian globin mRNA. Furthermore, the work of Imaizumi *et al.*, (1973) and Spohr *et al.*, (1974; 1976) is inconclusive, since it fails to demonstrate the presence of a discrete size-class precursor to avian globin mRNA.

This chapter describes hybridization experiments in which a 14S precursor to polysomal chicken globin mRNA is identified. It demonstrates that the 14S globin RNA species satisfies three essential criteria of a mRNA precursor. Firstly, it is a discrete sized, globin-containing RNA species that is higher in molecular weight than polysomal chicken globin mRNA. Secondly, the precursor is a newly-made RNA species that cannot be detected by techniques which identify steady-state RNA sequences. Thirdly, pulse-chase experiments suggest that the 14S precursor species is cleaved to polysomal globin mRNA.

7.2 THE SIZE OF GLOBIN mRNA SEQUENCES IN CHICKEN ERYTHROBLAST RNA

7.2.1 Hybridization with High Specific Radioactivity Globin cDNA

Newly-made globin mRNA sequences cannot be distinguished from steady-state globin mRNA in RNA-excess hybridization reactions. Nevertheless, this approach has one major advantage - it is sufficiently sensitive to detect globin RNA sequences present at levels as low as one molecule per cell (1-2 ng of globin RNA in 1 ml (5×10^9 cells) packed erythroblasts). In reconstruction experiments (see Section 6.3) keratin mRNA present at about this level was readily detected. Accordingly, RNA-excess hybridizations using high specific radioactivity globin cDNA, were used initially to probe for putative precursors to polysomal chicken globin mRNA.

Chicken erythroblast RNA was heated at 65°C for 10 minutes to ensure disaggregation of the RNA (Haines *et al.*, 1974) and centrifuged on sucrose gradients as shown in Fig. 7.1. The S value ranges of fractions "A" to "E", collected separately from this gradient, are listed in Table 7.1. Each of these size-classes was recentrifuged on a second set of sucrose gradients (Fig. 7.2 and 7.3). The second centrifugation was carried out in an attempt to enrich the higher molecular weight fractions for any putative globin mRNA precursor above the background of excess polysomal globin mRNA within the gradient. Fractions of 0.5 ml were collected separately from the second set of sucrose gradients, and the RNA from each fraction was hybridized to highly labelled globin cDNA. By hybridizing fractions from the recentrifugation of region "E", for a time sufficient to hybridize globin RNA sequences present

Figure 7.1

Isolation of Defined S Value Fractions of

Chicken Erythroblast RNA

Purified chicken erythroblast RNA was centrifuged on 10% - 40% sucrose gradients at 30,000 r.p.m. for 16 hr at 4° C. Defined S value fractions, as listed in Table 7.1, were collected and ethanol precipitated.

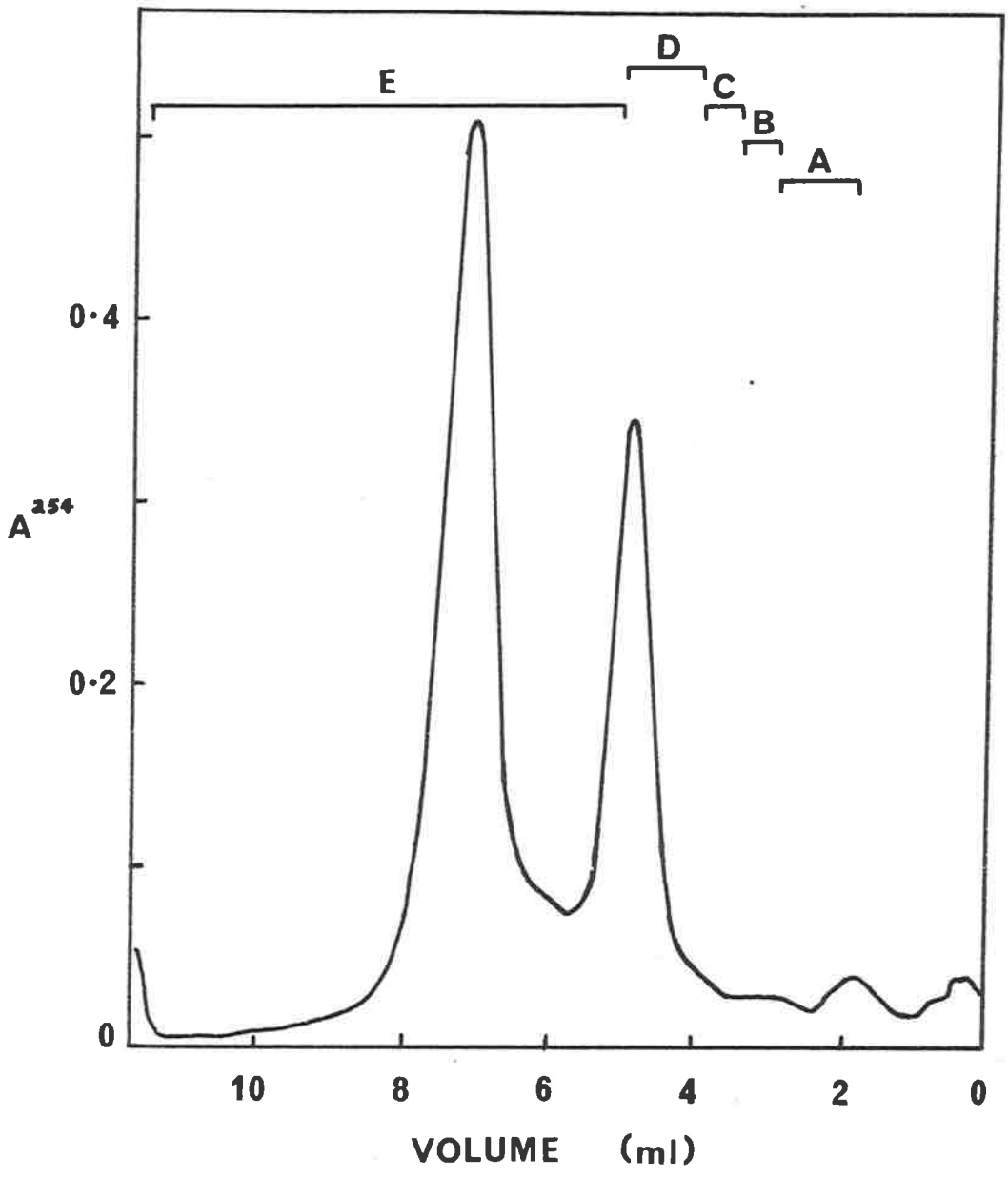

E
D
C
B
A
0·4
0·2
0
A254
10
8
6
4
2
0
VOLUME (ml)

Table 7.1

S Values of Chicken Erythroblast RNA Fractions from Fig. 7.1

fraction	S value range
A	5 - 10
B	10 - 12
C	12 - 15
D	15 - 20
E	20 - 50

at the level of one molecule per cell, it was possible to probe for globin sequences in the size range 20S to 50S. As shown in Fig. 7.2, no globin sequences between 20S and 50S in size were detected. Globin sequences were found only in the 10S region after the second centrifugation, and were therefore originally present at low levels in region "E" as a result of diffusion.

The conclusion from this experiment was that no globin mRNA precursor sedimenting from 20S to 50S could be detected at the one molecule per cell level, within chicken erythroblast RNA. Accordingly, no further attempt was made to locate globin RNA sequences within this size range.

The results of hybridizing fractions from the second centrifugation of regions "B", "C" and "D" (Fig. 7.3) were less clear cut. In each case peaks of hybridization were apparent in the 10S region only, and no enrichment for a larger molecular weight precursor was detected in the 10S to 20S region (Fig. 7.3).

Even when the "heavy" side of the 10S peaks in fractions "C" and "D" were selected from the second gradient, centrifuged for a third time and hybridized, no enrichment for 10S to 20S precursor globin RNA sequences was found (Fig. 7.4).

The results of these hybridizations in the 11S to 20S range were inconclusive. Although a precursor to polysomal globin mRNA was not detected in these gradients, hybridizing RNA was broadly distributed over the 10S to 15S region. Hence even if a globin mRNA precursor of this size was present, its detection may have been masked by excess polysomal globin mRNA. Rather than demonstrating

Figure 7.2

RNA Excess Hybridization of 20S - 50S Chicken Erythroblast RNA with Chicken Globin RNA

Fraction "E" (Fig. 7.1) was heated at 65° C for 10 minutes and centrifuged on 10% - 40% sucrose gradients. 0.5 ml fractions were collected and hybridized with high specific radioactivity chicken globin cDNA, for a time sufficient to detect 1 ng globin mRNA.

