APPROACHES TO THE PURIFICATION OF H5 mRNA SEQUENCES

A thesis submitted to the
University of Adelaide,
for the degree of Doctor of Philosophy

by

PAUL ANTHONY KRIEG, B.Sc. (Hons.)

Department of Biochemistry,
University of Adelaide,
South Australia.

October, 1980.

SUMMARY

Approaches to the Purification of H5 mRNA Sequences

1. The mechanism of control of gene expression is one of the major interests of modern molecular biology. The work described in this thesis was directed towards understanding a small part of this total problem, in particular, the relationship of the genes coding for the red-blood-cell specific histone H5 to the other histone genes, and the implications of this relationship on H5 gene expression.

In order to carry out these investigations, a pure probe for H5 mRNA sequences is required. Two main approaches were used. Firstly, purification of H5 mRNA was attempted so that complementary DNA probe could be prepared from this template. Secondly, as soon as the appropriate bio-hazard facilities were available, attempts were made to isolate pure H5 mRNA sequences using molecular cloning techniques.

2. Polysomal RNA was prepared from reticulocytes, since H5 is the only histone synthesised in these cells, and reticulocyte mRNA was used to optimise the wheat-germ cell-free translation system. Polysomal RNA was fractionated on the basis of poly A content using oligo-dT-cellulose and poly-U-Sepharose affinity columns. No useful enrichment of H5 mRNA could be achieved using these affinity chromatography techniques.

3. The translation and affinity chromatography techniques, developed for use with reticulocyte RNA, were applied to a 7-11S RNA fraction isolated from 5-day chicken embryos. Translation product profiles showed that the non-adenylated RNA fraction contained all five of the normal histone mRNAs, and no other detectable mRNA activity. cDNA probe prepared from this RNA fraction has been used to determine the reiteration frequency of the histone genes in the chicken genomal DNA. (Similar preparations of cDNA have been used by others in this laboratory to isolate clones containing genomic chicken histone DNA sequences).

4. Reticulocyte RNA was electrophoresed on several high resolution acrylamide gel systems. Although control experiments showed the gels to be capable of separating mRNA species of very similar molecular weights, no useful fractionation of globin and H5 mRNA sequences could be achieved. Resolution may have been less than optimal in this case due to the presence of poly-A-tracts on the globin mRNA, and so the mRNA was deadenylated by incubation with ribonuclease H in the presence of oligo-dT. Once again, no useful resolution of globin and H5 mRNA was observed.

5. Since it had proved impossible to isolate H5 mRNA by conventional physical techniques, recombinant DNA methods were used in an attempt to isolate cloned

H5 mRNA sequences. Double-stranded cDNA was prepared from reticulocyte polysomal RNA, inserted into a plasmid vector using the dC-dG tailing procedure and transformed into competent *E. coli* cells. The resultant recombinants were screened to eliminate those clones containing globin and ribosomal RNA sequences. Colonies containing H5 mRNA sequences should be found amongst those recombinants remaining after this screening.

6. Four recombinants potentially carrying H5 mRNA sequences were characterised in detail and the DNA-sequence of portions of the inserted DNA of each of the clones was determined. In no case was the nucleotide sequence determined consistent with that expected for a clone containing H5 mRNA sequences. Each of the four recombinants gave the same simple 'Southern blot' pattern to genomal DNA.

7. In view of the results of the recombinant DNA experiments, the translation products of reticulocyte RNA were examined under high stringency conditions. The results of these experiments indicated that H5 mRNA may be present in reticulocyte RNA at much lower concentrations than previously suggested.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge, it contains no material that has been previously published by any other person, except where due reference is made in the text.

PAUL ANTHONY KRIEG

ACKNOWLEDGEMENTS

I wish to thank the following people for the part they played in the production of this thesis:-

Dr. J.R.E. Wells, for his friendly and enthusiastic supervision.

Bob Crawford, Alan Robins, Richard Harvey, Rob Richards and Sue Clarke for their critical discussion, useful technical advice and friendship throughout the course of my experimental work.

Dr. J. Adams and members of his laboratory for their assistance and their generosity in allowing me to use the facilities in the Walter and Eliza Hall Institute for Medical Research.

Professor W.H. Elliott, for permission to work in the Department of Biochemistry, University of Adelaide.

During the course of this work I was supported by a Commonwealth Postgraduate Research Award.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

INTRODUCTION

The work presented in this thesis focusses on one aspect of differentiation. In particular it concerns the mRNA for a special histone, H5, which is expressed exclusively in certain nucleated red blood cells. While the experimental work has a narrow perspective, the introduction to this thesis considers the process of differentiation in a wider context. It reviews aspects of early embryology, of cellular determination, the role of cytoplasmic determinants on gene expression and the effects of cell position in the embryo on the subsequent fate of the cells. At the more molecular level, current knowledge of eukaryote gene structure and expression are described and detailed information about histone gene systems is presented. Finally, the properties of histone H5, and the possible relationship of the H5 gene to the other histone genes is discussed.

How gene expression is programmed to control the development of a single cell into a multi-cellular organism composed of many cell types, each expressing different specialised functions, is one of the fundamental questions of eukaryote biology. The developmental process can be split into two convenient, if somewhat arbitrary, parts, determination and differentiation. Determination is the acquisition, by a group of cells, of a committment that precisely defines the fate of their descendants, while differentiation is the process by which a precursor cell changes into a particular cell type in which characteristic specialised functions are expressed.

1.1 DETERMINATION

1.1.1 During development cells become committed to increasingly more specific developmental programmes

The fertilised egg has the opportunity to differentiate into all the cell types found in the organism, and so a single zygote will give rise to hundreds of different cell types during an organism's development. However, as development proceeds, the range of options available to a particular cell is narrowed. Early in embryogenesis a cell becomes committed to a broad path of development such as the formation of either ectodermal, (skin and nervous system), mesodermal (skeleton and muscle) or endodermal (respiratory and digestive tract) precursor cell lines. After further divisions, progeny cells will become involved in the formation of individual tissues or organs.

A good example of an increasingly specific developmental pathway is the production of haemoglobin synthesising red blood cells. After the original fertilised egg has divided a number of times (7-8 divisions in chick embryo), a determined haematopoietic stem cell is produced that is committed to the formation of the three major types of blood cells; white blood cells, platelet-producing cells and red blood cells (erythrocytes). One progeny cell of this precursor stem cell subsequently divides to produce an erythroid stem cell which, under the influence of the hormone erythropoietin, divides

and differentiates into mature haemoglobin synthesising erythrocytes, (review, Ingram, 1974).

The stage of embryonic development at which the reduction of a cell's developmental options is first observed differs from species to species. Any one of the 16 cells of the hydromedusa blastula has the ability to produce complete adult organisms, (Wilson, 1925), demonstrating that no committment of cells to specific developmental pathways has occurred at this stage. After the next division however, (the 32-cell stage), the cells are no longer equivalent and cannot give rise to adult organisms. On the other hand, in the nematode, the two cells resulting from the first division of the zygote already have a reduced range of developmental options, the anterior cell giving rise to the primary ectoderm while the posterior cell is the precursor of the germ-line tissue, (Nigan *et al.*, 1960; Hirsh and Vanderslice, 1976).

1.1.2 The initial determination of a cell may depend on its position in the embryo

During the first few divisions of the fertilised egg, the egg cytoplasm is distributed amongst the cells of the developing blastula. In many organisms, the developmental fate of a cell is at least partially determined by the portion of the egg cytoplasm that it receives. In *Drosophila*, for example, the germ line cells are

always derived from those cells that receive the cytoplasm from the posterior pole of the fertilised egg during the early divisions. Illmensee and Mahowald, (1974), showed that injection of posterior pole cytoplasm from the egg into the anterior end of developing embryos causes the production of germ line cells at this position. Genetic marker experiments show that, despite their abnormal starting position, these germ line cells can migrate through the embryo and take up their position in the developing gonad, giving rise to functional gametes, (Illmensee and Mahowald, 1976).

While the composition of the maternal cytoplasm received by the cells of the developing embryo has been shown to be important, there is evidence suggesting that, in the absence of any detectable partitioning of the egg cytoplasm, the position of the cell in the developing embryo plays a role in determining the future development of that cell. For example, in mammalian embryos, all blastomeres of the 4-cell embryo can give rise to either trophoblast tissue or inner-cell-mass tissue and so apparently have not been committed to a course of development by cytoplasmic composition. However, when these blastomeres are placed on the outside of other embryos, they always form trophoblast tissue and conversely, when the blastomeres are completely surrounded by other embryo cells they always from inner cell mass cells, (Hillman *et al.*, 1972; Stern 1973). These experiments indicate that in this case,

position is more important than cytoplasmic composition in determining the developmental fate of a cell.

It is probable that in most developing embryos both cytoplasmic composition and the position of the cell in the embryo are important. Classical embryology has established that precursor cells for various adult tissues occupy specific locations in the embryo and that patches of several cells can become determined simultaneously to follow a particular developmental pathway. This situation has been examined in great detail in experiments performed with *Drosophila* mosaics, (review, Postlethwait and Schneiderman, 1973).

Since *Drosophila* males are XY or XO and females are XX, the male is hemizygous for X-linked genes. A heterozygous female cell can thus be changed to a hemizygous male cell by the loss of an X-chromosome. Any recessive genes on the X-chromosome retained by the male cells may then be expressed, while previously their effect was hidden in the heterozygous female cells. In embryos developing from female zygotes containing one normal X-chromosome and one ring-X chromosome, certain nuclei sometimes lose the ring-X chromosome early in development, (Hall *et al.*, 1973). During subsequent cell division and differentiation these altered cells retain the same position, relative to other cells, that they occupied in the embryo. Using recessive X-chromosome markers it is possible to examine the adult fly and decide in which region of the embryo the original chromosome

loss occurred. The probability that two features on the adult fly will show different marker phenotypes is proportional to the distance between the embryonic cells that gave rise to these features, and so maps of the developmental potential of the different parts of the embryo can be prepared. These studies indicate that the precursor cells of the various structures of the adult fly always occupy precise positions in the embryo, once again showing that the position of a cell in the embryo determines the range of options available in its subsequent development.

The remarkable ability of the embryo to control the expression of cells contained within it, is illustrated by the experiments of Mintz and Illmensee, (1975), who injected differentiated mouse teratocarcinoma cells into blastocyst stage mouse embryos. These cells went on to divide and form functional tissues in the resulting adult mice, presumably becoming dedifferentiated in the embryonic environment and then differentiated once again, according to the instructions of the developing embryo.

1.1.3 The Stability of the Determined State

Certain aspects of *Drosophila* development illustrate clearly the stability of the determined state of cells. In *Drosophila*, most of the structural features of the adult fly are constructed during the pupal period from a set of 'imaginal discs' which form in the embryo and grow during larval development.

These imaginal discs have no apparent larval function and, while the disc cells show no sign of differentiation, they are committed rigidly to specific developmental pathways. For example, during metamorphosis the wing imaginal discs form the wing itself as well as parts of the thorax belonging to the wing segment.

These discs can be transplanted from one larva to almost any position on a second larva, where, during pupal development, they will differentiate into the particular structures that they would have formed in the original fly, (Ursprung, 1967). For example, if a wing disc is transplanted, the first larva will develop into an adult with only one wing, while the second larva will metamorphose into an adult with a third wing at the position of the transplanted disc.

The determined state of the original imaginal disc cells is inherited by all its descendants. While the number of determined cells comprising the imaginal disc rises from about 10-40 when they first appear in the late embryo, up to several thousand in the final larval stage, (Postlethwait and Schneiderman, 1973), each disc maintains its determined state and differentiates appropriately during metamorphosis. Furthermore, discs can be cultured for long periods by serial transplantation into the abdomens of adult flies, (where the cells proliferate but do not differentiate, (Bodenstein, 1943)), and when transplanted

into a larva the disc cells still respond to hormonal stimulation and undergo metamorphosis and differentiation along with the host, producing specific adult structures, (Hadorn, 1963).

The acquisition of a special state of determination seems to correlate with the ability of determined cells to recognise and associate with like cells, and to dissociate from unlike cells. Cells from a given imaginal disc, after dissociation *in vitro*, will reassociate and form organised patterns of cells. However, when cells from different types of disc are mixed, they will separate out from each other and reassociate with like cells to form separate patterns, (Nöthinger, 1964, Garcia-Bellido, 1966).

1.1.4 Changes in the Determined State

Mutations are sometimes observed in *Drosophila* in which a normal structure of the adult fly is replaced by other normal structures that would usually be found elsewhere in wild-type flies. For example, mutations in the *bithorax* gene transform the anterior part of the haltere into anterior wing, mutations in the *spineless-aristapedia* gene transforms the antenna into a leg and mutations in the *opthalmoptera* gene allow a wing to develop in the position where the eye would normally be. These mutations are called homoeotic mutations and they cause an imaginal disc committed to one developmental

programme to produce a tissue that is characteristic of a different programme.

The homoeotic genes are single loci which determine the developmental pathways taken by groups of cells, and the existence of, for example, the *bithorax* mutant, indicates that there is at least one time in the normal development of the haltere at which the alternative of whether to make a haltere or a wing is open. Specific disc cells require a continuous supply of the wild-type product of the homoeotic genes for the maintenance of their state of determination, from the late embryo stage when the discs first appear, right through to the final larval stage (Morata and Lawrence, 1975; Morata and Garcia-Bellido, 1976). If the activity of the homoeotic gene is interrupted the state of determination will change and the groups of cells will be switched from one developmental pathway to another.

1.2 DIFFERENTIATION

The process of cell differentiation is achieved through variable gene expression.

In this section evidence supporting the variable gene activity mechanism, and opposing other systems as the general mechanism of differentiation, will be presented and discussed.

1.2.1 The DNA composition of differentiated cells

A large body of evidence has accumulated

showing that the DNA composition of all cells of an organism is identical, and that specific irreversible changes in either the composition of the DNA or the arrangement of the sequence are not the mechanisms of differentiation.

To the limits of accuracy of the technique, DNA-DNA reassociation studies have shown that the same repeated and single-copy sequences are present in the DNA extracted from any tissue of an organism, (Davidson and Hough, 1971; Schultz *et al.*, 1973). For example, Kohne and Byers, (1973), have shown that DNA extracted from calf kidney cells reassociates with itself with precisely the same kinetics that it shows when bybridised to DNA extracted from calf brain, thymus and liver cells.

Even if these differentiated cells appear to have the same DNA content and composition it seems possible that stable changes in the DNA, (mutations), or in the organisation of the DNA, or amplification or deletion of specific sequences might cause the differentiation of cells. In general, these changes would not be detectable in total DNA hybridisation studies.

Nuclear transplantation experiments have demonstrated that if any changes are made in the composition or arrangement of the DNA in the course of differentiation, these changes are, at least, readily reversible. Gurdon, (1962), reported that *Xenopus* tadpoles could be produced from enucleated eggs injected with a nucleus derived from a differentiated tadpole intestinal cell. Subsequent experiments have shown

that nuclei derived from primary cultures of adult *Xenopus* lung, kidney, heart, testis and skin are able to give rise to differentiated tadpoles, when injected into enucleated eggs, (Laskey and Gurdon, 1970).

Experiments yielding similar conclusions have since been performed with *Drosophila*, (Illmensee, 1972; Okada *et al.*, 1974), and mouse, (Mintz and Illmensee, 1975), using genetically marked donor nuclei, making removal of the recipient genetic material unnecessary. Mintz and Illmensee, (1975), have shown that the DNA of cultured, malignant, teratocarcinoma cells gives rise to a large range of normal differentiated tissues when injected into normal blastocysts. The tissues so far detected include melanoblasts, hair follicle dermis, erythrocytes, leucocytes, liver, thymus, kidney and sperm. Progeny mice have been produced from eggs fertilised with these tumour-derived sperm. This ability of embryonal carcinoma cells to form normally functioning adult tissues demonstrates that even conversion to the differentiated malignant state does not involve irreversible structural changes in the genome, but rather a change in gene expression.

1.2.2 Gene Amplification

It seems possible that specialised cells might contain extra copies of certain genes, especially those required for intense, specialised activities. The ribosomal-RNA gene amplification in amphibian oocytes is a well-known case in which selective replication of the ribosomal genes is used to augment ribosomal RNA

synthesis when large amounts of this RNA are required during oogenesis, (Gall, 1968). While this mechanism is not unique to amphibia and is observed in a number of different phyletic groups, (Hourcade *et al.*, 1974), it is far from being a general biological mechanism for ribosomal RNA regulation. In the chicken, for example, DNA extracted from a wide range of tissues always contained the same number of ribosomal cistrons, even though the rates of ribosomal RNA synthesis differed markedly between these tissues, (Ritossa *et al.*, 1966).

Measurements have also been made on a number of genes coding for specialised protein products, including haemoglobin, (Harrison *et al.*, 1974), and ovalbumin, (Sullivan *et al.*, 1973). These genes were found to be present at only about one or two copies per haploid genome regardless of whether the DNA was extracted from cells actively synthesising the specific protein or from any other tissue.

Recently, Alt *et al.*, (1978), have shown that the genes for dihydrofolate reductase can be amplified up to 200-fold in tissue culture cells resistant to methoxytrexate, a drug that blocks purine synthesis by inhibiting dihydrofolate reductase. This gene amplification results in a corresponding 200-fold increase in dihydrofolate reductase mRNA and enzyme levels, effectively making the cells resistant to the drug. In most cases, when methoxy-trexate is removed from the culture medium the amount

of dihydrofolate reductase present decreases and this is paralleled by a drop in the gene reiteration. While this is an example of the amplification of a structural gene in order to regulate the levels of a required mRNA, the extreme selective pressure placed on the system suggests that this mechanism may not be generally used in the normal course of cell development.

1.2.3 Rearrangement of Genes

Hozumi and Tonegawa, (1976), using restriction and hybridisation analysis of mouse DNA, have demonstrated that the DNA coding for the lambda light chain immunoglobulin protein is not in the same arrangement in cells synthesising immunoglobulin as it is in germline DNA. These observations were verified by cloning and sequencing experiments, (Bernard *et al.*, 1978), and have since been extended to the DNA coding for kappa light chain proteins, (Max *et al.*, 1979; Sakano *et al.*, 1979).

The immunoglobulin light and heavy chain molecules consist of two regions, a constant (C) region and a variable (V) region and in each case the V region is involved in antigen recognition. Restriction enzyme and cloning experiments showed that the DNA sequences coding for the V and C regions of kappa and lambda chains were well-separated in DNA from germ-line tissue but close together, (as close as 1.2 kb in lambda light chain genes), in DNA from tissues synthesising antibodies. The final gene

still includes intervening sequences within the coding region of the lambda and kappa chains, (Brack and Tonegawa, 1977; Max *et al.*, 1979, respectively), however, the V and C regions are now apparently included in the same RNA transcript and the intervening sequences removed by RNA splicing, (review, Abelson, 1979). Thus, DNA rearrangement during development appears to be necessary to form the functional gene in the differentiated anti-body producing cell.

Evidence for movable elements in eukaryotic chromosomes has been available for many years. McClintock, (1956), working with maize, showed that drastic alterations of chromosome organisation, including inversions, deletions and duplications, occurred at sites containing 'controlling elements', and that the controlling elements were capable of transposition to new sites. While this particular system has not been characterised in molecular terms, evidence has recently been obtained for the presence of transposable elements in the DNA of *Drosophila melanogaster*, (Finnegan *et al.*, 1977) and yeast, (Cameron *et al.*, 1979). The 'copia' sequences of *D. melanogaster* and the 'Tyl' sequences of yeast are both moderately repeated, about 35 copies of each being dispersed throughout their respective genomes, and are terminated with short direct repeats, making overall structures extremely similar to those of some bacterial transposons,

(Ptashne and Cohen, 1975). Both the copia and Tyl sequences are transcribed and represent a significant proportion of the non-ribosomal RNA present in a cell. Copia sequences, for example, can comprise up to 3 percent of the poly-A-containing RNA from a number of *D. melanogaster* tissues. It has been suggested, (Cameron *et al.*, 1979), that the Tyl sequence may have an effect on the pattern of gene expression in yeast, and, because it is movable, it may have the ability to alter this pattern.

Despite this evidence for the presence of movable elements in eukaryotic DNA there is little reason to believe that rearrangement of the DNA plays a significant role in controlling the expression of specific structural genes. In a number of systems investigated, including rabbit globin, (Jeffreys and Flavell, 1977), chicken ovalbumin, (Breathnach *et al.*, 1977), and chicken keratin genes, (Saint, 1979), restriction enzyme studies have shown the arrangement of DNA sequences coding for the protein to be identical in tissues actively synthesising the protein and in tissues in which the protein is not produced. In these cases at least, expression must therefore involve the activation of pre-existing control regions rather than the insertion or rearrangement of control regions adjacent to the gene. These conclusions suggest that the DNA rearrangement observed for the immunoglobulin genes may have evolved specifically

as a means of generating at least part of the remarkable variability required by the immune system, and is not a feature common to most eukaryote structural genes.

1.2.4 Evidence for Transcriptional Control of Gene Activity

Evidence has been presented above, suggesting that the DNA content of all the somatic cells of an organism is identical, and that there is little evidence that gene amplification, or deletion, or rearrangement plays a significant part in the control of cell differentiation. Observations from a number of other studies, however, suggest that transcriptional control is the major mode of gene regulation.

Experiments have shown that only a small proportion of the genome is transcribed in differentiated cells, the RNA produced presumably being that required for maintenance of the cell and for the specific differentiated cell functions. For example, the total RNA extracted from metabolically active mouse liver cells only hybridises to 2-5 percent of mouse single-copy DNA, (Grouse *et al.*, 1972). In addition, when the RNA transcripts from mouse liver, spleen and kidney cells were compared, more than 70 percent of the RNA sequences were shown to be unique to each cell type, (Grouse *et al.*, 1972), indicating that overlapping but clearly distinct sets of RNA sequences are synthesised in various differentiated cell types.

Evidence for transcriptional control of gene expression can be obtained from the cytological examination of dipteran polytene chromosomes. The polytene chromosome puffs have been shown to be the sites of intense RNA synthesis, and these features can be localised to specific chromosomal regions, varying according to the state of differentiation, (Pelling, 1964; McKenzie *et al.*, 1975). A good example is the Balbiani rings of *Chironomus* salivary gland chromosomes, (Daneholt and Hosick, 1974). The synthesis of specific RNA and protein products has been associated with specific puffs and three of these chromosomal puffs are responsible for about 80 percent of the RNA synthesis in the cell, while apparently representing only a few percent of the total genomic DNA. Both the secretory peptides arising from the puff transcripts and the puffs themselves are specific to only certain salivary gland cells, indicating that the activity of a limited set of genes is responsible for the differentiated function of these cells.

A number of developmental systems have been studied in which the amounts of a certain protein rise from very low or undetectable levels to the very high levels characteristic of the mature, differentiated cell. Measurements made using complementary DNA probes to determine the levels of mRNA for, haemoglobin in the erythroid cell series, (Hunt, 1974) ovalbumin in the chicken oviduct, (Palmiter, 1973)

and keratin in the developing chick feather, (Powell *et al.*, 1976), have shown that during the process of differentiation, the increase in specific protein synthesis is due to the increase in the concentrations of the specific mRNAs. For example, in the developing embryonic feather, no keratin mRNA is detectable until day 12, when the first keratin protein is also observed. Maximum levels of keratin mRNA occur at about day 14 when maximal keratin protein synthesis is also observed. In the 11-14 day interval, the number of molecules of keratin mRNA increases over 600-fold with a corresponding increase in the amount of keratin protein, (Powell *et al.*, 1976).

1.2.5 Control of Transcription

While the evidence presented above adequately demonstrates that, in general, the expression of a specific gene is controlled at the transcriptional level, the problem of how transcription is controlled now arises.

A great deal of experimental evidence has shown that steroid hormones are capable of specifically altering the expression of certain genes. At the genetic level, the best understood steroid hormone is ecdysone, the hormone responsible for the transition from the larval to the pupal stage in insect development. In *Drosophila*, (Grossbach, 1973), for example, there are only about ten puffs visible in the salivary gland chromosomes during the larval stage.

However, when ecdysone is added, the pre-existing puffs begin to regress while a number of new puffs suddenly appear. During the transition from the larval to the pupal stages, about 120 new puffs eventually appear, each puff arising at a specific time after the addition of ecdysone and lasting for a very specific period. The first puffs arise within 5 minutes of ecdysone addition and appear to be direct responses to the hormone, since their appearance is not affected by inhibitors of protein synthesis. However, the late puffs, which only arise some hours after hormone treatment do not appear when protein synthesis is blocked and so it seems likely that protein products of some of the early puffs induce the formation of the late puffs.

The molecular biology of steroid hormone action has been most thoroughly investigated in the chicken oviduct system, where administration of oestrogen or progesterone immediately induces the synthesis of mRNAs coding for egg-white proteins, of which ovalbumin is a major component, (Harris *et al.*, 1975). Upon entering the cell, steroid hormones are initially bound to specific cytoplasmic receptor proteins. The hormone-receptor complex moves from the cytoplasm into the nucleus where it binds to sites on the target cell chromatin. This is followed by the activation of specific genes, resulting in the appearance of new species of mRNA. The progesterone receptor of chicken oviduct cells

has been purified and its properties extensively investigated, (Vedeckis *et al.*, 1978). The receptor has been shown to be a protein-dimer, one subunit locating the hormone-receptor complex to specific regions of the chromatin, while the other subunit appears to alter the structure or conformation of the specific chromatin-DNA sites so that initiation of new RNA synthesis can occur, (O'Malley *et al.*, 1978).

Transcription in eukaryotic cells is also controlled by different RNA polymerase molecules, (review, Chambon, 1975). Three types of RNA polymerase have been identified in eukaryotic cells and these can be distinguished by their different sensitivities to α-amanitin, an inhibitor of RNA synthesis. RNA polymerase type I transcribes the genes for 18S and 28S ribosomal RNA, RNA polymerase type III transcribes tRNA and 5S ribosomal RNA sequences and RNA polymerase II apparently transcribes most other DNA sequences, making it responsible, therefore, for mRNA synthesis. Whether additional controlling elements can bind to RNA polymerase type II and further affect the specificity of the sequences it transcribes, (as observed during the bacteriophage T_4 infection of prokaryote cells, (Schmidt *et al.*, 1970)) is yet to be determined.

The appearance of a functional protein product may be controlled at a number of other stages after the formation of the original RNA transcript. There may be control during the processing of

the original nuclear transcription products into mRNA or during the transport of mRNA sequences across the nuclear membrane. There is evidence, for example, suggesting that chicken globin pre-mRNA is processed at different rates in erythroblasts and reticulocytes, (Crawford and Wells, 1978).

Cases are known where control of gene expression is exercised at the translation level, for example, in amphibian oocytes, where maternal mRNA exists in the cytoplasm for some time until translation of these mRNAs is stimulated by fertilisation, (Smith and Ecker, 1965; 1970). Some proteins require post-translational modification to become functional. Examples are trypsinogen the precursor of trypsin and proinsulin the precursor of insulin. Activation of both these inactive precursor forms involves the excision of specific polypeptide fragments by proteolytic cleavage, (Steiner *et al.*, 1974).

1.2.6 Models of Transcriptional Control

While every possible mechanism of control is probably utilised at some stage during the development and differentiation of a complex organism, an overwhelming weight of evidence shows that the primary control of gene expression occurs at the level of gene transcription.

It seems likely that regardless of the mechanism of transcription control; via hormone-receptor action or different RNA polymerase specificities, the crucial

step will involve the recognition of a specific DNA control sequence. Since initiation of transcription of eukaryotic genes must occur at some point, it is likely that transcription control regions, perhaps similar to those detected in prokaryote systems, (for example, the *lac* operon system, (Jacob and Monod, 1961)), are present adjacent to the coding regions of eukaryotic genes. Britten and Davidson, (1969), have proposed a model for eukaryote gene control which is basically an extension of the bacterial operon model, and which attempts to explain how the enormous number of genes could be co-ordinated and controlled during eukaryote differentiation and development.

The model suggests that all genes required to be co-ordinately expressed have a common control element adjacent to those genes. Every set of co-ordinately expressed genes would need to have a different control sequence and so expression of a gene on more than one occasion would require more than one control element adjacent to it, each corresponding to the set of genes expressed at the different stages. Since these control regions will be present next to more than one gene, these sequences will appear as moderately repeated sequences in the eukaryote genome.

While there is no direct evidence for the Britten and Davidson model, there is considerable data showing that short, moderately repeated sequences

are interspersed between longer unique sequences in eukaryote genomes, (Davidson *et al.*, 1973; Schmid and Deininger, 1975), as predicted by the model. This model should be directly testable when the regions adjacent to co-ordinately expressed genes are isolated and examined in detail. Information on this problem may be available very soon, since three adjacent genes, all under the control of the hormone, oestrogen, and expressed in the chicken oviduct, have recently been isolated, (Royal *et al.*, 1979).

1.3 RECOMBINANT DNA

The use of restriction endonucleases to reproducibly generate defined fragments of DNA, and the resultant development of recombinant DNA technology has brought about a new era in biochemical genetic research. The application of recombinant DNA technology to the question of eukaryote gene structure and function has yielded a remarkable amount of information. Through the use of cloned mRNA sequences, to produce pure gene sequence probes, and the cloning of precisely defined fragments of genomic DNA, it has finally become possible to examine eukaryotic genes at the biochemical level.

It is important to realise, however, that many of the results of the recombinant DNA research would not have been possible without the availability of the recently-developed rapid DNA-sequencing techniques. The enzymatic, dideoxy chain termination

procedure of Sanger *et al.*, (1977), and the chemical, chain cleavage method of Maxam and Gilbert, (1977), are both ideally suited to the sequencing of the pure defined fragments of DNA produced by recombinant DNA techniques.

Some of the most significant discoveries concerning eukaryote gene structure, that have arisen as the result of recombinant DNA research, will be discussed in the following sections.

1.3.1 Intervening Sequences

Extra DNA sequences, contained within a region of DNA normally considered to be a gene, were first detected in the 28S ribosomal RNA gene of *Drosophila*, (Glover and Hogness, 1977: White and Hogness, 1977), and in the yeast tRNA genes, (Goodman *et al.*, 1977). Soon afterwards, intervening sequences were found within the protein-coding sequences of structural genes, for example, ovalbumin genes, (Breathnach *et al.*, 1977), globin genes, (Tilghman *et al.*, 1978 a), and immunoglobulin genes (Tonegawa *et al.*, 1978). The large number of reports of intervening sequences in structural genes, from a wide variety of organisms, now makes it appear likely that most eukaryote genes will contain intervening sequences. Some genes contain a remarkably large number of intervening sequences. The rat serum albumin gene contains at least 13 intervening sequences, (Sargent *et al.*, 1979), and the half of the sheep

collagen gene that has been examined in detail contains 12 to 16 intervening sequences, (Tolstoshev - pers. commun).

Currently, the best characterised set of intervening sequences are contained within the chicken ovalbumin gene. The sequence which appears in the cytoplasm as mRNA is interrupted in the genomic DNA by seven DNA sequences, varying in length from about 0.3 to 1.4 kilobase pairs. All but one of these sequences is contained within the coding region of the mRNA, the remaining inserted sequence appearing in the 5´ untranslated region, (Dugaiczyk *et al.*, 1978; Gannon *et al.*, 1979). These intervening sequences appear in the original precursor mRNA transcripts of the gene and are presumably excised by processing enzymes to yield the final cytoplasmic mRNA, (Roop *et al.*, 1978). The DNA sequence at the junctions between each structural and intervening sequence in the ovalbumin gene has been determined, (Catterall *et al.*, 1978) and the 'consensus' sequence obtained agrees with that observed at the junctions in the mouse lambda immunoglobulin gene (Tonegawa *et al.*, 1978) and the mouse beta-globin gene, (Tilghman *et al.*, 1978 b). This evidence suggests that the junction sequences may have been conserved during evolution.

Analysis of the beta-globin genes of mouse, (Konkel *et al.*, 1978) and rabbit, (van den Berg *et al.*, 1978) and human, (Lawn *et al.*, 1979), has shown that

the two intervening sequences occupy identical positions in these genes. While the inserts are found in the same place in the genes, their lengths and sequences have been found to be quite dissimilar. Within a species, the mouse beta-globin (major) insert sequence has been shown to have diverged considerably from the mouse beta-globin (minor) sequence, (Miller *et al.*, 1978). When mouse beta-globin (major) intervening sequence was used as a probe back to restricted genomal DNA, only the beta-globin genes were detected, showing that the insert sequence is unique.

It has been suggested that the intervening sequences flank regions of the gene that code for domains in the protein product that have a particular functional role. During evolution, new proteins might be constructed very quickly by bringing together pre-formed DNA-sequences coding for domains in the final protein, (reviewed by Blake, 1979). While the idea seems consistent with the arrangement of intervening sequences in the immunoglobulin genes, (Sakano *et al.*, 1979), it does not explain the existence of an intervening sequence in the 5´ untranslated region of the ovalbumin gene.

Perhaps one of the most interesting consequences of the existence of intervening sequences is that they make complete eukaryote genes extremely large; far larger than would be expected from the size of the cytoplasmic RNAs. For example, the genomic ovalbumin gene, and the primary RNA transcript, are more than four times

longer than the ovalbumin mRNA. Obviously much of the total genomic DNA, and especially unique sequence DNA, can be accounted for by intervening sequences.

1.3.2 Eukaryotic Promoters

By analogy with prokaryote genes, transcription of eukaryote genes should be initiated at a well-defined promoter region, located somewhere before the start of the mRNA sequence. The 5′ regions of a number of eukaryote genes have been sequenced in the hope of finding a common sequence which might act as the eukaryotic promoter. Analysis of the *Drosophila* histone genes has shown the presence of a putative promoter sequence about 70 bases to the 5′ side of the initiation codons of each of the histone sequences. This sequence shows similarities to the 'Pribnow box' prokaryote promoter sequence. Like the prokaryote promoter, this sequence is very A-T rich, and has been called the 'Hogness box', (review, Proudfoot, 1979). The 5′ end regions of other eukaryote genes, including sea-urchin histones, silk-worm fibroin, mouse and rabbit alpha-globin and yeast isocytochrome C, each contain sequences showing partial homology with the Hogness box sequence. Recently, Gannon *et al.*, (1979), have sequenced the 5′ end of the chicken ovalbumin gene, and, once again, a sequence bearing similarity to the Hogness box sequence has been detected about 25 bases away from the start of the mRNA sequence.

While the evidence for the eukaryotic promoter sequence is currently based on a relatively few sequences, data from many more eukaryotic genes and from RNA polymerase binding and transcription studies can be expected to confirm or question the proposed sequence, in the near future.

1.3.3 Regulatory Sequences

DNA-sequencing experiments have shown the presence of highly conserved sequences at the 3´ end of each of the five histone genes in the sea urchin *Psammechinus miliaris*, (Busslinger *et al.*, 1979). Incomplete sequence data from another sea-urchin species, *Strongylocentrotus purpuratus*, shows the presence of a remarkably similar sequence at the 3´ end of, at least, the H2A, H2B and H3 histone genes. The longer conserved sequence in these genes is palindromic and if transcribed into mRNA would form a stable hairpin structure. It has been proposed that a signal structure such as this may play some role in the maturation or termination of histone RNA transcripts, (Busslinger *et al.*, 1979).

Other putative regulatory sequences have been detected in eukaryotic genes. The sequence AAUAAA has been detected towards the 3´ end of all eukaryote mRNAs sequenced, except those of the sea urchin histone genes, (Proudfoot and Brownlee, 1976). The absence of this sequence in the histone mRNAs, (Sures *et al.*, 1978; Grunstein and Grunstein, 1977; Schaffner *et al.*, 1978), shows that it is not a signal sequence essential

for all eukaryotic mRNAs, and, since the histone mRNAs are unusual in that they are not, in general, polyadenylated, the AAUAAA sequence may act as a polyadenylation signal. The DNA sequences of more eukaryotic genes, and in particular the histone genes of different species, will be required before this proposition can be accepted with confidence.

1.3.4 Gene Movement

The application of restriction enzymes and recombinant DNA techniques to the question of gene rearrangement has been discussed earlier, (section 1.2.3), with particular reference to the mouse immunoglobulin genes, (Hozumi and Tonegawa, 1976; Bernard *et al.*, 1978), and to the transposable copia and Tyl genes of the *Drosophila* and yeast genomes respectively.

Like so many of the recently discovered features of the eukaryote genome, the precise function of these transposable elements is unknown.

1.3.5 Future developments

Many of the studies currently in progress into the composition and arrangement of the eukaryote genome will be continued and extended in the future and it is likely that the information available on the detailed structure and sequence of specific eukaryote genes will be vastly increased in the next few years.