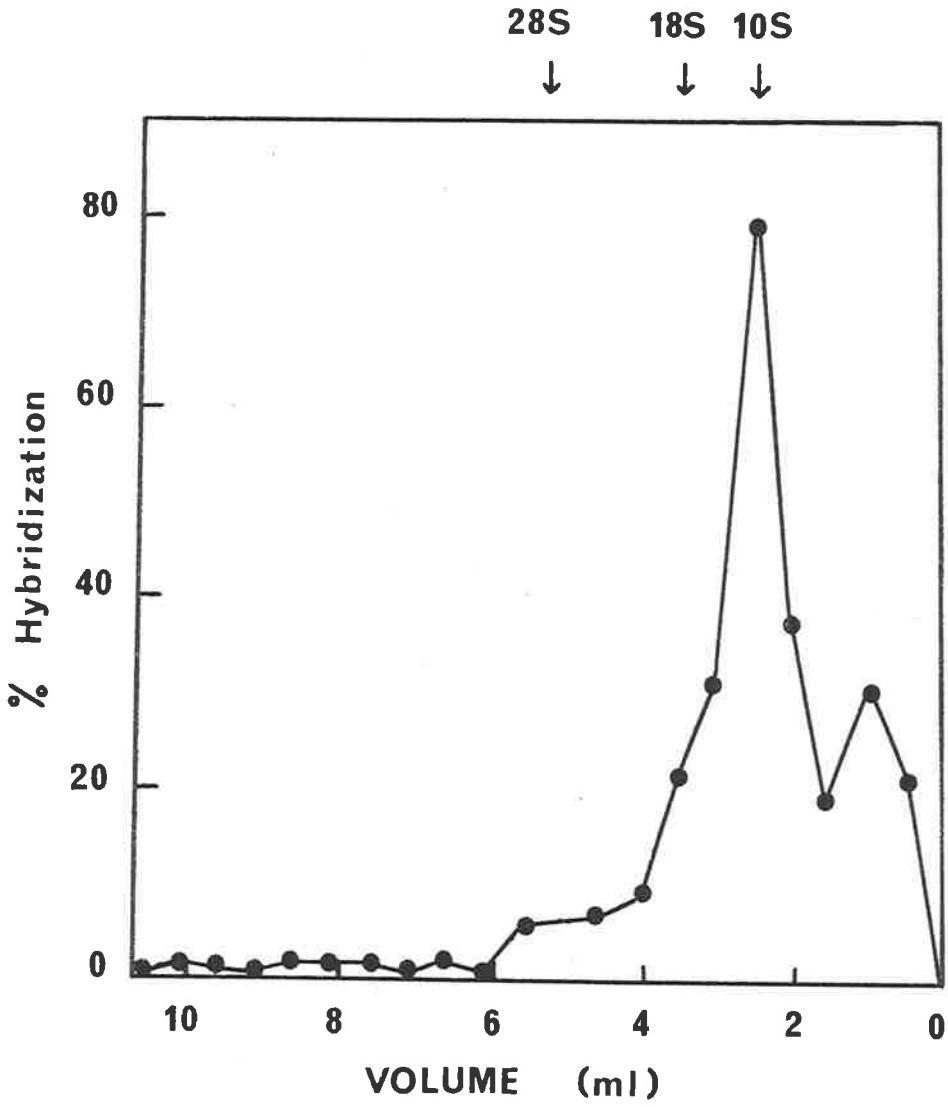

28S
18S
10S
80
60
40
20
0
% Hybridization
10
8
6
4
2
0
VOLUME (ml)

Figure 7.3

RNA Excess Hybridization of 10S - 20S Chicken Erythroblast With Chicken Globin cDNA

Fractions "B", "C" and "D" (Fig. 7.1) were heated at 65° C for 10 minutes, and recentrifuged on separate 5% - 20% sucrose gradients at 38,000 r.p.m. for 16 hr at 4° C. 0.5 ml fractions were collected and hybridized with high specific radioactivity chicken globin cDNA for times sufficient to detect 1 ng of globin mRNA.

The "heavy" side of the globin RNA peaks in gradients "C" and "D" were isolated as shown by brackets, and recentrifuged as shown in Fig. 7.4.

B fraction "B"

C fraction "C"

D fraction "D"

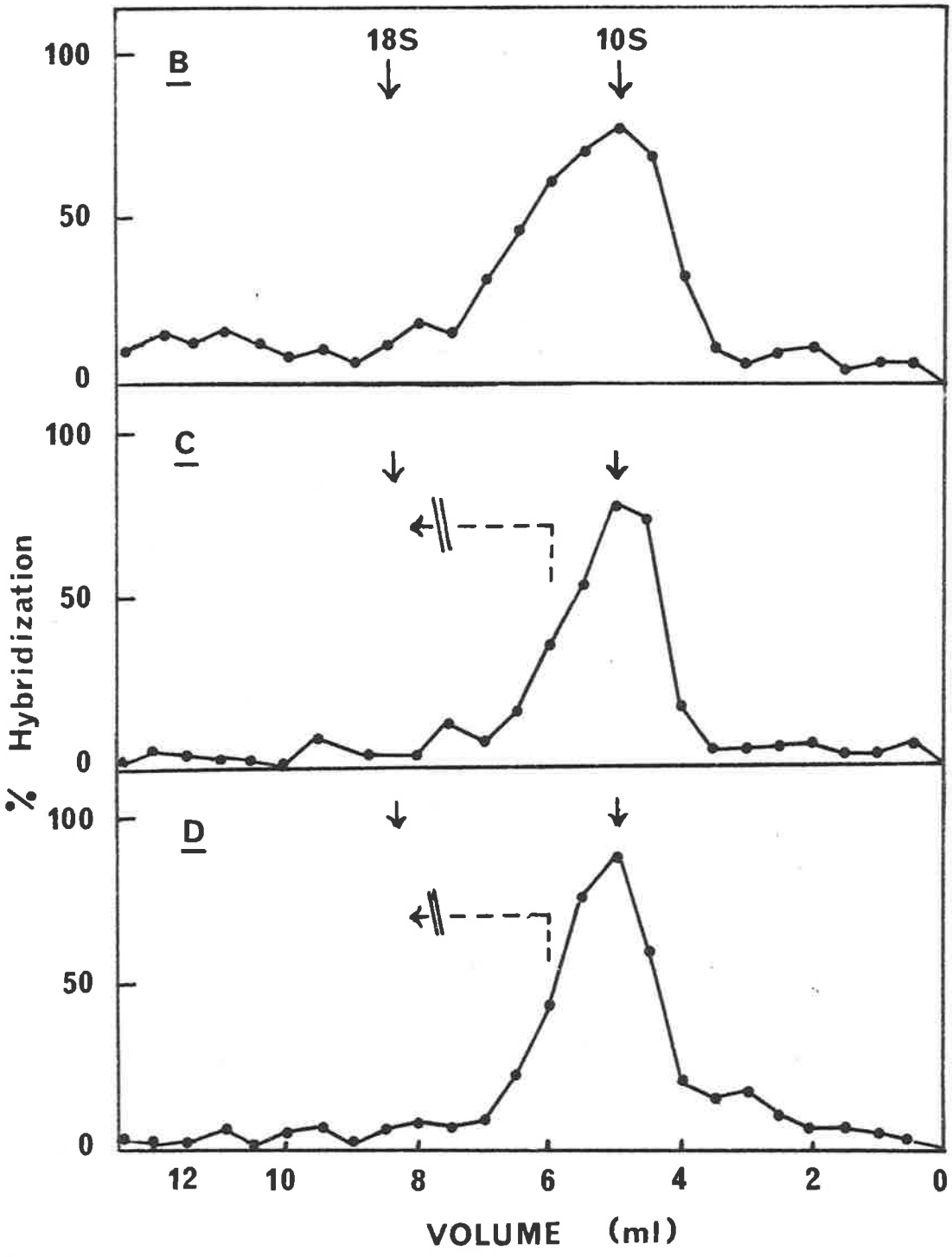

B
18S
10S
100
50
0
C
100
50
0
D
100
50
0
% Hybridization
12
10
8
6
4
2
0
VOLUME (ml)

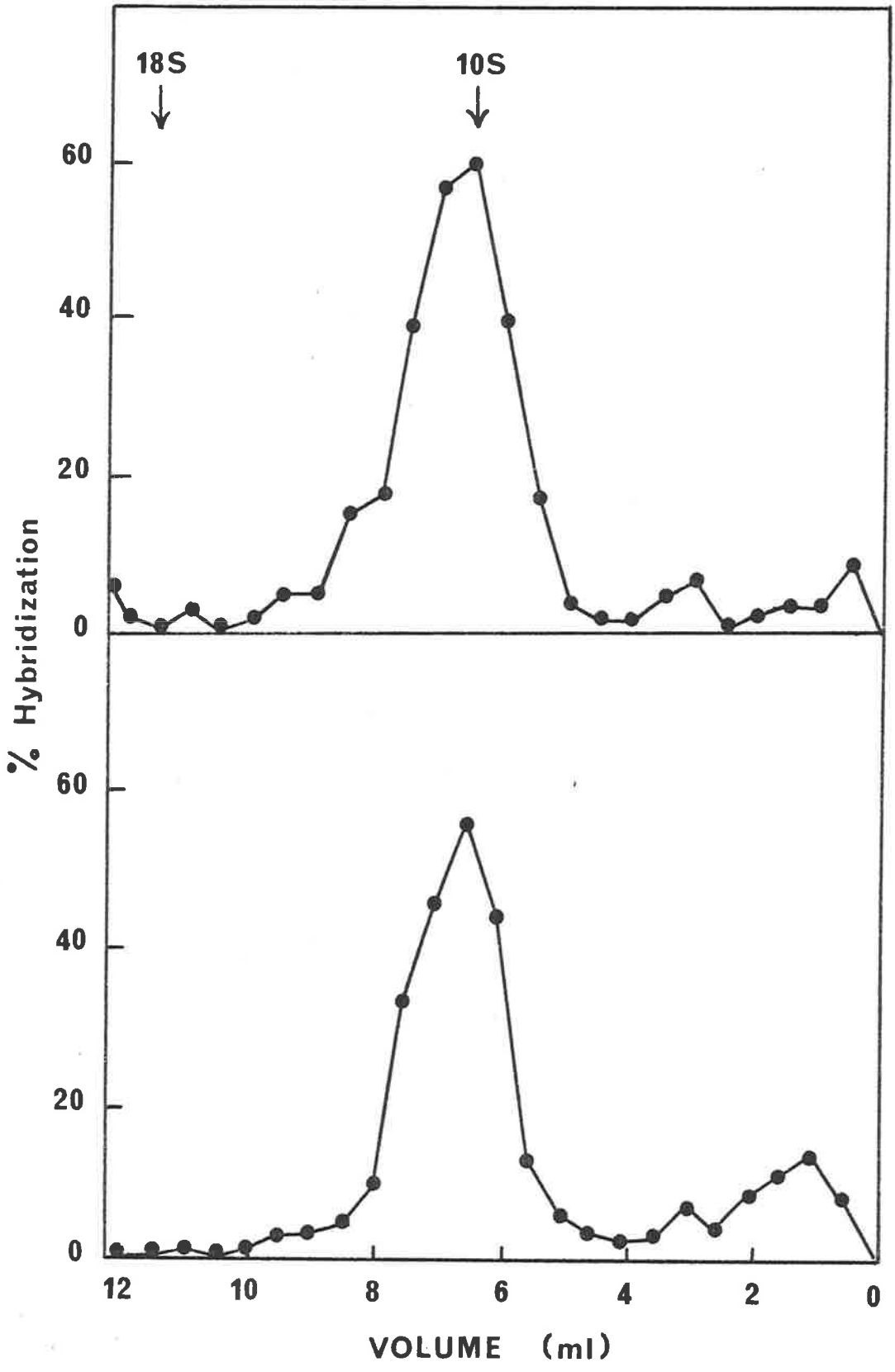
Figure 7.4

Chicken Globin cDNA Hybridization to 12S - 20S Chicken Erythroblast RNA, Further Fractionated by Sucrose Gradient Centrifugation

RNA from the "heavy" sides of the globin RNA peaks in gradients "C" and "D" (Fig. 7.3) was isolated and recentrifuged on 5% - 20% sucrose gradients. Hybridization was carried out as described in Fig. 7.3.

Top panel: Hybridization of RNA from "C"

Bottom panel: Hybridization of RNA from "D"


18S
10S
60
40
20
0
% Hybridization
60
40
20
0
12
10
8
6
4
2
0
VOLUME (ml)

the absence of an 11S to 20S globin RNA precursor, these experiments have illustrated the difficulties in attempting to locate a precursor that is only marginally larger than polysomal globin mRNA, when total erythroblast RNA and RNA-excess hybridizations are used. To overcome these problems, cDNA-excess hybridizations were carried out in order to test for the presence of a globin mRNA precursor sedimenting between 11S and 20S.

7.2.2 Globin cDNA-excess Hybridization with Avian Erythroblast RNA Labelled *in vivo*

When erythroblasts are incubated with a short pulse of ^{3}H-uridine, the newly-synthesized RNA is labelled exclusively. It follows that the amount of ^{3}H-RNA in hybrid form after hybridization of the erythroblast RNA with excess unlabelled globin cDNA, is a relative measure of newly-synthesized globin mRNA. The vast majority of polysomal mRNA is already present in the cells prior to pulse labelling, and is therefore unlabelled, and not detected by this technique. In this way, the presence of low levels of newly-synthesized RNA is not masked by a vast excess of steady-state RNA.

Accordingly, erythroblasts were incubated with ^{3}H-uridine for 5 minutes, and the RNA isolated from these cells was sedimented on sucrose gradients. Fractions were collected drop-wise from the bottom of the tube to ensure maximum resolution of different size-class globin RNA species in the 10S to 20S size range. In this way, fractions containing possible higher-molecular weight precursors to globin mRNA are collected before the polysomal globin mRNA, so that contamination of higher molecular weight fractions with 10S

polysomal RNA is minimised.

Fractions sedimenting between 5S and 22S were hybridized separately with 0.5 - 1.0 µg of globin cDNA. Control incubations containing no globin cDNA were also carried out. The level of RNA'ase resistant ^{3}H c.p.m. in both the cDNA and control hybridizations was determined by digesting with RNA'ase, under conditions in which RNA in hybrid form is undigested (Ross, 1976). The results of these assays are shown in Fig. 7.5.

Ideally, the difference between the cDNA and control plots, directly measures the number of ^{3}H c.p.m. in globin RNA within each fraction. However, the number of background resistant ^{3}H c.p.m. in larger molecular weight fractions is greater than those in the 7S to 15S fractions (Fig. 7.5). Hence experimental errors can cause large absolute ^{3}H c.p.m. differences between cDNA and control assays of larger molecular weight fractions, although the relative difference might be small. Accordingly, the number of ^{3}H c.p.m. in each cDNA assay was calculated relative to a constant background, in an attempt to normalise the influence of experimental error. The background resistant ^{3}H c.p.m. in the 10S fraction was chosen as the constant background value, to ensure that the absolute number of ^{3}H c.p.m. in globin RNA within each fraction could be estimated. The difference between the calculated number of ^{3}H c.p.m. in the cDNA assay, and the constant background value, indicated the level of ^{3}H incorporation into globin RNA.

The data from Fig. 7.5 was plotted in this manner, and the

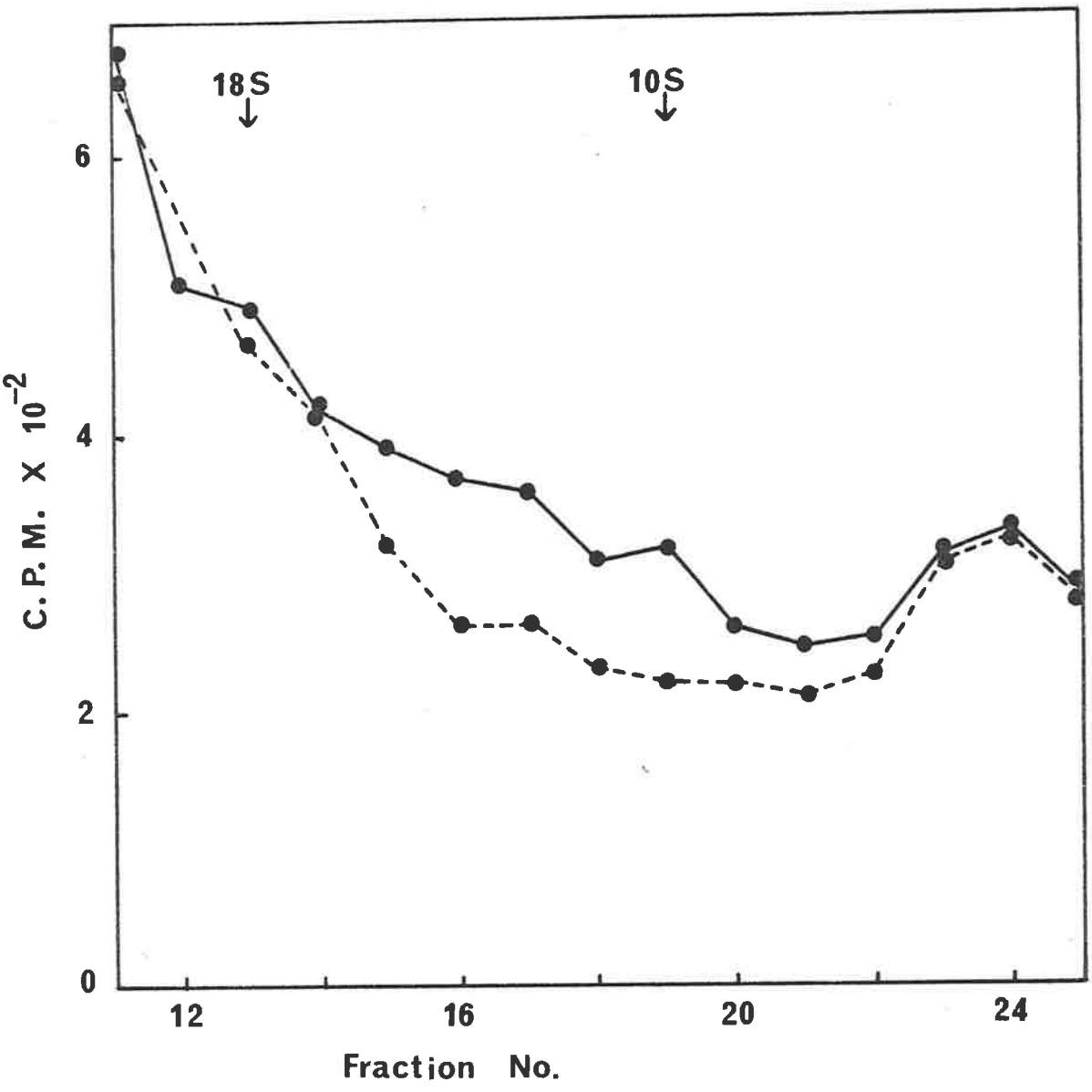
Figure 7.5

Globin cDNA Excess Hybridization to Pulse Labelled Chicken Erythroblast RNA

Chicken erythroblasts (0.25 ml packed cells) were pulse-labelled with 4 mCi ^{3}H-uridine in 4 mls Eagles medium for 5 minutes. RNA was isolated from the washed cells and centrifuged on 10% - 40% sucrose gradients at 38,000 r.p.m. for 16 hr at 4° C. 0.4 ml fractions were collected drop-wise from the bottom of the gradient, and the 18S and 28S ribosomal RNA located by measuring the number of A^{260} units in each fraction. Fractions that contained RNA sedimenting between 5S and 22S were ethanol precipitated, and hybridized separately to 0.5 µg chicken globin cDNA in a 70 µl hybridization volume. Control hybridizations contained no chicken globin cDNA. Labelled RNA sequences hybridizing to globin cDNA were detected by RNA'ase assay (Ross, 1976).

RNA'ase resistant ^{3}H c.p.m. in presence of cDNA •————•

RNA'ase resistant ^{3}H c.p.m. in absence of cDNA o--------o


18S
10S
6
4
2
0
C.P.M. X 10^{-2}
12
16
20
24
Fraction No.

distribution of ^{3}H-labelled globin RNA sequences is shown in Fig. 7.6. After incubating erythroblasts for 5 minutes with ^{3}H-uridine, about 40% - 45% of the ^{3}H-labelled globin RNA sequences sedimented as a 14S species, while 60% sedimented at 10S.