In addition, however, studies involving the combined transcription and translation of specific eukaryote genes, under precisely defined conditions,

should contribute a great deal to the understanding of how the eukaryote gene is controlled. Work along these lines is already in progress.

Experiments have shown that yeast tRNA precursor molecules are correctly processed into functional tRNA when injected into *Xenopus* oocytes, (De Robertis & Olson, 1979), indicating that *Xenopus* is able to recognise and excise the yeast intervening sequences and then rejoin the RNA to form the final tRNA product. As a next step, the transcription and processing of RNA from clones containing the yeast tRNA genes is currently being investigated in a *Xenopus* cell-free system, (Hall, cited by Wolfe, 1979). Hinnen *et al.*, (1978), have shown that bacterial plasmids containing cloned DNA sequences may be used to transform yeast. In recent experiments, yeast have been transformed with a plasmid containing the rabbit globin genes and associated sequences, (Beggs, cited Wolfe, 1979). While globin sequences were transcribed in the yeast system, transcription stopped in the second intervening sequence and the first intervening sequence was not excised from the transcript. This result suggests that despite the success of the *Xenopus* cell free system in processing yeast precursor molecules, this may not be a general case, and it may be necessary to study transcription and translation in an homologous system where possible.

1.4 THE HISTONE GENE SYSTEM

1.4.1 The Histone Proteins

The histones are a set of five small basic proteins. They are the fundamental structural proteins of chromatin and are found in all eukaryotic organisms. Four of the histones, H2A, H2B, H3 and H4 interact with DNA and with each other to form the nucleosome core, the fundamental unit of chromatin structure. The fifth histone, H1, apparently plays a role in the higher order structure of chromatin and is found in chromatin at about half the molar concentration of the other histones, (Kornberg, 1974; Noll and Kornberg, 1977).

The primary sequence of the histone proteins has been highly conserved during evolution. The amino acid sequence of histone H4 is considered to be the most stringently conserved protein sequence known, the H4 of peas and cows differing in only 2 of 102 amino acid residues. Recently, however, sequencing of histone H4 of *Tetrahymena*, (Glover and Gorovsky, 1978), has shown considerable differences between the amino acid sequences of this, and other H4 proteins. The other histones show a high degree of primary sequence conservation in several regions of the protein and it is presumably these regions that are critical for molecular interactions within or outside the nucleosome.

1.4.2 Histone mRNA

Histone protein synthesis is closely coupled to DNA replication in the somatic cells of animals, (Robbins and Borun, 1967), yeast, (Moll and Wintersberger, 1976) and protozoans, (Prescott, 1966), and the available evidence shows that there is a rapid loss of histone mRNA activity from the cytoplasm and polyribosomes when DNA replication is stopped, (Robbins and Borun, 1967; Gallwitz and Mueller, 1969).

The mature histone mRNAs are about 9S in size and in general are not polyadenylated, (Adesnick and Darnell, 1972). Since the poly-A-tract has no function in protein synthesis, (Williamson *et al.*, 1974; Bard *et al.*, 1974), it has been proposed that it acts as a nuclease inhibitor, (Levy *et al.*, 1975) and could therefore effect the stability of the mRNA. The demonstration that other naturally occurring mRNAs that lack poly-A-tails are also rapidly processed, (Milcarek *et al.*, 1974; Nemer *et al.*, 1974), supports this view. This could explain the rapid turn-over of cytoplasmic histone mRNA sequences at the end of S-phase, (Gallwitz, 1975), which is not observed for most of the other cytoplasmic mRNAs. Recently, evidence has accumulated showing that at least a subfraction of histone mRNAs may be polyadenylated, (Borun *et al.*, 1977; Ruderman and Pardue, 1978), and it has been proposed that

these differences may be related to stage specific switches of histone mRNA sequences. For example, although considerable amounts of histone mRNA from *Xenopus* oocytes are found in the poly-A containing fraction of RNA, after fertilisation most of the histone mRNA activity is found in the poly-A-minus fraction, (Ruderman and Pardue, 1978).

While it seems generally agreed that the regulation of histone mRNA activity is controlled at the transcriptional level, the evidence for this is not convincing, with conflicting results being reported from different studies. Melli *et al.*, (1977a), find putative histone mRNA sequences in the nuclear RNA of HeLa cells throughout the cell cycle and so suggest that processing or transport of mRNA from the nucleus to the cytoplasm may be more important than transcription in controlling histone protein synthesis. Stein *et al.*, (1977), on the other hand, only detect histone mRNA sequences in HeLa cells during DNA replication, and so conclude that the mechanism of control is transcriptional. Further studies with pure mammalian histone probes will probably be required before this question is resolved.

There is evidence that histone mRNA may be transcribed in a high molecular weight precursor form, at least twice as long as the final mRNA, in sea urchin gastrulae, (Kunkel *et al.*, 1978),

but no such precursor molecules have been detected in cleavage stage embryos, (Childs *et al.*, 1979). Evidence for a high molecular weight precursor of histone mRNA has also been found in HeLa cells, (Melli *et al.*, 1977b; Hackett *et al.*, 1978), but these studies disagree considerably on the size of the initial RNA transcript. As yet, there is no reason to believe that any of these precursors are polycistronic.

1.4.3 The Histone Genes

1.4.3(i) Sea Urchin

The histone genes of sea urchin are perhaps the best characterised of all eukaryotic structural genes, with almost complete DNA sequences available for two species, *Strongylocentrotus purpuratus*, (Sures *et al.*, 1978) and *Psammechinus miliaris*, (Schaffner *et al.*, 1978).

The genes for the five different histones, H1, H2A, H2B, H3 and H4 are clustered into a unit that is tandemly repeated in the genome. Restriction analysis of sea urchin genomal DNA, (for example, *P. miliaris*), showed the presence of an approximately 6 kilobase long repeat unit, in which genes coding for each of the five histone proteins were interspersed with one another, and that spacer DNA was located between each of the coding regions (Cohn *et al.*, 1976; Schaffner *et al.*, 1976, Kedes, 1976). The repeat

unit was cloned into a phage vector and thermal denaturation was used to examine A-T rich spacer regions under the electron microscope, (Portmann *et al.*, 1976).

Gross *et al.*, (1976a), isolated the individual sea urchin histone mRNA species and then, using restriction enzymes, limited exonuclease digestion of cloned DNA and hybridisation with purified histone mRNAs, was able to demonstrate that the polarity of the histone gene cluster in *P. miliaris* is,

$$5' \xrightarrow{H1} \xrightarrow{H4} \xrightarrow{H2B} \xrightarrow{H3} \xrightarrow{H2A} 3'$$

with all mRNAs transcribed from the same DNA strand, (Gross *et al.*, 1976b). Similar studies on *S. purpuratus* using restriction, (Cohn *et al.*, 1976), and electron microscopic, (Wu *et al.*, 1976), techniques, demonstrated that the polarity of the histone repeat and the position of A-T rich spacer regions are conserved between these two species.

Direct DNA-sequencing of cloned histone gene repeats, (Schaffner *et al.*, 1978; Sures *et al.*, 1978), has indicated that the coding sequences are colinear with the amino acid sequences and so no intervening sequences are present within these histone genes. The A-T rich spacer regions are made up of relatively simple nucleotide arrangements, but do not show any evidence of internally repetitive sequences.

1.4.3.(ii) *D. melanogaster* histone genes

The histone genes of *D. melanogaster* are the only ones other than sea urchin that have been extensively characterised, and they are present at about 100 copies per haploid genome. A number of clones of *D. melanogaster* DNA containing histone mRNA coding sequences have been isolated, (Lifton *et al.*, 1977). These were selected from a collection of cloned sequences by their ability to hybridise to sea urchin histone mRNA.

Two major types of repeating unit were found. In both cases all five histone coding sequences were contained within a repeat, and the coding regions were separated by A-T rich spacer DNA. The two types of repeat are 4.8 and 5.0 kilobase pairs long and they only differ by the presence, in the longer repeat, of a piece of DNA inserted into a spacer region. Analysis of the position and polarity of the genes has shown that, contrary to the sea urchin situation, the direction of transcription is not the same for all the histone mRNA species, and that mRNA is transcribed from both strands of the DNA. The arrangement of the genes is,

$$5' \xrightarrow{} H1 \xleftarrow{} H2B \xrightarrow{} H2A \xleftarrow{} H4 \xrightarrow{} H3 \; 3',$$

requiring at least two independent sites for initiation of transcription, (Lifton *et al.*, 1977). In the absence of any evidence for

polycistronic precursor RNA it seems likely that each of the mRNA transcripts is independently initiated. This is in contrast to the situation in sea urchin where, in theory at least, only one initiation event would be required for coordinate expression of the histone genes.

1.4.3.(iii) Other Invertebrate Species

Very little data is currently available on the distribution and arrangement of the histone genes in species other than *D. melanogaster* and sea urchin. Restriction and hybridisation techniques have been used to determine the size of the histone gene repeat unit for a number of invertebrates (Freigan *et al.*, 1976), including the horseshoe crab, (4.1 kb), clam (4.5 kb), oyster (6.3 kb) and worm (5.2 kb). Hybridisation with different histone probes suggests that, once again, the genes coding for each of the histone proteins are clustered within a repeat unit, in these species.

Recently recombinant DNA techniques have been used to isolate yeast DNA clones containing histone coding sequences, (Hereford and Fahner, cited Kedes, 1979). Two clones isolated from different regions of the yeast genome contain H2A and H2B genes only, and these are on opposite strands of the DNA. This shows that in yeast, at least, the histone genes are not all contained in an ordered unit, as has been found in sea urchins and *D. melanogaster*.

1.4.3.(iv) Vertebrate Histone Genes

Until very recently, the only information available on vertebrate histone genes concerned the reiteration frequency of the genes and some *in situ* hybridisation data. However, the wide variation in arrangement of the histone genes in those species already investigated makes the organisation of the histone genes in higher eukaryotes a particularly interesting area of research.

DNA excess hybridisation experiments have shown the reiteration frequency of the histone genes to be 20-50 fold for *Xenopus laevis*, (Jacob *et al.*, 1976), 10-20 fold for mouse, (Jacob *et al.*, 1976) and 30-40 fold for human DNA (Wilson and Melli, 1977). These numbers are considerably smaller than the reiteration frequencies of the histone genes of several lower eukaryotes, for example, 100-fold in *D. melanogaster*, (Lifton *et al.*, 1977) and 300-400 fold for sea urchin, (Kedes and Birnstiel, 1971).

In situ hybridisation studies using human histone H4 mRNA and cloned sea urchin histone genes have shown that the human histone coding sequences are located on the long arm of chromosome 7, (Yu *et al.*, 1978; Chandler *et al.*, 1978). Since identical results were obtained using probes from different sea urchin genes, it is likely that

the human histone genes are all incorporated into a repeating unit, and these are tandemly arrayed in the genome.

Recently, work in this laboratory has produced some preliminary data on the chicken histone genes.

Histone mRNAs have been isolated from 5-day old chicken embryos and complementary DNA probe prepared from this mRNA has been used to determine the reiteration frequency and the size of the histone gene repeat unit in the chicken genome, (Crawford *et al.*, 1979). (The characterisation of the histone mRNA, and the determination of the histone gene reiteration frequency are described in detail in this thesis).

Harvey and Wells, (1979), have selected a clone containing histone genes from a library of chicken DNA sequences inserted into the lambda phage vector, Charon 4A, (Blattner *et al.*, 1977). This clone contains two regions of histone coding sequences separated by at least 7 kilobases of spacer DNA, and limited DNA-sequencing information and restriction mapping have shown that the H1 and H2A genes are adjacent in this clone. The H1 and H2A genes are adjacent in sea urchin, but separated in the *D. melanogaster* histone gene repeats.

The two DNA strands of the chicken histone clone have been separated on low-percentage agarose gels and then probed with labelled cDNA

prepared from histone mRNA, (Bruschi, 1979). Auto-radiography clearly shows hybridisation to both DNA strands, indicating that histone mRNA is transcribed from both strands of the DNA, a situation similar to that observed for *D. melanogaster*, but different from sea urchin, where all transcription is from the same strand of DNA.

Together, the information from the gene order and the transcription experiments suggests that the organisation of the chicken histone genes differs from that observed in either sea urchin or *D. melanogaster*.

Experiments are in progress to determine the order and polarity of the histone coding sequences. DNA-sequencing will then be carried out so that a detailed comparison may be made between the chicken histone genes and those of other species.

1.4.4 Tissue-specific Histone H5

The nucleated erythrocytes of birds, reptiles, amphibians and fish, contain a basic histone H5 that partially, but not completely, replaces histone Hl in the mature erythrocyte, (Neelin *et al.*, 1964); Champagne *et al.*, 1968; Micki and Neelin, 1975). The histones Hl and H5 show many similarities to each other, and they differ from the other histones in many respects. They are both considerably higher in molecular weight than any of the other histone proteins and they have a much higher lysine content, (Johns, 1971). Despite their very basic

nature, however, they are the first fraction to be extracted from chromatin on raising the ionic strength, (Ohlenbusch *et al.*, 1967). Neither H1 nor H5 is present in the nucleosome core structure.

H1 and H5 both show considerable sequence heterogeneity. H1 is the most variable of the histones, and there are several distinct subtypes of H1 with tissue-specific and species-specific differences in amino-acid sequence, (review, Cole, 1977). Despite the variations in size and sequence observed in the H1 protein, the molecule contains regions in which the sequence has been highly conserved and these are presumably the sites of interaction with the nucleosome core or with DNA. H5 also exhibits sequence variation both within species and between species. H5 isolated from chicken erythrocytes may contain either a glutamine or an arginine residue at amino acid position 1S, (Greenaway and Murray, 1971), and the histone H5 proteins isolated from goose and chicken show differences at 22 of the 160 amino acid positions that have been determined for both molecules, (Seligy *et al.*, 1976).

Conformational studies have shown that H1 and H5 are very similar, with each molecule showing the presence of three well-defined domains, an N-terminal disordered region, a compact globular central region and a long, disordered C-terminal chain. The major difference between the H1 and H5 proteins is that the N-terminal domain of H5 is shorter by about

12 residues and has a much lower net positive charge, (Aviles *et al.*, 1978).

These lysine-rich histones have the ability to act as chromatin cross-linking agents and may in this way be able to effect transcription. For example, chromatin gels in water can be induced to shrink to about 10 percent of their original volume by dialysis against salt solutions, but only in the presence of histone H1, (Bradbury *et al.*, 1973). Pure H1-DNA complexes react in a similar way and NMR data suggests that the H1 molecule undergoes changes in its physical structure during this process, (Bradbury *et al.*, 1974a).

H1 has also been implicated in the contraction of chromatin at metaphase, (Bradbury *et al.*, 1974b). There is evidence that the modification of histones, particularly the phosphorylation of H1, may be responsible for at least some aspects of this contraction. Specific sites on the H1 protein are phosphorylated, (Adler *et al.*, 1971), and this reduces the tendency of the histone to bind to DNA, (Adler *et al.*, 1972). Similarly, histone H5 is also phosphorylated at specific sites and undergoes a conformational change that effects the strength of binding to DNA, (Tobin and Seligy, 1975). This effect may increase histone-histone interactions in the chromatin and so facilitate contraction.

The similarities between the H1 and H5 molecules and their involvement in higher order chromatin structure has prompted the suggestion, (Johns, 1971), that, in species containing nucleated erythrocytes, H5 has evolved to suppress all RNA synthesis and hence protein synthesis; a situation equivalent to the enucleation process seen in mammals. RNA synthesis decreases in red blood cells as they mature, (Attardi *et al.*, 1970), and the chromatin changes from a diffused to a highly condensed state, (Brasch *et al.*, 1971). During this process, H5 is the only histone made in the non-dividing reticulocyte, (Appels and Wells, 1972), until in the mature erythrocyte, about 80 percent of the H1 protein is replaced by H5. Thus H5 may be responsible for a gross form of gene control in erythroid cells by causing a change in the chromatin structure, making it inaccessible to RNA polymerase, and thus terminating all transcriptional activity. It should be noted, however, that this repression is readily reversible. When terminally differentiated red blood cells are fused with HeLa cells, the condensed chromatin rapidly becomes dispersed, (the nuclear volume increases 20-30 fold), and the erythrocyte nucleus recommences RNA and DNA synthesis, (Harris, 1967; Bolund *et al.*, 1969).

Recently, as protein sequences for large portions of H1 (Cole, 1977; Macleod, 1977) and H5, (Sautiere *et al.*, 1975, 1976; Champagne - pers. commun), have become available, a great amount of sequence homology has been detected in these two histones. Yaguchi *et al.*, (1977), have shown that, when the first 111 known residues of H5 histones are compared to H1 sequences, from different species, 63 of these residues correspond to identical residues in at least one of the H1 species. At least 8 other conservative substitutions are found in this region of the H1 and H5 sequences indicating that the homology between the two molecules is perhaps even more extensive. A gap of 12 residues was introduced into the H5 sequence to facilitate the alignment of amino acid residues in the two sequences, and it was observed that the region of this gap corresponds to a homologous repeating sequence in H1. This suggested a deletion or duplication in this region, and that the genes for H1 and H5 histones may have evolved from a common ancestral gene, (Yaguchi *et al.*, 1977).

In the only two systems examined in detail, sea urchin and *D. melanogaster*, the H1 gene is interspersed with the other histone genes, and preliminary evidence suggests that this is also true in the chicken histone gene system, (Harvey and Wells, 1979). Scott and Wells, (1976), have

shown that the H5 gene is reiterated about 10-fold in the chicken genome and Crawford *et al.*, (1979), have shown that the other histone genes are reiterated to about the same extent. The arrangement of the H5 genes, and the relationship between the H5 genes and the genes coding for the other histone proteins, is therefore, a question of considerable interest. It remains to be determined whether the H5 genes are included in the histone gene cluster or whether they are linked to the other histone genes at all. While the normal histone genes appear to lack intervening sequences it will be interesting to find out whether the H5 gene has acquired intervening sequences since the H1 and H5 genes have diverged. Research into these questions and other related questions on gene evolution and regulation of gene expression will require a pure probe for H5 gene sequences.

1.5 Aims of the Project

Using indirect immunoprecipitation techniques, Scott and Wells, (1976), were able to prepare small amounts of pure H5 mRNA, and cDNA copied from this RNA template was used to determine the reiteration frequency of the H5 genes in the chicken genome. In repeated experiments, however, the immunoprecipitation procedure proved to be extremely unreliable, and H5 mRNA could not be reproducibly obtained using this method. The aim of the project

described in this thesis is, therefore, to develop a straightforward physical procedure for the isolation of pure H5 mRNA sequences. Probe prepared from these sequences would be used to examine the relationship of H5 to the other histone genes.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

Chemicals. All chemicals were of analytic reagent grade or of the highest available purity.

Column chromatography materials. Oligo-dT-cellulose (T3 grade) was purchased from Collaborative Research. Poly-U-Sepharose and Sephadex G-50 and Sephadex G-150 were purchased from Pharmacia.

Nitrocellulose filter paper, was obtained from Sartorius.

Enzymes. The enzymes used in the course of this work were obtained from the sources listed below.

Creatine phosphokinase : Sigma.

RNA-dependent DNA-polymerase, (reverse transcriptase), was a gift from J.W. Beard and the N.I.H. Cancer Program.

E. coli DNA-polymerase I : Boehringer, Mannheim.

S_1 nuclease : Boehringer, Mannheim.

Ribonuclease A : Sigma.

E. coli Deoxyribonuclease I : Sigma.

Proteinase K : E. Merck, Darmstadt.

Calf-thymus terminal deoxynucleotidyl transferase, was supplied by D.J. Kemp.

Bacterial alkaline phosphatase, was supplied by R. Richards.

Polynucleotide kinase : Boehringer, Mannheim.

Restriction Enzymes. The restriction endonuclease Eco RI was prepared using basically the method of Yoshimori, (1971). Other restriction enzymes were

obtained from the sources listed below.

Hind III, was supplied by R. Harvey and J.R.E. Wells.

Pst I, was supplied by C.P. Morris.

Hae III : New England Biolabs.

Hpa II : New England Biolabs.

Alu I : New England Biolabs.

Cloned Sequences. The α and β-globin cDNA clones used in the screening experiments were supplied by R. Richards. The keratin cDNA clone, pK23, was supplied by R. Saint.

2.2 METHODS

2.2.1 Preparation of Erythroid Cells

Reticulocytes were obtained from the circulation of White Leghorn-Australorp pullets which had been rendered anaemic by five daily injections of 2.5% phenylhydrazine in 47.5% ethanol, (v/v), pH 7.0. All blood was obtained by heart puncture into NKM, (0.14 M NaCl, 5 mM KCl, 2 mM $MgCl_2$), containing 1mg/ml of heparin. The red cells were washed several times by centrifugation (1000g, 5 minutes), and resuspension into ice-cold NKM. The white-cell layer was removed during this operation by aspiration.

2.2.2 Isolation of RNA

All procedures involving RNA were carried out at $0^{o}C$ using sterile solutions and glassware.

Preparation of chicken reticulocyte 10S RNA.

Basically the method of Pemberton *et al.*, (1972). Washed reticulocytes were lysed by the addition of 10 volumes of 2 mM $MgCl_2$ and vortex mixing for about 5 minutes. 1/20 volume of 10 x TKM, (TKM = 0.01 M Tris-Cl pH 7.0, 0.2 M KCl, 2 mM $MgCl_2$), was then added to restore tonicity, and the nuclei were removed by centrifugation, (10,000g for 10 minutes). This step was repeated to remove any remaining nuclei. Polysomes were collected from the supernatant by centrifuging, (210,000g for 60 minutes) through a pad of 50% (w/v) sucrose in TKM, in a Beckman Ti50 rotor. The polysome pellet was rinsed twice with TK (10 mM Tris-Cl pH 7.0, 15 mM KCl). The polysomes were resuspended in TK buffer and 1/10 volume of 0.3 M EDTA pH 7.6 was added. This dissociated the polysomes into ribonucleoprotein, (RNP) particles which were fractionated by centrifugation, (160,000g, 16 hours), through linear 10-40% (w/v) sucrose gradients in TK, in the Beckman SW41 rotor. The 20S RNPs were collected and recovered from solution by ethanol precipitation. They were then dissolved in 10 mM Tris pH 7.6, 0.2% SDS, heated to 65^{o}C for 10 minutes and fractionated once more on sucrose gradients in NET buffer, (0.1 M NaCl, 10 mM Tris-Cl pH 7.6, 1 mM EDTA). The 10S RNA was collected and stored as an ethanol precipitate.

Preparation of 7-11S RNA from chicken embryos. Basically the method of Seeburg *et al.*, (1977a). About 30 five-day-old chick embryos were snap-frozen in liquid nitrogen, then homogenised in 7 M guanidinium-Cl, 20 mM Tris-Cl pH 7.5, 1 mM EDTA, 1% (w/v) Sarkosyl in a Dounce homogeniser in a final volume of 30 mls. Total RNA from this homogenate was recovered as material centrifuged through 5.7 M CsCl. The clear RNA pellets were resuspended in 10 mM Tris-Cl pH 7.5, 1 mM EDTA, 5% Sarkosyl and 5% phenol, then made 0.1 M in NaCl and extracted with an equal volume of phenol:chloroform (1:1 v/v). RNA from the aqueous phase was collected after ethanol precipitation, resuspended in 10 mM Tris-Cl pH 7.5, 1 mM EDTA, 0.5% SDS, heated at 65^{o}C for 5 minutes, chilled and centrifuged on 10-40% sucrose gradients, (210,000g for 16 hours). The 7-11S RNA was collected and then refractionated on a second 10-40% sucrose gradient.

Preparation of total RNA from chicken red blood cells. (John Brooker - personal communication). Cells were homogenised in a Dounce homogeniser in 6 M guanidinium-Cl, 0.2 M Na-acetate pH 5.2 and 1 mM 2-mercapto-ethanol using about 10 mls of this solution per 1 ml of packed cells. The homogenate was passed through a French pressure cell at 12000 psi and then an equal volume of ethanol was added. After standing at -20^{o}C for at least one hour the suspension was centrifuged at 10,000g for 15 minutes

to recover the insoluble material. The pellet was suspended in half the original volume of 6 M guanidinium-Cl, 0.2 M Na-acetate pH 5.2, 10 mM EDTA and again precipitated by addition of an equal volume of ethanol, (-20^{o}C, 1 hour). The insoluble material was recovered by centrifugation (10,000g for 15 minutes) and the pellet resuspended in 1/4 of the original volume of 7 M urea, 0.1 M Tris-Cl pH 8.5, 0.1 mM EDTA, 0.1% SDS, at room temperature, (to prevent crystallisation of the urea). The suspension was extracted twice with an equal volume of water-saturated phenol/chloroform (1:1) and then the aqueous phase extracted with chloroform. The aqueous phase was adjusted to 0.1 M K-acetate pH 5.0 and the nucleic acid precipitated by the addition of two volumes of ethanol. The pellet was resuspended in 5 mls of ice-cold 2M LiCl. DNA and low molecular weight RNA, (less than about 7S), are soluble in this solution, while the high molecular weight RNA remains insoluble. The RNA is recovered by centrifugation, (15,000g for 15 minutes), washed with ethanol and stored as an ethanol precipitate.

2.2.3 *In vitro* Translation of RNA : Wheat-germ cell-free system

The wheat-germ translation extract was prepared from commercial wheat germ, (Adelaide Milling Co.), using the method of Marcu and Dudock, (1974).

Cell-free translations were carried out using the conditions described by Roberts and Paterson, (1973). The 50 µl reaction mix contained 20 mM HEPES-KOH pH 7.5, 75 mM KCl, 3 mM Mg-acetate, 2 mM DTT, 1 mM ATP, 20 µM GTP, 8 mM creatine phosphate, 4 µg/ml creatine phosphokinase, 5 µCi of ^{3}H-leucine, ^{3}H-lysine or ^{35}S-methionine, 25 µM of the other unlabelled amino-acids and 25 µl of the wheat-germ extract. About 0.1 to 0.5 µg of mRNA was added to each 50 µl translation mix and incubations were carried out at 25^{o}C for 1 hour.

For determination of incorporated radioactivity, aliquots of the translation mix were spotted onto GF/A filters, washed in several changes of 20% TCA and then rinsed with ethanol and ether. After drying, toluene scintillant was added to the filters and the radioactivity determined in a Packard scintillation spectrometer.

For analysis of the labelled products of the cell-free system, the translation mix was precipitated by the addition of 3 volumes of acetone:ether, (3:1), and washed 3 times by centrifugation and resuspension in the same solution. The protein was dissolved in the appropriate loading buffer and analysed on polyacrylamide gels.

2.2.4 Analysis of Labelled Translation Products

Preparation of ^{14}C-labelled marker protein.

Electrophoretically-pure samples of globin, H5 and

total chicken histone were labelled, *in vitro*, with ^{14}C using potassium ^{14}C-cyanate. 10 µCi of K-^{14}C-cyanate was added to 1 mg of protein dissolved in 100 µl of 10 mM Tris-HCl pH 8.8. This was incubated at 45°C for 3 hours and then dialysed against several changes of 10 mM ammonium bicarbonate pH 8.5. The specific activity of the resulting ^{14}C-protein was about 2000 dpm per µg.

SDS-urea tube gels. (Swank and Munkres, 1971). The 0.6 cm diameter by 8 cm long tube gels contained 10% acrylamide, 6 M urea, 0.1% SDS in Tris buffer pH 6.8 and were electrophoresed at 1.5 mA per gel for 16 hours. Electrophoresis buffer was 0.1 M Tris-H_2PO_4 pH 6.8, 0.1% SDS.

Low pH-urea tube gels. (Panyim and Chalkley, 1969). Gels contained 15% acrylamide, 2.5 M urea in 0.9% acetic acid, (v:v), and were electrophoresed at 1.75 mA per gel for 3 hours with the anode on top. Electrophoresis buffer was 0.9% acetic acid.

Analysis of tube gel profiles. For the analysis of labelled proteins on either tube gel system, the gels were frozen with dry ice and sliced into 1 mm slices with a Mickle gel slicer. The slices were incubated at 37°C, overnight, in 0.2 ml of 0.2% SDS solution. 2 ml of toluene-triton scintillation fluid, (62.5% toluene, 0.2% PPO, 0.02% POPOP, 37.5% triton X-114), was added and radioactivity was determined in a Packard scintillation spectrometer.

Discontinuous Slab Gels. (Laemmli, 1970).

Slab gels, (14 cm X 14 cm X 0.3 cm), contained a 3% acrylamide stacking gel and a 12.5% acrylamide separating gel in 0.375 M Tris-HCl pH 8.8, 0.1% SDS. These were electrophoresed for about 2½ hours at 100 V. Electrophoresis buffer was 25 mM Tris-glycine pH 8.3, 0.1% SDS.

Labelled products separated on slab gels were examined by fluorography. The gels were prepared for fluorography essentially as described by Bonner and Laskey, (1974), except that 25% naphthalene, 1% PPO replaced PPO/POPOP as the fluorescing agent. The dried gel was placed into contact with X-ray film in the presence of an intensifying screen, (Ilford, fast tungstate), and exposed at -80^{o}C.

2.2.5 Affinity Chromatography

Oligo-dT-cellulose. 0.2 g of oligo-dT-cellulose was packed into a small water-jacketted column. After flushing with 0.1 N NaOH to destroy ribonuclease, the column was equilibrated with high-salt buffer, 0.3 M NaCl, 10 mM Tris-HCl pH 7.0 and 1 mM EDTA. After heat disaggregation, (80^{o}C for 3 minutes), RNA was loaded onto the column in high-salt buffer and the A_{254} elution profile was followed using a Uvicord column monitor. The unbound RNA was reapplied to the column, re-eluted and stored. The bound RNA was then eluted with 10 mM Tris-HCl pH 7.0, 1 mM EDTA. In general chromatography was carried out at 20^{o}C although

some experiments were performed at 0°C. All RNA fractions were recovered from the elution buffers by ethanol precipitation.

Poly-U-Sepharose. (Lindberg and Persson, 1972). 0.2 g of poly-U-Sepharose was packed into a small, water-jacketted column and equilibrated with 0.3 M NaCl, 10 mM Tris-HCl pH 7.4. After heat disaggregation, RNA was loaded onto the column in high-salt buffer and the elution profile followed with a column monitor. The unbound RNA was re-passaged through the column and collected. The bound RNA was eluted by flushing the column with 90% de-ionised formamide, 1 mM EDTA adjusted to pH 7.4 with Tris base. 3 volumes of water were added to the bound RNA fraction to dilute the formamide and the RNA was then recovered by ethanol precipitation.

In general, the RNA fractions obtained by affinity chromatography procedures were centrifuged on sucrose gradients prior to use in translation experiments.

2.2.6 *In vitro* Synthesis of Labelled DNA

Oligo-dT-primed reverse transcription. Oligo-dT-primed reverse transcription was carried out in a 20 µl reaction mix containing up to 2 µg of mRNA, 1 mM each of dATP, dGTP and dTTP, about 0.1 mM α-^{32}P-dCTP, (10-50 Ci/mmole), 50 mM Tris-HCl pH 8.3, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol and 20 µg/ml

of oligo-dT_{10}. 1 µl of reverse transcriptase was added and the solution incubated at 42°C for 30 minutes. The RNA template was removed by alkaline hydrolysis with 0.3 N NaOH for 30 minutes at 65°C, and the solution neutralised by the addition of HCl to 0.3 M and Tris-HCl pH 7.5 to 0.1 M. The mix was extracted with an equal volume of phenol : chloroform, (1:1) and the aqueous phase loaded onto a 0.4 cm X 4 cm Sephadex G-50 column and eluted with 10 mM Tris-HCl pH 7.6, 1 mM EDTA, to remove the unincorporated nucleotides.

Random-primed reverse transcription. Priming of RNA lacking a 3´ poly-A tract was achieved by the random hybridisation of oligo-nucleotides of salmon-sperm DNA, prepared as described by Taylor *et al.*, (1976). Conditions for the synthesis of this cDNA were exactly as described for the oligo-dT-primed reaction, except that oligo-dT_{10} was replaced by a 2 mg/ml final concentration of oligo-nucleotide, and the mix was incubated at 37°C for 60 minutes. The cDNA synthesised was isolated as described for oligo-dT-primed synthesis.

Nick-translation of double-stranded DNA. Labelling double-stranded DNA using *E. coli* DNA polymerase I was carried out essentially as described by Maniatis *et al.*, (1975). The 50 µl incubation mix contained 50 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 50 µg/ml bovine serum albumin, 5 µM each of ^{32}P-dCTP and ^{32}P-dGTP, (100-200 Ci/mmole)

and unlabelled dATP and dTTP. The DNA was nicked by the addition of 20 pg of *E. coli* DNAase I and the reaction was started by the addition of 2 units of *E. coli* polymerase I. The solution was incubated at 15°C for 90 minutes, phenol/chloroform extracted and the unincorporated nucleotides removed as described for oligo-dT-primed reverse transcription. If the labelled DNA was to be used as a hybridisation probe, the DNA strands were separated by boiling the solution for 2 minutes and then snap-cooling.

2.2.7 Generation of $C_o t$ Curves

5 µl of a 10 mg/ml solution of sonicated DNA in hybridisation buffer, (0.18 M NaCl, 10 mM Tris-HCl pH 7.0, 1 mM EDTA, 0.1% SDS), and containing about 5000 dpm of the appropriate labelled hybridisation probe was sealed into 20 µl micro-capillary tubes. After boiling for 2 minutes to separate the DNA strands, hybridisation was commenced by immersing the capillary tubes in water at 65°C. The capillary tubes were removed from the 65°C water-bath and snap-cooled at times corresponding to appropriate $C_o t$ values. The percentage of the labelled-DNA in double-stranded form was determined using the single-strand specific nuclease S_1, (Vogt, 1973), under the S_1 nuclease assay conditions described by Kemp, (1975). A $C_o t$ curve was usually generated from 12 to 15 separate $C_o t$ point determinations.

2.2.8 Gel-electrophoresis of RNA

Agarose-acrylamide-urea gels, (Schuerch *et al.*, 1975). The 0.6 cm X 8 cm tube gels contained 2.1% acrylamide, 0.6% agarose, 6 M urea in Leoning gel buffer, (36 mM Tris, 30 mM Na-H_2PO_4, 1 mM EDTA, pH 7.8). Electrophoresis was carried out at 4°C for 4 hours at 3 mA per gel. The gels were stained with a 0.01% solution of ethidium bromide in 20 mM Na-acetate and the RNA visualised under UV light.

Acrylamide-urea gels, (Sanger and Coulson, 1975). The tube gels contained either 4% or 6% acrylamide and 8 M urea in 25 mM Tris-glycine pH 8.3. Electrophoresis was carried out at room temperature for 4 hours at 2 mA per gel using 25 mM Tris-glycine pH 8.3 electrophoresis buffer. The gels were stained with toluidene-blue and destained in distilled water, in the dark at 4°C, to prevent degradation of the RNA, (Popa and Bosch, 1969).

Formamide gels, (Pinder *et al.*, 1974). The tube gels were 4% or 6% acrylamide, 20 mM diethyl-barbituric acid pH 9.0, (adjusted with NaOH), in de-ionised formamide. These were overlaid with formamide buffered to pH 9.0 with 20 mM diethyl-barbituric acid and the RNA sample layered under this in buffered formamide containing 5% sucrose. Electrode buffer was 0.02 M NaCl which was circulated constantly to maintain the pH. Gels were electrophoresed at 4 mA per gel for 4 hours at 4°C, stained

with ethidium bromide and destained in water. The RNA was visualised under UV light.

2.2.9 Electrophoretic elution of RNA from acrylamide gels.

During the course of the work concerned with the fractionation of RNA species on polyacrylamide gels an efficient and simple method was developed to electrophoretically elute RNA from slices of polyacrylamide gel.

Slices of polyacrylamide gel, containing RNA, were supported on a piece of gauze located about 5/6 of the way down a perspex gel tube. The bottom of the tube was sealed with 18/32 dialysis membrane to permit the flow of current through the gel tube but to prevent the migration of eluted RNA out of the tube. Electro-elution was carried out at $4^{o}C$ for 30 minutes using a current of 8 mA per gel tube. The electrode buffer was 40 mM Tris-acetate pH 7.8, 30 mM Na-acetate, 20 mM EDTA containing 0.2% SDS. The electro-eluted RNA was recovered from the volume of buffer between the gel slice and the dialysis membrane. This volume was approximately 0.4 ml.

Experiments with ^{32}P-labelled RNA showed that, using these electro-elution conditions, greater than 80% of the RNA could be recovered from the gel slice in intact form. Longer elution times or removal of SDS from the elution buffer resulted in increased breakdown of the eluted RNA.