Separate hyrbidizations, using the same RNA fractions and high specific radioactivity ^{32}P globin cDNA, were carried out to determine the distribution of steady-state globin mRNA within this gradient. As shown in Fig. 7.6, the distribution of steady-state globin mRNA within this gradient, is centered around 10S.

Hence two types of hybridizations have demonstrated that two separate globin RNA species are apparent in erythroblast RNA after sucrose gradient centrifugation. One of these species sediments at the same S value (10S) as polysomal globin mRNA. However, the properties of the 14S globin RNA species partly satisfy the criteria for a higher molecular weight precursor to polysomal globin mRNA. Firstly, it sediments at a higher molecular weight S value than polysomal globin mRNA, and secondly, it is a newly-synthesized RNA species.

Although these experiments have detected globin mRNA sequences in two separate size-classes of RNA, they have not shown that the globin RNA sequences in the 14S sepcies are processed to 10S polysomal length globin mRNA. In other words, they have not demonstrated that the 14S RNA species is a functional precursor to polysomal globin mRNA.

Figure 7.6

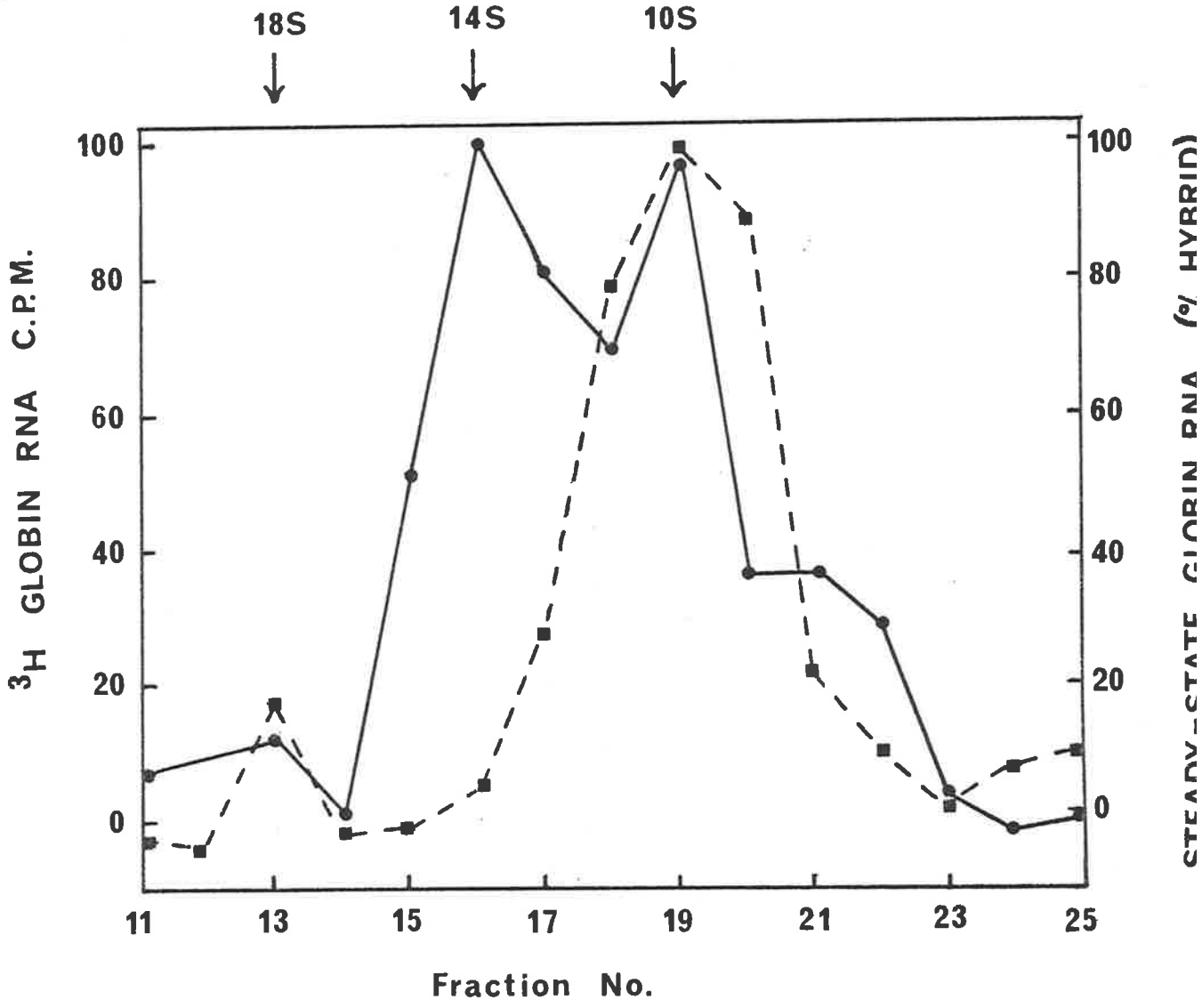
Distribution of ^{3}H-Labelled Globin RNA, and Steady-State Globin mRNA in Pulse-Labelled Chicken Erythroblase RNA

From the data plotted in Fig. 7.5, the number of ^{3}H c.p.m. in globin RNA was calculated for each fraction. The background resistant ^{3}H c.p.m. in the 10S region (Fraction 19) was used as a constant background value, and the number of RNA'ase resistant ^{3}H c.p.m. in (+ cDNA) hybridizations was calculated relative to this value. The difference between the calculated number of (+ cDNA) resistant ^{3}H c.p.m., and the constant background value, directly indicated the number of ^{3}H c.p.m. in globin RNA within each fraction.

The distribution of steady-state globin mRNA within this gradient was determined by hybridizing aliquots from each fraction with high specifi radioactivity ^{32}P globin cDNA. The percentage hybridization was determinec by S1 nuclease assay.

^{3}H labelled globin RNA ●——————●

steady-state globin RNA ■----------■


18S
14S
10S
100
80
60
40
20
0
11
13
15
17
19
21
23
25
Fraction No.
^{3}H GLOBIN RNA C.P.M.
STEADY-STATE GLOBIN RNA (% HYRRID)

To determine whether the 14S globin RNA satisfies the criteria of a functional precursor, pulse-chase experiments were carried out. Erythroblasts were pulsed for 3 minutes with ^{3}H-uridine, and RNA was isolated immediately from half the cells in the incubation. The ^{3}H-uridine in the remaining half of the incubation was diluted out with a 100 fold excess of unlabelled uridine, to minimise further incorporation of ^{3}H into erythroblast RNA. The erythroblasts in this half of the medium were incubated for a further 30 minutes, prior to RNA isolation.

RNA from both pulsed cells, and pulse-chased cells was centrifuged on sucrose gradients, and separate size-class fractions were hybridized with excess unlabelled globin cDNA to locate ^{3}H-labelled globin RNA sequences. As shown in Fig. 7.7, RNA isolated from pulsed cells, contained about equal amounts of 14S and 10S labelled globin RNA, whereas the distribution of labelled globin RNA in pulse-chased RNA was centered around 10S.

If the 14S globin RNA is a functional precursor to polysomal mRNA, and not simply turned-over rapidly during the chase incubations, then the ^{3}H c.p.m. in 14S globin RNA on the pulse gradient, would sediment at 10S after the chase period. In other words, the sum of the ^{3}H c.p.m. in 14S globin RNA and 10S globin RNA on the pulse gradient, would be equivalent to the number of ^{3}H c.p.m. in 10S polysomal globin RNA after the chase incubation. As shown in Table 7.2, there were equal numbers of ^{3}H c.p.m. in globin RNA in the pulsed RNA and pulse-chased RNA preparations - no nett loss of ^{3}H c.p.m. in globin

Figure 7.7

Globin cDNA Excess Hybridization to Chicken Erythroblast RNA Pulsed, and Pulse-Chased with ^{3}H-Uridine

Erythroblasts (0.25 ml packed cells) were pulsed for 3 minutes with 4 mCi ^{3}H-uridine. RNA was isolated immediately from half these cells. To the remainder of the cells, a 100 fold excess of unlabelled uridine was added, and incubation carried on for a further 30 minutes prior to RNA isolation.

Labelled globin RNA sequences were detected in both pulsed, and pulse-chased RNA as described in Figs. 7.5 and 7.6.