The 0.4 ml of RNA solution recovered from the

bottom of the elution tube was extracted with an equal volume of phenol/chloroform (1:1) and centrifuged hard, (12000g for 15 minutes), to separate the phases and to sediment any contaminating acrylamide. The RNA was recovered from the aqueous phase by ethanol precipitation and aliquots of the RNA were used in translation experiments.

2.2.10 Restriction and Analysis of DNA

Restriction digest conditions. Restriction endonuclease digestion of DNA with either commercial or non-commercial preparations of enzyme were carried out using the conditions for the appropriate enzyme detailed in the New England Biolabs catalogue. All reactions were stopped by the addition of EDTA to a final concentration of 25 mM. The reaction mix was extracted with an equal volume of phenol/chloroform, (1:1), and the aqueous phase ethanol precipitated.

Agarose gel electrophoresis. Electrophoresis of DNA for analytical purposes or for transfer to nitrocellulose was carried out on 14 cm X 14 cm X 0.3 cm slab gels containing 1% agarose. Electrophoresis buffer consisted of 40 mM Tris-acetate pH 8.2, 20 mM Na-acetate, 1 mM EDTA and electrophoresis was carried out at 60 mA for about 3 hours. DNA was visualised by staining with 0.02% ethidium bromide solution for 15 minutes and examination under UV light.

Polyacrylamide gel electrophoresis. Electrophoresis of DNA species less than about 1 kilobase in length was carried out on vertical 14 cm X 14 cm X 0.3 cm gels containing between 4% and 8% acrylamide. Electrophoresis buffer was 0.09 M Tris-borate pH 8.3, 2.5 mM EDTA and electrophoresis was performed at 150V for 2 hours. DNA was visualised by ethidium bromide staining.

Transfer of DNA to nitrocellulose and hybridisation with a labelled probe. Restricted DNA fractionated on 1% agarose slab gels was transferred to nitrocellulose filter paper using the method of Southern, (1975), as modified by Wahl *et al.*, (1979). In this procedure the rapid transfer of DNA from the gel to the nitrocellulose is facilitated by a partial hydrolysis of the DNA in the gel with 0.25 N HCl, and the transfer is complete within about 2 hours.

The prehybridisation, hybridisation and washing conditions for the nitrocellulose filter in the Southern blot experiments were exactly as described by Wahl *et al.*, (1979). The washed, dried nitrocellulose filter was placed in contact with X-ray film and exposed at -80^{o}C in the presence of an intensifying screen.

2.2.11 Preparation and Tailing of double-stranded cDNA.

Synthesis of the first strand. Random-primed synthesis of the first strand on the 10S RNA template

was carried out as described in section 2.2.6, except that the concentration of the labelled nucleotide was increased to 1 mM with cold nucleotide, to ensure maximum cDNA synthesis. After alkaline hydrolysis of the RNA template and neutralisation, the mix was passaged through a Sephadex G-50 column to remove the unincorporated nucleotides and the oligonucleotide primer sequences. The cDNA was ethanol precipitated in preparation for the second strand synthesis reactions.

Synthesis of the second strand. The second strand synthesis reaction was carried out with *E. coli* DNA polymerase I, using basically the method of Seeburg *et al.*, (1977b). The first strand cDNA was resuspended in a 50 µl reaction volume containing 50 mM Tris-HCl pH 8.1, 20 mM KCl, 7 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.1 mM EDTA, 400 µM each of dATP, dCTP, dGTP and dTTP and 15 units of DNA polymerase I. Unlike Seeburg *et al.*, the reaction was carried out at 15°C for 6 hours rather than at 40°C for 2 hours.

S_1 cleavage of ds cDNA. The second strand reaction mix was diluted directly into a 200 µl final volume containing 30 mM Na-acetate pH 4.6, 0.3 M NaCl, 4.5 mM $ZnCl_2$ and 130 units of S_1 nuclease. The S_1 digestion mix was incubated at 37°C for 30 minutes and then extracted with an equal volume of phenol/chloroform, (1:1). The aqueous

phase was passaged through Sephadex G-50 and the ds cDNA collected.

Size selection and tailing of ds cDNA. The ds cDNA eluate from the Sephadex G-50 column was loaded directly onto a 10-40% sucrose gradient and centrifuged at 160,000 g for 16 hours. 0.5 ml fractions were collected across the gradient and different size classes of ds cDNA were pooled and ethanol precipitated.

Poly-dC nucleotide tails were added to the ds cDNA using calf thymus terminal deoxynucleotidyl transferase. 500 pmoles of ^{3}H-dCTP was dried down and resuspended in a 25 µl volume containing 0.1 M Na-Cacodylate pH 7.4, 40 mM Tris-HCl pH 7.4, 1 mM $CoCl_2$, 0.2 mM DTT and 0.07 pmoles of ds cDNA. 0.5 µl of terminal deoxynucleotidyl transferase was added and the reaction, at 4^{o}C, followed by the conversion of the ^{3}H-dCTP to a trichloroacetic acid-insoluble form. When an average of 10-20 nucleotides per end had been added, the reaction was stopped by the addition of EDTA to a final concentration of 5 mM.

2.2.12 Annealing Tailed cDNA and Vector DNA and Transformation.

Annealing to vector. 0.04 pmoles of pBR 322 DNA, cleaved with Pst I and tailed with deoxyguanosine, as described above, was annealed to an equimolar amount of deoxycytosine-tailed cDNA in 0.2 M NaCl, 10 mM Tris-HCl pH 8.2 by heating for 10 minutes at

65°C, incubating for 1 hour at 45°C and finally allowing the solution to cool slowly to 4°C. The annealed DNA was stored at this temperature.

Transformation of *E. coli*. *E. coli* strain ED8654:$r_k^- m_k^+$ *sup E sup F trp R* was grown overnight at 37°C in Luria broth and then diluted 1/50 into fresh Luria broth and grown to an A_{600} of 0.6-0.8. The cells were chilled on ice for 30 minutes, pelleted by centrifugation and resuspended in 1/2 volume of ice-cold 0.1 M $MgCl_2$. The cells were pelleted immediately and resuspended in 1/20 of the original volume of ice-cold 0.1 M $CaCl_2$. The cells were kept on ice for at least one hour. 0.2 ml of these competent cells was added to 0.1 ml of the DNA in 0.1 M Tris-HCl pH 7.4 and stood on ice for 30 minutes, with occasional stirring. The cells were heat-shocked at 42°C for 2 minutes, kept on ice for a further 30 minutes and then warmed slowly to room temperature. 0.5 ml of L-broth was added to the transformed cells and incubated at 37°C for 20-30 minutes. The transformed cells were mixed with 3 mls of 0.7% L-agar and plated on 1.5% L-agar plates containing 15 µg/ml tetracycline. These were incubated overnight at 37°C.

2.2.13 Detection and Examination of Recombinants

Colony Screening. (Grunstein and Hogness, 1975). Colonies from a transformation were transferred by

toothpick to a master plate and to a sheet of nitrocellulose that had been boiled three times in distilled water and lain onto an L-agar plate containing 15 µg/ml tetracycline. The colonies were grown overnight on the nitrocellulose at 37°C, and the colonies lysed by transferring the nitrocellulose sequentially onto 3 MM paper saturated with 0.5 N NaOH for 7 minutes, 1 M Tris-HCl pH 7.4 for 2 minutes, 1 M Tris-HCl pH 7.4 for 2 minutes and 1.5 M NaCl, 0.5 M Tris-HCl pH 7.4 for 4 minutes. The nitrocellulose filter was washed in 95% ethanol and then baked at 80°C, under vacuum, for 2 hours. Hybridisation and washing conditions were as described for Southern blot experiments. In some cases, after the initial hybridisation and detection of colonies, annealed probe was removed from the filters by boiling for 10 minutes in 2 changes of distilled water. The hybridisation procedure was then repeated using a different labelled hybridisation probe.

Miniscreen examination of plasmid recombinants. Isolation of plasmid DNA from small cultures was carried out as follows. 1.5 ml cultures of each recombinant were grown overnight in L-broth containing 15 µg/ml tetracycline. The cells were pelleted by centrifugation for 30 seconds in an Eppendorf centrifuge, washed in 10 mM Tris-HCl pH 9.0, 1 mM EDTA, pelleted again and resuspended in 150 µl of 15% sucrose, 50 mM Tris-HCl pH 9.0 and 50 mM EDTA.

50 µl of freshly prepared 4 mg/ml lysozyme solution was added and the solution incubated at room temperature for 15 minutes and at 0°C for 30 minutes. 200 µl of ice-cold water was added and the solution heated to 72°C for 15 minutes. The solution was centrifuged at 30,000 g for 20 minutes and the supernatant removed from the white, filamentous pellet and ethanol precipitated. The plasmid DNA was resuspended in water and an aliquot removed for linearisation with Eco RI restriction enzyme in a 200 µl digestion mix. The DNA was extracted with an equal volume of phenol/chloroform, ethanol precipitated and electrophoresed on 1% agarose slab gels as described.

2.2.14 Large-scale Preparation of Recombinant Plasmid DNA.

500 ml cultures of recombinant cells were grown in L-broth to an A_{600} of 1.0 and then chloramphenicol was added to a final concentration of 150 µg/ml. The cells were incubated overnight at 37°C to allow amplification of the plasmid DNA, (Clewell, 1972). The cells were pelleted by centrifugation, (10,000 g for 5 minutes), and resuspended in 10 mls of 15% sucrose, 50 mM Tris-HCl pH 9.0, 50 mM EDTA. 2 mls of 6 mg/ml lysozyme solution was added and the cells were incubated at room temperature for 15 minutes and at 0°C for 30 minutes. 10 mls of boiling-hot 50 mM Tris-HCl pH 9.0, 50 mM EDTA was added with vigorous mixing and the solution was heated at 72°C

for 15 minutes. The lysate was centrifuged at 30,000 g for 30 minutes, the supernatant removed and treated with DNAase-free RNAase A, (20 µg/ml), for 30 minutes at 37°C and Proteinase K, (50 µg/ml), for 30 minutes at 37°C. The solution was extracted with an equal volume of phenol/chloroform and the aqueous phase ethanol precipitated. The plasmid DNA was resuspended in 2 mls of water, adjusted to 0.2 M NaCl and chromatographed through a 2.5 cm X 20 cm Sephadex G-150 (medium) column equilibrated with 0.2 M NaCl, 10 mM Tris-HCl pH 7.6, 1 mM EDTA, to separate high-molecular weight DNA from degraded RNA. The material eluting at the exclusion volume was collected and examination on gels showed this to be plasmid DNA free of RNA and containing not greater than 10% contamination with bacterial chromosomal DNA. Using this procedure it was routinely possible to prepare more than 1 mg of plasmid DNA from 500 ml of bacterial culture.

2.2.14 Preparation of DNA fragments for Sequencing and DNA Sequence Determination.

All DNA fragment preparation and DNA-sequencing procedures were carried out exactly as described by Maxam and Gilbert, (1977, 1979), with the exception of the (G+A)-specific chain cleavage reaction which was performed with 100% formic acid, (R. Richards - personal communication), rather than with pyridinium formate pH 2.0, and the C and (C+T) stop reactions

which were carried out by adding 200 µl of 0.1 mM EDTA and 200 µl of 0.3 M NaCl, 0.1 mM EDTA respectively, (Busslinger *et al.*, 1979). End-labelling was achieved using polynucleotide kinase and γ-^{32}P-ATP that had been prepared by R. Richards. DNA labelled at a single site, and therefore suitable for sequencing, was obtained by strand separation rather than by cleavage of the DNA fragment with a second restriction enzyme.

2.2.15 Containment facilities.

All work involving recombinant DNA was carried out under C3/EK1 containment conditions for work involving viable organisms and CO containment conditions for work not involving viable organisms, as defined and approved by the Australian Academy of Science Committee On Recombinant DNA and by the University Council of the University of Adelaide.

CHAPTER 3

AFFINITY CHROMATOGRAPHY FRACTIONATION OF GLOBIN AND H5 mRNAS

3.1 INTRODUCTION

It had been shown previously that although all histones were expressed in dividing erythroblasts, H5 was the only histone synthesised in non-dividing reticulocytes, (Appels and Wells, 1972). Using the indirect immunoprecipitation procedure of Shapiro *et al.*, (1974), Scott obtained a polysome fraction from reticulocytes which, on the basis of *in vitro* translation studies, contained H5 mRNA, (Scott and Wells, 1976). Because the immunoprecipitation approach lacked reproducibility, (Scott and Wells, 1978), and was not suited to routine isolation of H5 mRNA, an alternate physico-chemical method for purification of H5 mRNA was considered essential.

With the microgram quantities of H5 mRNA available, Scott was able to show that it was 10S in size and lacked a poly-A tract on the 3´ end of the RNA. The absence of poly-A sequences on the RNA was demonstrated in two ways, firstly by the failure of the mRNA to bind to oligo-dT-cellulose affinity columns, and secondly by the failure of the mRNA to act as a template for cDNA synthesis when oligo-dT was used as a primer. In this respect H5 mRNA resembles other histone mRNAs, since histone mRNAs isolated from a number of species do not bind to oligo-dT cellulose and therefore apparently lack any substantial poly-A tract, (Adesnick and Darnell, 1972; Destree *et al.*, 1977; Burckhardt and Birnstiel,

1978). The oligo-dT-cellulose chromatography evidence is supported by experiments which demonstrate the absence of ribonuclease-resistant poly-A-containing oligonucleotides in radioactively-labelled histone mRNA, (Grunstein *et al.*, 1976).

The isolation of pure H5 mRNA from reticulocytes is complicated by the fact that H5 mRNA and globin mRNA are both 10S in size. The major protein product of the reticulocyte is globin, which constitutes greater than 90 percent of all protein synthesised in the cell, (Sadgopal and Kabat, 1969). Isolated 10S mRNA from reticulocytes yields greater than 95 percent globin protein when translated in cell-free systems, (Scott, 1975; Knöchel, 1975), inferring that by far the largest proportion of mRNA present in the 10S fraction is globin mRNA. It has been estimated, (Scott and Wells, 1975), that only about 4 percent of the polysomal mRNA consists of H5 mRNA sequences.

Many eukaryotic mRNAs have a polyadenylic acid tract at their 3´ terminus, (Lee *et al.*, 1971; Edmonds *et al.*, 1971; Darnell *et al.*, 1971), and it has been shown that the majority of globin mRNA sequences, isolated from chicken reticulocytes, bind to oligo-dT affinity columns, due to the presence of poly-A tracts on the RNA, (Knöchel, 1975; Scott, 1975).

The available information suggests that a very useful enrichment for H5 mRNA sequences would be achieved by fractionating chicken reticulocyte 10S RNA according to its poly-A content. While all the H5 mRNA should be found in the poly-A minus fraction, the majority of the globin mRNA should be present in the poly-A plus fraction. While an absolute purification of H5 mRNA sequences is most unlikely to be achieved using affinity chromatography techniques alone, the non-polyadenylated RNA should be considerably enriched for H5 mRNA, representing a useful first step in a purification protocol.

Before attempting to enrich for H5 mRNA using affinity chromatography, it is necessary to prepare large amounts of 10S RNA from chicken reticulocytes.

Methods for preparation of mRNA fall, broadly, into two classes either involving the isolation of total cellular nucleic acids from whole cells, prior to a size selection of RNA, or involving the preparation of polysomes before selecting mRNA. Since H5 mRNA is to be isolated from the non-polyadenylated RNA fraction, which will also include ribosomal RNA and ribosomal RNA breakdown products, it is essential that the 10S RNA used in these experiments should contain the minimum possible amounts of degraded RNA. While a number of preparative techniques were tried, the highest yields of undegraded RNA were

obtained using the method of Pemberton *et al.*, (1972). In this approach, ribonuclease degradation of mRNA is minimised by only partially dissociating polysomes, into ribonucleoprotein particles, in the first step of the procedure. Ribonucleoprotein particles containing mRNA are then selected from a sucrose gradient before being dissociated with SDS and recentrifuged so that specific size-classes of RNA can be isolated. Using this method, large yields of intact 10S RNA could routinely be obtained from chicken reticulocytes.

Since H5 mRNA is only a minor mRNA species in the chicken reticulocyte, a sensitive and efficient assay system is required if H5 mRNA is to be detected in unfractionated 10S RNA and in the various RNA subfractions produced by affinity chromatography. The wheat germ cell-free translation system of Roberts and Paterson, (1973), was chosen for use in all routine assays because it can be programmed with very small amounts of mRNA and has a very low background, essential if minor translation products are to be detected and quantitated easily. In addition, the wheat germ extracts are simple to prepare and retain full activity even after prolonged storage.

The wheat germ translation system was optimised for translation of chicken 10S mRNA and analysis of the labelled translation products

was carried out on polyacrylamide gels. The assay of the labelled translation products was performed by slicing and counting polyacrylamide tube gels rather than fluorography of slab gels for two main reasons, firstly because the detection of small amounts of tritium-labelled protein may require an extremely long exposure time using fluorography, while being readily detectable by liquid scintillation counting techniques and, secondly, because it is very difficult to quantitate the distribution of a protein product amongst various fractions using fluorography. On the other hand, quantitation of even small amounts of labelled protein is relatively simple using radioactivity profiles.

Having established a reproducible method for the preparation of reticulocyte 10S RNA and a reliable cell-free translation system, affinity chromatography fractionation of the reticulocyte RNA was investigated.

A number of techniques have been developed utilising the poly-A segment as a specific method for the selection of mRNA. Materials used include Millepore filters, (Lee *et al.*, 1971), unmodified cellulose, (Schutz *et al.*, 1972), poly-U bound to filters, (Kates, 1973), to cellulose, (Kates, 1970) or to Sepharose, (Lindberg and Persson, 1972) and oligo-dT bound to cellulose, (Aviv and Leder, 1972).

The more efficient methods involve the formation of stable base-pairing between the poly-A tail of the mRNA and complementary nucleotide bases, (poly-U or oligo-dT), attached to an inert support.

Since the original observation that H5 mRNA was not polyadenylated was made using oligo-dT-cellulose chromatography, (Scott and Wells, 1976), and since oligo-dT cellulose is a reliable and efficient chromatography material, this was chosen for use in the initial fractionation experiments.

While in theory the separation of polyadenylated globin mRNA from nonpolyadenylated H5 mRNA should be a simple procedure, in practice no well-defined localisation of H5 mRNA to the unbound fraction could be achieved. Despite repeated attempts, using a number of different chromatography conditions, it was not possible to fractionate H5 and globin mRNA sequences using oligo-dT cellulose.

It had been suggested, (G. Partington - pers. commun.), that poly-U-Sepharose was a more sensitive and reliable chromatography material than oligo-dT cellulose, and so the fractionation experiments were repeated using this material. The results obtained using poly-U-Sepharose closely resembled those obtained with oligo-dT cellulose and, once again, no effective fractionation of H5 and globin mRNA sequences could be achieved.

This chapter describes the preparation of 10S RNA from reticulocytes, the optimisation of the wheat germ cell-free translation system for reticulocyte 10S RNA and the experiments performed using affinity chromatography techniques in an effort to isolate a fraction of RNA enriched for H5 mRNA sequences.

3.2 RESULTS

3.2.1 Preparation of 10S RNA from reticulocytes

Reticulocytes were obtained from the blood of chickens, made anaemic by phenylhydrazine treatment, and 10S RNA was isolated from these cells using essentially the method of Pemberton *et al.*, (1972). A number of small modifications have been made to the original method and these are described in detail in chapter 2.

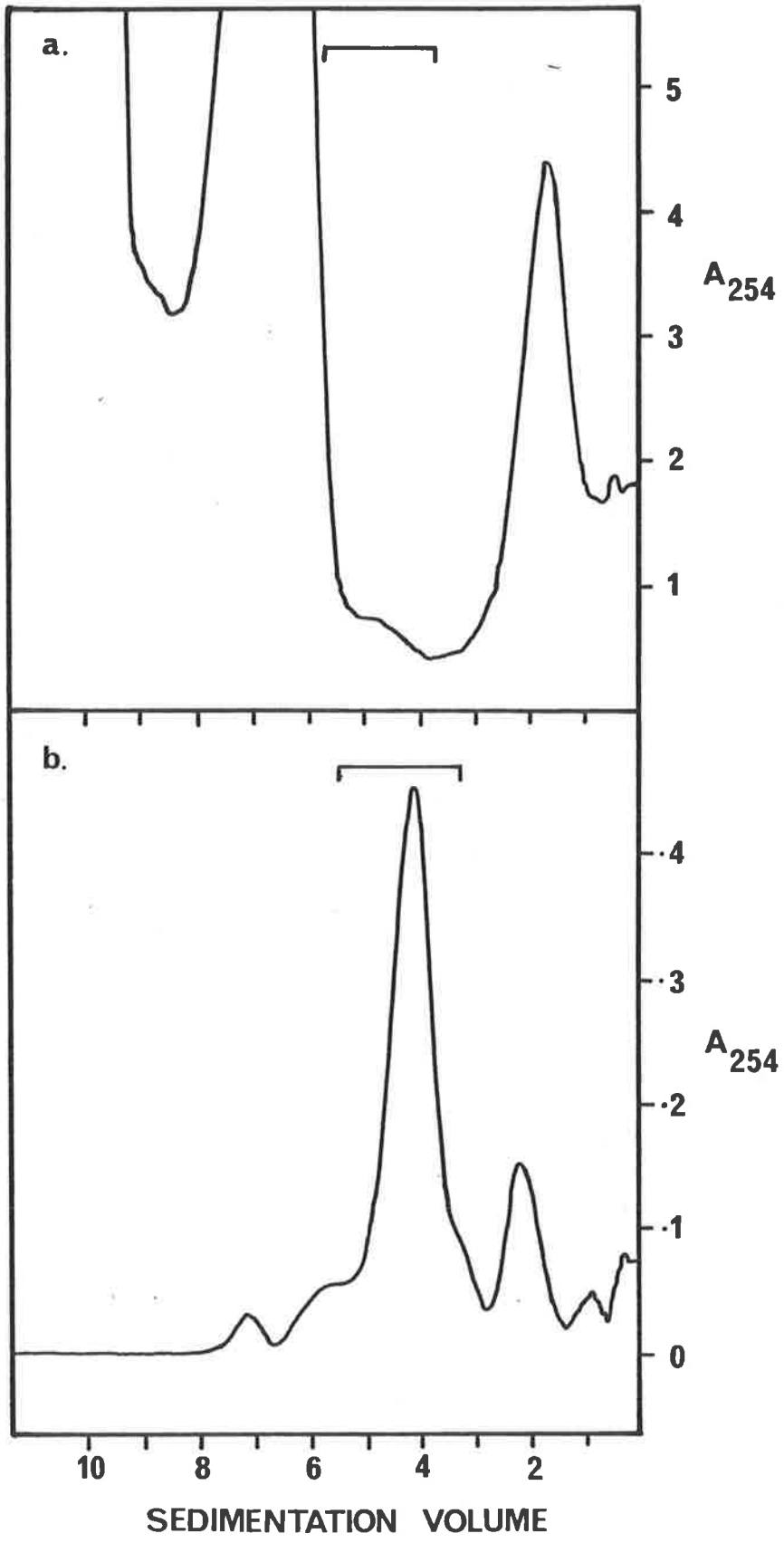
Absorbance profiles obtained when ribonucleoprotein particles and isolated RNA are fractionated on sucrose gradients are shown in figure 3.1a and 3.1b respectively. The 10S RNA represents a single sharp peak on the sucrose gradient profile and this material can be collected, effectively free of tRNA, 5S RNA and ribosomal RNA contaminants. Using this RNA preparation procedure, yields of 0.5 to 1.0 micrograms of 10S RNA per millilitre of anaemic blood could routinely be obtained.

Figure 3.1

Preparation of 10S reticulocyte RNA from ribonucleoprotein particles.

a. Fractionation of RNP-particles on a 10-40% sucrose gradient.

b. Fractionation of RNA obtained from 20S RNPs on a 10-40% sucrose gradient.


a.
5
4
3
2
1
A_{254}
b.
·4
·3
·2
·1
0
A_{254}
10
8
6
4
2
SEDIMENTATION VOLUME

3.2.2 Cell-Free Translation and Analysis of Labelled Protein Products

All *in vitro* translations were carried out using the wheat germ cell-free translation system of Roberts and Paterson, (1973). The wheat germ extract itself was prepared as described by Marcu and Dudock, (1974).

3.2.2(i) Translation of reticulocyte 10S RNA

Addition of 10S reticulocyte RNA to the wheat germ translation system stimulated the incorporation of ^{3}H-leucine into high molecular weight protein products.

Labelled protein was recovered from the translation reaction mix by precipitation with acetone/ether 3:1, and was washed several times before resuspending in the appropriate gel loading buffer. Control experiments using purified ^{14}C-labelled H5 and globin protein had shown that, in the presence of carrier protein, 100 percent of the H5 and globin protein could be recovered from solution using this method.

Figure 3.2 shows the radioactivity profiles obtained when 10S RNA translation products were electrophoresed on SDS-urea gels, (Swank and Munkres, 1971) and on low pH-urea gels, (Panyim and Chalkley, 1969), in the presence of ^{14}C-labelled H5 marker protein.

In both profiles, the major translation product is globin protein, (Scott, 1975). The

Figure 3.2

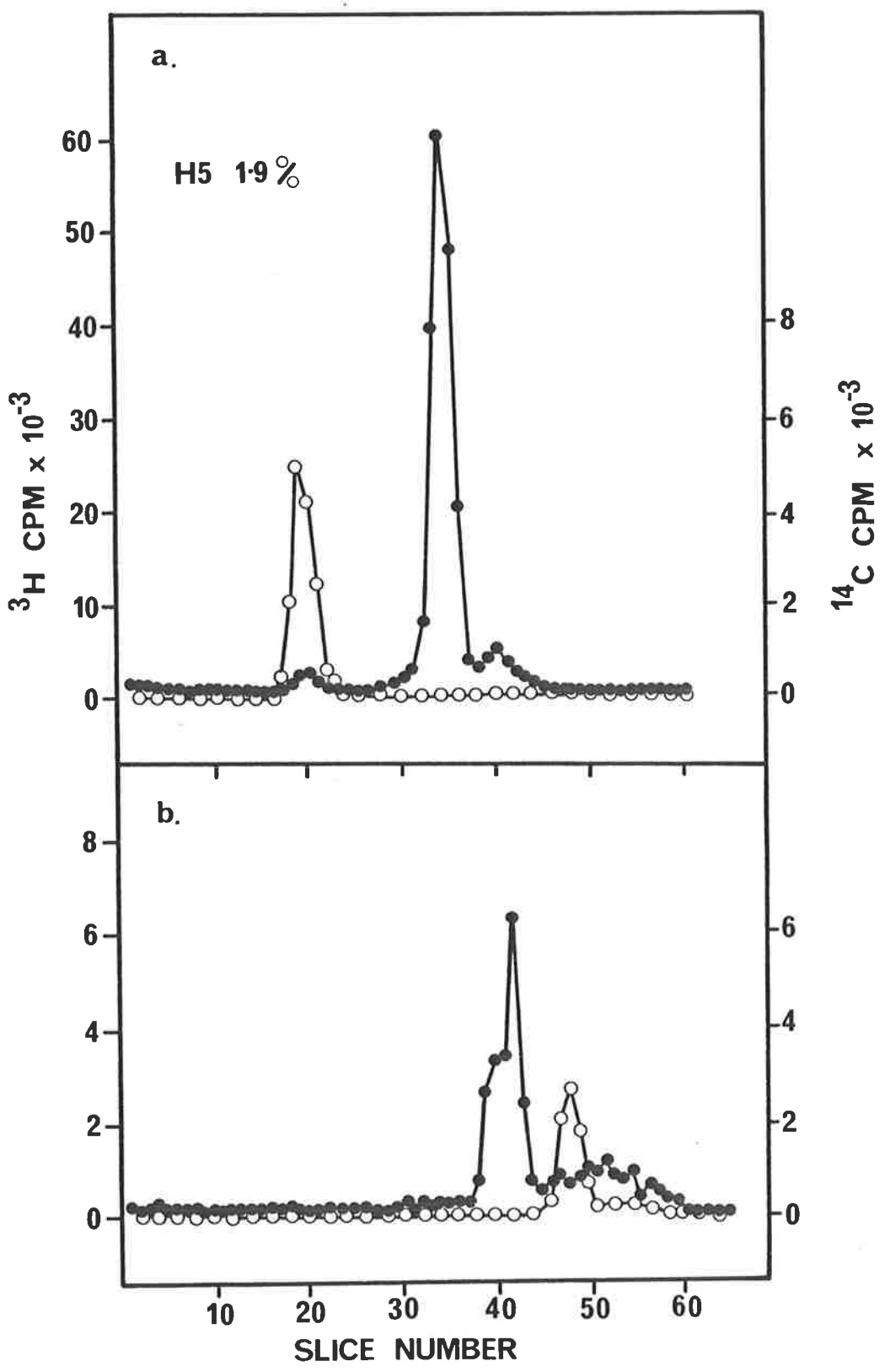
Analysis of the *in vitro* translation products of reticulocyte 10S RNA. mRNA was translated in the wheat-germ cell-free system and the ^{3}H-leucine labelled products examined by electrophoresis on,

a. SDS-urea polyacrylamide gels.

b. low pH-urea polyacrylamide gels.

●—● ^{3}H-labelled translation products.

o—o ^{14}C-labelled H5 standard.


a.
H5 1·9 %
b.
^{3}H CPM x 10^{-3}
^{14}C CPM x 10^{-3}
SLICE NUMBER

SDS-urea gel, (figure 3.2a), shows the presence of a second, small peak, comprising about 2 percent of the labelled products, which corresponds in position to the H5 marker protein. The shoulder of labelled material on the leading edge of the major globin peak probably consists of premature termination products of the globin protein chains, however, centrifugation of the translation mix, to remove polysomes prior to preparation of the sample, does not significantly effect the size of this peak. The nature of this material is therefore uncertain.

The low pH-urea gel, (figure 3.2b), again shows a major peak of globin protein, but also shows a number of smaller peaks electrophoresing at about the position of the H5 marker. These peaks are probably also prematurely terminated globin chains and, unfortunately, are always present in sufficient amounts to make detection of small amounts of H5 protein impossible using this gel system.

When reticulocytes are pulse-labelled with ^{14}C-leucine, the radioactivity recovered in H5 indicates that the amount of H5 protein synthesised is about 4 percent that of globin, (Scott and Wells, unpublished). Although Scott was unable to detect H5 in cell-free translation products from un-enriched 10S RNA, the translation system used in these experiments lacked sensitivity, (see section 3.2.3). With

the more sensitive assay system used here, it was possible to detect a translation product from 10S RNA, constituting about 2 to 4 percent of the total *in vitro* translation products, (for effect of labelled amino-acid see section 3.2.2(ii)), which co-electrophoresed with marker H5 protein in the SDS-urea gel system. This was entirely consistent with the pulse-labelling results mentioned above. When detected in the translation products of enriched RNA fractions, Scott, (1975), identified this protein peak as H5.

Therefore, since H5 protein is readily detectable amongst the translation products of 10S RNA when these are analysed on SDS-urea gels, and because H5 protein electrophoreses well away from other unidentified peaks on this gel system, SDS-urea gels were used to assay cell-free translation products in all further experiments, (following the routine procedure of Scott, (1975)).

3.2.2(ii) Labelled amino-acid

^{3}H-leucine was the labelled amino-acid used in the preliminary translation experiments, and, because the amino-acid compositions of globin and H5 are not identical, the relative areas under the peaks on the radioactivity profile do not necessarily indicate the relative amounts of protein translated. Table 3.1 shows the amino-acid composition of H5 and of chicken α and β globin. The globin chains contain about twice as many leucine residues,

Table 3.1

Amino acid composition of chicken globin chains and chicken histone H5. The amino acid composition data was obtained from the following sources:

chicken H5; Sautiere *et al.*, (1975).

chicken α_A globin; Matsuda *et al.*, (1971)

chicken α_D globin; Takei *et al.*, (1975)

chicken α_S globin; Richards and Wells, (1980)

chicken β globin; Matsuda *et al.*, (1973)

Amino acid	H5	GLOBIN CHAIN			
		α_S	α_A	α_D	β
Asp	2	6	7	6	6
Thr	6	11	8	7	7
Ser	26	6	7	8	7
Glu	3	4	6	9	7
Pro	13	6	5	5	5
Gly	9	11	10	7	8
Ala	29	18	17	17	16
Val	8	11	11	11	12
Met	1	1	1	4	1
Ile	6	7	7	2	7
Leu	8	15	17	15	18
Tyr	3	4	3	5	2
Phe	1	7	8	7	8
Lys	44	13	13	11	10
His	3	10	9	6	7
Arg	21	2	3	4	6
Trp	-	-	-	1	4
Cys	-	2	2	1	3
Gln	5	3	2	8	5
Asn	1	4	5	7	7
Total	189	141	141	141	146

per mole of protein, as histone H5 and therefore, the labelled translation product profiles underestimate, by about half, the amount of H5 protein translated, relative to globin.

H5 contains a much higher molar ratio of lysine than globin, (Table 3.1), and so ^{3}H-lysine would appear to be a better label than leucine for detecting small amounts of H5 protein. When ^{3}H-lysine was used in the wheat germ translation system, (figure 3.3), the relative amounts of radioactivity appearing in the H5 and globin peaks changed, approximately as expected, but the total number of counts incorporated into translation products was much less than that obtained when using ^{3}H-leucine, and so the sensitivity of the system for detecting H5 protein was actually reduced. Consequently, ^{3}H-leucine was used in all further translation experiments.

3.2.2(iii) Determination of Optimum Salt Concentration

It has been reported that different mRNAs require different salt concentrations for optimal translation in cell-free systems, (Efron and Marcus, 1973; Schmeckpeper *et al.*, 1974). This has been observed even for mRNAs that would normally be expressed in the same cell, (Palmiter, 1974). Consequently, the salt optimum for total 10S RNA translation was determined, (figure 3.4),

Figure 3.3

Analysis of the ^{3}H-lysine labelled translation products of reticulocyte 10S RNA on SDS-urea polyacrylamide gels.