a) pulse-chased erythroblast RNA

b) pulse labelled erythroblast RNA

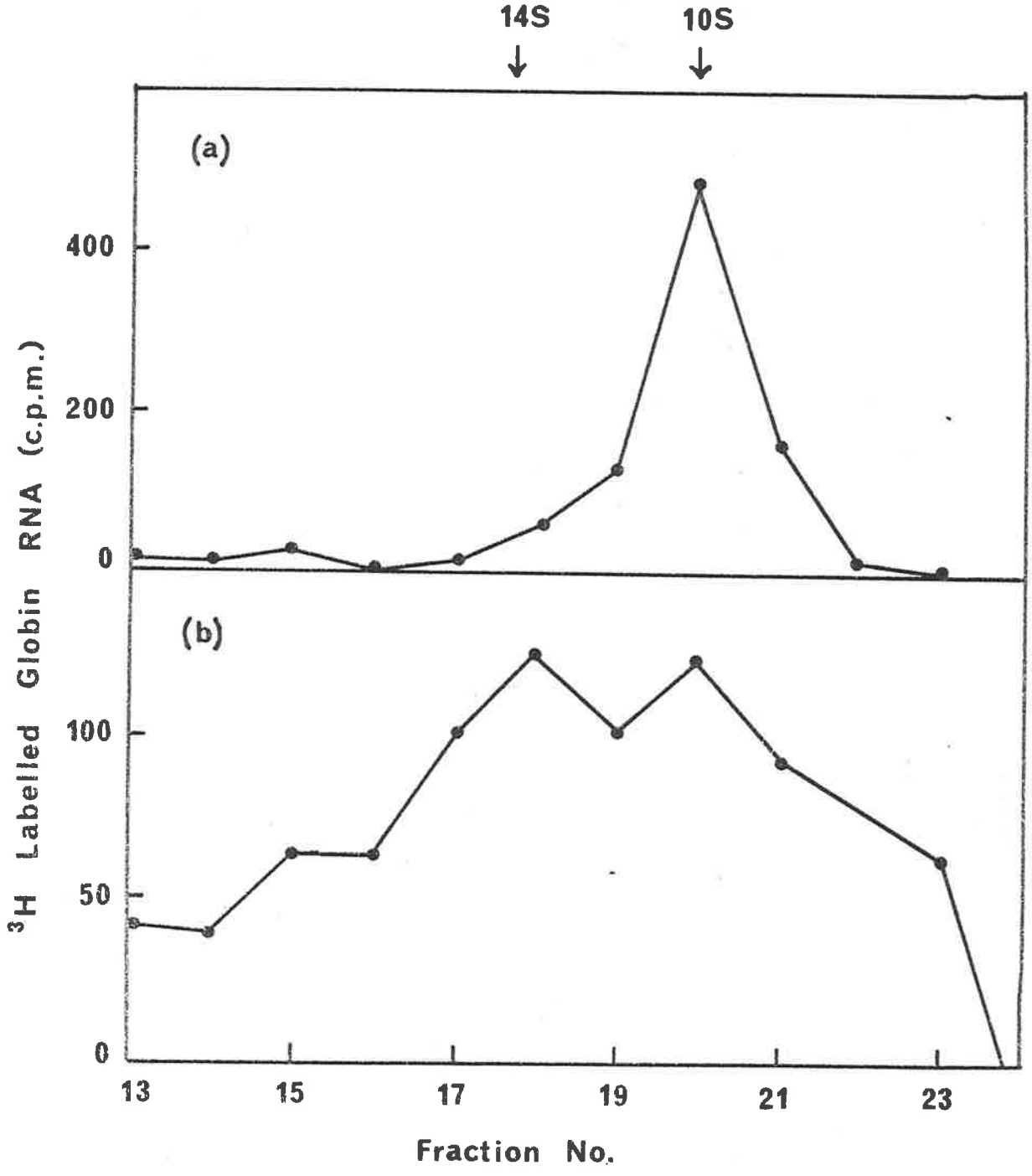

14S
10S
(a)
400
200
0
(b)
100
50
0
^{3}H Labelled Globin RNA (c.p.m.)
13
15
17
19
21
23
Fraction No.

Table 7.2

^{3}H c.p.m. Incorporated into Globin RNA After Pulse and Pulse-Chase Incubation of Erythroblasts

	^{3}H c.p.m. in globin RNA
pulse erythroblast RNA	810
pulse-chased erythroblast RNA	836

The number of ^{3}H c.p.m. in globin RNA in pulsed, and pulse-chased erythroblast RNA was determined from the data plotted in Fig. 7.7.

RNA had occurred during the chase incubation. On this basis, the 14S globin RNA species can be defined as a functional precursor to 10S polysomal globin mRNA.

As well as detecting a 14S precursor to polysomal globin mRNA, these experiments have enabled the half-life of the precursor to be estimated. After pulsing erythroblasts with ^{3}H-uridine for 3 minutes, 50% to 60% of the labelled globin RNA sedimented at 14S (Fig. 7.7), whereas after 5 minutes, about 40% was detected in the 14S species (Fig. 7.6). Hence the half-life of the precursor is about 3 to 4 minutes. This estimate for the half-life of the precursor is the maximum possible time taken for half the precursor population to be processed into 10S polysomal RNA, since it makes no allowance for the time taken for ^{3}H-uridine uptake into erythroblasts, and its conversion to ^{3}H-UTP.

7.3 DISCUSSION

This chapter has described two different hybridization approaches that were used to probe for a higher molecular weight precursor to polysomal globin mRNA. RNA excess hybridizations, using trace amounts of globin cDNA labelled to high specific radioactivity, failed to detect any globin mRNA precursor that was larger than 20S. Because of the sensitivity of this hybridization, where globin RNA sequences present at about the level of one molecule per cell would be readily detected, this result was interpreted as showing that no precursor to globin mRNA, larger than 20S was present in isolated erythroblast RNA. Accordingly, no further hybridizations were carried out with erythroblast RNA larger than 20S.

The results of RNA excess hybridizations, using RNA that sedimented between 10S and 20S were not conclusive. Fractions in this size range were repeatedly sedimented on sucrose gradients in an attempt to enrich the fractions for a putative globin mRNA precursor. However, RNA excess hybridizations did not detect a discrete globin RNA species that sedimented at a value greater than 10S. Instead, globin RNA appeared to be broadly distributed from 10S to 15S, suggesting that attempts to selectively reduce the steady-state globin RNA content of 11S to 15S fractions were ineffective. It was therefore probable that a precursor within this size range, would not be present in sufficient amounts to contribute significantly to this distribution. Hence any precursor may have been hidden under a sizeable excess of steady-state globin mRNA when this hybridization approach was used.

Accordingly, an alternative hybridization approach, using excess globin cDNA, was used to look for a precursor to globin mRNA in the 10S to 20S size range. The nature of this type of hybridization, where only the labelled (newly-made) globin RNA sequences are detected, meant that the detection of a larger precursor should not be masked by the presence of a vast excess of steady-state (pre-formed) globin mRNA. Using this hybridization approach, a 14S globin RNA species was detected within erythroblast RNA. This RNA class had properties that satisfied the criteria for a functional precursor to polysomal globin mRNA. Firstly, it was larger in molecular weight than 10S polysomal globin mRNA. The distribution of steady-state globin mRNA within this RNA preparation was essentially at 10S, indicating that the 14S species was genuinely higher in molecular weight. General globin RNA aggregation is not the reason that this putative

precursor sediments at 14S, since globin sequences generally, are distributed around 10S. The 14S species, which is prominent in the distribution of the ^{3}H-labelled globin RNA population, does not contribute to this overall distribution. Furthermore, data from the pulse-chase experiments is consistent with the view that this 14S species is not a specific aggregation of newly made globin RNA.

Secondly, it is a newly-made RNA species. Thirdly, pulse-chase experiments showed that the 14S globin RNA was processed into 10S polysomal globin mRNA. The rate at which the labelled 14S globin RNA is processed into 10S polysomal mRNA was also measured in 5 minute and 3 minute ^{3}H-uridine pulse experiments. The precursor appears to be relatively short-lived, having a half-life of 3 or 4 minutes.

CHAPTER 8.

FINAL DISCUSSION

CHAPTER 8. FINAL DISCUSSION

8.1 INTRODUCTION

This thesis has made two principal points. Firstly it has described work that led to an estimation of the length of the 3' non-coding region (adjacent to the poly(A)) of chicken globin mRNA. Based on this estimation, a model for the organisation of sequences in chicken globin mRNA was proposed.

The other major point that this thesis has established is that polysomal chicken globin mRNA is derived from a 14S RNA precursor. The half-life of this precursor has also been estimated.

In this final discussion, these two principal findings are related to work already published on mRNA structure and biogenesis, and future studies are also suggested.

8.2 POLYSOMAL CHICKEN GLOBIN mRNA

Data from the heterologous hybridizations described in Chapter 5 was used to estimate that 90 nucleotides is the minimum nucleotide length of the 3' non-coding region of chicken globin mRNA. This estimate for the number of nucleotides in the 3' non-coding region of chicken globin mRNA is almost identical to that of rabbit globin mRNAs, as determined by nucleotide sequencing data. The 3' non-coding regions of rabbit α and β globin mRNAs contain 89 ± 2 and 95 ± 2 nucleotides respectively (Proudfoot, 1976).

It is likely that the different functional regions within globin

mRNAs from various species, are similar in length. All globin mRNAs so far studied contain a similar number of total nucleotides (Lewin, 1975b), and the same number of coding nucleotides (Croft, 1973). Hence the close agreement between the heterologous hybridization data described here, and the sequencing data for rabbit globin mRNA, supports the validity of the heterologous hybridization approach, at least in this case.

Despite the similarities that exist between different globin mRNAs, it is now clear that the 3' non-coding regions of globin mRNAs from different species, need not be identical in length. Whereas the 3' non-coding regions of rabbit α and β globin mRNA contain 89 ± 2, and 95 ± 2 nucleotides respectively, these regions in human α and β globin mRNAs consist of about 112 and at least 109 nucleotides respectively (Proudfoot and Longley, 1976).

The functions of the 3' non-coding region of eukaryote mRNAs are unknown. It may be involved in the termination of both transcription and translation, or serve as a recognition site for poly(A) polymerase. These nucleotides may also play a part in other processes, such as mRNA stability, the derivation of polysomal mRNA from higher molecular weight precursors, or the packaging of mRNA into ribonuclear protein particles.

It is possible that the two nucleotides immediately adjacent to the poly(A) sequence of several mRNA species, are involved in poly(A) polymerase recognition, although poly(A) polymerase from Hela cells at least, does not seem to recognise a specific 3' sequence *in vitro* (Brown and Sugimoto, 1973). This sequence in chicken globin mRNA appears to be ... G.C. poly(A), and is identical to that in rabbit and human globin mRNAs,

and chicken ovalbumin mRNA (Proudfoot and Brownlee, 1976). This sequence is not universal however. By studying a mixture of Hela cell mRNAs, Nichols and Eiden (1974) demonstrated some heterogeneity in the nucleotide sequences immediately adjacent to poly(A). More specifically, this sequence in light chain immunoglobulin mRNA is ... U.G. poly(A) (Proudfoot and Brownlee, 1976), so that it also differs in this region from those mRNAs discussed above.

Sequence data published since the work in this thesis was completed, by Proudfoot and Brownlee (1976), Proudfoot and Longley (1976), and Cheng *et al.*, (1976) has allowed the evolutionary stability of this region within eukaryote mRNAs to be examined, to perhaps give some clue to its biological role. In particular, all eukaryote mRNAs so far investigated, contain the hexanucleotide sequence AAUAAA positioned about 15 to 25 residues in from the poly(A) tract. This sequence has been found in the 3' end of mRNAs as diverse as early mRNAs from SV40 (Dhar *et al.*, 1974). Hence the role of these six nucleotides, although unknown, appears to depend rigidly on this precise sequence, and its function appears to be essential for the overall biological activity of the mRNA. It is probable therefore, that this sequence is contained within chicken globin mRNAs, and that it is a general feature of poly(A)-containing eukaryote mRNAs at least. Whether it is contained within non poly(A) mRNAs, such as histone mRNAs, has yet to be determined.

It is not clear whether the evolutionary stability of the remainder of the 3' non-coding regions of eukaryote mRNAs is essential to its overall function. Although sequence data is incomplete, a large number of the

nucleotides within this region of rabbit and human α globin mRNAs, and rabbit and human β globin mRNAs appear to be homologous. On average about 83% of the sequences so far determined in the 3' untranslated end of rabbit and human β globin mRNAs, and rabbit and human α globin mRNAs, have been conserved (Proudfoot and Longley, 1976) suggesting that the function of this region depends on a major portion of this sequence being maintained. Other data however, suggests that this need not be so. There is little nucleotide homology evident in this region when rabbit α and β globin mRNAs are compared, yet presumably the role of these nucleotides is similar in both mRNAs. Similarly, the 3' untranslated regions of mRNAs from other tissues show very little homology with similar regions of either rabbit or human globin mRNAs. For example, when the sequence data of rabbit α or β globin mRNA is compared to that of chicken ovalbumin mRNA, a much lower level of nucleotide homology is apparent in the 70 nucleotides adjacent to poly(A). Rabbit β globin mRNA contains no more than 30% homology with this region of ovalbumin mRNA, whereas that within rabbit α globin mRNA is no more than 10%.

This type of comparison may be more relevant in attempting to assess the degree of nucleotide homology needed to retain the overall function of this region, than the comparison between α globin mRNAs from different species. An even more useful comparison might be to compare mRNAs from different tissues of the same species. For example, it may be useful to compare the 3' non-coding regions of chicken globin mRNA with ovalbumin mRNA. The result of such a comparison may be predicted. By analogy with rabbit globin mRNAs, it is possible that chicken α and β globin mRNAs contain little

sequence homology in their 3' non-coding regions, so that they could not both share a large number of homologous nucleotides with the 3' untranslated region of ovalbumin mRNA.

The overall conclusion from this discussion is that strict overall nucleotide homology within the 3' non-coding region of eukaryote mRNAs is not essential to the biological function of the mRNAs. Assuming that the role of this region is similar in all eukaryote mRNAs, this function may be performed by regions that vary in their nucleotide sequence, with the probable exception of the hexanucleotide fragment referred to previously.

The results of the work described in this thesis, suggest that there is greater nucleotide divergence in the 3' non-coding regions of rabbit and chicken globin mRNAs, than the coding region. This is compatible with the view that strict sequence conservation is not essential to maintain the function of the 3' untranslated region, although it is in contrast to the data on rabbit and human globin mRNAs, where an average of about 83% of the nucleotides in this region are homologous. However, the evolutionary distance between birds and mammals is much larger than that between two separate mammalian species. Hence it might be expected that the degree of homology in the 3' region of rabbit and chicken globin mRNAs is not as great as that in rabbit and human globin mRNAs. It is also relevant that, while the experiments described in this thesis might superficially indicate that there is little homology in the 3' untranslated region of rabbit and chicken globin mRNAs, sequence data is required to establish this point. In the heterologous hybridizations, ^{3}H-dC labelled globin cDNA was assayed with S1 nuclease and the only homologous sequences that were scored as such,

were those that contain guanosine. Those homologous nucleotide lengths not containing guanosine would not have been detected as such by this assay. Secondly, the stability of the hybrids after S1 nuclease digestion was such that hybrids of less than 10 to 15 nucleotides in length were scored as non-homologous. It is therefore more accurate to state from these results, that the pattern of nucleotide sequence divergence within the 3' non-coding region of chicken and rabbit globin mRNAs, is different to that of the coding regions.

It has been proposed that the 3' non-coding region of different eukaryote mRNAs, might form hair-pin loop structures (Proudfoot and Brownlee, 1974). However, the stability of these structures depends on only a few nucleotide-pairs, and such structures may not be relevant *in vivo*. Furthermore, some mRNAs (e.g. rabbit α globin mRNA) recently sequenced do not appear to have the capacity to form these hair-pin structures.

It is apparent that sequence data alone is not sufficient to provide precise information on the role of the 3' non-coding regions of eukaryote mRNAs. The *in vivo* significance of specific nucleotide sequences may only become apparent when information from biologically-orientated experiments is also available. For example, it may be possible to determine whether this region of eukaryote mRNAs is involved in translation, by altering the nature of the 3' non-coding region, and testing the altered mRNA by using *in vitro* translation systems.

The model proposed for the organisation of sequences within chicken globin mRNA (Fig. 5.5) predicts that there is a non-coding region of about 40 to 70 nucleotides at the 5' end of globin mRNA. All globin mRNAs so far

investigated, contain a 5' terminal cap with the structure m^7 Gppp X (Shatkin, 1976). Hence chicken globin mRNA is also likely to contain an identical 5' terminal cap.

It is likely that this 5' non-coding region includes a ribosome binding sequence. Ribosome binding sites have been demonstrated at the 5' end of prokaryote mRNAs, where a short sequence of mRNA nucleotides may base-pair with the 3' sequence of 16S ribosomal RNA (Shine and Dalgarno, 1974). A similar complementarity between the 5' end of rabbit globin mRNAs, and the 3' end rabbit 18S ribosomal RNA has recently been demonstrated (Baralle, 1977). Hence this mechanism of ribosome binding to mRNAs may be universal.

8.3 DETECTION OF A 14S PRECURSOR TO POLYSOMAL CHICKEN GLOBIN mRNA

Part of the work described in this thesis, was aimed at determining whether polysomal chicken globin mRNA is derived from a higher molecular weight precursor. Firstly, total erythroblast RNA was isolated, and the precursor to 18S and 28S ribosomal RNAs was detected as a discrete 40S - 45S RNA species, indicating that undegraded nuclear RNA was present in the isolated RNA.

Hybridizations with globin cDNA detected a precursor to globin mRNA, sedimenting as a 14S RNA species in aqueous sucrose gradients. The properties of this 14S species satisfied three criteria of a larger precursor to polysomal globin mRNA.

1) it is newly synthesized

2) it is higher in molecular weight

3) the globin coding nucleotides in the 14S species, can be chased into 10S polysomal globin mRNA.

The other significant property of this precursor, is that its half-life is 3 to 4 minutes. In comparison, the half-life of the 14S precursor to mouse foetal globin mRNA is about 45 minutes (Ross, 1976) and that of the 15S precursor found in induced Friend cells appears to be about 8 to 10 minutes (Curtis and Weissman, 1976). Hence the processing time of the 14S chicken globin RNA precursor is faster than that for both these precursors to mouse globin mRNAs.

It is unlikely that the chicken globin RNA sediments at 14S because of aggregation. Firstly, the steady-state 10S polysomal globin mRNA, which serves as an ideal internal control, shows no evidence for aggregation (Fig. 7.6). Secondly, no 14S globin RNA species was detected after a 30 minute chase experiment (Fig. 7.7). Thirdly, heat treatment of the RNA sample prior to sucrose gradient centrifugation is sufficient to disrupt RNA aggregates (Haines *et al.*, 1974).

On this basis, the 14S value of this chicken globin mRNA species, reflects its true molecular integrity. This precursor is therefore about twice as large as polysomal globin mRNA.

The size of this precursor is approximately equivalent to the discrete globin mRNA precursors found in mouse erythroid cells (Ross, 1976; Curtis and Weissman, 1976). In each case the precursor sediments at about 14S to 15S relative to ribosomal RNAs in aqueous sucrose gradients. Hence each of these precursor species is about twice the molecular weight of polysomal globin mRNA.

A nuclear globin RNA isolated from duck erythroid cells, and

sedimenting at about 14S in sucrose gradients in 85% formamide was reported earlier (Macnaughton *et al.*, 1974). The methods of RNA preparation and detection were different to those employed successfully in this thesis. Erythroid cells from the circulation of anaemic ducks (defined in this thesis as reticulocyte cells) were used in those experiments as the source of RNA, whereas erythroblasts from the bone-marrow of anaemic chickens were used here. The method of RNA isolation was another major difference between the two approaches. Macnaughton *et al.* prepared nuclear RNA from isolated reticulocyte nuclei, whereas this thesis uses total RNA (nuclear RNA and cytoplasmic RNA) from erythroblasts. In our hands it is not possible to specifically isolate nuclear RNA in an undegraded form, from chicken erythroblast nuclei, although it may be possible to do so with reticulocytes, where the nuclease levels are relatively reduced.

Even though Macnaughton *et al.* isolated RNA from erythroid nuclei, it is surprising that no polysomal length (9.5S) globin mRNA was detected in the nuclear RNA preparation, since it was estimated that the larger globin RNA species was present at about 0.1% to 0.2% the level of cytoplasmic globin mRNA. This was despite the extreme sensitivity of the RNA-excess hybridization approach used, where globin RNA present in amounts as low as 0.02% of the total globin mRNA should be readily detected.

These differences apart, the putative precursor for duck globin mRNA sediments at about 14S in formamide gradients, which according to the authors, corresponds to a nucleotide length of about three times that of polysomal mRNA. If this putative precursor is equivalent to the 14S RNA detected in chicken erythroid cells, then this apparent difference in nucleotide length, may be due to a difference in the relative sedimentation

rate of ribosomal RNA markers in the two systems.