●—● ^{3}H-labelled translation products

○—○ ^{14}C-labelled H5 marker

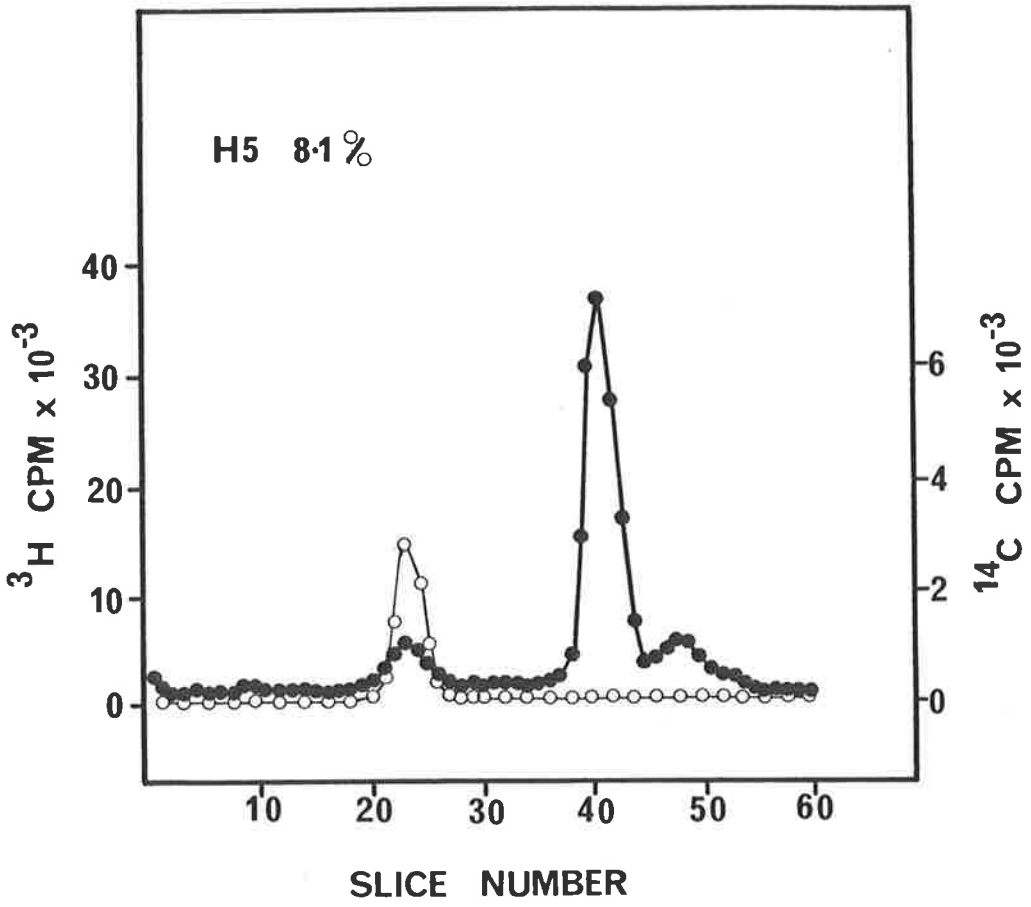
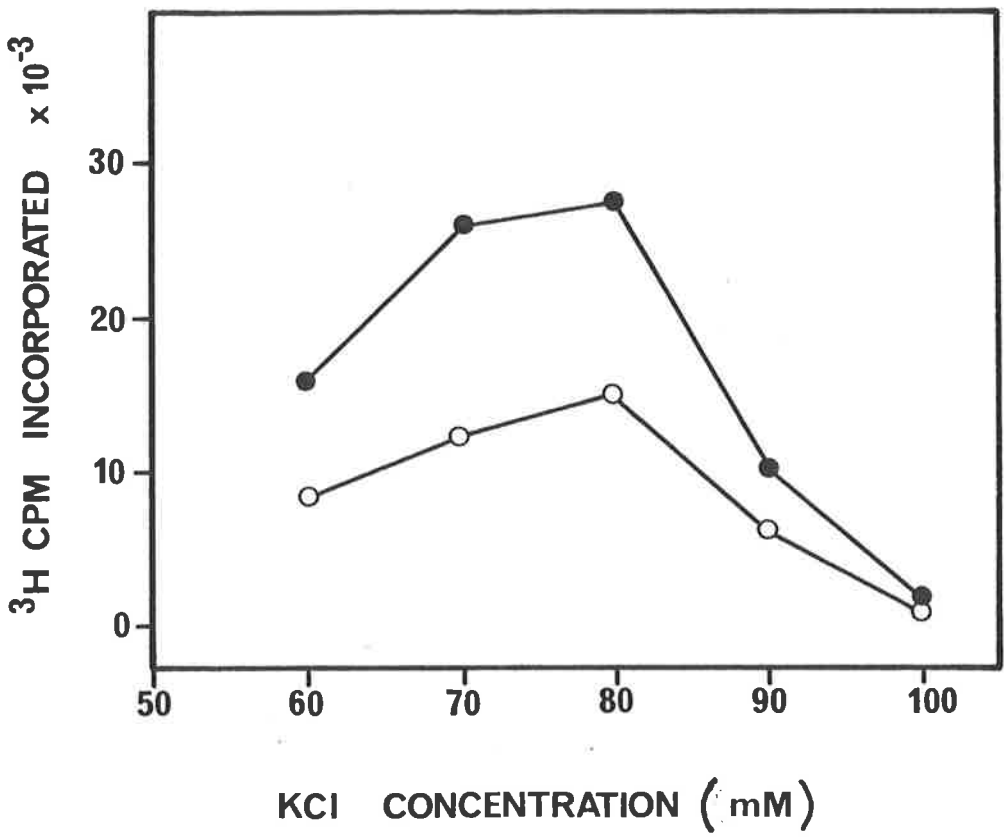

H5 8·1 %
40
30
20
10
0
^{3}H CPM x 10^{-3}
6
4
2
0
^{14}C CPM x 10^{-3}
10
20
30
40
50
60
SLICE NUMBER

Figure 3.4

Effect of varying KCl concentration on the efficiency of translation of reticulocyte 10S RNA. The RNA was translated with a constant $MgCl_2$ concentration, (3 mM), but at varying KCl concentrations.

●—● Wheat-germ preparation 1.

○—○ Wheat-germ preparation 2.


^{3}H CPM INCORPORATED x 10^{-3}
30
20
10
0
50
60
70
80
90
100
KCl CONCENTRATION (mM)

and then the translation products synthesised at each salt concentration were examined independently on SDS-urea gels, to determine whether H5 and globin had different salt optima. These translation product profiles, (shown in figure 3.5), indicate that full-length globin and H5 proteins were produced at all salt concentrations and that the relative proportion of H5 to globin remained constant, suggesting that the salt optimum was the same for both mRNAs. Therefore, all subsequent translations were performed at the optimum salt concentration for total 10S RNA to ensure the maximum incorporation of label. In addition, the levels of mRNA used in translation experiments were non-saturating, (less than about 0.5 microgram of mRNA per translation mix), to minimise the effects of any possible differential translation efficiencies of the mRNAs.

In the final, optimised wheat germ translation system, using 5 μCi of ^{3}H-leucine per 50 microlitre incubation mix, it was routinely possible to incorporate about 2000 d.p.m. of ^{3}H-leucine into labelled protein per nanogram of added 10S RNA.

In all cases, it is most important to remember that the relative amount of each protein synthesised in a translation system will only reflect the relative amount of each mRNA added

Figure 3.5

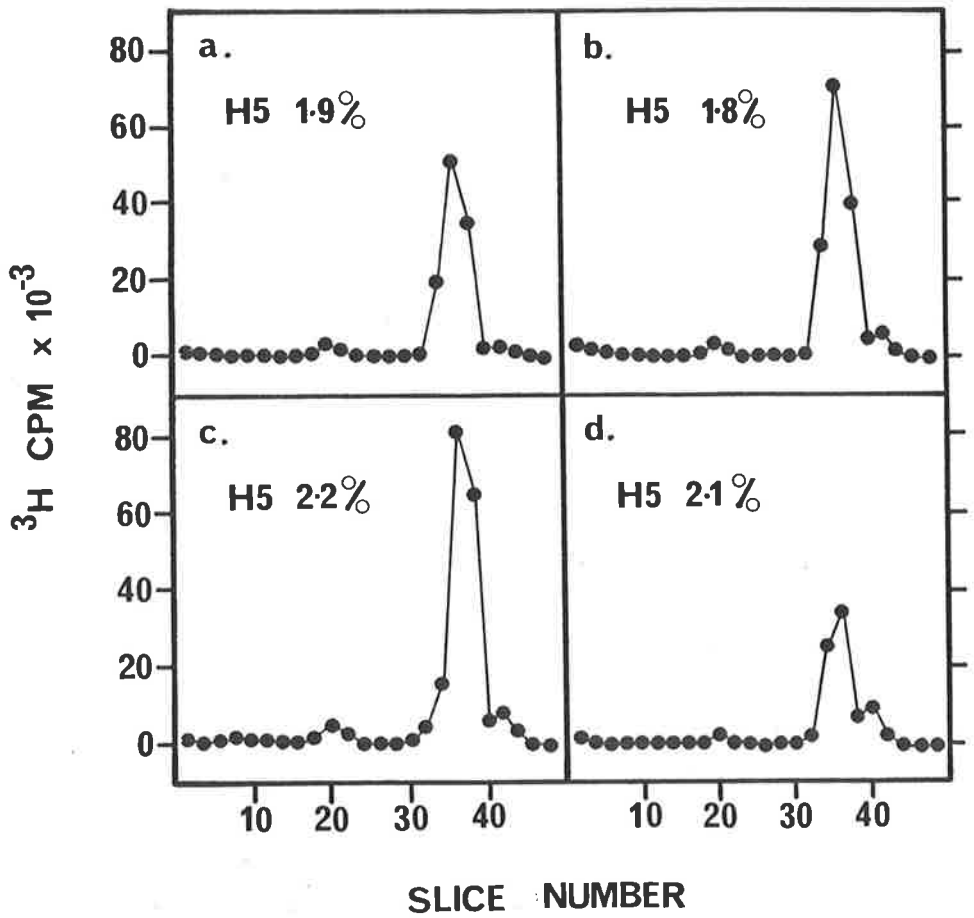
Analysis of the *in vitro* translation products synthesised at different KCl concentrations, (see figure 3.4).

a. 60 mM KCl

b. 70 mM KCl

c. 80 mM KCl

d. 90 mM KCl


a.
H5 1·9%
b.
H5 1·8%
c.
H5 2·2%
d.
H5 2·1%
80
60
40
20
0
10
20
30
40
^{3}H CPM x 10^{-3}
SLICE NUMBER

to the system, if all the mRNAs have identical translational activity. In fact, there are many instances where mRNAs that are normally expressed in the same cell have different translational activities in cell-free systems, for example, rabbit α globin mRNA translates more efficiently than rabbit β globin mRNA in the wheat germ system, (Roberts and Paterson, 1973) and ovalbumin mRNA always translates more efficiently than conalbumin in the reticulocyte lysate system, (Palmiter, 1974). Therefore, the relative amounts of protein translated in cell-free systems is, at best, only an indication of the relative amounts of mRNA present.

3.2.3 Oligo-dT-Cellulose Chromatography

Total 10S RNA from reticulocyte polysomes was prepared as described previously and then passaged through an oligo-dT-cellulose column as described in chapter 2. A typical column elution profile is shown in figure 3.6, and shows the presence of a small peak of RNA that passes straight through the column without binding, and a larger peak of RNA that is initially bound to the oligo-dT-cellulose. The bound and unbound fractions were then centrifuged on sucrose gradients, (figure 3.7), and samples of RNA from each of the fractions used to programme the wheat germ translation system.

Figure 3.6

Uvicord elution profile of reticulocyte 10S RNA chromatographed on oligo-dT-cellulose. (The A_{254} scale is logarithmic and so the areas under the peaks do not indicate the relative amount of RNA in each fraction). Bound and unbound material was collected as indicated.

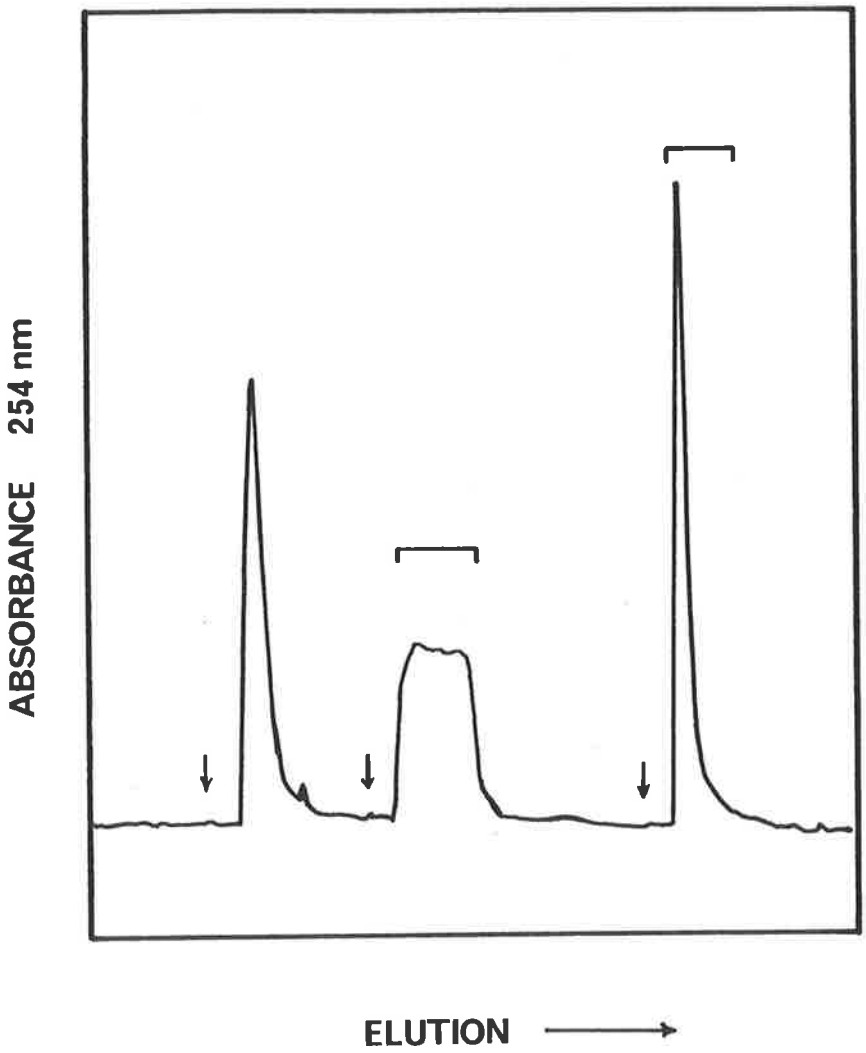
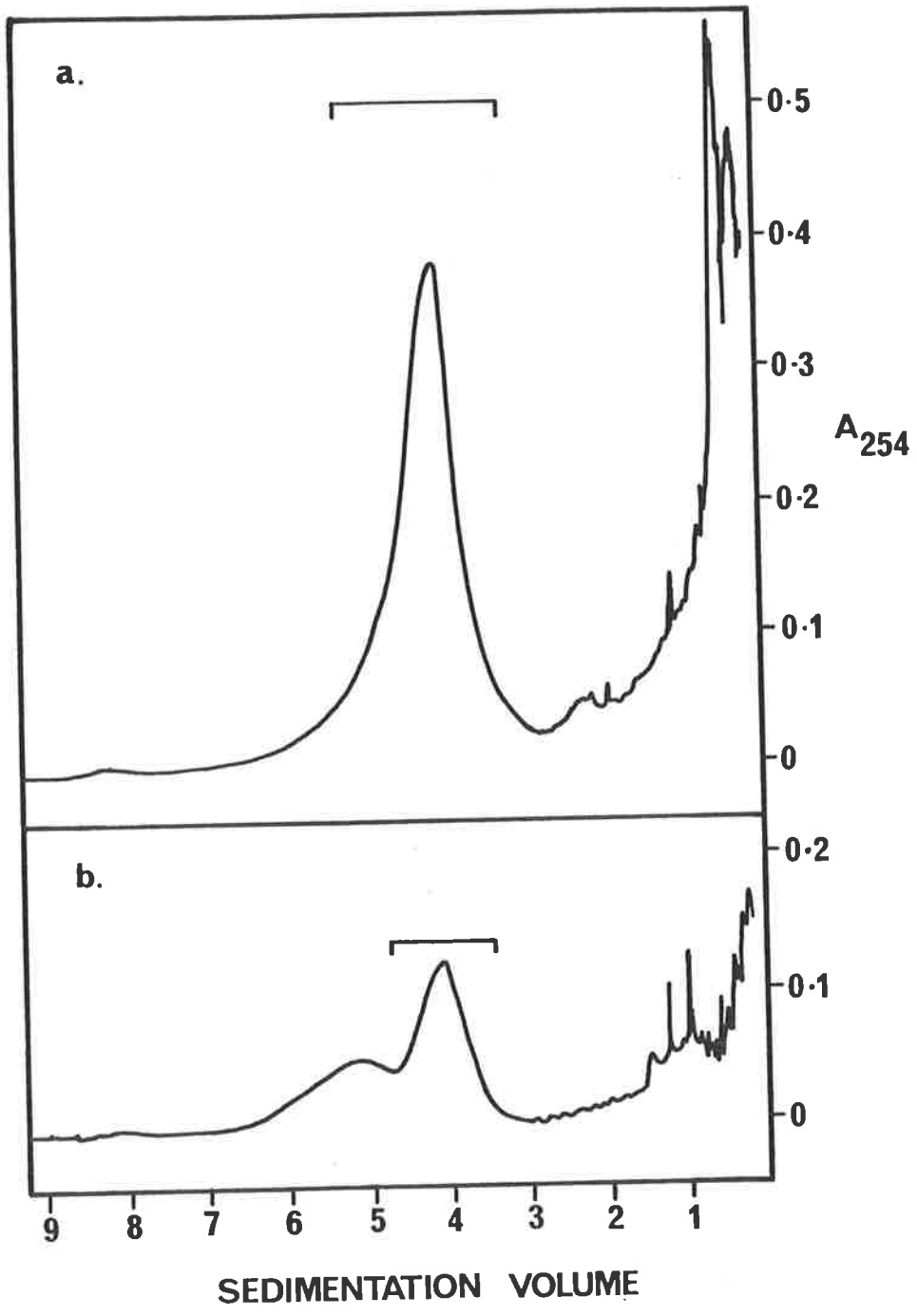

ABSORBANCE 254 nm
ELUTION

Figure 3.7

Sucrose gradient examination of oligo-dT-cellulose fractions.

a. RNA bound to the column.

b. RNA unbound to the column.


a.
0·5
0·4
0·3
0·2
0·1
0
A254
b.
0·2
0·1
0
9
8
7
6
5
4
3
2
1
SEDIMENTATION VOLUME

When labelled translation products were electrophoresed on gels, (figure 3.8), the major product observed in both fractions was globin, but H5 protein was also detected in both the bound and unbound fractions. The relative proportion of H5 to globin differs in the two profiles, about 4 percent of the translation products of the unbound fraction being H5, while only about 2 percent of the protein products of the bound material ran at the H5 position. Scott, (1975), however, reports the presence of the H5 peak only in the protein products of the RNA unbound to oligo-dT-cellulose. This was almost certainly due to the insensitivity of the translation assay system used, which was unable to detect any products other than globin when total 10S RNA was translated. (The sensitivity of the system used here was almost 100-fold higher than that used by Scott).

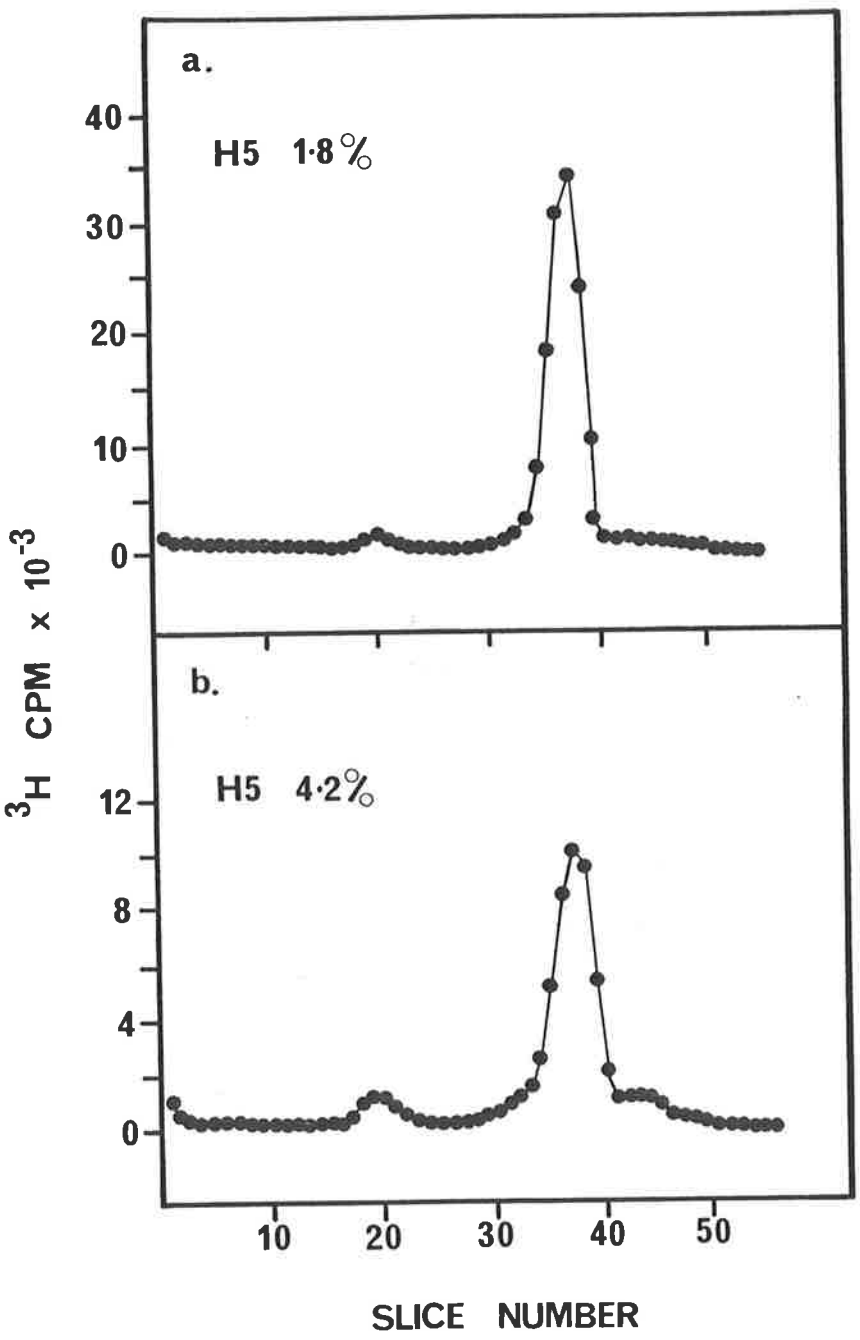
The presence of H5 mRNA in the fraction of RNA bound to oligo-dT-cellulose suggests that either a proportion of the H5 mRNA is polyadenylated, (and its adsorption to the oligo-dT is a consequence of this), or that the H5 mRNA appears in the bound fraction due to aggregation or other non-specific effects. Such non-specific binding has been noted by Suzuki *et al.*, (1972), and Bantle *et al.*, (1976) who showed that non-adenylated ribosomal RNA may appear together with poly-A containing RNA in the fraction bound to the column.

Figure 3.8

SDS-urea acrylamide gel analysis of the *in vitro* translation products of 10S reticulocyte RNA fractionated by oligo-dT-cellulose chromatography. Equal amounts of RNA, (0.2 µg) were added to each translation mix.

a. ^{3}H-labelled products of RNA bound to oligo-dT-cellulose.

b. ^{3}H-labelled products of RNA unbound to oligo-dT-cellulose.


a.
H5 1·8%
40
30
20
10
0
b.
H5 4·2%
12
8
4
0
10
20
30
40
50
SLICE NUMBER
^{3}H CPM x 10^{-3}

A number of experiments were performed to determine whether the binding of H5 mRNA to the oligo-dT-cellulose column was specific, or the result of aggregation. These experiments can be grouped as follows;

a) Heat disaggregation of RNA prior to chromatography. The oligo-dT-cellulose column was jacketed in an ice/water slurry and all solutions were held at $0^{o}C$. The low temperature should aid the efficiency of the column by stabilising any A-T hybrids formed. The RNA samples were disaggregated by heating to $65^{o}C$ for 10 minutes or $80^{o}C$ for 3 minutes and then snap-cooled to $0^{o}C$ before loading onto the column.

b) Alteration of chromatographic solutions. In an attempt to minimise aggregation and non-specific binding to the column, loading solutions were reduced from 0.5 M NaCl to 0.3 M NaCl and then to 0.1 M NaCl. Some loss of poly-A containing RNA from the bound fraction is observed when binding is carried out at 0.1 M NaCl, (Bantle *et al.*, 1976).

c) Chromatography in the presence of excess ribosomal RNA. In all affinity chromatography experiments performed with 10S RNA, the fraction of RNA bound to the column was always the larger, and it seemed possible that some H5 mRNA was always co-eluting with the larger fraction, possibly due to non-specific interactions. In order to preclude non-specific binding of H5 mRNA to the polyadenylated

RNA fraction, a five-fold excess of ribosomal RNA was added to the 10S RNA prior to chromatography, so that when passaged through oligo-dT-cellulose, the unbound RNA was the much larger fraction. In this experiment, the denaturation and chromatography conditions were exactly those used in (a) above.

d) Batch of oligo-dT-cellulose used. Two commercial and one non-commercial batch of oligo-dT-cellulose were compared using the standard chromatography procedure described in chapter 2, together with heat denaturation prior to chromatography. The non-commercial preparation of oligo-dT-cellulose was the same one used by Scott and Wells, (1976).

e) Multiple passages through columns. Fractions obtained when heat-denatured 10S RNA was chromatographed through oligo-dT-cellulose were adjusted to the appropriate salt concentration, heat denatured, and repassaged through the same column. In these circumstances, none of the RNA that is unbound to oligo-dT-cellulose binds on a second passage through the column, and greater than 90 percent of the RNA bound to oligo-dT-cellulose on the first passage through the column rebinds when passed through the column a second time. A small proportion of the RNA previously bound to oligo-dT now appears as an unbound fraction.

For each of the experiments detailed above, samples of RNA from each fraction were translated and the labelled translation products identified on gels. In every case, the RNA fraction bound to oligo-dT-cellulose contained H5 mRNA and H5 protein never represented less than 2 percent of the bound fraction translation products. The unbound fraction always showed H5 protein when translated, however this never amounted to more than about 5 percent of the total protein synthesised. The small RNA fraction that did not bind when re-passaged through oligo-dT-cellulose, ((e) above), appeared similar to normal unbound RNA and showed about 5 percent H5 protein when translated.

To summarise the results obtained using oligo-dT-cellulose, the RNA fraction unbound to oligo-dT is 2-3-fold enriched for H5 sequences over unfractionated RNA, while the bound fraction has an H5 mRNA content very similar to that of unfractionated RNA. None of the experimental approaches employed was able to localise H5 mRNA to the unbound fraction as was expected if H5 mRNA was not polyadenylated. Despite the small enrichment for H5 sequences, because the unbound RNA is the smaller fraction, and because this fraction also contains non-messenger RNA species, such as ribosomal RNA breakdown products, that will

not be detected by translation assays, it is likely that the unbound fraction contains less of the total H5 mRNA than the bound fraction. The combination of H5 mRNA loss to the bound fraction and globin mRNA contamination of the unbound fraction, has resulted in very little useful fractionation of H5 mRNA sequences using oligo-dT-cellulose chromatography.

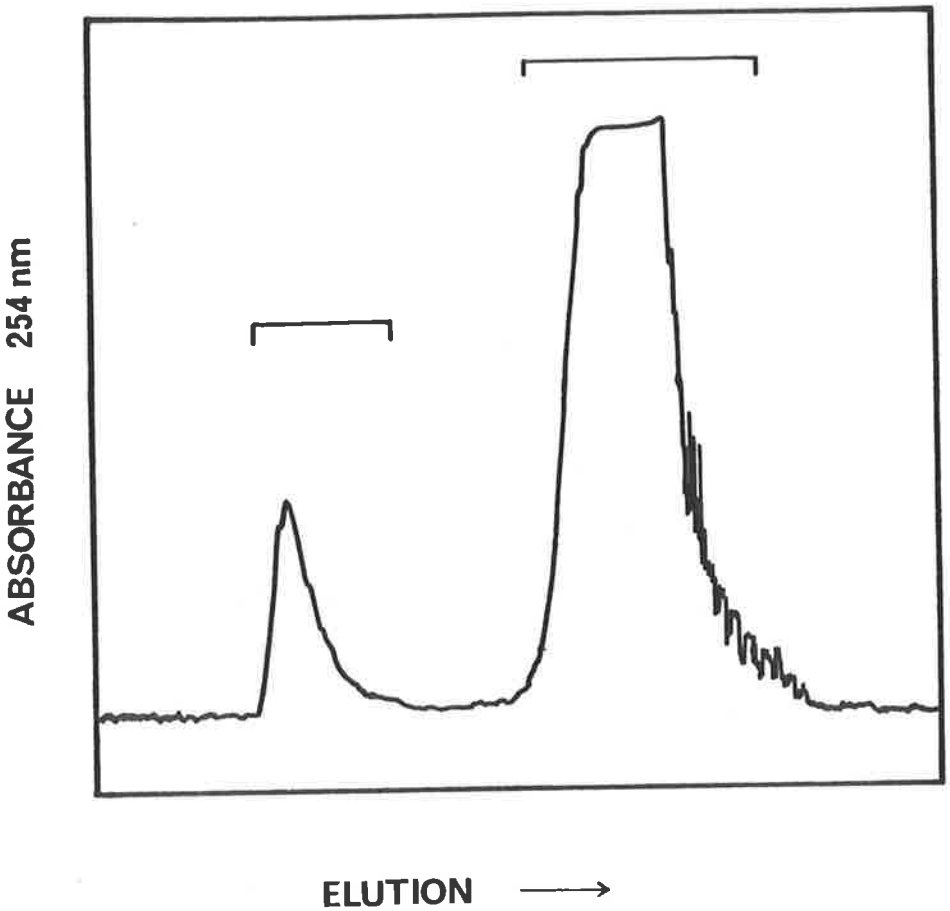
3.2.4 Poly-U-Sepharose Chromatography

Poly-U linked to Sepharose beads has been used successfully to separate poly-A containing RNA from non-polyadenylated RNA species, (Lindberg and Persson, 1972). As non-specific binding of RNA to oligo-dT-cellulose has been attributed to contaminants in the cellulose, (Sullivan and Roberts, 1973; Bantle *et al.*, 1976), the use of an inert agarose support might avoid these problems. Therefore, a series of experiments was undertaken to determine whether poly-U-Sepharose could be used to efficiently separate globin and H5 mRNA sequences.

Total 10S RNA was heat-denatured and fractionated on a poly-U-Sepharose column using the procedure described in chapter 2. Figure 3.9 shows a typical loading and elution profile, the very large second peak of absorbance being due to the presence of formamide in the elution buffer. After dialysis, to remove the formamide, the absorbance of the fractions was read and the

Figure 3.9

Uvicord elution profile of reticulocyte 10S RNA chromatographed on poly-U-Sepharose. The very large second peak of absorbance is due to the presence of formamide in the elution buffer.


ABSORBANCE 254 nm
ELUTION

distribution of RNA between the bound and unbound fractions determined. Typically about 75 percent of the 10S RNA bound to poly-U-Sepharose, compared to only about 60 percent binding when the same RNA preparation was applied to oligo-dT-cellulose.

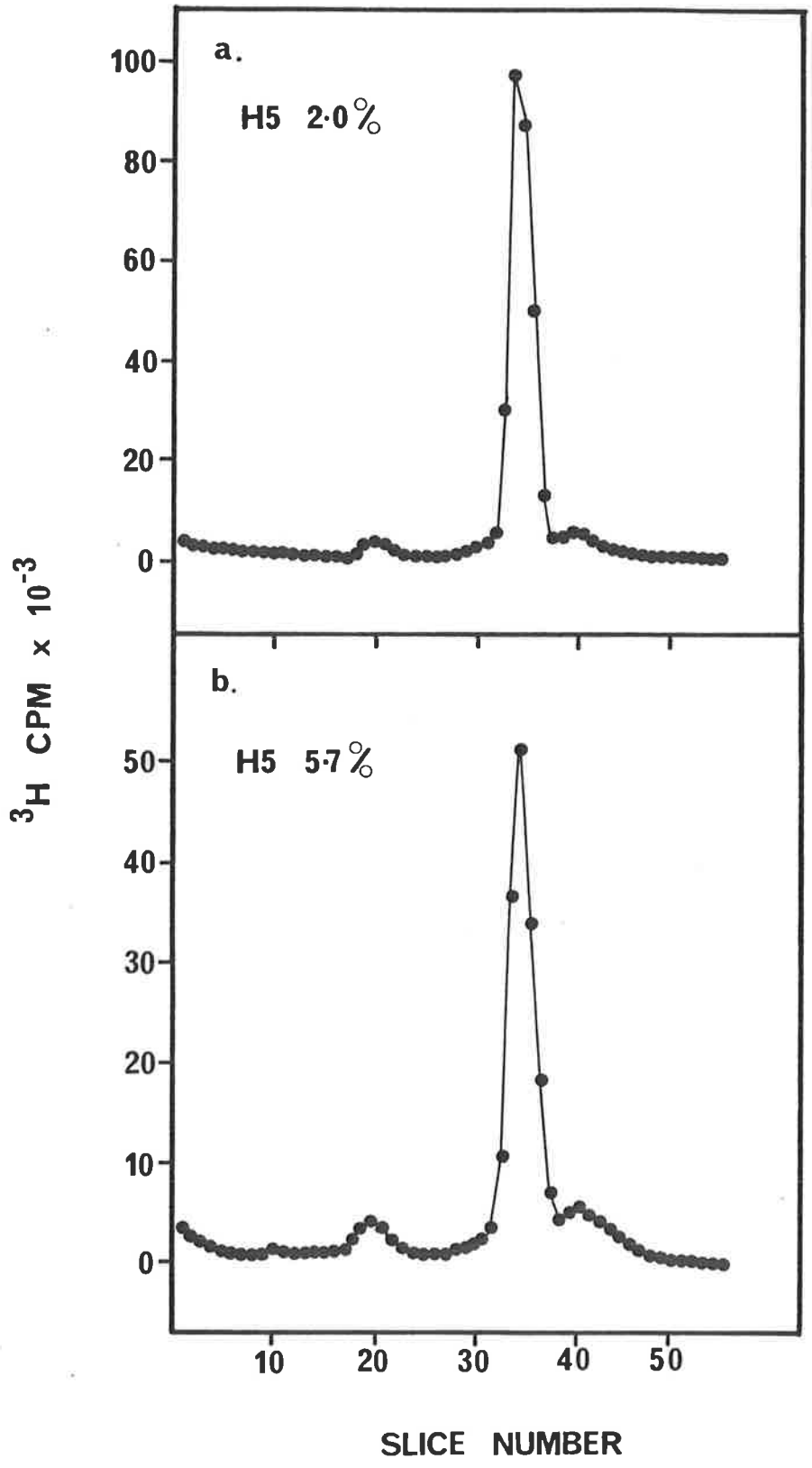
Samples of the RNA from the fractions bound and unbound to poly-U-Sepharose were translated and the products examined on SDS-urea gels, (figure 3.10). The major product observed in each fraction is globin, and H5 comprises about 2 percent of the protein products of the bound RNA and about 6 percent of the protein products of the unbound RNA fraction.

In repeated experiments, the results using poly-U-Sepharose chromatography were essentially identical to those obtained using oligo-dT-cellulose. The use of a new affinity material and a new, inert support matrix has not significantly altered the distribution of H5 mRNA sequences and has not resulted in the removal of H5 mRNA from the bound fraction. The increased sensitivity of the poly-U-Sepharose affinity column has merely resulted in an increased fraction of the total H5 mRNA being found in the bound fraction. Therefore, the ability to bind to poly-U-Sepharose seems to be a real property of H5 mRNA and not merely the result of aggregation with poly-A containing RNA or non-specific interaction with the column material.

Figure 3.10

SDS-urea acrylamide gel analysis of the *in vitro* translation products of 10S reticulocyte RNA fractionated on poly-U-Sepharose. Equal amounts of RNA were added to each translation mix.

a. products of RNA bound to poly-U-Sepharose.
b. products of RNA unbound to poly-U-Sepharose.


a.
H5 2·0%
b.
H5 5·7%
100
80
60
40
20
0
50
40
30
20
10
0
10
20
30
40
50
^{3}H CPM x 10^{-3}
SLICE NUMBER

3.2.5 Double-column Chromatography

Since poly-U-Sepharose binds about 75 percent of added 10S RNA, while oligo-dT-cellulose only binds about 60 percent, poly-U binds a proportion of the 10S reticulocyte RNA that is not bound to oligo-dT-cellulose, (perhaps RNA carrying an intermediate length poly-A tract). In an experiment designed to determine the nature of this intermediate RNA fraction, RNA that was unbound to oligo-dT-cellulose, (and had previously been translated to give 5 percent H5 protein), was passaged through poly-U-Sepharose and the two resulting RNA fractions collected. Approximately half of the RNA formerly unbound to oligo-dT-cellulose was found to bind to poly-U-Sepharose. Samples of these bound and unbound RNA fractions were translated and the labelled products analysed on gels.