There are other reports that avian globin genes are transcribed into higher molecular weight RNA products (Imaizumi *et al.*, 1973; Spohr *et al.*, 1976). These reports can be criticized however, since discrete size-classes of globin RNA species were not detected. Instead, a gradation in the levels of globin RNA was detected, ranging from lowest amounts (about 0.01% of the total globin RNA) in the highest molecular weight fractions, up to high levels (about 95% - 99% of the total globin RNA) in the 10S fractions of the gels. Even though both sucrose gradient centrifugation and gel electrophoresis were carried out under denaturing conditions, diffusion could cause such a gradation in globin mRNA levels. Furthermore, if the globin gene primary transcript is much larger than 14S, and 14S is one of the processed products, then it should still be possible to detect the larger primary transcript, and any intermediate sized products as discrete RNA species.

It is not possible to say yet, whether the 14S globin RNA species is the primary product of the globin genes. If the globin genes are transcribed into an even higher molecular weight product, then the half-life of this RNA species must be extremely short. There are two possible ways that the half-life of such a primary transcript, might be prolonged sufficiently to permit its detection. By incubating erythroblasts at a lower temperature than chicken body temperature (41° C), the time taken to process any primary transcript may be prolonged. Alternatively, cordycepin (Penman *et al.*, 1970) may block the processing of the globin gene primary transcript, so that a build-up in this product

may occur within the cells. The detection of a globin RNA species larger than 14S, in RNA isolated from cells treated in this manner, would suggest that the 14S species was not a primary transcription product.

There are other questions concerning this 14S globin RNA, that might be readily answered. Firstly, it should be possible to determine whether it is polyadenylated, by using oligo(dT)-cellulose chromatography. It might also be possible to synthesize cDNA to this precursor by using polymerase I globin cDNA as the primer (see Section 4.5). The synthesis of a full length copy of this precursor would immediately show that the polysomal globin mRNA sequences are located at the 3' end of the precursor molecule. The availability of a cDNA to this RNA species would make it possible to clone the precursor sequence. It would then be possible to answer a wide range of questions relevant to eukaryote gene control. For example

1) are the sequences adjacent to the polysomal nucleotides repeated in the chicken genome?

2) are the extra polysomal sequences identical for chicken α and β globin 14S RNAs?

3) what is the nucleotide sequence of these extrapolysomal regions?

For example, if the sequences adjacent to the 5' end of the polysomal globin mRNA sequences were reiterated within the genome, and identical for both α and β globin genes, then the expression of the globin genes might be coordinated in a manner similar to that proposed by Davidson and Britten (1973).

BIBLIOGRAPHY

Adams, J. M. and Corey, S. (1975) *Nature* 255, 28.

Adesnik, M., Salditt, M., Thomas, W. and Darnell, J. E. (1972) *J. Mol. Biol.*, 71, 81.

Artman, M. and Roth, J. S. (1971) *J. Mol. Biol.*, 60, 291.

Axel, R., Feigelson, P. and Schutz, G. (1976) *Cell*, 7, 247.

Baralle, F. E., (1977) *Nature*, 267, 279.

Benyajati, C. and Worcel, A. (1976) *Cell*, 9, 393.

Both, G. W., Banerjee, A. K. and Shatkin, A. J. (1975) *Proc. Natl. Acad. Sci., U.S.A.* 72, 1189.

Brandhorst, B. P. and McConkey, E. H. (1974) *J. Mol. Biol.*, 85, 451.

Braunitzer, G., Best, J. S., Flamm, U. and Schrank, B. (1966) *Hoppe-Seyler's Z. Physiol. Chem.* 347, 207.

Brawerman, G., Mendecki, J. and Lee. Y., (1972) *Biochemistry*, 11, 637.

Britten, R. J. and Kohne, D. E. (1968) *Science*, 161, 529.

Britten, R. J. and Davidson, E. H. (1969) *Science*, 165, 349.

Brown, D. D. and Sugimoto, K. (1973) *J. Mol. Biol.* 78, 397.

Burdon, R. H., and Clason, A. E. (1969) *J. Mol. Biol.* 39, 113.

Burgess, R., Travers, A. A., Dunn, J. J. and Bautz, E. K. F. (1969) *Nature*, 222, 537.

Cech, T. R. and Hearst, J. E. (1975) *Cell*, 5, 429.

Chadwick, P., Pirotta, V., Steinberg, R., Hopkins, W. and Ptashne, M. (1970) *Cold Spring Harbor Symp. Quant. Biol.* 35, 283.

Chamberlin, M., McGrath, J. and Waskell, L. (1970) *Nature*, 228, 227.

Chambon, P., Gussinger, F., Kedinger, C., Mandel, J. L., Meillhac, M. and Huret, P. (1972) *'Gene Transcription in Reproductive Tissue'* (Ed. Diezfalusy) *5*, 222 Karolinska Institute, Stockholm.

Cheng, C. C., Brownlee, G. G., Carey, N. H., Doel, M. T., Gillam, S. and Smith, M. (1976) *J. Mol. Biol. 107*, 527.

Cohen, S. N. (1976) *Nature, 263*, 731.

Crick, F. (1976) *New Scientist, June 17th* p. 625.

Croft, L. R. (1973) *Handbook of Protein Sequences*, (Joynson-Bruvvers, Oxford).

Curtis, P. J. and Weissman, C. (1976) *J. Mol. Biol. 106*, 1061.

Darnell, J. E., Jelinek, W. R. and Molloy, G. R. (1973) *Science, 181*, 1215.

Dasgupta, R., Shih, D. S., Saris, C. and Kaesberg, P. (1975) *Nature, 256*, 624.

Davidson, E. H. and Britten, R. J. (1973) *Quart. Rev. Biol. 48*, 565.

Davidson, E. H., Hough, B. R., Klein, W. H. and Britten, R. J. (1975) *Cell, 4*, 217.

DeCrombrugghe, B., Chen, B., Anderson, W. B., Parkes, J., Nissley, S. J., Straub, M., Goltesman, M. E. and Perlman, R. L. (1971) *Nature New Biol. 231*, 139.

DeLange, R. J., Fambrough, D. M., Smith, E. L. and Bonner, J. (1969) *J. Biol. Chem., 244*, 5669.

Derman, E., and Darnell, J. E. (1974) *Cell, 3*, 255.

Derman, E., Goldberg, S. and Darnell, J. E. (1976) *Cell, 9*, 465.

Dhar, R., Weissman, S. M., Pan, J. and Subramanian, K. (1974) *Proc. Natl. Acad. Sci., U.S.A. 71*, 371.

Edmonds, M., Vaughan, M. H. and Nakazato, H. (1971) *Proc. Natl. Acad. Sci., U.S.A.* *68*, 1336.

Efstradiatis, A., Maniatis, T., Kafatos, F. C., Jeffrey, A. and Vournakis, J. N (1975) *Cell*, *4*, 367.

Egházi, E. (1976) *Cell*, *7*, 507.

Eron, L., Arditti, R., Zubray, G., Connaway, S. and Beckwith, J. R. (1971) *J. Mol. Biol.* *68*, 215.

Firtel, R. A. and Lodish, H. F. (1973) *J. Mol. Biol.* *79*, 295.

Flamm, U., Best, J. S. and Braunitzer, G. (1971) *Hoppe-Seyler's Z. Physiol. Chem.*, *352*, 885.

Furuichi, Y., LaFiandra, A. and Shatkin, A. J. (1977) *Nature*, *266*, 235.

Galau, G. A., Britten, R. J. and Davidson, E. H. (1974) *Cell*, *2*, 9.

Georgiev, G. P. and Mantieva, V. L. (1972) *Biochim, Biophys. Acta*, *61*, 153.

Getz, M. J., Birnie, G. D., Young, B. D., MacPhail, E. and Paul, J. (1975) *Cell*, *4*, 121.

Gilbert, W. and Müller-Hill, B. (1966) *Proc. Natl. Acad. Sci., U.S.A.* *58*, 1891.

Gilmour, R. S. and Paul, J. (1970) *FEBS Letts.* *9*, 242.

Gilmour, R. S. and Paul, J. (1973) *Proc. Natl. Acad. Sci., U.S.A.* *70*, 3440.

Gilmour, R. S., Windass, J. D., Affaras, N. and Paul, J. (1975) *J. Cell Physiol.*, *58*, 449.

Gross-Bellard, M., Oudet, M. and Chambon, P. (1973) *Eur. J. Biochem.*, *36*, 32.

Grouse, L., Chilton, M. D. and McCarthy, B. J. (1972) *Biochemistry*, *11*, 798.

Gurdon, J. B. (1968) *Essays in Biochemistry*, *4*, 26.

Gurdon, J. B. and Woodland, H. R. (1970) *Current Topics in Developmental Biology*, *5*, 39.

Haines, M. E., Carey, N. H. and Palmiter, R. D. (1974) *Eur. J. Biochem.*, *43*, 549.

Hamlyn, P. H. and Gould, H. J. (1975) *J. Mol. Biol.*, *94*, 101.

Harris, H. (1970) *'Nucleus and Cytoplasm'* Clarendon Press, Oxford.

Harris, S. E., Means, A. R., Mitchell, M. M. and O'Malley, B. W. (1973) *Proc. Natl. Acad. Sci., U.S.A.*, *70*, 3776.

Herman, R. C., Williams, J. G. and Penman, S. (1976) *Cell*, *7*, 429.

Hozumi, N. and Tonegawa, S. (1976) *Proc. Natl. Acad. Sci., U.S.A.*, *73*, 3628.

Huang, R. C. C. and Bonner, J. (1965) *Proc. Natl. Acad. Sci., U.S.A.*, *54*, 960.

Huez, G., Marbaix, G., Hubert, E., Leclercq, M., Nudel, U., Soreq, H., Salomon, R., Lebleu, B., Revel, M. and Littauer, U.Z. (1974) *Proc. Natl. Acad. Sci., U.S.A.*, *71*, 3143.

Ilan, J. and Ilan, J. (1971) *Dev. Biol.*, *25*, 280.

Imaizumi, T., Diggelmann, H. and Scherrer, K. (1973) *Proc. Natl. Acad. Sci., U.S.A.*, *70*, 1122.

Jacob, F. and Monod, J. (1961) *J. Mol. Biol.*, *3*, 318.

Kacian, D. L., Spiegelman, S., Bank, A., Terada, M., Metafora, S., Dow, L. and Marks, P. A. (1972) *Nature New Biol.*, *235*, 167.

Kacian, D. L. and Spiegelman, S. (1974) *Methods Enzumol.*, *39*, 150.

Kedes, L. H., Chang, A. C. Y., Houseman, D. and Cohen, S. N. (1975) *Nature*, *255*, 533.

Kemp, D. J. (1975) *Nature*, *254*, 573.

Kern, H. (1975) *Analty. Biochem.*, *67*, 147.

Leder, P., Ross, J., Gielen, J., Packman, S., Ikawa, Y., Aviv, J. and Swan, D. (1973) *Cold Spring Harbor Symp. Quant. Biol.*, *38*, 753.

Lee, Y., Mendecki, J. and Brawerman, G. (1971) *Proc. Natl. Acad. Sci., U.S.A.*, *68*, 1331.

Lewin, B. (1975a) *Cell*, *4*, 11.

Lewin, B. (1975b) *Cell*, *4*, 77.

Lizardi, P. M. (1976) *Cell*, *7*, 239.

Loeb, J. E. and Creuzet, C. (1969) *FEBS Letts.* *5*, 37.

Macnaughton, M., Freeman, K. B. and Bishop, J. O. (1974) *Cell*, *1*, 117.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975) *Proc. Natl. Acad. Sci., U.S.A.*, *72*, 1184.

Marbaix, G., Huez, G., Burny, A., Clenter, Y., Hubert, E., Leclerq, M., Chantrenne, H., Soreq, H., Nudel, U., and Littauer, U. Z. (1975) *Proc. Natl. Acad. Sci., U.S.A.*, *72*, 3065.

Matsuda, G., Takei, J., Wu, K. C. and Shiozawa, T., (1971) *Int. J. Protein Res.*, *3*, 173.

Matsuda, G., Maita, T., Mizuno, K. and Ota, J. (1973) *Nature New Biol.*, *244*, 244.

Maxwell, C. R., Kamper, C. S. and Rabinovitz, M. (1971) *J. Mol. Biol.*, *58*, 317.

McKnight, G. S. and Schimke, R. T. (1974) *Proc. Natl. Acad. Sci., U.S.A.*, *71*, 4327.

Morrison, M. R., Merkel, C. G. and Lingrel, J. B. (1973) *Mol. Biol. Reports*, *1*, 55.

Moss, B. A. and Thompson, E. O. P., (1969) *Aust. J. Biol. Sci.*, *22*, 1455.

Nichols, J. L. and Eiden, J. J. (1974) *Biochemistry*, *13*, 4629.

Nissley, S. J., Anderson, W. B., Gottesman, M. E., Perlman, R. L. and Pastan, T., (1971) *J. Biol. Chem.*, *246*, 4671.

O'Malley, B. W., McGuire, W. L. and Koreman, S. G. (1967) *Biochim. Biophys. Acta*, *145*, 204.

Pagoulatos, G. N. and Darnell, J. E. (1970) *J. Mol. Biol.*, *54*, 517.

Parker, C. S. and Roeder, R. G., (1977) *Proc. Natl. Acad. Sci., U.S.A.*, *74*, 44.

Pemberton, R. E., Housman, D., Lodish, H. F. and Baglioni, C. (1972) *Nature New Biol.*, *235*, 99.

Penman, S., Rosbach, M. and Penman, M. (1970) *Proc. Natl. Acad. Sci., U.S.A.*, *67*, 1878.

Perry, R. P. (1962) *Proc. Natl. Acad. Sci., U.S.A.*, *48*, 2179.

Perry, R. P. (1976) *Ann. Review Biochem.*, *45*, 928.

Perry, R. P., Greenberg, J. R. and Tartof, K. D. (1970) *Cold Spring Harbor Symp. Quant. Biol.*, *35*, 577.

Perry, R. P., Kelley, D. E. and LaTorre, J. (1974) *J. Mol. Biol.*, *82*, 315.

Perry, R. P., Kelley, D. E., Friderici, K. H. and Rottman, F. M. (1975) *Cell*, *6*, 13.

Pinder, J. C., Staynov, D. Z. and Gratzer, W. B. (1974) *Biochemistry*, *13*, 5373.

Platz, R. D., Kish, V. M. and Kleinsmith, L. J. (1970) *FEBS Letts.* *12*, 38.

Pogo, B., Allfrey, V. G. and Mirsky, A. E. (1966) *Proc. Natl. Acad. Sci., U.S.A.*, *56*, 550.

Price, R. P., Ransom, L. and Penman, S. (1974) *Cell*, *2*, 253.

Proudfoot, N. J. (1976) *J. Mol. Biol.*, *107*, 491.

Proudfoot, N. J. and Brownlee, G. G. (1974a) *FEBS Letts*, *38*, 179.

Proudfoot, N. J. and Brownlee, G. G. (1974b) *Nature*, *252*, 359.

Proudfoot, N. J. and Brownlee, G. G. (1976) *Nature*, *263*, 211.

Proudfoot, N. J. and Longley, J. I. (1976) *Cell*, *9*, 733.

Ptashne, M. (1967) *Proc. Natl. Acad. Sci., U.S.A.*, *57*, 306.

Rabbitts, T. H. and Milstein, C. (1975) *Eur. J. Biochem.*, *52*, 125.

Ramirez, F., Gambino, R., Maniatis, G. M., Rifkind, R. A., Marks, P.A. and Banks, A (1975) *J. Biol. Chem.*, *250*, 6054.

Roeder, R. G. and Rutter, W. J. (1969) *Nature*, *224*, 234.

Rose, J. K. and Lodish, H. F. (1976) *Nature*, *262*, 32.

Rosen, J. M., Liarakos, C. D. and O'Malley, B. W. (1973) *Biochemistry*, *12*, 2803.

Ross, J. (1976) *J. Mol. Biol.*, *106*, 403.

Ryfell, G. U. and McCarthy, B. J. (1975) *Biochemistry*, *14*,]379.

Schachat, F. H. and Hogness, D. S. (1973) *Cold Spring Harbor Symp. Quant. Biol.*, *38*, 371.

Scherrer, K., Latham, H. and Darnell, J. E. (1963) *Proc. Natl. Acad. Sci., U.S.A.*, *49*, 240.

Scherrer, K. and Marcaud, L. (1968) *J. Cell Physiol.*, *72*, 181.

Schmid, C. W., Manning, J. E. and Davidson, N. (1975) *Cell*, *5*, 159.

Scott, A. C. (1975) Ph.D. thesis, University of Adelaide.

Shatkin, A. J. (1976) *Cell*, *9*, 645.

Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci., U.S.A.*, *71*, 1342.

Sibitani, E., DeKloet, S. R., Allfrey, V. G. and Mirsky, N. E. (1962) *Proc. Natl. Acad. Sci., U.S.A.*, *48*, 471.

Smith, K. D., Church, R. B. and McCarthy, B. J. (1969) *Biochemistry*, *8*, 4271.

Soiero, R., Vaughan, M., Warner, J. R. and Darnell, J. E. (1968) *J. Cell Biol.*, *39*, 112.

Soiero, R. and Darnell, J. E. (1969) *J. Mol. Biol.*, *44*, 551.

Spelsberg, T. C., Hnilica, L. S. and Ansevin, A. T. (1971) *Biochim. Biophys. Acta*, *228*, 550.

Spelsberg, T. C., Mitchell, W. M., Chytil, F., Wilson, E. M. and O'Malley, B. W. (1973) *Biochim. Biophys. Acta*, *312*, 765.

Spohr, G., Imaizumi, T. and Scherrer, K. (1974) *Proc. Natl. Acad. Sci., U.S.A.*, *71*, 5009.

Spohr, G., Dettori, G. and Manzari, V. (1976) *Cell*, *8*, 505.

Spradling, A., Penman, S., Campo, M.S. and Bishop, J. O. (1974) *Cell*, *3*, 23.

Stedman, E. and Stedman, E. (1950) *Nature*, *166*, 780.

Stein, G. S., Park, W. D., Thrall, C. L., Mans, R. J. and Stein, J. L. (1975) *Nature*, *257*, 764.

Taylor, J. M., Illmensee, R. and Summers, J. (1976) *Biochim. Biophys. Acta*, *442*, 324.

Teng, C. S., Teng, C. T. and Allfrey, V. G. (1971) *J. Biol. Chem.*, *246*, 3597.

Tolstochev, P. and Wells, J. R. E. (1974) *Biochemistry*, *13*, 103.

Travers, A. A., Kramen, R. I. and Schleif, R. F. (1970) *Nature*, *228*, 748.

Verma, I. M., Temple, G. F., Fan, H. and Baltimore, D. (1972) *Nature New Biol.*, *235*, 163.

Ward, S., Wilson, D. L. and Gillian, J. J. (1970) *Analyt. Biochem.*, *38*, 90.

Warocquier, R. and Scherrer, K. (1969) *Eur. J. Biochem.*, *10*, 362.

Williams, A. F. (1970) Ph.D. thesis, University of Adelaide.

Wobus, U., Panitz, R. and Serfling, E. (1971) *Experentia*, *27*, 1202.

Zimmerman, S. B. and Sandeen, G. (1966) *Analyt. Biochem.*, *14*, 269.