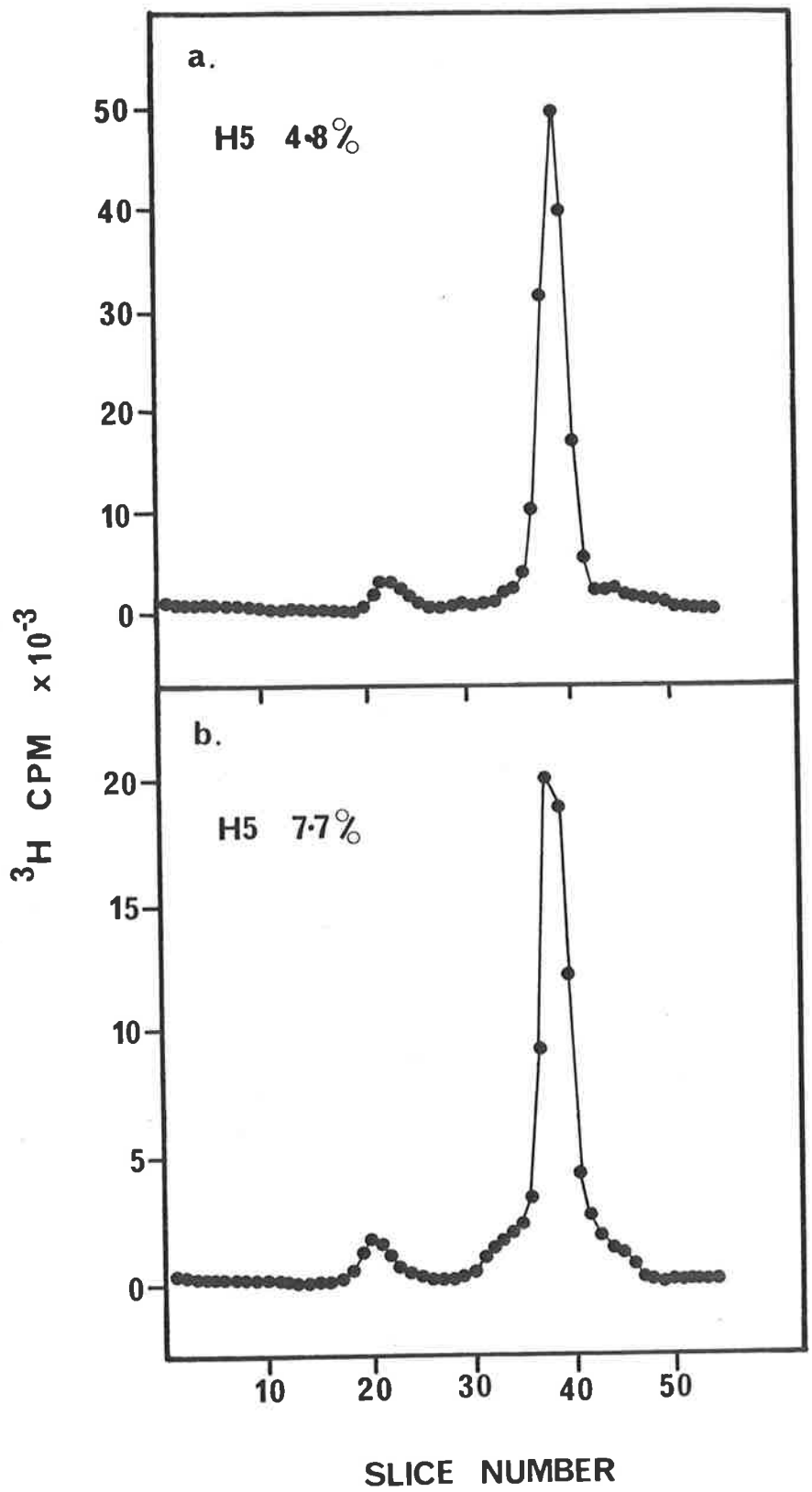
As shown in figure 3.11, of the labelled translation products, H5 protein was present as 5 percent and 8 percent respectively of the bound and unbound fractions. The RNA fraction that passes through oligo-dT-cellulose but binds to poly-U-Sepharose has about the same H5 content as the unbound RNA obtained from a single-column fractionation using either chromatography material, however, in comparison, it should be essentially free of contaminating ribosomal RNA breakdown

Figure 3.11

SDS-urea gel analysis of *in vitro* translation products. 10S reticulocyte RNA that was not bound to oligo-dT-cellulose was passaged through poly-U-Sepharose and the bound and unbound fractions collected. Equal amounts of RNA, (0.1 µg), were added to each translation mix.

a. ^{3}H-labelled translation products of RNA unbound to oligo-dT-cellulose and bound to poly-U-Sepharose.

b. ^{3}H-labelled translation products of RNA unbound to oligo-dT-cellulose and unbound to poly-U-Sepharose.


a.
H5 4·8%
b.
H5 7·7%
50
40
30
20
10
0
20
15
10
5
0
10
20
30
40
50
^{3}H CPM x10^{-3}
SLICE NUMBER

products. These and other non-polyadenylated sequences should be localised exclusively to the unbound fraction.

While far from achieving a purification of H5 mRNA, the combination of oligo-dT-cellulose and poly-U-Sepharose affinity chromatography techniques has made possible the isolation of a fraction of 10S polysomal RNA that is enriched for H5 sequences and that is relatively free from contaminating non-polyadenylated sequences. Unfortunately, this fraction represents only a very small proportion of the original 10S polysomal RNA preparation, but use of this RNA fraction in association with other separation techniques may make a useful enrichment for H5 mRNA sequences possible.

CHAPTER 4

CHARACTERISATION OF 7-11S RNA FROM CHICKEN EMBRYOS

4.1 INTRODUCTION

Investigation of the relationship between the H5 genes and the other, normally expressed, histone genes requires molecular probes for both sets of genes. The preparation of cDNA probes and the construction of dscDNA clones require a suitable, enriched, mRNA template, and so a series of experiments was carried out in order to isolate histone mRNAs from chicken cells.

Histone mRNA has been isolated, in varying degrees of purity, from a number of species including *Xenopus*, (Destree *et al.*, 1977), *Drosophila*, (Burckhardt and Birnstiel, 1978), human (HeLa) cells, (Stephens *et al.*, 1977) and sea urchin, (Gross *et al.*, 1976). In all cases except sea urchin, where histone mRNA accounts for more than 30 percent of the polysomal RNA in the rapidly dividing embryos, (Moav and Nemer, 1971), histone mRNA has been isolated from tissue-culture cells. Because, apparently, histone mRNA is only synthesised while DNA replication is in progress, (Stein *et al.*, 1977), it has been found necessary to synchronise the cultured cells so that maximum yields of histone mRNA could be extracted from S-phase cells. It was reasoned however, that, by analogy with the sea urchin case, rapidly developing chicken embryos should contain significant amounts of total histone mRNA,

and so efforts were directed towards the isolation of histone mRNA from this source.

Total RNA was prepared from 5-day chicken embryos using a guanidinium-Cl method, (Seeburg *et al.*, 1977a), and 7-11S RNA isolated by sucrose-gradient centrifugation. The techniques of cell-free translation and affinity chromatography, developed while attempting to enrich reticulocyte RNA fractions for H5 mRNA, were applied to the 7-11S RNA fraction in order to determine the nature of any mRNA sequences present.

The experiments described in this chapter show that the 7-11S RNA contained histone mRNA, and that after passage through a poly-U-Sepharose column, the unbound fraction contained all five histone mRNAs but no other detectable mRNA activities.

After removal of ribosomal sequences, cDNA prepared from the unbound RNA fraction was used to estimate the reiteration frequency of the histone genes in the chicken genome.

4.2 RESULTS

4.2.1 Preparation of 7-11S RNA from chicken embryos

Total RNA was prepared from 5-day chicken embryos using the guanidinium-Cl method of Seeburg *et al.*, (1977a), and this material was centrifuged on sucrose gradients. The RNA running between the 7 and 11S positions on the gradient

was selected and re-centrifuged. The sucrose gradient profiles of the total cellular RNA and of the re-centrifuged 7-11S RNA are shown in figure 4.1, parts a and b respectively. In an average preparation, using 36 chicken embryos, yields of about 50 micrograms of 7-11S RNA were routinely obtained.

4.2.2 Translation of 7-11S RNA

Addition of samples of the 7-11S RNA to the wheat germ cell-free translation system stimulated the incorporation of ^{3}H-leucine into high molecular weight products, indicating that the RNA preparation contained mRNA. The ^{3}H-labelled translation products were co-electrophoresed with ^{14}C-histone marker proteins on two different gel systems, and the resulting translation product profiles are shown in figure 4.2. Clearly the 7-11S RNA contains mRNA species which translate to yield protein products, some of which co-migrate with authentic histones on two gel systems.

While histone proteins seem to be present amongst the translation products, it is quite clear that other proteins are also present. Especially notable is the large peak of labelled protein that runs at the position of H1 on the low pH-urea gel, (figure 4.2a), and at the position of H4 on the SDS-urea gel system, (figure 4.2b). The anomalous migration of this major protein product on the two gel systems suggests that the protein is not a histone.

Figure 4.1

Preparation of 7-11S RNA from 5-day old chick embryos, (see section 2.2.2).

a. 10-40% sucrose gradient fractionation of total cellular RNA.

b. Rerun of 7-11S RNA fraction selected from the first gradient.

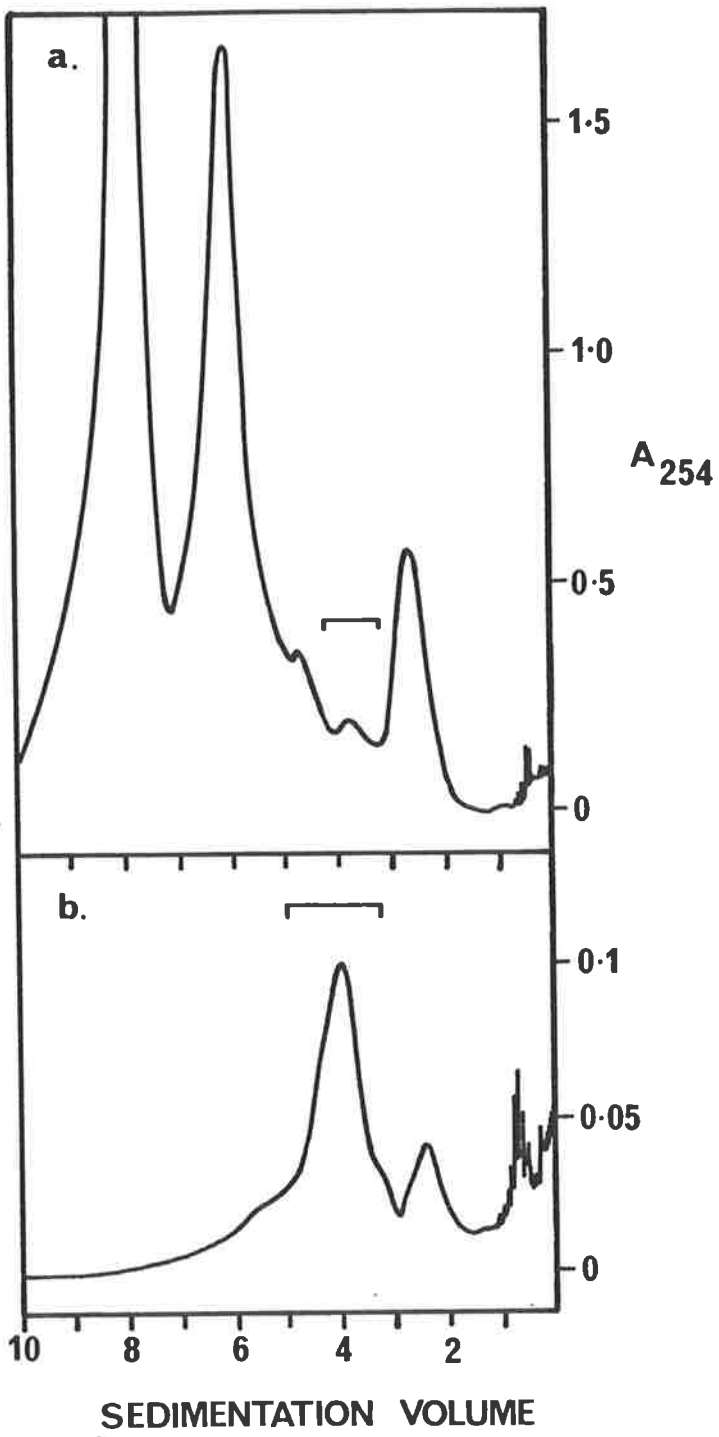

a.
1·5
1·0
A254
0·5
0
b.
0·1
0·05
0
10
8
6
4
2
SEDIMENTATION VOLUME

Figure 4.2

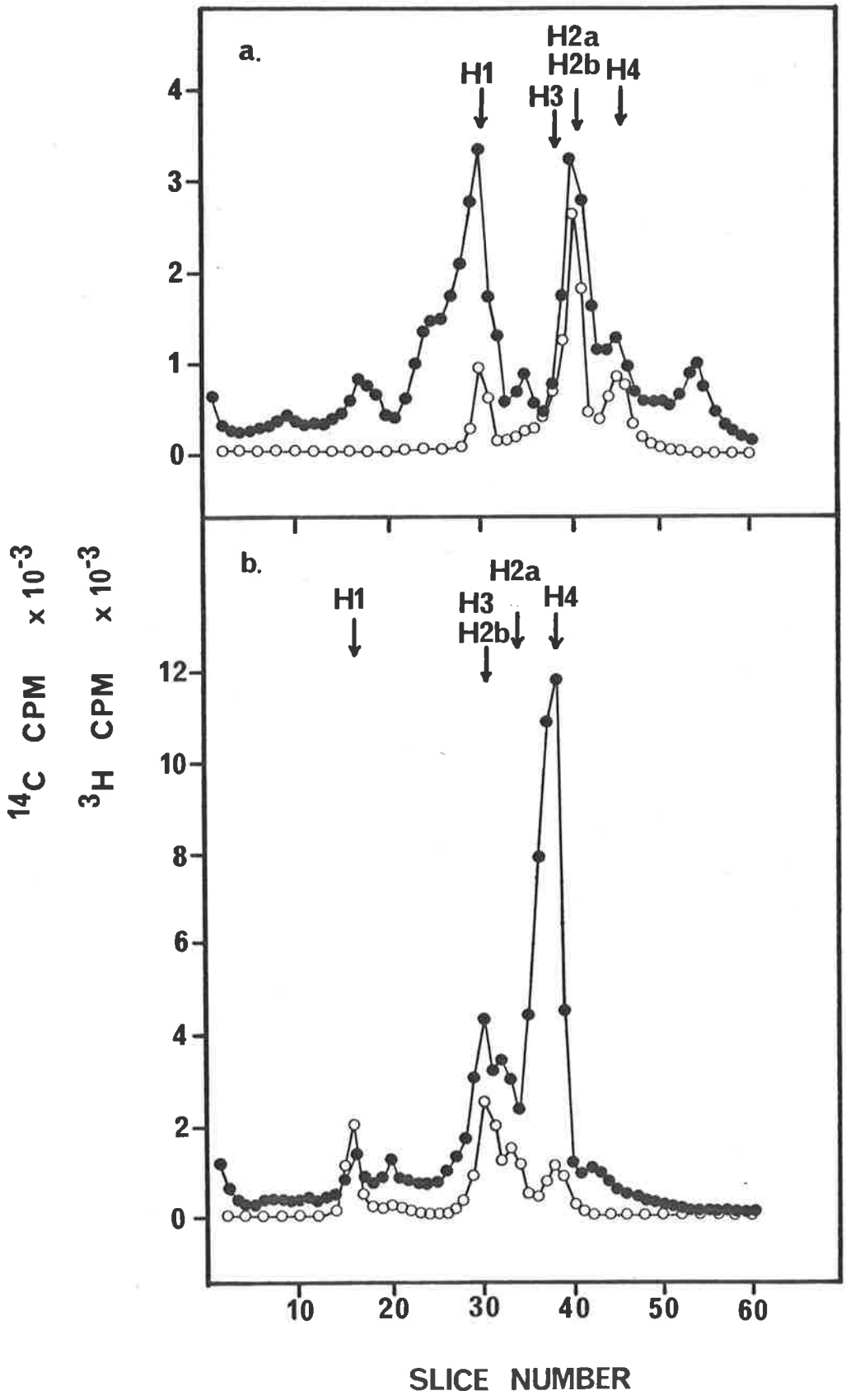
Analysis of the *in vitro* translation products of 7-11S RNA isolated from 5-day chick embryos. mRNA was translated in the wheat-germ cell-free system and the ^{3}H-leucine labelled products examined by electrophoresis on polyacrylamide gels in the presence of ^{14}C-labelled histone markers.

a. low pH-urea gel.

b. SDS-urea gel.

●—● ^{3}H-labelled translation products

○—○ ^{14}C-labelled histone standards


a.
H1
H2a
H2b
H4
H3
b.
H1
H3
H2b
H2a
H4
^{14}C CPM x 10^{-3}
^{3}H CPM x 10^{-3}
SLICE NUMBER

Using hybridisation techniques, it has been shown that up to 10 percent of the embryonic 7-11S RNA fraction consists of globin sequences, (Crawford - pers. commun.), and it is known that globin protein co-electrophoreses with histone H1 on low pH-urea gels, (Appels and Wells, 1972). If the 7-11S RNA preparation does contain globin mRNA or any other polyadenylated mRNAs, chromatography of the RNA through a poly-U-Sepharose column will remove most of these contaminating species, while leaving the histone mRNA in the unbound fraction.

4.2.3 Poly-U-Sepharose chromatography of 7-11S RNA

After heat denaturation and snap-cooling, the embryonic 7-11S RNA was passed through a poly-U-Sepharose column and the bound and unbound RNA fractions collected. Only about 10 percent of the RNA loaded bound to the column. Samples of the bound and unbound RNA were translated in the wheat germ cell-free system and the labelled products analysed on polyacrylamide gels. Figure 4.3 shows the radioactivity profiles resulting when the translation products of the bound and unbound RNA were electrophoresed on low pH-urea gels and figure 4.4 shows the radioactivity profiles of the two fractions on SDS-urea gels.

On both gel systems, the translation product profiles of the bound and unbound fractions differ greatly from each other. The large peak of

Figure 4.3

Low pH-urea gel analysis of the *in vitro* translation products of 7-11S chick embryo RNA fractionated by poly-U-Sepharose chromatography. Equal amounts of RNA, (150 ng), were added to each translation mix.

a. RNA bound to poly-U-Sepharose

b. RNA unbound to poly-U-Sepharose

●—● ^{3}H-labelled translation products

○—○ ^{14}C-labelled histone markers

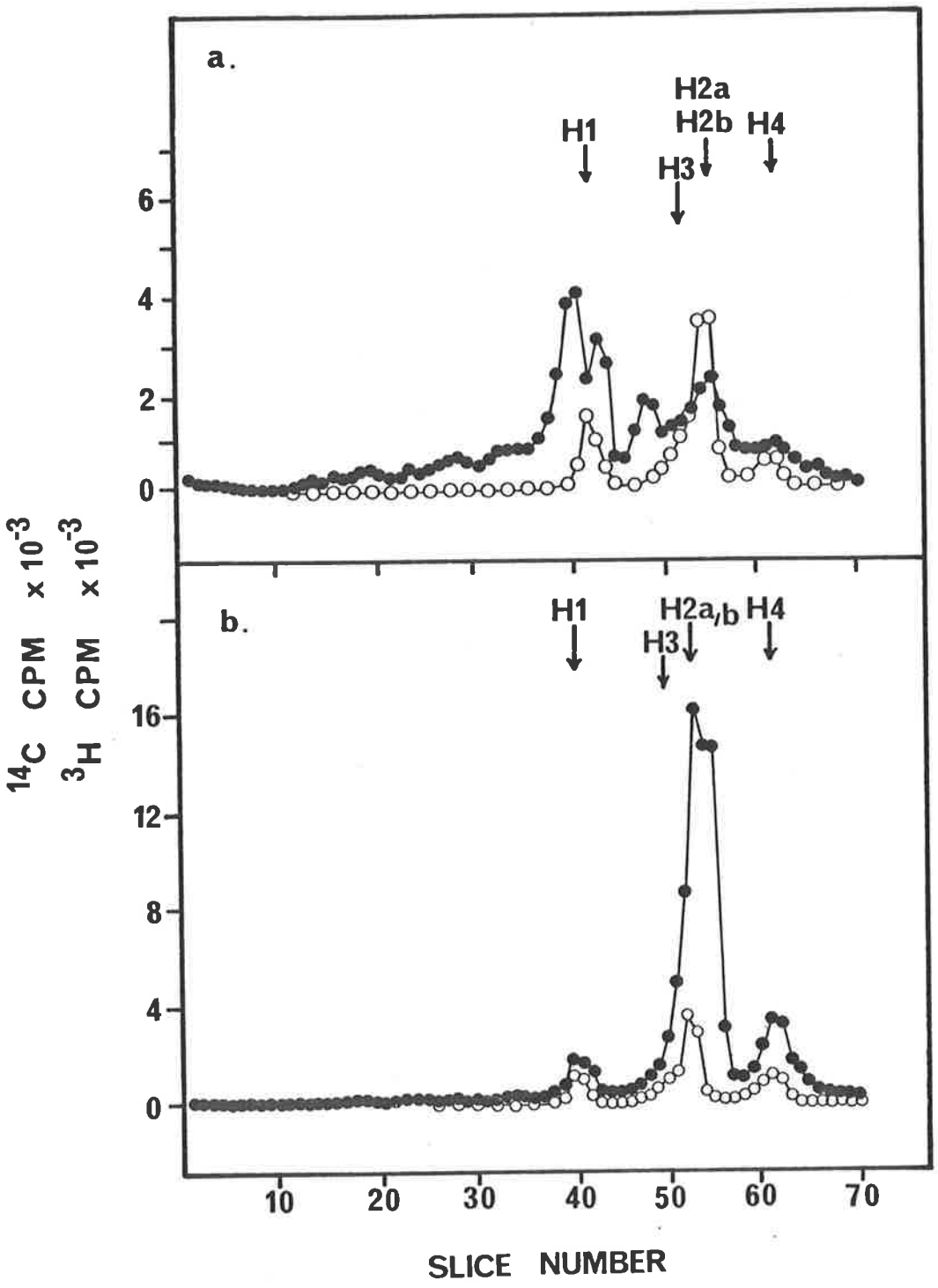

a.
H1
H2a
H2b
H4
H3
6
4
2
0
b.
H1
H2a/b
H4
H3
16
12
8
4
0
10
20
30
40
50
60
70
SLICE NUMBER
14C CPM x10-3
3H CPM x10-3

Figure 4.4

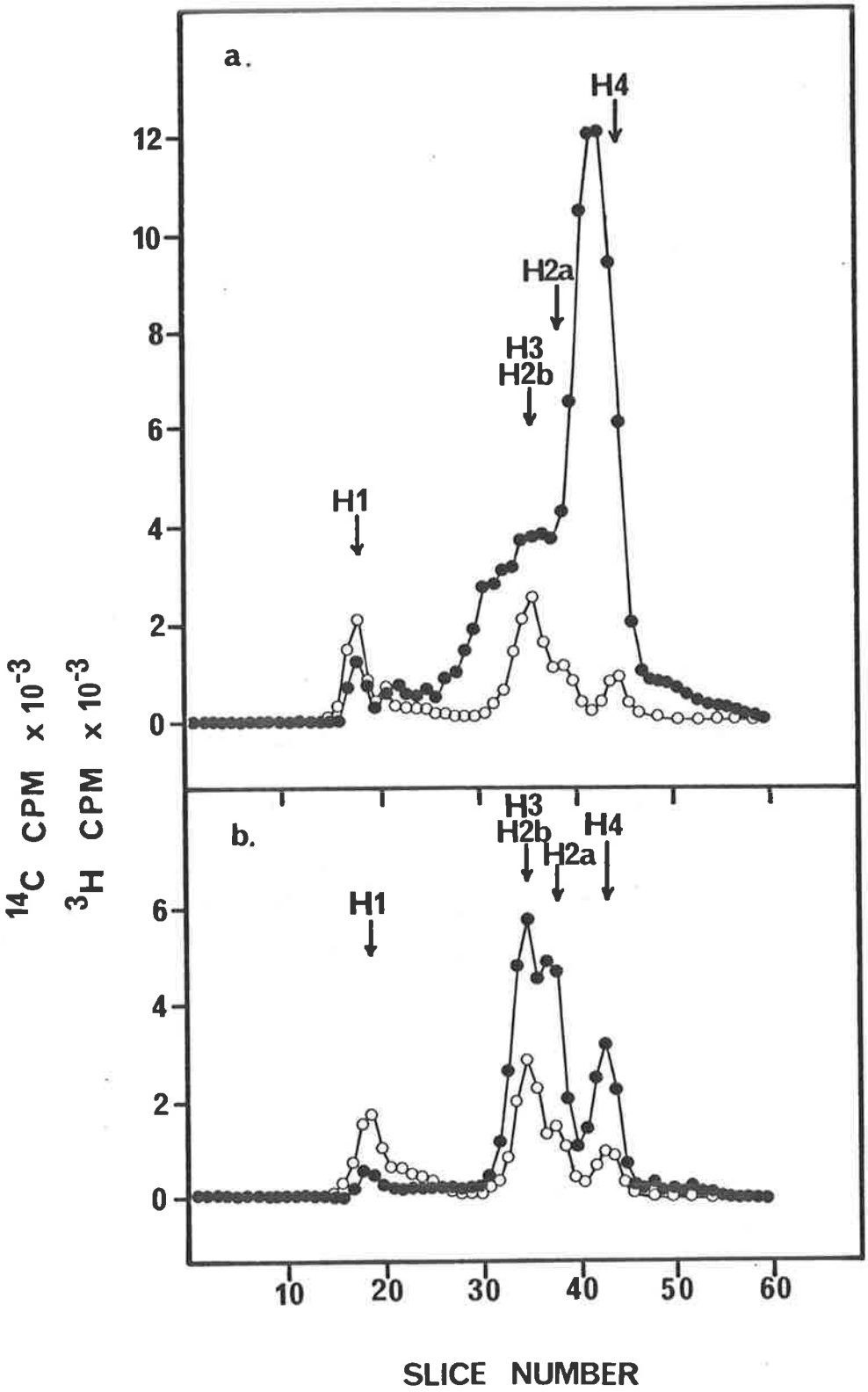
SDS-urea gel analysis of the *in vitro* translation products of 7-11S chick embryo RNA fractionated by poly-U-Sepharose chromatography. 0.1 µg of RNA was added to each translation mix.

a. RNA bound to poly-U-Sepharose.

b. RNA unbound to poly-U-Sepharose.

●—● ^{3}H-labelled translation products.

○—○ ^{14}C-labelled histone markers.


a.
H4
H2a
H3
H2b
H1
b.
H3
H2b
H4
H2a
H1
12
10
8
6
4
2
0
6
4
2
0
10
20
30
40
50
60
^{14}C CPM $\times 10^{-3}$
^{3}H CPM $\times 10^{-3}$
SLICE NUMBER

non-histone protein observed in the translation products of unfractionated RNA, (figure 4.2), is now located exclusively in the products of the bound RNA fraction, and a number of other minor protein peaks which do not co-electrophorese with histones also appear in the bound fraction profile. The translation profiles of the unbound material however, mimic exactly the profiles of the marker histones.

Reference to the two gel systems shows that all five histones are represented amongst the translation products of the unbound RNA, and that there are no other detectable proteins present in this material. While histone H1 appears to be poorly represented amongst the translation products, relative to the other histone proteins, H1 mRNA has been reported to translate poorly in the wheat germ system, and in all other cell-free translation systems, (Gallwitz *et al.*, 1978).

The hybridisation and translation data suggest that globin mRNA is present in the total 7-11S RNA, and this is extremely likely since globin mRNA is approximately 10S in size, and the embryo has a well-developed circulatory system at the 5-day stage. It is interesting to note that in these experiments, apparently 100 percent of the globin mRNA and other polyadenylated mRNAs were bound to the poly-U-Sepharose column, while in the

experiments described in the previous chapter, (section 3.2.4), only 70-80 percent of the total globin mRNA activity could be bound. This infers that the newly synthesised embryonic mRNA carries longer poly-A tracts, on average, than the mRNA isolated from adult reticulocytes.

In addition to the putative globin protein peak, the poly-U-Sepharose bound RNA translation profile shows the presence of a number of other minor protein peaks, some of which correspond in position to the histone markers. Especially noticeable is the peak of protein co-electrophoresing with histone H1 marker on the SDS-urea gel system, (figure 4.4). It is possible that a proportion of the histone mRNA is bound to the poly-U-Sepharose, since polyadenylated subfractions of total histone mRNA have been reported in *Xenopus* (Levenson and Marcu, 1976), HeLa tissue culture cells, (Borun *et al.*, 1977) and sea urchin, (Ruderman and Pardue, 1978).

While the translation assay is limited, and may not reflect the actual frequency of an mRNA species in a population, (see sections 3.2.2(ii) and 3.2.2(iii)), the results presented above show that the 7-11S RNA fraction unbound to poly-U-Sepharose contains all five histone mRNAs, and is free of other detectable mRNA contaminants.

4.2.4 Preparation of histone cDNA

While the translation experiments have shown that 7-11S embryonic RNA unbound to poly-U-Sepharose contains histone mRNA but is free from other detectable mRNA species, this RNA fraction is certain to be contaminated with ribosomal RNA breakdown products. Therefore, cDNA prepared using this RNA as template will contain sequences complementary to ribosomal RNA. If this cDNA preparation is to be used to determine an unambiguous reiteration frequency for the chicken histone genes, it is desirable that the cDNA be free of any contaminating ribosomal sequences. Therefore, the following approach was used to remove the ribosomal sequences from the cDNA preparation.

cDNA was synthesised from the unbound 7-11S RNA fraction using the random-primed method of Taylor *et al.*, (1976). (In this technique non-polyadenylated RNA may be copied into cDNA using short, random oligo-nucleotides as primers in place of the usual oligo-dT primer). The cDNA was then hybridised to a large excess of 18S and 28S ribosomal RNA. Hybridisation was continued until all ribosomal cDNA had annealed to the ribosomal RNA molecules, (effective $R_o t=1$), and the hybridisation mix was loaded onto a sucrose gradient. The sucrose gradient solutions were all

in 0.4 M NaCl to ensure the stability of the RNA-cDNA hybrids during centrifugation.

The A_{260} and radioactivity profiles resulting after sucrose-gradient centrifugation are shown in figure 4.5. About 30-40 percent of the radioactive label is associated with the ribosomal RNA, while the remainder of the labelled material is at the top of the gradient at about the 5S position. When the 5S cDNA was used to detect sequences in restriction digests of chicken genomal DNA, the bands corresponding to the ribosomal genes were not detected, (Crawford *et al.*, 1979), showing the 5S cDNA to be free of ribosomal cDNA sequences.

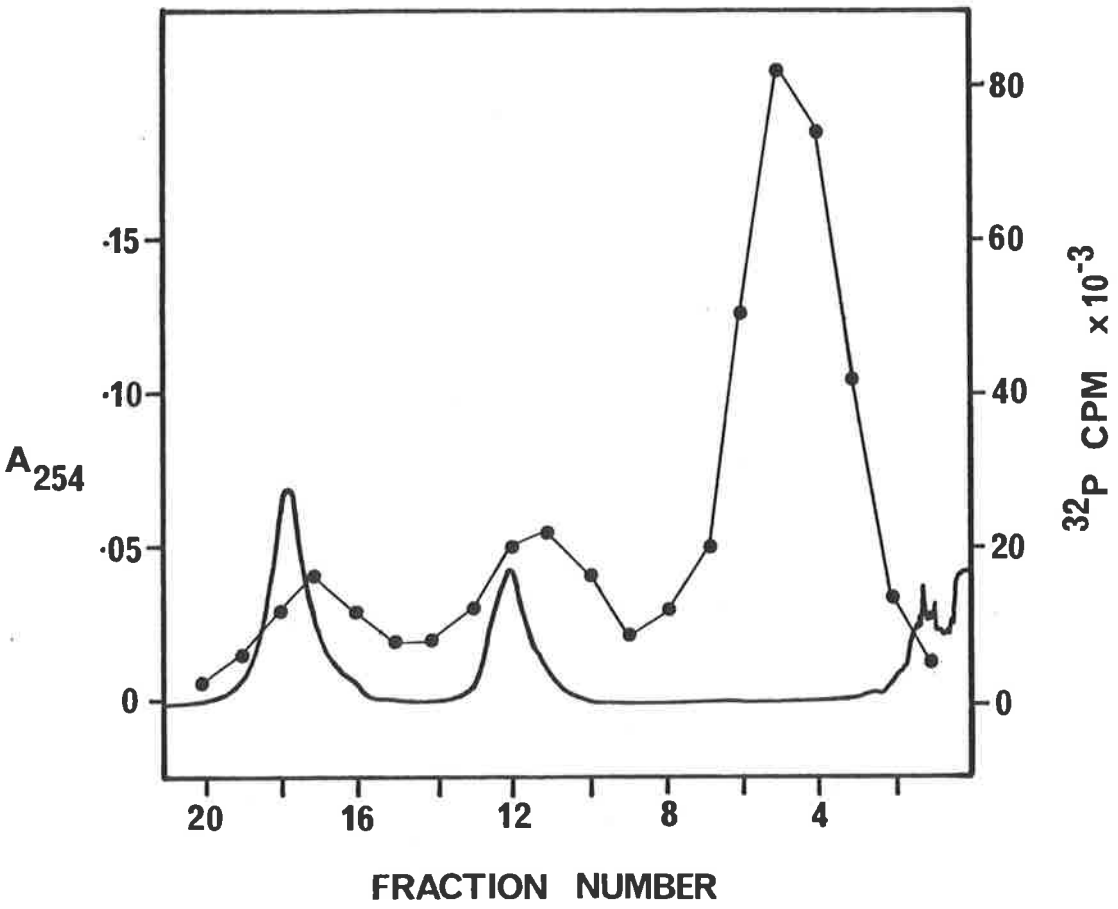
However, the 5S cDNA did hybridise to cloned sea urchin histone genes, (*Echinus esculentus* histone genes: λ clone 55, gift from K. Murray), and detected the histone genes, specifically, in restriction digests of total sea urchin genomal DNA, (Crawford *et al.*, 1979), thus proving that the 5S cDNA contained histone gene sequences.

4.2.5 Reiteration frequency of the histone genes in the chicken genome

^{32}P-labelled 5S histone cDNA, prepared as described above, was mixed with a vast excess of chicken genomal DNA that had been sonicated to an average length of about 300 bases, (approximately

Figure 4.5

Sucrose gradient fractionation of cDNA prepared to 7-11S chick embryo RNA. cDNA was prepared using the random priming procedure, (section 2.2.6), and then hybridised to an excess of purified chicken 18S and 28S ribosomal RNA in 200 µl of 0.18 M NaCl, 10 mM Tris-HCl pH 7.0, 1 mM EDTA, 0.5% SDS, for 2 hours at 65°C. The hybridisation mix was loaded directly onto a 10-40% sucrose gradient and centrifuged at 160,000 g for 16 hours. The cDNA sedimenting at about 5S was pooled and used as a histone gene probe.


A254
·15
·10
·05
0
80
60
40
20
0
32P CPM x10-3
20
16
12
8
4
FRACTION NUMBER

the same length as the cDNA probe). A small amount of ^{3}H-labelled chicken genomal DNA was included as a marker for the reassociation of unique sequences, which comprise the major portion of the genomal DNA, (Sullivan and Roberts, 1973). Identical hybridisation mixes were set up in capillary tubes, allowed to hybridise at 65^{o}C for various time intervals and then the percentage of DNA in double-stranded form was assayed using the single-strand specific nuclease S_1, (Vogt, 1973). The $C_o t$ curve generated is shown in figure 4.6.

The unique sequences reassociated with a $C_o t_{\frac{1}{2}}$ value of 1.2×10^3, while the 5S histone cDNA hybridised to the total chicken DNA in a single sharp transition with a $C_o t_{\frac{1}{2}}$ value of about 120. Although factors such as the length, and G + C content, (Britten and Kohne, 1968), of the reassociating DNA are known to effect the rate of hybridisation, they are unlikely to influence the $C_o t_{\frac{1}{2}}$ figure to any significant extent. This $C_o t_{\frac{1}{2}}$ value of 120, compared to the unique sequence value of 1.2×10^3 indicates that the hybridisation of chicken histone sequences to total DNA occurs about ten times faster than the unique rate and that, therefore, the histone gene sequences are reiterated about ten-fold in the chicken genome.

Included in figure 4.6 is the $C_o t$ curve for the hybridisation of ribosomal cDNA sequences

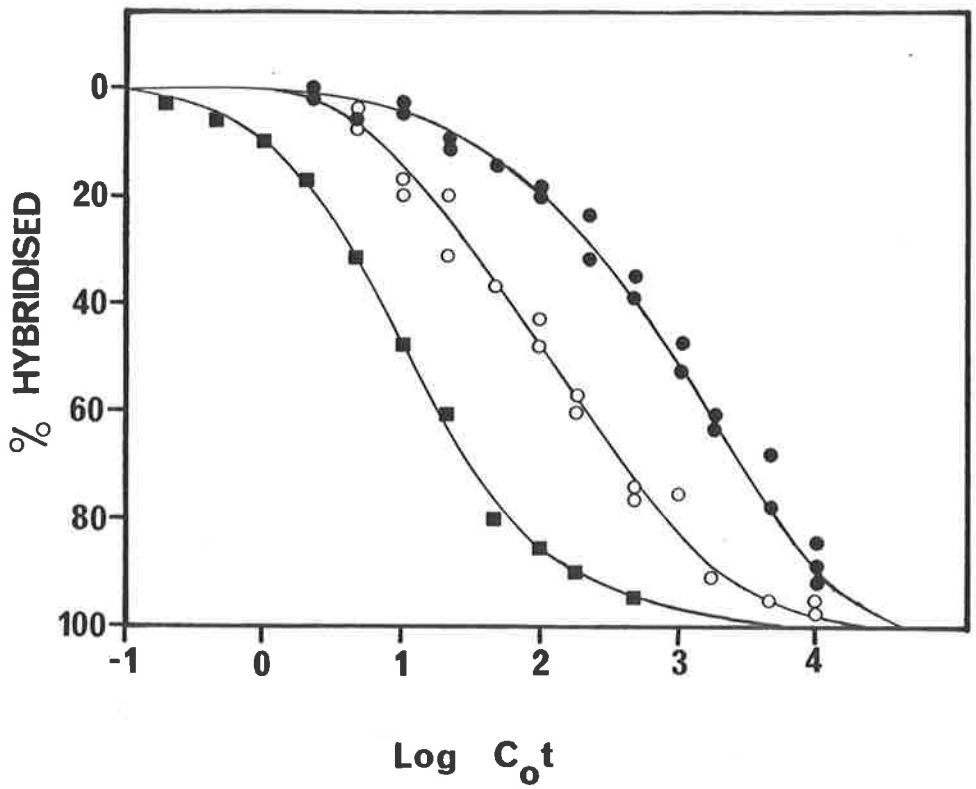
Figure 4.6

Estimation of the histone gene reiteration frequency in the chicken genome. Reassociation reactions of ^{3}H-labelled total chicken genomal DNA in the presence of either ^{32}P-labelled histone cDNA or ribosomal cDNA were assayed at various times using nuclease S_1. At the highest $C_o t$ values the maximum level of hybridisation was about 70%; the points have been renormalised to 100%.

●—● ^{3}H-labelled chicken genomal DNA.

○—○ ^{32}P-labelled histone cDNA.

■—■ ^{32}P-labelled ribosomal cDNA.


% HYBRIDISED
0
20
40
60
80
100
-1
0
1
2
3
4
Log $C_o t$

to the chicken genomal DNA. The $C_o t_{\frac{1}{2}}$ value of the ribosomal gene curve is 12, indicating a reiteration frequency of about 100 for the ribosomal genes in the chicken genome. Clearly the $C_o t$ curves produced by ribosomal cDNA and the 5S histone cDNA are distinctly different, once again showing that the histone cDNA preparation is free of contaminating ribosomal cDNA sequences.

The reiteration frequency value of 10 for the histone genes in chicken is similar to the value of 10-20 observed in mouse, (Jacob, 1976) and 30-40 observed in man, (Wilson and Melli, 1977), but far less than the histone gene reiteration value of several hundred observed in sea urchin, (Kedes and Birnstiel, 1971). Scott and Wells, (1976), obtained a reiteration frequency of 10 for the tissue specific histone H5, in the chicken genome. Either the genes for H5 and the other histones are unlinked, and merely happen to have a similar reiteration frequency, or the H5 and normal histone genes share the ten-fold reiteration because they are physically associated. With the limited information available it is not possible to decide which of these two options is correct.

In summary, histone mRNA has been isolated from 5-day chick embryos and the mRNA identified

by *in vitro* translation. cDNA enriched for histone sequences has been used to obtain an estimate of the reiteration frequency of these genes. Similar preparations of histone cDNA have been used to select genomal histone genes from a chicken library, (Harvey and Wells, 1979).

CHAPTER 5

SIZE FRACTIONATION OF GLOBIN AND H5 mRNAS

EXPLANATORY NOTE

Some of the results presented in this chapter were obtained in association with other students.

All electrophoresis of RNA on acrylamide gels and subsequent elution of RNA from the gel was carried out in association with Alan Robins, who analysed the reticulocyte RNA translation products for their α and β globin composition.

The ribonuclease-H enzyme was prepared by Richard Harvey, who also carried out the de-adenylation of the chicken and rabbit reticulocyte RNA.

5.1 INTRODUCTION

While H5 mRNA has previously been purified, and shown to centrifuge at the 10S position on sucrose gradients, (Scott and Wells, 1976), the purified mRNA was only available in microgram quantities and so it did not seem profitable to electrophorese the whole preparation on formamide gels in order to determine the size of the mRNA more accurately. From the sucrose gradient size estimate, one can calculate that the molecular weight of H5 mRNA should be in the range of about 200,000 to 230,000 daltons. Unfortunately, the globin mRNAs so far isolated also fall in the molecular weight range 200-230,000 daltons. (Bishop *et al.*, 1972; Hamlyn and Gould, 1975; Lewin , 1975).

Unless the size estimate for H5 mRNA is greatly in error, it seems unlikely that H5 mRNA would be resolvable from globin mRNA on sucrose gradients. Preliminary experiments by Scott, (1975), suggested that the two mRNAs could not be separated on sucrose gradients, but the lack of sensitivity of the translation assay used, hampered the interpretation of these results. Nevertheless, the inherently low resolution of sucrose gradients is likely to offer, at best, only a partial purification of the H5 mRNA.

The most powerful technique for isolating individual species of RNA molecules from complex mixtures, is polyacrylamide gel electrophoresis. Recently there have been a number of reports of successful fractionation of RNA species that have only very small differences in molecular weight. For example, polyacrylamide gel electrophoresis has been used to separate the five individual histone mRNAs, from the complex mixture of RNA isolated from *Drosophila* tissue culture cells, (Burckhardt and Birnstiel, 1978) and from sea-urchin embryos, (Gross *et al.*, 1976a). Rabbit and mouse, (Morrison *et al.*, 1974) and human, (Nudel *et al.*, 1977), α and β globin mRNAs have been separated on gels even though the molecular weights of the two mRNA species are very similar, about 200,000 and 230,000 daltons for α and β sequences respectively. When separating mRNAs, the presence of heterogeneous length poly-A tracts on the mRNA further complicates the separation, because each species of RNA no longer electrophoreses as a single discrete entity on the gel.

A number of acrylamide gel systems were available for investigation and each was claimed to yield excellent resolution of single-stranded nucleic acids. There is little point, however, in separating different mRNA species on an acrylamide gel if the RNA cannot be recovered

from the gel in a biologically active form, and the literature contains a number of reports of the difficulty of recovering RNA from acrylamide gels in active form and in adequate yields, (Hamlyn and Gould, 1975; Nudel *et al.*, 1977).

Schuerch *et al.*, (1975), described a gel system that not only provided excellent resolution of single-stranded RNA species of up to about $3x10^6$ daltons, but also allowed recovery of greater than 50 percent of the RNA from the gel in translatable form. It was claimed that the agarose-acrylamide-urea gel system described was superior in resolution to equivalent acrylamide-urea and acrylamide-formamide gel systems. A series of experiments was undertaken to determine the effectiveness of the agarose-acrylamide-urea gel system for the separation of globin and H5 mRNA sequences. In the course of this work, a very useful general technique for the electrophoretic elution of RNA from gels, in high yields, was developed, and this procedure was used in all subsequent polyacrylamide gel experiments. (The technique is described in detail in chapter 2). In practice, the results obtained using the agarose-acrylamide gels were disappointing. When the RNA was visualised in the gel using ethidium bromide staining, it was visible as a single band spread out over a broad region of the gel. Translation experiments showed

that no separation of globin and H5 mRNA had been achieved. Overall, the results obtained suggested that this gel system was not really suitable for low molecular weight RNA molecules.

In an effort to achieve much better resolution of low molecular weight single-stranded RNA, another gel system was investigated. The DNA-sequencing gels of Sanger and Coulson, (1975), were reported to have excellent resolving power for single-stranded nucleic acids. 8M urea is included in the gels in an attempt to ensure complete denaturation of secondary structure. This is essential if nucleic acids differing in length by only a single nucleotide base are to be clearly resolved, as required in sequencing experiments. While maximum resolution is obtained using high percentage acrylamide gels, this would make the recovery of RNA from the gels very difficult, and so as a compromise 4 percent acrylamide gels were used in the initial experiments. When 10S reticulocyte RNA was electrophoresed on these gels and visualised by staining, two sharp bands of RNA were observed, apparently analogous to the bands observed when rabbit and mouse globin mRNA were electrophoresed on gels, (Morrison *et al.*, 1974). The results of translation experiments however, showed the RNA to be broadly distributed on the gel and gave no evidence for a resolution of H5 and globin mRNA sequences.

Assuming that secondary structure in the mRNA molecules was at least one of the factors contributing to the broad distribution of RNA across the gel, and preventing resolution of H5 and globin mRNA sequences, fractionation was attempted using 98 percent formamide gels, (Pinder *et al.*, 1974). While this type of gel had previously been used to separate rabbit α and β globin mRNAs, (Hamlyn and Gould, 1975) and human α and β globin mRNAs, (Nudel *et al.*, 1977), and so had the proven ability to separate mRNAs of very similar size, problems in recovering translatable RNA from these gels had been reported, (Hamlyn and Gould, 1975). It was hoped that the ethidium bromide staining procedure and the electro-elution techniques used in the previous experiments would allow the recovery of translatable RNA from these gels. 10S chicken reticulocyte RNA was electrophoresed on these gels but, once again, no resolution of H5 and globin mRNA sequences, or resolution of the different globin mRNA species from one another, was observed. In a control experiment, rabbit globin mRNA was electrophoresed on formamide gels. Translation assays indicated that a significant fractionation of the rabbit α and β globin mRNA sequences had been achieved, and this result suggested that the problems experienced in attempting to fractionate H5 mRNA from globin mRNA, and chicken

α and β globin mRNAs from each other, were problems associated with the chicken mRNA system and not problems with the gel technique or RNA handling and assay procedures.

One of the reasons for the low resolution of mRNAs on gels is the presence of poly-A sequences, which are known to be heterogeneous in length, (Vournakis *et al.*, 1974), causing a broadening of the size distribution of any single RNA species. Vournakis *et al.*, (1975), were able to separate a number of previously unresolvable mRNAs, coding for silk moth chorion proteins, by removing the poly-A tails from the RNA, (using calf thymus ribonuclease H), and so making each mRNA species a discrete length. These poly-A minus mRNAs ran as tight bands on acrylamide gels and migrated faster than the poly-A plus RNA because they were, on average, about 60 bases shorter.

Removing the poly-A tracts from chicken 10S RNA should prove useful in two ways. Firstly, the globin mRNAs will each become a discrete length and should therefore, run as tight bands on gels, so increasing resolution between the species, and secondly, since H5 mRNA is not polyadenylated, (Scott and Wells, 1976), the poly-A minus globin mRNAs should now migrate further into the gel, leaving the H5 mRNA at its original position. The expected reduction in length of about 50 bases, (Scott, 1975; Crawford, 1977), should cause a

sufficient difference in migration to achieve at least a partial resolution of H5 mRNA from globin mRNA, provided that H5 mRNA really does remain unaltered in size.

Samples of both rabbit and chicken 10S RNA were de-adenylated by treatment with ribonuclease H, and the RNA electrophoresed on formamide gels. The de-adenylated RNA samples migrated further into the gel and gave sharper bands than the untreated RNA samples electrophoresed on identical gels. The de-adenylated rabbit globin mRNA showed extremely good resolution of the α and β globin mRNA species, but the de-adenylated chicken α and β globin mRNA species were only partially resolved. Even after ribonuclease H treatment, the H5 mRNA and globin mRNA sequences continued to coelectrophorese.

This chapter describes the efforts made to utilise any small difference in size that might exist between the globin and H5 mRNAs to bring about a separation of the two mRNA species on polyacrylamide gels. Despite the excellent results obtained using rabbit globin mRNA in control experiments, no useful enrichment of H5 mRNA sequences could be achieved using any of the high resolution polyacrylamide gel systems.

5.2 RESULTS

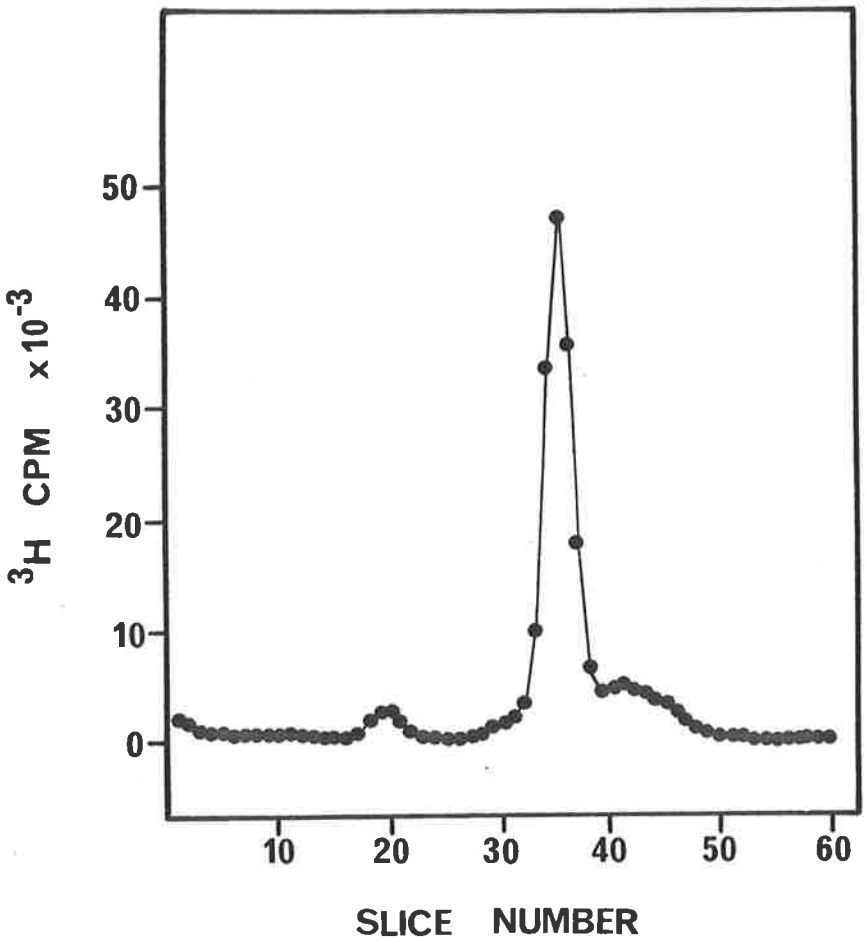
5.2.1 Electrophoresis of 10S Reticulocyte RNA on agarose-acrylamide-urea gels

A simple electrophoretic elution procedure was developed to recover RNA from polyacrylamide gels. The characterisation experiments were performed using ^{32}P-labelled *E. coli* ribosomal RNA and it was possible to demonstrate approximately 100 percent recovery of intact ribosomal RNA from agarose-acrylamide-urea gels, (data not shown). Before attempting to interpret the results of experiments using electro-eluted mRNA however, a control experiment was performed.

A sample of 10S chicken RNA was loaded onto an agarose-acrylamide-urea gel and electrophoresed a short distance into the gel. When visualised by ethidium bromide staining under UV light, the RNA was visible in a single sharp band. This section of the gel was excised and the RNA electro-eluted. A sample of the RNA was translated in the wheat germ cell-free system and the products analysed on an SDS-urea gel. The radioactivity profile produced, (figure 5.1), was identical to that obtained when total 10S RNA, (that had not been electrophoresed or electro-eluted), was translated, (see, for example, figure 3.2). This result demonstrated that translatable mRNA could readily be recovered from gels and that neither the H5 nor the globin

Figure 5.1

Analysis of the *in vitro* translation products of RNA electrophoretically eluted from an agarose-acrylamide-urea gel. The ^{3}H-leucine labelled translation products were examined on an SDS-urea tube gel.


50
40
30
20
10
0
^{3}H CPM x10^{-3}
10
20
30
40
50
60
SLICE NUMBER

mRNA was preferentially broken down in the gel or preferentially electro-eluted out of the gel. Therefore, the distribution of mRNA, as indicated by translation experiments, will faithfully represent the distribution of mRNA in the gel.

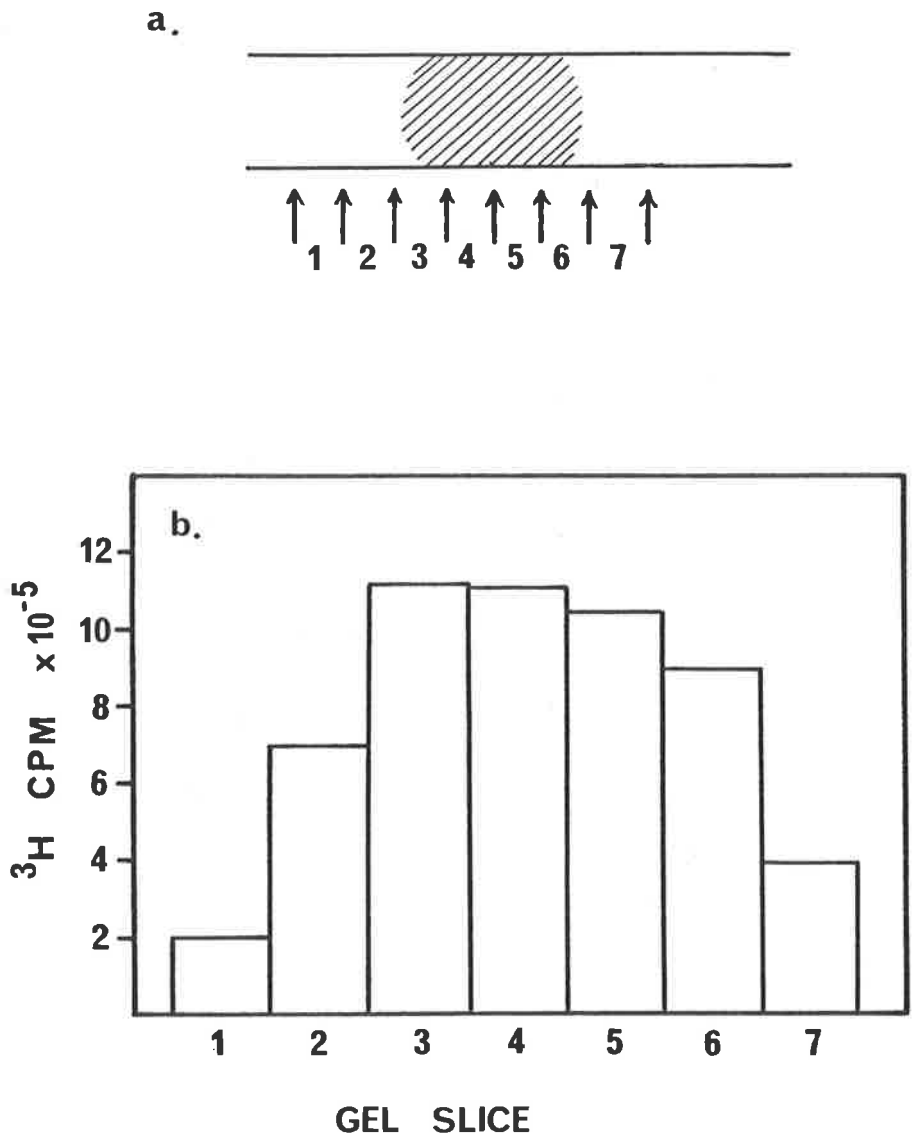
With the successful completion of the control experiment, an attempt was made to separate the H5 and globin mRNA sequences. Total 10S RNA isolated from reticulocytes was heated to 80^{o}C for 3 minutes to disaggregate the RNA, snap-cooled and then loaded onto a 0.6 percent agarose-2.1 percent acrylamide-6M urea tube gel, (Schuerch *et al.*, 1975). In a standard experiment, 5 micrograms of 10S RNA was loaded per gel. After electrophoresis, the gel was removed from the tube and the RNA visualised under UV light by ethidium bromide staining. The RNA was visible as a broad fluorescent band about 4 centimetres down from the top of the gel. The two bands observed when chicken globin mRNA was electrophoresed analytically on 98 percent formamide gels, (Crawford, 1977), were apparently not resolved under these electrophoresis conditions. The region of the gel containing RNA was sliced with a sterile scalpel blade and seven approximately equal sections of gel were removed. As shown in figure 5.2a, these slices included sections of the gel in which no RNA at

Figure 5.2

Electrophoresis of reticulocyte 10S RNA on agarose-acrylamide-urea tube gels.

a. General appearance of ethidium bromide stained RNA on the gel. The arrows mark the position of the slices.

b. Efficiency of translation of the eluted RNA fractions in the wheat-germ cell-free system.


a.
1
2
3
4
5
6
7
b.
12
10
8
6
4
2
^3H CPM $\times 10^{-5}$
1
2
3
4
5
6
7
GEL SLICE

all was visible. The RNA was eluted from each gel slice and samples were translated in the wheat germ cell-free system. As illustrated in figure 5.2b, RNA eluted from every slice was able to stimulate translation to a level significantly above background. This was an encouraging result because it showed that even very small amounts of RNA could be recovered from the gel and detected using the translation assay.

The labelled translation products, programmed by RNA derived from each of the seven gel slices, were individually analysed on SDS-urea gels and the amount of radioactivity present in H5 protein was determined for each fraction. The number of H5 protein counts in each fraction, expressed as a percentage of the total number of counts in H5 protein in all fractions combined, was calculated, and the distribution of H5 mRNA across the gel represented on this basis. The same analysis was applied to globin translation products. The H5 and globin mRNA distributions are presented in figure 5.3a. If the two mRNA species had been clearly resolved by the gel, the two distributions shown in the diagram would not overlap. Reference to the diagram shows that although the H5 mRNA sequences might

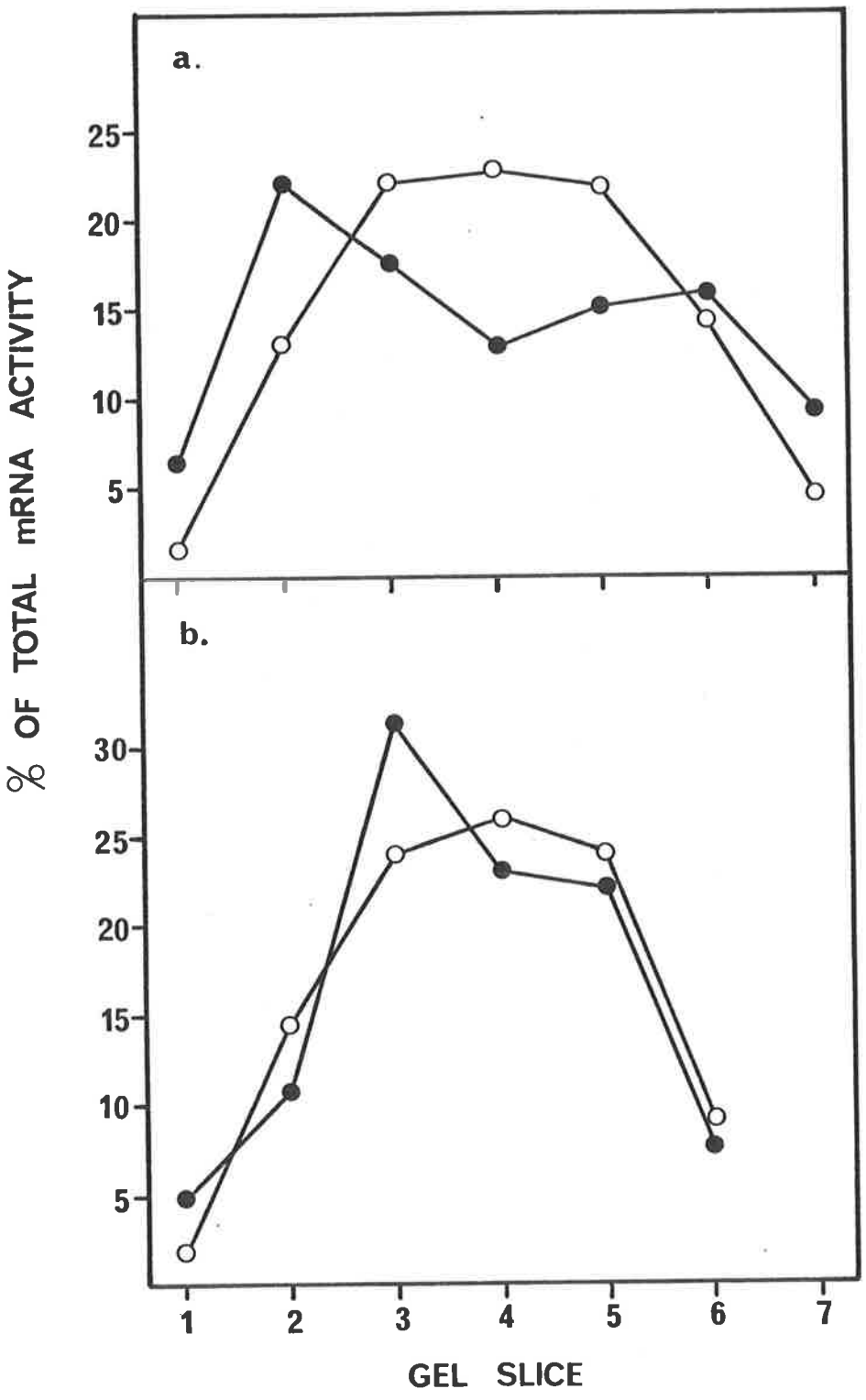
Figure 5.3

Distribution of total globin and H5 protein detected in the translation products of RNA eluted from agarose-acrylamide-urea tube gels. In each case the main band of RNA had been electrophoresed about 4 cm into the gel.

a and b. Results of duplicate experiments.

●—○ globin protein

●—● H5 protein


a.
b.
25
20
15
10
5
30
25
20
15
10
5
1
2
3
4
5
6
7
% OF TOTAL mRNA ACTIVITY
GEL SLICE

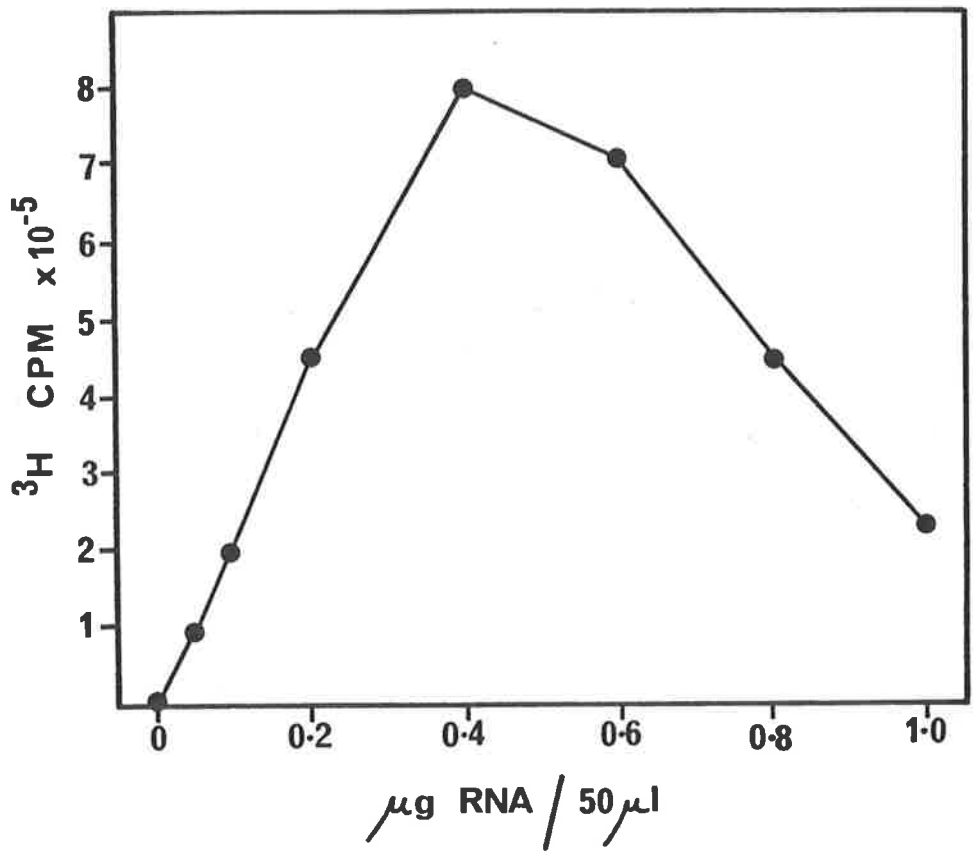
show a slight displacement towards the top of the gel, the H5 mRNA also appears to be spread over a greater length of the gel than the globin mRNA.

Figure 5.3b shows the results of a second experiment using the agarose-acrylamide gels. Once again the RNA appeared as a broad featureless band when stained with ethidium bromide and again, the distributions of H5 and globin mRNA overlap almost completely.

While the majority of the translatable RNA was eluted from the stained regions of the gel, a considerable amount of mRNA was recovered from slices of the gel well removed from the stained RNA bands. This seems to indicate that the RNA was very broadly distributed and not localised to sharp bands at all. However, there is an alternate explanation for the apparent broadness of the RNA distribution. The wheat germ preparation used in these translation experiments was saturated by the addition of about 0.5 μg of globin mRNA, (figure 5.4), and above this quantity the amount of label incorporated into protein products drops markedly. Because of the uncertainties of recovery of translatable RNA from gels, it is possible that the aliquot of RNA from the peak fractions exceeded the saturating level of the translation system, and so the number of counts incorporated under-

Figure 5.4

Effect of increasing amounts of RNA on the efficiency of translation in the wheat-germ cell-free system. Standard 50 µl reaction mixes, (see section 2.2.3), contained the indicated amounts of 10S reticulocyte RNA.


^{3}H CPM x10^{-5}
8
7
6
5
4
3
2
1
0
0·2
0·4
0·6
0·8
1·0
µg RNA / 50µl

estimates the amount of RNA present. (In these experiments, if 100 percent of the RNA was extracted from the gel in translatable from, the peak fractions could contain up to about 1.2 micrograms of mRNA). The amount of RNA eluted from the unstained slices, however, is sure to be sub-saturating. This interpretation suggests that the actual distribution of RNA across the gel may be much sharper than indicated by the translation assay.

Due to the lack of sharp visual RNA bands on the gels, the spread of translatable RNA over a broad area and the lack of any effective resolution of H5 and globin mRNA sequences, experiments using the agarose-acrylamide gels were discontinued. The results obtained suggested that this gel system was not really suitable for the fractionation of low molecular weight RNA molecules and so a series of experiments were carried out using the high resolution acrylamide-urea gels used for DNA-sequencing.

5.2.2 Electrophoresis of 10S reticulocyte RNA on acrylamide-urea gels

As a compromise between maximum resolution, (requiring a high percentage acrylamide), and ease of recovery of RNA from the gels, (low percentage acrylamide), tube gels with a final

concentration of 4 percent acrylamide were used in these experiments. 4 percent acrylamide-8M urea tube gels were prepared, exactly as described by Sanger and Coulson, (1975), and about 5 micrograms of heat-denatured 10S RNA was loaded onto the gels. In parallel experiments, rabbit globin mRNA and chicken 10S RNA were electrophoresed about 2 centimetres into the acrylamide-urea gels before staining with toluidene-blue and destaining in sterile water. All staining and destaining operations were carried out in total darkness to prevent the light-catalysed degradation of RNA by toluidene-blue stain, (Popa and Bosch, 1969).

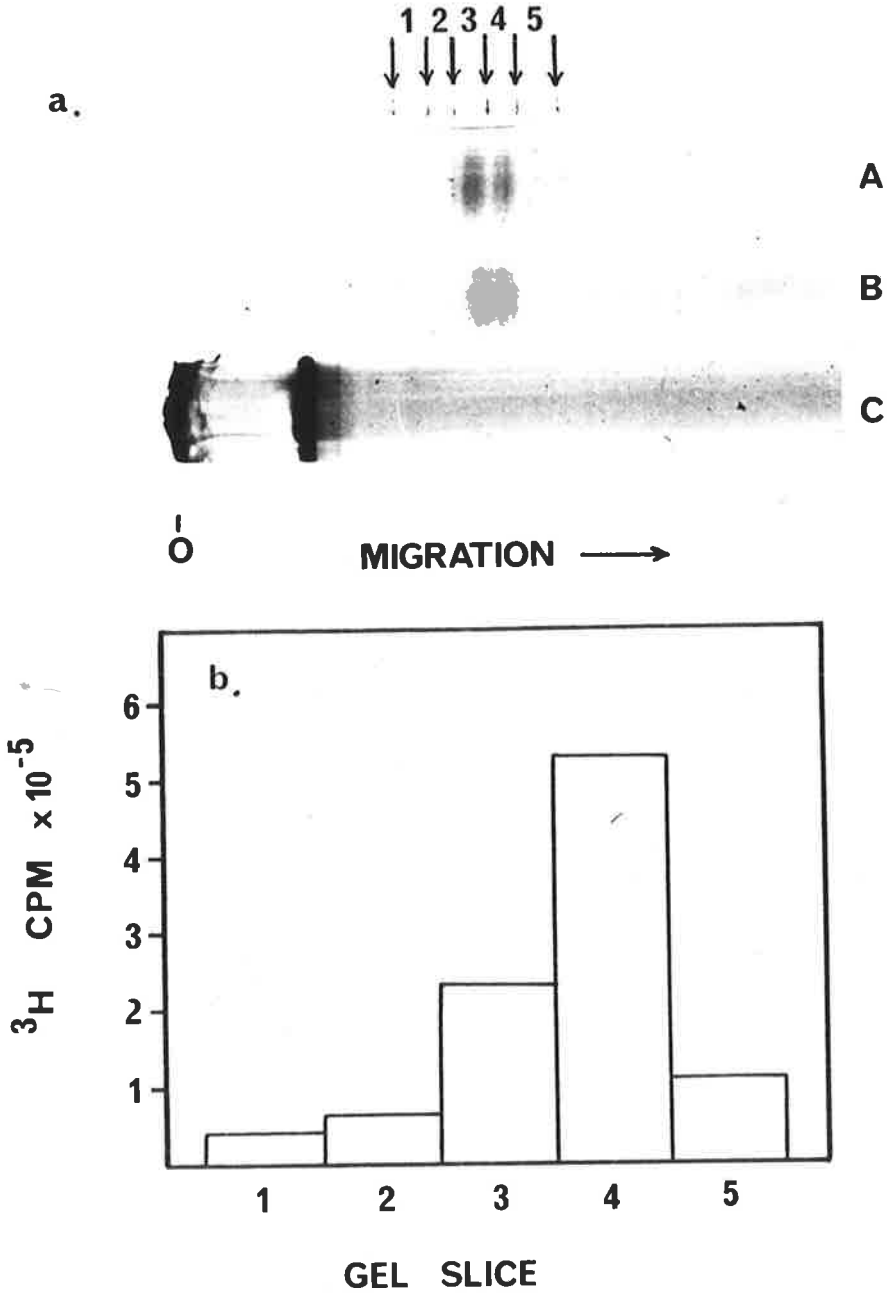
Figure 5.5a shows the appearance of chicken 10S reticulocyte RNA and of 10S rabbit globin mRNA on a gel after destaining. In both cases, two slightly diffuse stained bands can be observed, with the chicken RNA bands apparently better separated than those of the rabbit RNA. The gel containing chicken 10S RNA was sliced, (as shown in figure 5.5a) and the RNA electro-eluted and translated. RNA from all gel slices stimulated translation in the wheat germ system, as indicated in figure 5.5b, and, once again, fractions containing little or no visible RNA were able to stimulate the translation system to a level many times above background.

Figure 5.5

Electrophoresis of reticulocyte 10S RNA on acrylamide-urea tube gels.

a. Appearance of toluidene-blue stained rabbit globin mRNA and chicken reticulocyte 10S RNA, (2 µg of each), electrophoresed for 4 hours at 2 mA per gel on acrylamide-urea tube gels.
A, chicken 10S reticulocyte RNA, B, rabbit globin mRNA, C, 18S and 28S chicken ribosomal RNA markers.

b. Efficiency of translation of eluted RNA fractions in the wheat-germ cell-free system.


a.
1 2 3 4 5
A
B
C
O
MIGRATION
b.
^3H CPM $\times 10^{-5}$
6
5
4
3
2
1
1
2
3
4
5
GEL SLICE

The translation products of each fraction were examined independently on SDS-urea gels and the proportions of H5 and globin protein calculated. The distribution of H5 and globin coding sequences is illustrated in figure 5.6. While the distribution of mRNA sequences is not as broad as those observed in the previous experiments, (figure 5.3), this is probably at least partly due to the fact that the RNA was not electrophoresed as far into the gel. The centres of the H5 and globin mRNA distributions once again, coincide, showing that there has been no effective resolution of the H5 and globin mRNA sequences.

When the fractionation of chicken α and β globin coding sequences was tested, (Allan Robins), there was found to be no resolution of the different species of mRNA, despite the clear separation of two RNA bands on the gels. It seems that the high visual resolution apparently afforded by the 4 percent acrylamide-8M urea gels was largely artifactual, since these gels were no more effective than the agarose-acrylamide gels in achieving a separation of H5 mRNA from globin mRNA or of α and β globin mRNA sequences from each other.

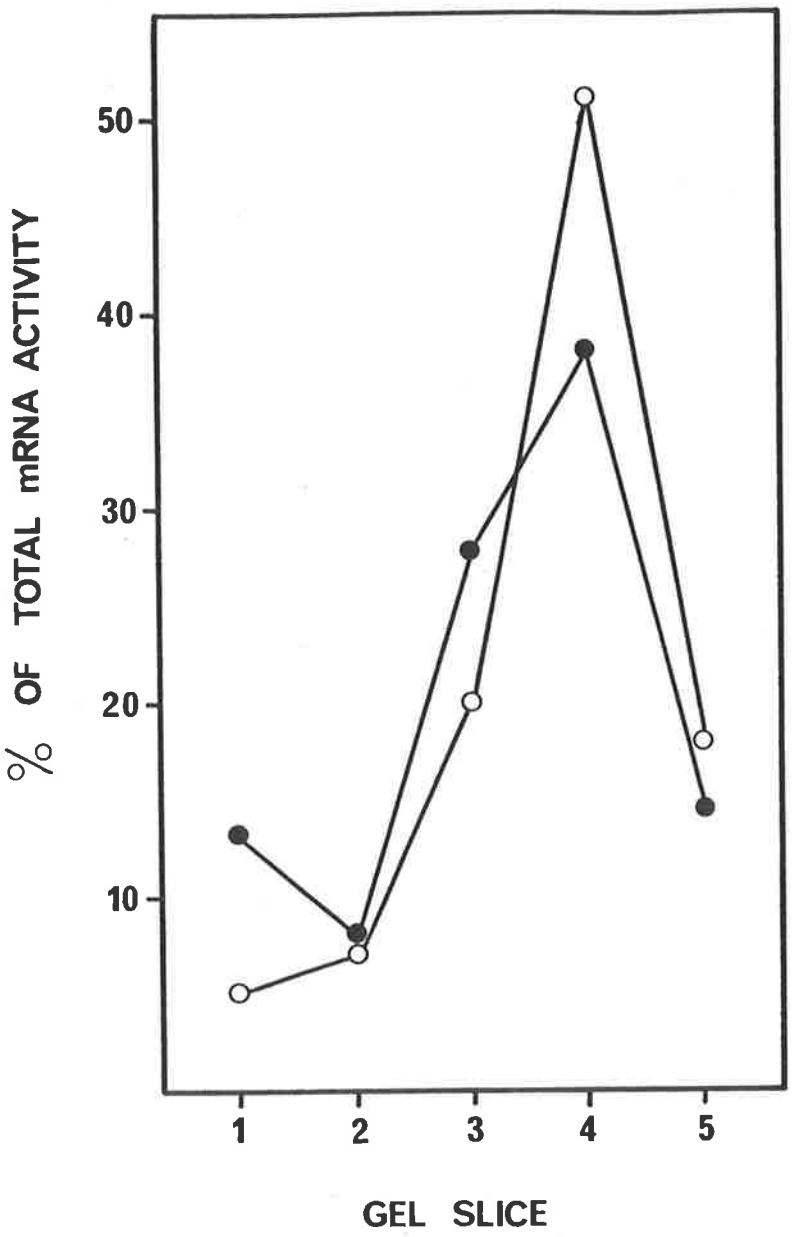
The lack of resolution of the acrylamide-urea gels may have been due to residual secondary structure in the RNA, since Sanger and Coulson,

Figure 5.6

Distribution of total globin and H5 protein detected in the translation products of RNA eluted from acrylamide-urea tube gels. The gel slices are shown in figure 5.5a.

o—o globin protein

●—● H5 protein


% OF TOTAL mRNA ACTIVITY
50
40
30
20
10
1
2
3
4
5
GEL SLICE

(1975), suggested that these gels were most denaturing when electrophoresis was carried out at elevated temperatures. Rather than raise the temperature of electrophoresis of the acrylamide-urea gels, and risk degradation of the RNA, secondary structure effects could largely be eliminated by electrophoresis of the RNA in gels containing 98 percent formamide. Because formamide gels had already been used to separate α and β-globin mRNA sequences, (Hamlyn and Gould, 1975; Nudel *et al.*, 1977), this gel system was used in all further experiments.

5.2.3 Electrophoresis of 10S reticulocyte RNA on formamide gels

4 percent acrylamide gels containing 98 percent formamide (Pinder *et al.*, 1974), were prepared in 16 centimetre long tubes. These gels were twice as long as those used in previous experiments since it had been shown that migration over a greater distance permitted greater separation of individual RNA species. For example, Longacre and Rutter, (1977), reported better separation of chicken globin mRNA species when these were electrophoresed long distances into the gel.

5 micrograms of 10S reticulocyte RNA was heat denatured in 80 percent formamide, loaded onto the gel and electrophoresed overnight.

When visualised by ethidium bromide staining the RNA was visible as two broad diffuse bands about 10 centimetres down from the top of the gel. The gels were sliced, as indicated in figure 5.7a, the RNA electro-eluted and samples translated in the wheat germ cell-free system. The amount of translation stimulated by RNA from each slice is illustrated in figure 5.7b. Translated material was electrophoresed on SDS-urea gels and the proportion of H5 and globin protein present was estimated for each fraction. The distribution pattern of the two mRNA species across the six gel slices is shown in figure 5.8.

Unfortunately, it seems that the use of formamide gels has not solved the resolution problem. While fractions 1 and 2 appear to be somewhat enriched for H5 mRNA sequences, the peak of the H5 mRNA distribution still lies directly under the peak of the globin mRNA distribution. If anything, the H5 distribution is broader than the globin distribution and it is this factor that has led to the enrichment of the H5 mRNA in the top two slices. Simultaneous experiments, in which the distributions of the different globin species were examined, showed that there was no separation of the chicken α and β globin mRNAs on these long formamide gels, (A. Robins - unpublished data).

Figure 5.7

Electrophoresis of 10S reticulocyte RNA on formamide tube gels.

a. General appearance of ethidium bromide stained gel. Arrows mark the position of the slices.

b. Efficiency of translation of the eluted RNA fractions in the wheat-germ cell-free system.

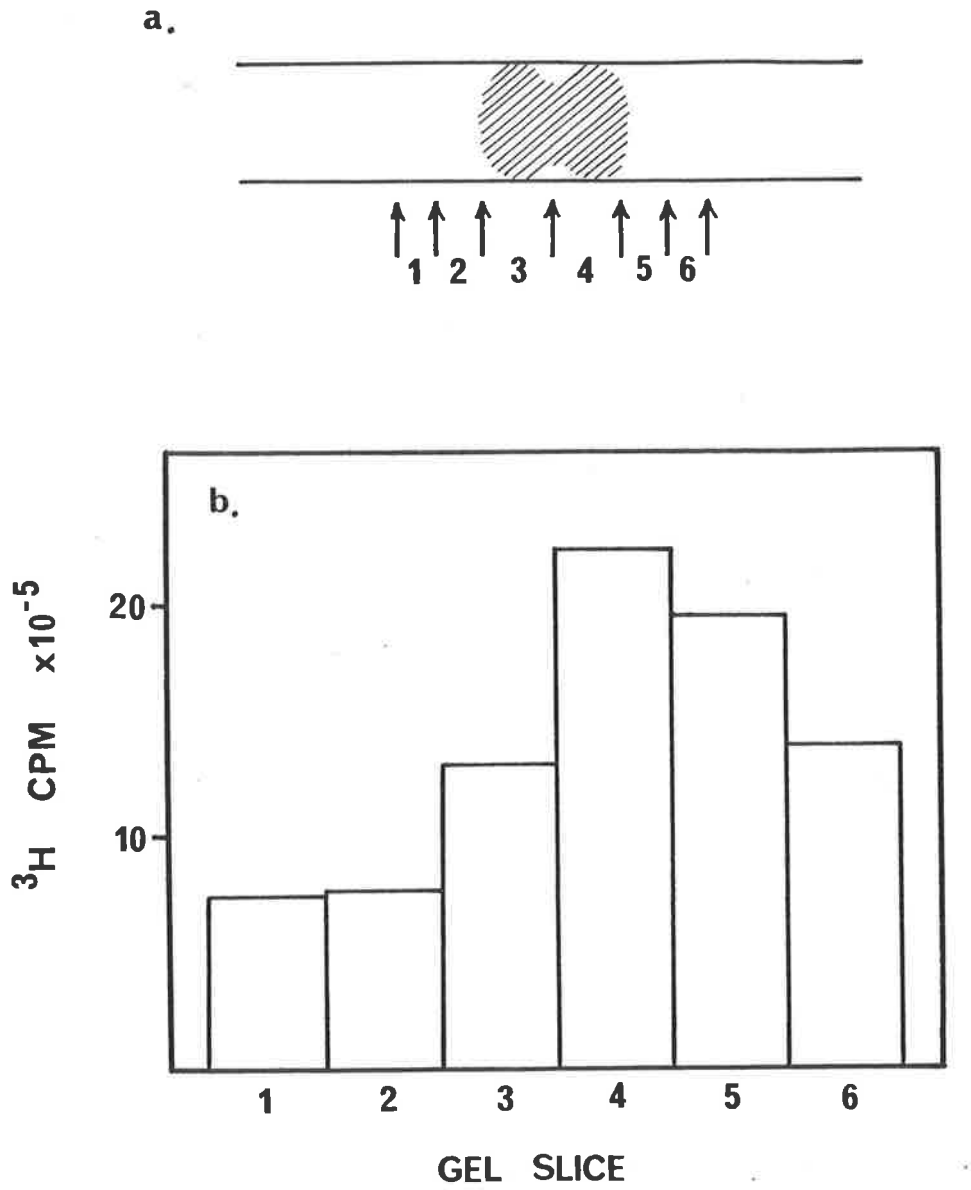
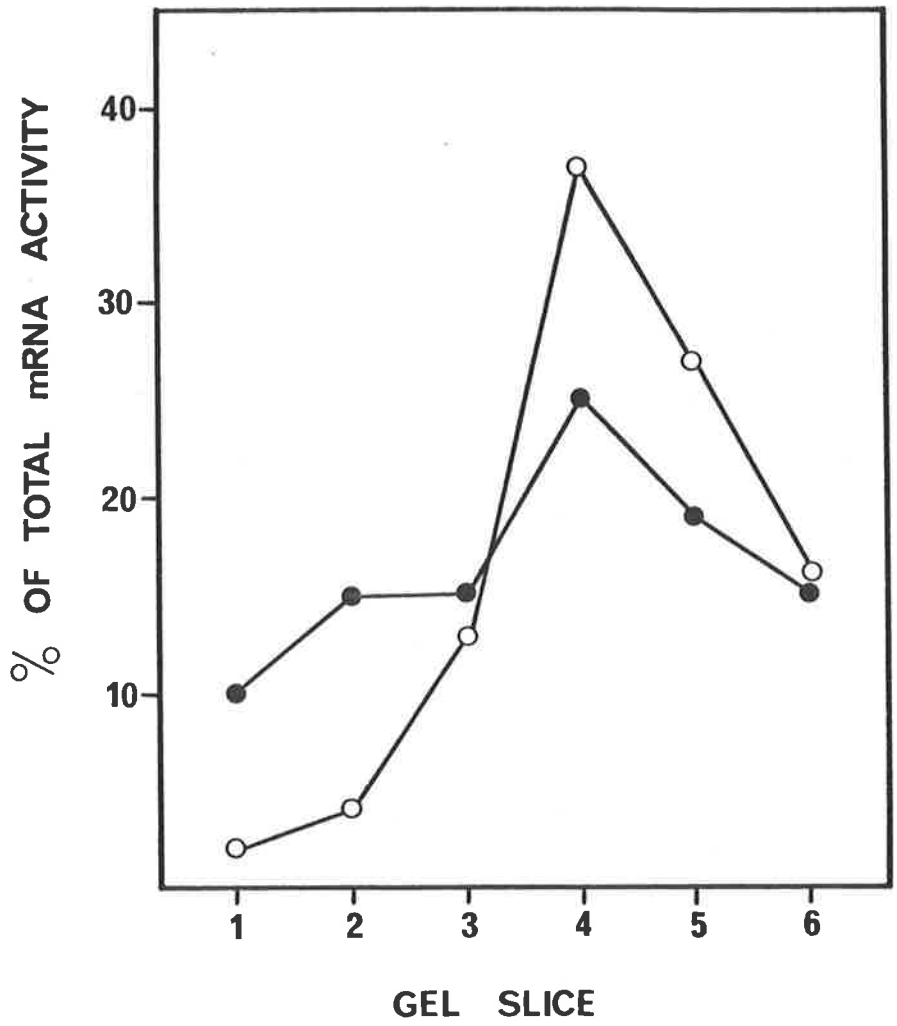

a.
1 2 3 4 5 6
b.
20
10
^{3}H CPM x10^{-5}
1 2 3 4 5 6
GEL SLICE

Figure 5.8

Distribution of total globin and H5 protein in the translation products of RNA eluted from formamide tube gels.

o—o globin protein

●—● H5 protein


% OF TOTAL mRNA ACTIVITY
40
30
20
10
1
2
3
4
5
6
GEL SLICE

These results were discouraging, since others had reported little difficulty in separating mammalian α and β globin mRNAs on formamide gels, (Morrison *et al.*, 1974; Hamlyn and Gould, 1975; Nudel *et al.*, 1977). To show that the formamide gel system was performing acceptably and that the techniques of electro-elution and translation, as used, were adequate to demonstrate a fractionation of mRNAs if it occurred, an attempt was made to separate rabbit globin mRNA into its α and β components.

5.2.4 Electrophoresis of rabbit globin mRNA on formamide gels

5 micrograms of rabbit globin mRNA was loaded onto a 4 percent acrylamide-98 percent formamide tube gel, and electrophoresed until the mRNA had migrated about 4 centimetres into the gel. Two slightly diffuse RNA bands were visible when the gel was stained with ethidium bromide. The gel was sliced as shown in figure 5.9a and RNA eluted from the gel segments and translated, (figure 5.9b). The labelled translation products were separated on gels containing Triton X-100, (Zweidler and Cohen, 1972). As shown in figure 5.10a, these gels were able to resolve the two rabbit globin proteins. The translation product profile of unfractionated rabbit globin mRNA is presented

Figure 5.9

Electrophoresis of rabbit globin mRNA on formamide tube gels.

a. General appearance of the ethidium bromide stained gel. Arrows mark the position of the slices.

b. Efficiency of translation of the eluted RNA fractions in the wheat-germ cell-free system.

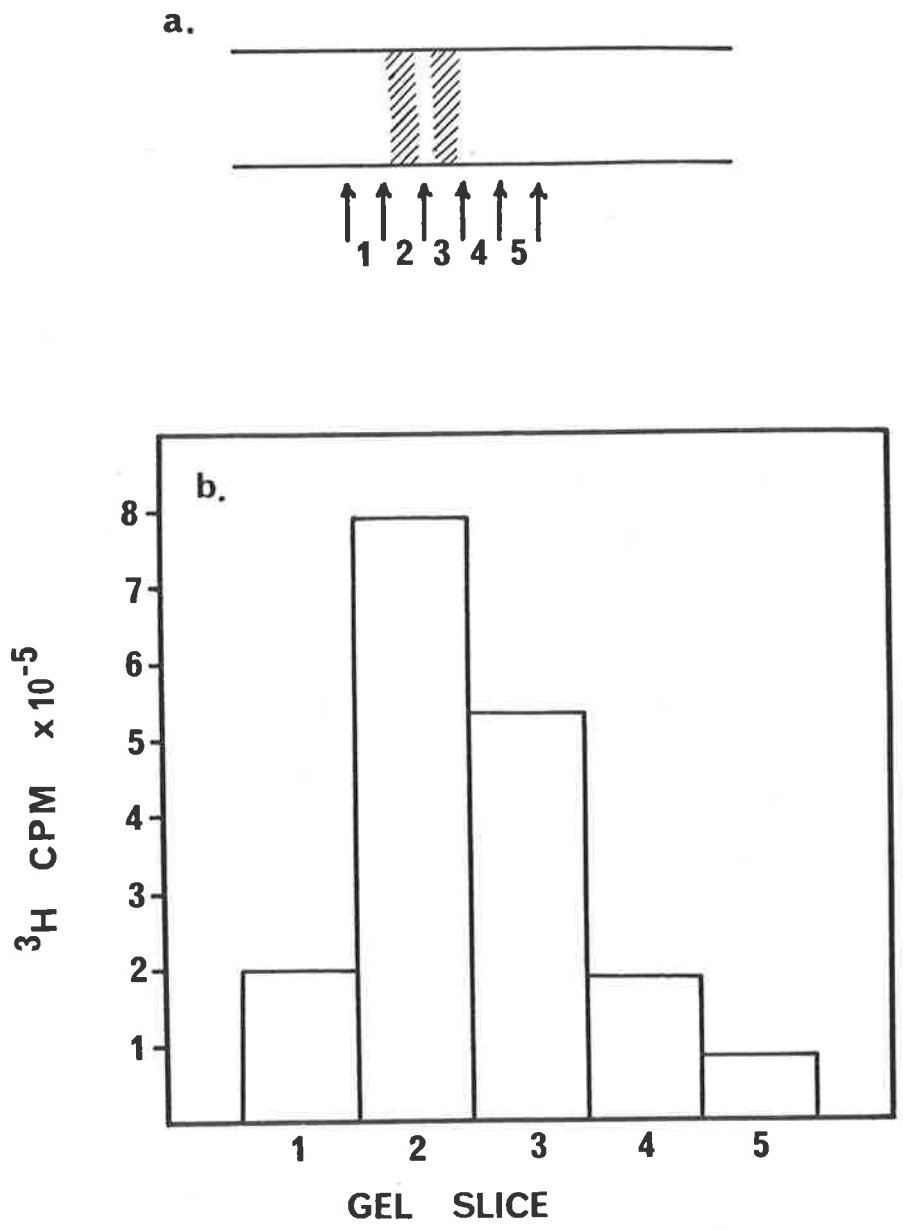

a.
1 2 3 4 5
b.
8
7
6
5
4
3
2
1
^{3}H CPM x10^{-5}
1 2 3 4 5
GEL SLICE

Figure 5.10

Analysis of rabbit globin protein on acrylamide gels containing Triton X-100, (Zweidler and Cohen, 1972).

a. Appearance of rabbit α and β globin protein electrophoresed on a Triton slab gel and stained with Coomassie-blue. The relative position of the α and β proteins was not determined.

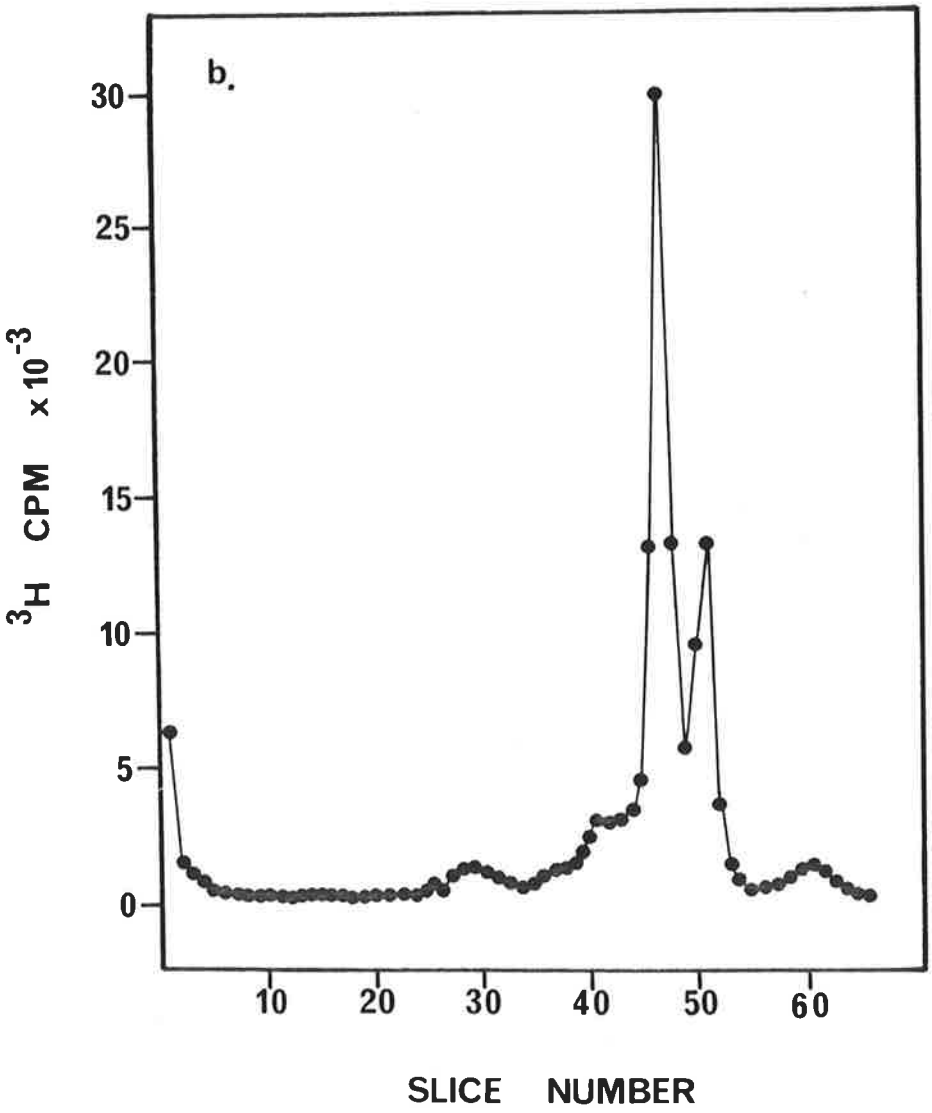
b. Electrophoresis of ^{3}H-leucine labelled translation products of rabbit globin mRNA on Triton tube gels.

a.

O

MIGRATION ⟶

b.



in figure 5.10b and the translation profiles of RNA eluted from slices 1, 2 and 3 of the formamide gel are shown in figure 5.11.

Obviously there is a very marked difference between the proteins synthesised by different RNA fractions across the gel, especially the adjacent fractions 2 and 3. Hamlyn and Gould, (1975), have shown that the β-globin mRNA is the slower migrating species on formamide gels and so the major protein product of this RNA band will be β-globin. When RNA from the slower migrating band, (slice 2), was eluted and translated, the major product synthesised corresponds in position to the trailing peak on the Triton gel, (figure 5.10). Therefore, it may be concluded that the leading and trailing peaks on the Triton gel are α and β globin respectively.

RNA eluted from the top slice of the formamide gel, (slice 1, figure 5.9), yields a single peak of β-globin protein when translated and therefore contains apparently pure β-globin mRNA. Slice 2 yields protein products showing a major peak of β-globin and a minor α-globin peak, while slice 3 RNA translated primarily into α-globin protein. Reference to the translation product profile of unfractionated rabbit globin mRNA, (figure 5.10b), allows the calculation of the relative amounts of α and β-globin mRNA in

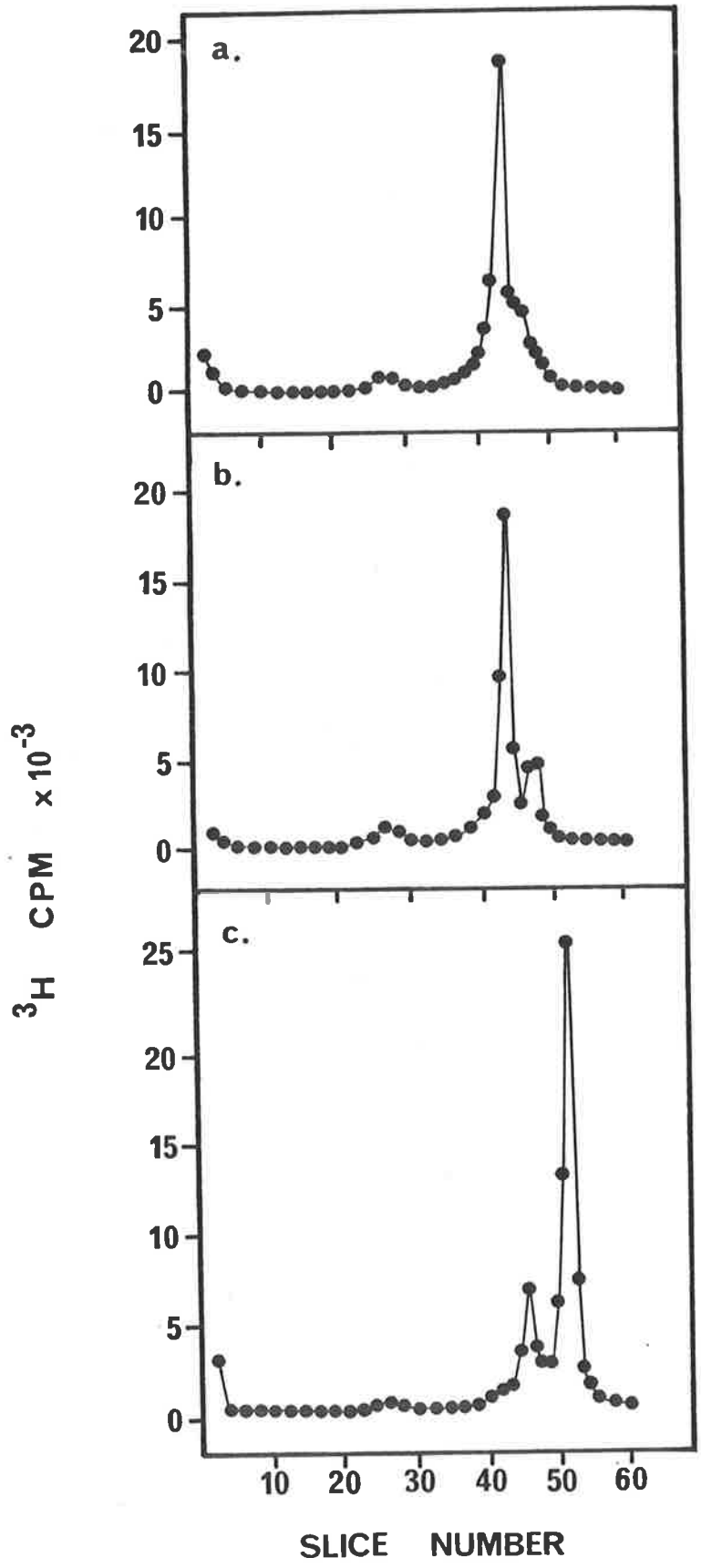
Figure 5.11

Analysis of the *in vitro* translation products of rabbit globin mRNA eluted from a formamide tube gel, (sliced as shown in figure 5.9a).

a. ^{3}H-labelled products of RNA eluted from slice 1.

b. ^{3}H-labelled products of RNA eluted from slice 2.

c. ^{3}H-labelled products of RNA eluted from slice 3.


a.
b.
c.
20
15
10
5
0
25
10
20
30
40
50
60
^3H CPM $\times 10^{-3}$
SLICE NUMBER

each fraction, if we assume that the two mRNAs are present in equal amounts initially, (Hamlyn and Gould, 1975). On this basis, fraction 2 RNA is 65 percent β-globin mRNA, while the adjacent fraction 3 contains nearly 90 percent α-globin mRNA. Clearly, a very significant fractionation of the two mRNA species has been achieved, and, in fact, these results are almost identical to those reported by Hamlyn and Gould, (1975). The success achieved using rabbit globin mRNA and the formamide gel system, suggests that the problems experienced in attempting to fractionate H5 mRNA from globin mRNA and chicken α and β globin mRNAs from each other, are problems associated with the chicken globin mRNA system and not problems with the gel technique or RNA handling and assay procedures.

One observation made during the course of this work is worth mentioning. For reasons already described, the trailing peak on the Triton gels has been ascribed to β-globin protein. This means that the major protein peak observed when unfractionated rabbit globin mRNA was translated in the wheat germ system, (figure 5.10b), was also β-globin. This result is not consistent with other reports, (Roberts and Paterson, 1974; Longacre and Rutter, 1977), which suggests that α-globin is preferentially translated in the wheat-germ system. While this question was of some interest, it was not pursued since it did not

effect the basic conclusion of the experiment; that the two rabbit globin mRNAs could readily be separated using techniques that would not resolve the components of chicken 10S mRNA.

5.2.5 Ribonuclease H treatment of 10S reticulocyte RNA

In an attempt to increase separation of H5 and globin mRNA sequences and to improve the resolution of the globin mRNA species, chicken 10S RNA was treated with ribonuclease H to remove the heterogeneous-length poly-A tracts from the mRNA molecules. A control experiment using rabbit globin mRNA was also included.

Ribonuclease H was prepared from *E. coli* using the method described by Berkower *et al.*, (1973). Rabbit and chicken 10S RNA were deadenylated in the following manner in a duplicate experiment, (performed by R. Harvey).

10 micrograms of globin mRNA was suspended in 600 microlitres of ribonuclease H digestion buffer and divided into two 300 microlitre fractions. To one of these fractions, 2 micrograms of oligo-dT_{10} was added, and the mixture was incubated at room temperature for 15 minutes to facilitate hybridisation to poly-A tracts. Ribonuclease H enzyme was added to both fractions, which was incubated at 37^{o} for 10 minutes and then phenol extracted. Phenol extraction was carried out at pH 9.0, since specific loss of poly-A containing RNA has been reported to occur

at lower pH, (Brawerman *et al.*, 1972).

After ethanol precipitation the RNA fractions were electrophoresed on 4 percent acrylamide-98 percent formamide gels. These were stained with ethidium bromide and examined under UV light. Photographs of the stained gels are presented in figure 5.12. In the case of both the chicken and rabbit RNAs the differences caused by ribonuclease H treatment are noticeable. The diffuse bands have become sharper and the ribonuclease H treated RNA has migrated further into the gel, indicating that it is shorter than untreated RNA.

Four, 3 millimetre slices were excised from each of the gels shown in figure 5.12, one slice containing each visible RNA band, and one slice above and below the bands. The RNA was electro-eluted from all slices and translated. Translation products from all four gel slices were analysed on Triton gels to determine the efficiency of α and β globin mRNA fractionation, (A. Robins), while the translation products of ribonuclease H treated and untreated chicken RNA fractions were electrophoresed on SDS-urea gels and the amounts of H5 protein and globin protein determined for each gel slice. The H5 and globin mRNA distributions are presented in figure 5.13, for both adenylated and deadenylated chicken RNA.

Figure 5.12

Electrophoresis of ribonuclease H-treated and untreated RNA on formamide tube gels.

a. Chicken reticulocyte 10S RNA.
   A. treated with ribonuclease H
   B. untreated control RNA

b. Rabbit globin mRNA.
   A. treated with ribonuclease H
   B. untreated control RNA

a.

b.

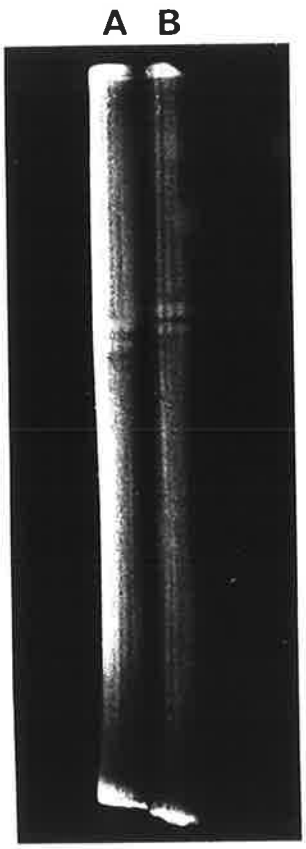


Figure 5.13

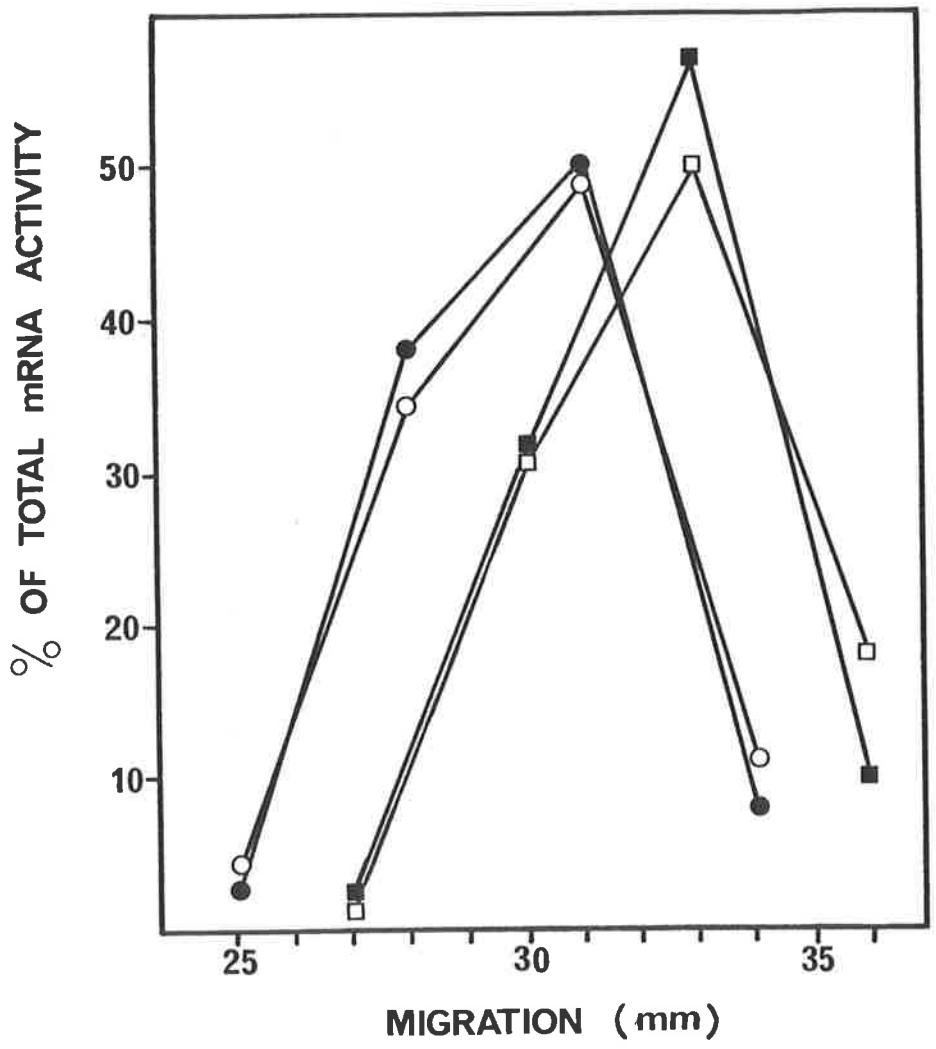
Distribution of total globin and H5 protein detected in the translation products of ribonuclease H treated and untreated reticulocyte 10S RNA eluted from formamide gels.

○—○ globin protein - untreated RNA

●—● H5 protein - untreated RNA

□—□ globin protein - ribonuclease H-treated RNA

■—■ H5 protein - ribonuclease H-treated RNA


% OF TOTAL mRNA ACTIVITY
50
40
30
20
10
25
30
35
MIGRATION (mm)

In both cases the H5 and globin mRNA distributions coincide exactly. This is the expected result for adenylated RNA since it has been observed before, (figure 5.7), but it is not consistent with the predictions for ribonuclease H-treated RNA. If H5 mRNA was unaffected by ribonuclease H treatment, as expected, the H5 mRNA should have migrated exactly the same distance into the gel as it did in the untreated RNA experiment. This would have placed the peak of the H5 mRNA distribution about 2 millimetres behind the peak of the globin distribution, (since globin mRNA has migrated about 2 millimetres further into the gel due to the removal of the poly-A tails).

In the control experiment, using rabbit globin mRNA, the ribonuclease H treatment achieved a further improvement in the separation of the α and β globin mRNAs. Adjacent slices 2 and 3 of the ribonuclease H-treated rabbit globin mRNA showed 85 percent pure β-globin mRNA and 99 percent pure α-globin mRNA respectively. The chicken α and β globin mRNAs remained effectively unfractionated after deadenylation, adjacent slices yielding 55 percent pure β-globin mRNA and 65 percent pure α-globin mRNA.

Two explanations are available for the failure of deadenylation to separate H5 and

globin mRNAs. Firstly, there may be interactions between globin and H5 mRNA sequences such that even 98 percent formamide is not sufficient to achieve complete disaggregation. Secondly, contrary to previous observations, H5 mRNA may be polyadenylated. If H5 mRNA carried a poly-A tract with a length similar to that carried by globin mRNA, ribonuclease H treatment would shorten the H5 mRNA by about the same amount that globin mRNA was shortened and so the two mRNA species would continue to coelectrophorese on gels.

5.3 DISCUSSION

The results of this chapter show that it has not been possible to separate H5 mRNA from globin mRNA using polyacrylamide gel electrophoresis. While a number of gel systems have been studied, each reported to yield excellent resolution in specific cases, none of these has proved to have sufficient resolution to separate H5 and globin mRNA.

In the control experiments using 98 percent formamide gels it was shown that rabbit globin mRNA could readily be fractionated into its α and β mRNA components. This demonstrated that both the gel electrophoresis and the assay systems were working well, and that the problems of separating H5 from globin mRNA and α-globin from β-globin mRNA were associated with the chicken

RNA itself and not the techniques employed.

The precise coincidence of the H5 and globin mRNA distributions, on all gel systems investigated, suggests that H5 mRNA must be very similar in size to the chicken globin mRNAs. Indeed, the broad distribution of H5 mRNA on the gels suggests that the H5 mRNA might be heterogeneous in length, either due to polyadenylation or to different length mRNA transcripts from the different copies of the H5 genes, (Scott and Wells, 1976). Since polyadenylated α-globin mRNA has a molecular weight of about 200,000 and β-globin mRNA a molecular weight of around 230,000, the molecular weight of H5 mRNA probably lies very close to these values.

Treatment of rabbit globin mRNA with ribonuclease H produced a marked improvement in the fractionation of the α and β-globin mRNAs, but had little effect on the separation of chicken 10S RNA sequences. The question of polyadenylation of the H5 mRNA is raised again by the results of the ribonuclease H digestion experiments. If H5 mRNA is not polyadenylated, the deadenylation of globin mRNA and its consequent increased migration on denaturing gels should have, at least, started to separate the H5 and globin mRNAs. Instead, the two mRNA species coelectrophoresed exactly, suggesting

that the H5 mRNA had also been shortened. Since the ribonuclease H treatment, as performed, was specific for poly-A sequences, it seems very likely that the observed increase in the migration of H5 mRNA was due to the removal of a poly-A tract.

While the failure to resolve the H5 and globin mRNA sequences and the α and β globin mRNA sequences using polyacrylamide gels was disappointing, it is reassuring to note that this is probably because the fractionation is inherently difficult. While a number of attempts have been made to fractionate the chicken globin mRNAs, (Knöchel and Grundmann, 1977; Longacre and Rutter, 1977), a successful fractionation has not been reported.

CHAPTER 6

MOLECULAR CLONING OF RETICULOCYTE 10S RNA SEQUENCES

6.1 INTRODUCTION

Chapter 3 described the use of affinity chromatography techniques as a means for enriching an RNA fraction for H5 mRNA sequences. While there seemed to be a tendency for H5 mRNA to fractionate with the non-adenylated RNA, no clear fractionation could be achieved, and, at best, an RNA fraction containing about 8 percent H5 mRNA was obtained, assuming approximately equal efficiencies of translation of H5 and globin mRNAs.

Chapter 5 described attempts made to fractionate H5 and globin mRNA sequences by the use of high resolution polyacrylamide gel electrophoresis. Although a control experiment utilising rabbit globin mRNA demonstrated that the gel system was working well, no resolution of H5 and globin mRNAs could be achieved.

The results of these previous experiments make it quite clear that a purification of H5 mRNA sequences by conventional physical methods will be very difficult to achieve. Furthermore, it will always be difficult to obtain H5 mRNA in reasonable amounts, due to both the low starting concentration of H5 mRNA and to the large losses of H5 mRNA incurred during the enrichment procedures.

The isolation of pure mRNA sequences from a complex mixture of RNA species can be achieved by the use of recombinant DNA techniques. This procedure was first used in the purification of

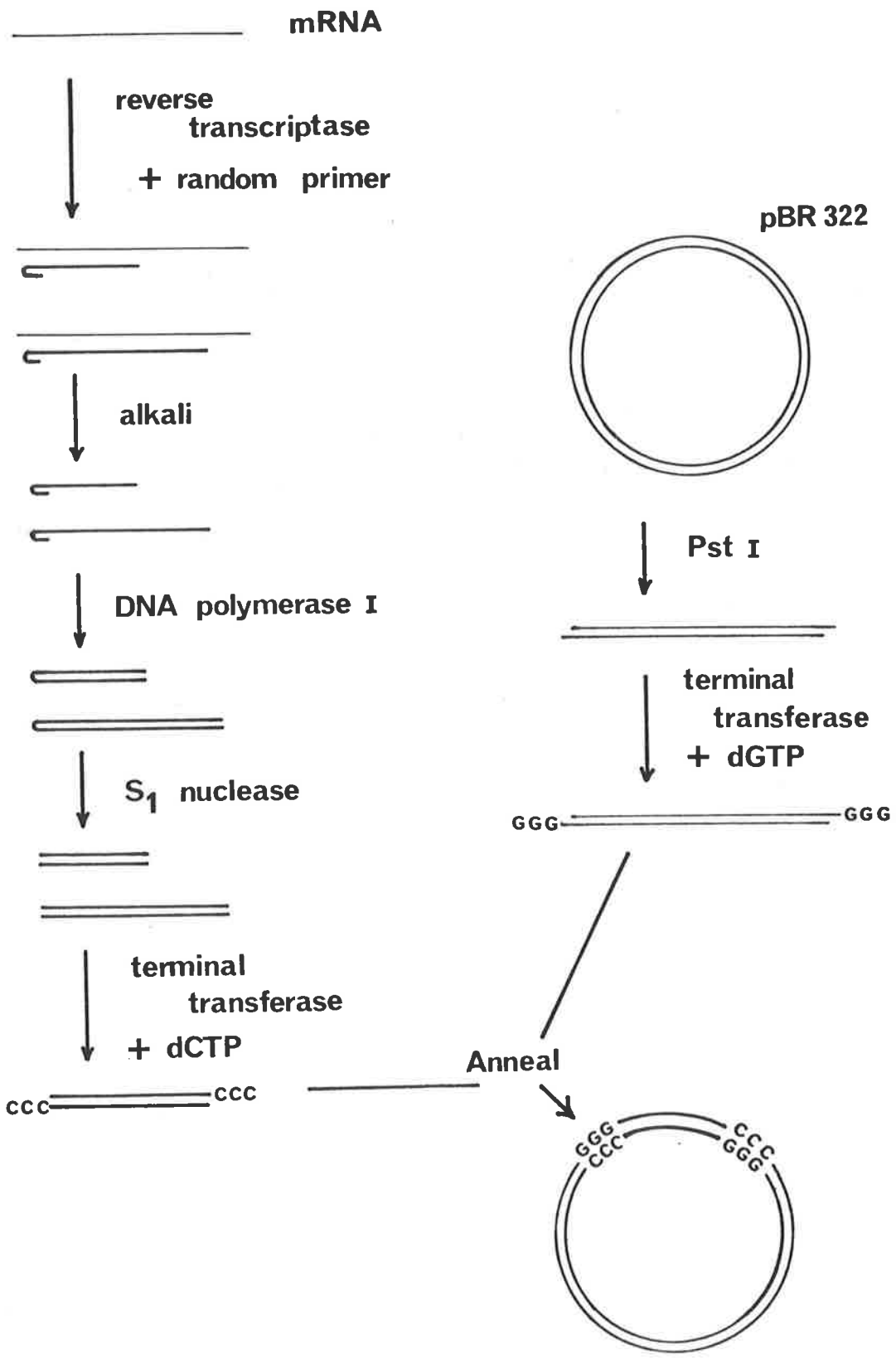
rabbit β-globin mRNA sequences, (Rougen and Mach, 1976; Maniatis *et al.*, 1976; Rabbitts, 1976).

All approaches to the molecular cloning of mRNA species have used the same general procedure, involving the synthesis of a double-stranded complementary DNA molecule (ds cDNA), the sequence of which is derived directly from the starting mRNA. Each ds cDNA molecule is then joined to a vector DNA molecule, usually a plasmid, that has the ability to replicate autonomously in the bacterial cell. These recombinant molecules are used to transform competent *E.coli* cells which are then plated out for single colonies. Since each colony is derived from a single cell, and conditions are used such that each cell accepts only a single recombinant molecule, each colony carries copies of only a single mRNA sequence. If all steps in the cloning procedure are general, and do not tend to select for, or discriminate against, certain nucleotide sequences, then the final population of recombinants will reflect exactly the starting population of RNA molecules.

The general procedure used in the cloning experiments is shown in figure 6.1. Synthesis of cDNA on the RNA template using avian myeloblastosis virus, (AMV), reverse transcriptase is the first step in the ds cDNA synthesis procedure. The most commonly used

Figure 6.1

Diagramatic representation of the procedure used to clone reticulocyte 10S RNA sequences into plasmid pBR 322 vector, using the tailing procedure.


mRNA
reverse transcriptase
+ random primer
alkali
DNA polymerase I
S1 nuclease
terminal transferase
+ dCTP
ccc
ccc
pBR 322
Pst I
terminal transferase
+ dGTP
GGG
GGG
Anneal
GGG
CCC
CCC
GGG

primer for this reaction is oligo-dT_{10}, which binds to the 3´ poly-A tract that is present in most mRNA molecules. Poly-A sequences have been added to histone mRNAs, (Thrall *et al.*, 1974), and to non-coding RNA species such as ribosomal RNA, (Hagenbüchle *et al.*, 1978), both of which lack 3´ poly-A tracts *in vivo*, so that oligo-dT primer may be used for first-strand cDNA synthesis.

A potential problem with oligo-dT priming from the 3´ end of the RNA has been reported by Hagenbüchle *et al.*, (1978), who have shown that cDNA synthesis is blocked at a point less than 20 bases from the end of 18S ribosomal RNA isolated from several eukaryote species. Therefore, long cDNA molecules cannot be synthesised when 18S ribosomal RNA is primed from the 3´ end. Taylor *et al.*, (1976), have reported an alternative method for priming cDNA synthesis. Oligonucleotides of 10-15 bases length, formed by the DNAase I digestion of a complex population of DNA sequences, (for example, salmon sperm DNA), have been shown to prime cDNA synthesis from a large number of sites on an RNA molecule, thus generating a distribution of different length cDNA molecules. Using this method, no *in vitro* polyadenylation of RNA molecules is required, and specific blocks to cDNA elongation may be avoided. Although it introduces short lengths of foreign sequence into the cDNA molecule, it was decided to use the random-priming method in attempts to clone H5 mRNA sequences, because it was the most general approach available.

AMV reverse transcriptase leaves a short hairpin loop at the 3´ end of the newly synthesised cDNA, (figure 6.1), and this segment of DNA is able to act as a self-primer, (Efstratiadis *et al.*, 1975). The second step in the ds cDNA synthesis involves extending the cDNA back along the molecule, using either *E. coli* DNA polymerase I, (Efstratiadis *et al.*, 1976), or AMV reverse transcriptase, which can also act as a DNA-dependent DNA-polymerase, (Rougen and Mach, 1976). It has been reported that although mouse immunoglobulin light chain cDNA can be copied-back efficiently using *E. coli* DNA polymerase I, the same is not true for AMV reverse transcriptase, which will not form the second strand of cDNA, (Rougen and Mach, 1976). It seems therefore, that, using AMV reverse transcriptase, the efficiency of the second-strand synthesis reaction is not the same for all nucleotide sequences, and, if this is the case, all sequences in a population of RNA molecules will not be equally represented in a population of ds cDNA molecules. For this reason, *E. coli* DNA polymerase I was used in preference to AMV reverse transcriptase in all second-strand synthesis reactions.

Because the two strands of ds cDNA are covalently joined by the primer loop at one end, the final step in the synthesis is the cleavage

of this end-loop, and the removal of any other single-stranded portions of the molecule, with the single-strand specific nuclease S_1, (Vogt, 1973). The resulting double-stranded DNA molecule has free 3′ hydroxyl and 5′ phosphate groups at both ends.

The ds cDNA molecule may be joined to the vector DNA molecule in one of two ways. The first is by the addition of homopolymeric tails to the 3′ end of the DNA using terminal deoxynucleotidyl transferase, (Jackson *et al.*, 1972; Lobban and Kaiser, 1973). When complementary base-pairing nucleotides are added to the ds cDNA and vector DNA respectively, the tails anneal to form a circular molecule that will transform competent *E. coli* cells, (figure 6.1). A-T base-pairing has been used by Maniatis *et al.*, (1976), and Rabbitts, (1976), for the cloning of rabbit β-globin mRNA sequences, and G-C tailing for the cloning of *Xenopus* globin mRNA sequences, (Humphries *et al.*, 1978) and rat preproinsulin mRNA sequences, (Villa-Komaroff *et al.*, 1978).

The second method utilises the blunt-end ligation properties of T_4-ligase to add a synthetic oligonucleotide linker sequence to the ends of the ds cDNA. These linkers contain the recognition sequence for a specific restriction endonuclease and so cleavage with

a restriction enzyme yields complementary base-pairing sequences that may be annealed to similar sequences in the vector DNA molecule. After ligation the closed circular recombinant molecule is used to transform competent *E. coli* cells. This approach has been used by Ullrich *et al.*, (1977), for the cloning of rat insulin mRNA sequences.

It was decided to use the tailing approach for the insertion of ds cDNA into the vector molecule for two reasons. Firstly, it is a very straightforward procedure involving minimal handling of small amounts of ds cDNA. Secondly, as demonstrated by Bolivar *et al.*, (1977), if a poly-dG tail is added to Pst I cut vector DNA and this is joined to poly-dC tailed ds cDNA, it is possible to regenerate the Pst I cleavage site, allowing convenient excision of the inserted ds cDNA sequences.

The plasmid pBR322, (Bolivar *et al.*, 1977), was chosen as the vector molecule for use in the ds cDNA cloning experiments because it is small, well-characterised and carries two anti-biotic resistance genes which aid in the selection of recombinant molecules.

Using the synthesis techniques described, it should be possible to prepare a population of recombinant DNA molecules representing the starting population of RNA species in 10S chicken reticulocyte

RNA. A certain proportion of the recombinants should contain H5 mRNA sequences.

This chapter describes the preparation of ds cDNA from chicken 10S RNA sequences and the molecular cloning of these sequences.

6.2 RESULTS

6.2.1 Preparation of ds cDNA from 10S RNA sequences

6.2.1.(i) Synthesis of the first strand of cDNA

It seemed advisable to use the RNA fraction most enriched for H5 mRNA sequences as the starting material for ds cDNA synthesis. In chapter 3, the isolation of a fraction of 10S RNA enriched for H5 mRNA sequences and supposedly free of ribosomal RNA sequences was described. Synthesis of cDNA from this 10S RNA template was carried out using AMV reverse transcriptase and the random-priming method of Taylor *et al.*, (1976). For preparative synthesis, 5 micrograms of RNA was used as template, yielding about 2 micrograms of cDNA.

6.2.1.(ii) Synthesis of the second strand of cDNA

Both *E. coli* DNA polymerase I and AMV reverse transcriptase may be used to extend the cDNA into double-stranded form. Although experiments were carried out with both enzymes, for reasons explained previously, *E. coli* DNA polymerase I was used in preference to reverse transcriptase, so that all sequences copied into

cDNA would also be represented in the ds cDNA.

The reaction conditions used for the DNA polymerase I extension were those of Seeburg *et al.*, (1977b), except that the reaction was carried out for 6 hours at 15°C, (Efstratiadis *et al.*, 1976), rather than for 2 hours at 40°C as recommended by Seeburg *et al.*

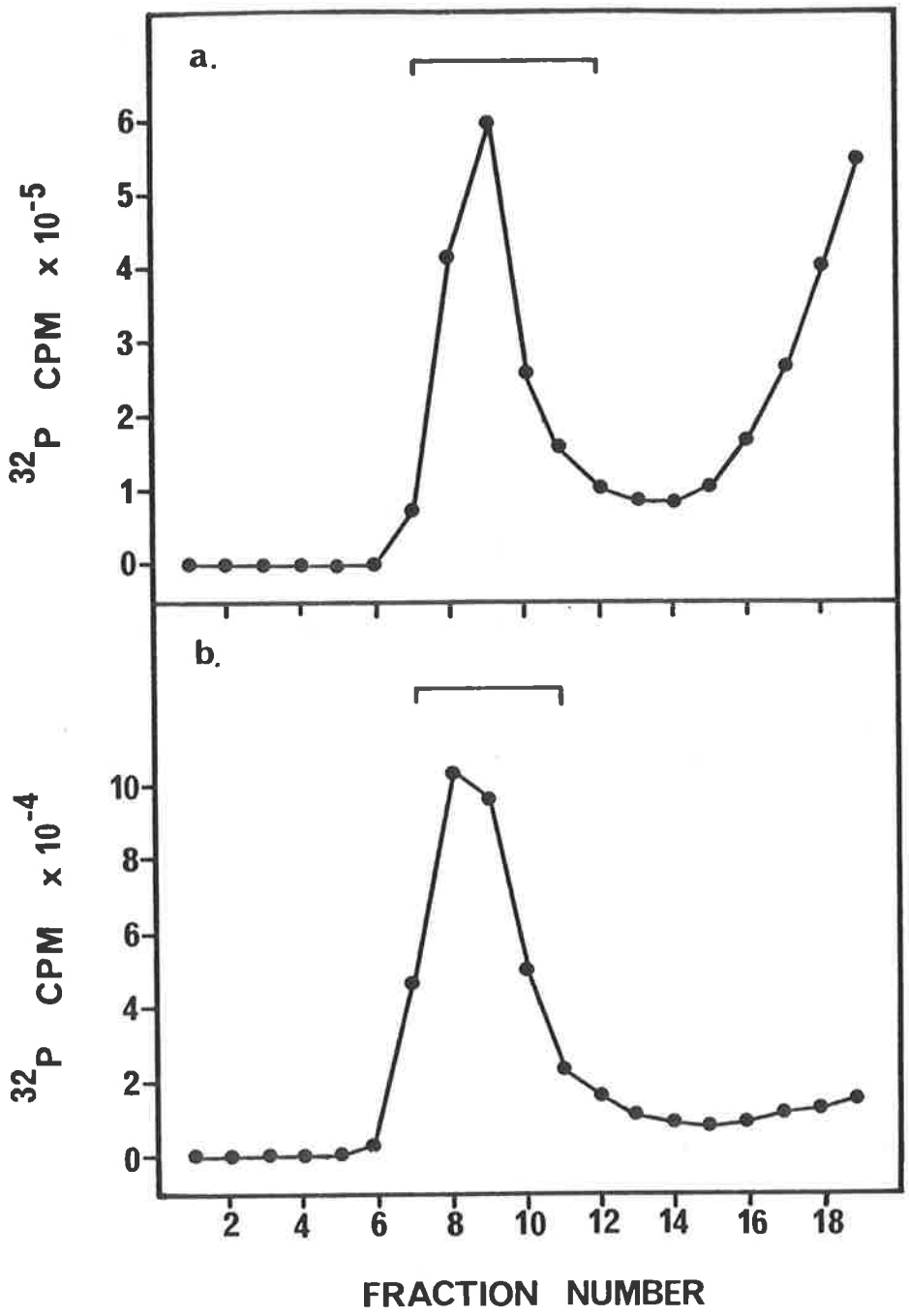
The double-stranded cDNA formed was diluted into 200 microlitres of digestion buffer and cleaved with S_1 nuclease, (Vogt, 1973). When the second strand synthesis reaction was carried out at 40°C, only about 10 percent of the radioactivity incorporated into the first strand remained in high molecular weight material after S_1 cleavage, indicating about 10 percent back-copy. However, when the reaction was carried out for 6 hours at 15°C, about 40 percent of the single-stranded cDNA was converted to the double-stranded form. Therefore, in an experiment starting with 2 micrograms of single-stranded cDNA, about 1.6 micrograms of ds cDNA remained after S_1 nuclease treatment. Typical Sephadex G-50 elution profiles of cDNA, and cDNA after DNA polymerase I extension and S_1 digestion, are shown in figure 6.2. Gel electrophoresis demonstrated that the great majority of the ds cDNA molecules were greater than 200 bases in length, and that a small proportion represented almost full-length copies.

Figure 6.2

Preparation of double-stranded cDNA suitable for cloning.

a. Sephadex G-50 elution profile. After alkaline hydrolysis and neutralisation the first-strand cDNA reaction mix was passaged through a 0.4 cm x 4 cm Sephadex G-50 column to remove unincorporated nucleotides.

b. Sephadex G-50 elution profile. After DNA polymerase I back-copy and nuclease S_1 treatment the reaction mix was passed through the same Sephadex column. About 20% of the label contained in the original cDNA was present in ds cDNA after this step.


a.
b.
^{32}P CPM x 10^{-5}
^{32}P CPM x 10^{-4}
0
1
2
3
4
5
6
0
2
4
6
8
10
2
4
6
8
10
12
14
16
18
FRACTION NUMBER

The ds cDNA was fractionated according to size on 10-40 percent sucrose gradients. While most of the label appeared in high molecular weight ds cDNA, (greater than 200 bases long), it was most important to remove as much of the low molecular weight DNA as possible, because these molecules may have been in a very great molar excess over the high molecular weight molecules. The presence of a large number of short ds cDNA molecules causes difficulties in estimating the number of residues added during tailing reactions, and results in the cloning of many very short ds cDNA sequences.

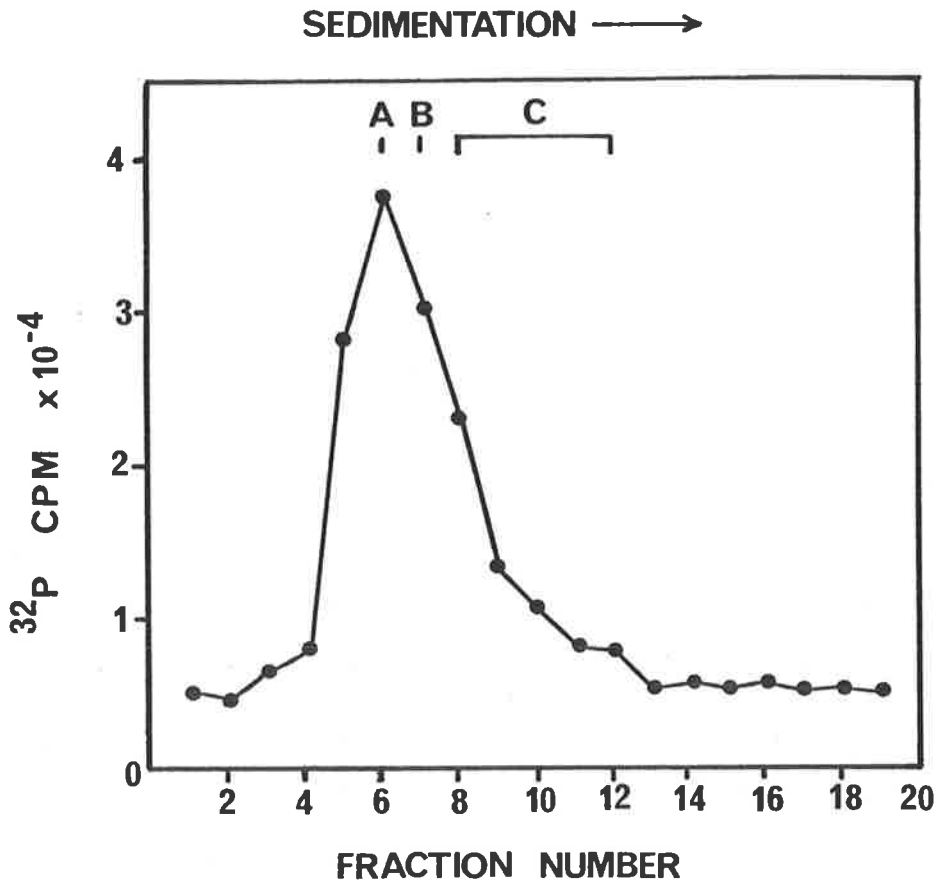
The sucrose gradient profile of the final preparative ds cDNA preparation is shown in figure 6.3. Three high molecular weight classes of ds cDNA were selected, precipitated by the addition of two volumes of ethanol, washed with ice-cold ethanol and then resuspended in water. The final yield of the highest molecular weight fraction of ds cDNA was about 100 nanograms, after starting with 5 micrograms of 10S RNA.

6.2.1.(iii) Tailing of ds cDNA

Homopolymer tailing of the ds cDNA was performed using calf-thymus terminal deoxynucleotidyl transferase and dCTP as substrate. Although it allows tailing from single-stranded nicks in the ds cDNA molecules, $CoCl_2$ was used in the tailing solution rather

Figure 6.3

Size fractionation of ds cDNA on sucrose gradients. The ds cDNA eluted from the Sephadex G-50 column, (figure 6.2b), was loaded directly onto a 10-40% sucrose gradient and centrifuged at 160,000 g for 16 hours. Three size classes, A, B and C were selected, of which only B and C were used in the preparation of recombinants.


SEDIMENTATION ⟶
A
B
C
4
3
2
1
0
^{32}P CPM x 10^{-4}
2
4
6
8
10
12
14
16
18
20
FRACTION NUMBER

than $MgCl_2$ because the reaction is more efficient in the presence of the cobalt ion, (Roychoudhury *et al.*, 1976). Figure 6.4 shows a typical time course for the addition of nucleotides to the ds cDNA. When an average of 10-30 residues had been added to each end of the molecule, the reaction was stopped by the addition of the chelating agent EDTA.

6.2.2 Tailing of vector plasmid DNA

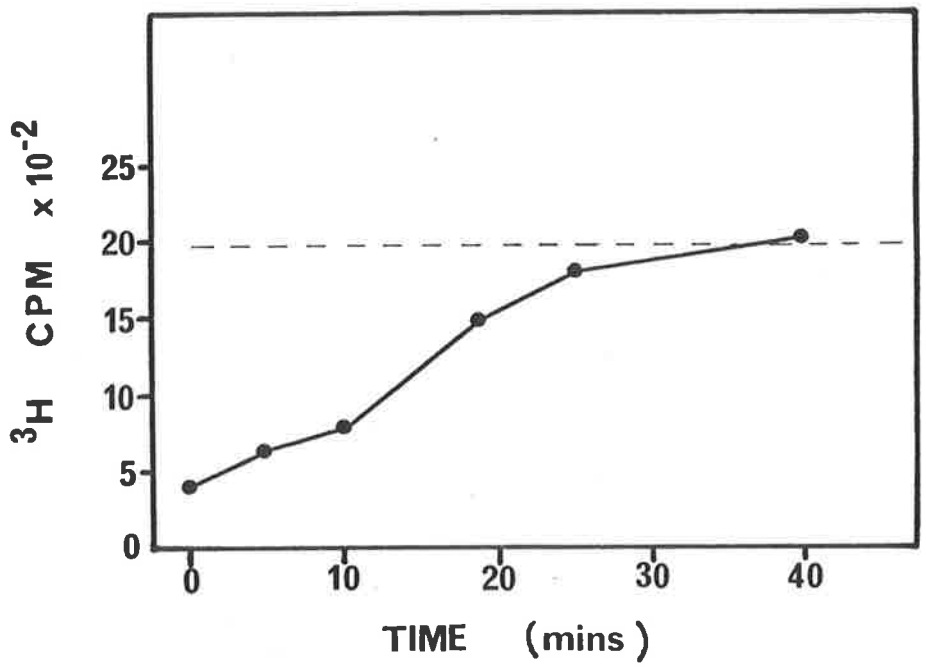
Plasmid pBR322 DNA was linearised by two consecutive Pst I digestions to ensure that cleavage was complete. Deoxyguanosine tails were added to the 3´ OH groups of the linear plasmid molecule, but in this reaction $MgCl_2$ was used in the solutions, so that tailing occurred only at the Pst I site and not at internal single-stranded nicks. The reaction was stopped by the addition of excess EDTA when about 25 dG residues per end had been added.

6.2.3 Annealing and transformation

Equimolar amounts of dG-tailed pBR322 and dC-tailed ds cDNA were annealed, using a high-salt hybridisation mix to stabilise G-C base-pairing, and the resulting circular molecules were used to transform competent *E. coli* cells. Two transformation procedures and a number of strains of *E. coli* host cells were tested to determine which combination gave the greatest efficiency of transformation. The most effective

Figure 6.4

Poly-dC-tailing of ds cDNA. The ds cDNA was incubated with terminal deoxynucleotidyl transferase and ^{3}H-dCTP as described, (section 2.2.11). 1 µl samples of the reaction mix were taken and the amount of acid-insoluble radioactivity determined. The dotted line indicates the number of counts incorporated equivalent to the addition of an average of 20 dC residues per 3´ end of the molecules.


^3H CPM x 10^{-2}
25
20
15
10
5
0
0
10
20
30
40
TIME (mins)

system, (the Edinburgh procedure - Ken Murray, pers. commun.), resulted in a transformation efficiency of 2×10^7 colonies per microgram of pBR322 DNA. Under the same conditions, the Pst I digested, dG-tailed pBR322 transformed with an efficiency of only 10^3 colonies per microgram. When 3 nanograms of pBR322, annealed to tailed ds cDNA, was transformed into *E. coli*, about 300 colonies containing the recombinant sequences resulted. This is approximately 100-fold greater than the background and equivalent to 10^5 transformants per microgram of pBR322 annealed to ds cDNA.

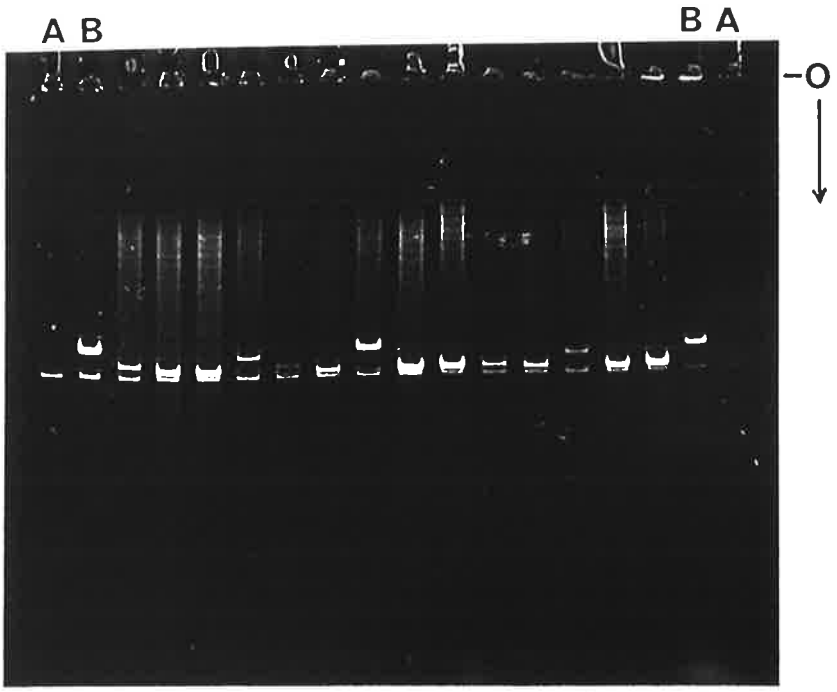
The colonies resulting from transformation experiments using recombinant DNA were transferred to fresh agar plates using sterile toothpicks. After overnight growth at 37°, these 'master plates' were stored at $4^{\circ}C$.

6.2.4 Detection of recombinants using the 'miniscreen' procedure

As a preliminary check that the colonies resulting from the transformation procedures actually contained recombinant plasmids, small amounts of the plasmids were prepared using the 'miniscreen' procedure and samples were electrophoresed on 1 percent agarose gels. Figure 6.5 shows the appearance of a number of miniscreen preparations of plasmid that have been linearised

Figure 6.5

Miniscreen analysis of recombinant plasmids. Small amounts of plasmid DNA were prepared, from a random selection of colonies, by the miniscreen procedure, (section 2.2.13), linearised by Eco RI digestion and electrophoresed on 1% agarose gels. Linearised parental pBR 322 was included in all tracks as a size marker. Tracks labelled A contain linear pBR 322, tracks labelled B a β-globin recombinant with a total insert length of 1200 base pairs, and all other tracks contain samples of linearised recombinant plasmid DNA.


A B
B A
O

with Eco RI restriction enzyme. Linear pBR322 was included in each track as a marker. The presence of a higher molecular weight band in a track indicated the presence of a recombinant plasmid molecule carrying a ds cDNA insert. The molecular weight markers used on this gel were linearised pBR322 recombinants containing either inserted α-globin mRNA sequences or a double insert of β-globin mRNA sequences. These inserts increase the length of the linear pBR322 DNA by about 600 and 1200 bases respectively, (Rob Richards - pers. commun.). The double-stranded DNA inserts observed in these preparations seem to range in size from about 200 base pairs up to approximately 450 base pairs in length.

The results presented in this chapter show that it has been possible to form ds cDNA in reasonable yields from 10S RNA. Every effort has been made to ensure that all sequences present in the original 10S RNA population will also be represented in the final ds cDNA preparation.

The ds cDNA was joined to vector plasmid molecules and transformed into *E. coli* at a very high efficiency, each nanogram of the final, tailed ds cDNA population yielding approximately 1000 molecular clones. The screening of these recombinants for H5 mRNA sequences is described in the next chapter.

CHAPTER 7

SCREENING OF ds cDNA RECOMBINANTS PREPARED FROM RETICULOCYTE 10S RNA

7.1 INTRODUCTION

Most recombinants so far isolated have been selected by hybridisation of the cloned sequences to a cDNA probe made from an mRNA preparation highly enriched for the desired sequence. Examples are the selection of recombinants containing immunoglobulin, (Seidman *et al.*, 1978), ovalbumin, (Humphries *et al.*, 1977) and procollagen, (Lehrach *et al.*, 1978), sequences.

Despite the attempts, reported in previous chapters, to achieve a significant enrichment for H5 mRNA sequences, it has not been possible to prepare an RNA fraction containing more than a few percent H5 mRNA, and so a direct, positive probe for cloned H5 sequences is not available. In other cases where a pure probe has not been available, special procedures have been undertaken to ensure that the ds cDNA cloned was highly enriched for the desired sequences, so that a minimum of screening was required to identify the recombinants. For example, Shine *et al.*, (1977), cloned a specific restriction fragment of ds cDNA copied from mRNA containing human chorionic somatomammotropin sequences. This restriction fragment had previously been shown to contain the desired sequence by DNA-sequencing of the equivalent single-stranded cDNA fragment, (Seeburg *et al.*, 1977a).

DNA bound to DBM-cellulose, (Noyes and Stark, 1975), has been used to isolate complementary

sequences from a complex mixture with a high degree of specificity, (for example the isolation of *Drosophila* histone mRNA by Lifton *et al.*, (1977)). In theory, globin cDNA sequences could be removed from total 10S cDNA probe by prior hybridisation to cloned globin sequences bound to DBM-cellulose, thus enriching the remaining cDNA for H5 mRNA sequences. While it may have been possible to enrich cDNA probe for H5 sequences in this way, some special features of the reticulocyte 10S RNA system suggested a straightforward alternative approach.

When total 10S mRNA is translated, only two protein products are observed on SDS-urea gels, H5 protein and globin protein, (figure 3.2a). The reticulocyte 10S RNA appears to contain only these two major mRNA activities, and so the recombinants formed from total 10S RNA should contain only H5 and globin mRNA sequences and some ribosomal RNA sequences. Since pure globin and pure ribosomal probes can be obtained, it should be possible to identify recombinants containing these sequences. If H5 mRNA is indeed the only other mRNA species present in the 10S RNA, the recombinants remaining when globin and ribosomal clones have been eliminated will contain H5 mRNA sequences.

Because H5 mRNA represents only a small percentage of the population of sequences copied into ds cDNA, a large number of colonies will need to be

screened to be certain of detecting H5 mRNA recombinants. The colony hybridisation procedure developed by Grunstein and Hogness, (1975), is ideally suited to this purpose. Plasmid recombinants are detected by growing colonies directly onto nitrocellulose filters placed in contact with nutrient plates, lysing the cells *in situ*, binding the released, denatured DNA to the filter and then probing with radioactive sequences complementary to those of the desired recombinant.

Globin sequences will be the most common sequences in a population of recombinant molecules formed from reticulocyte 10S RNA. Elimination of those recombinants containing globin sequences will therefore represent a considerable enrichment for H5 mRNA sequences amongst the remaining clones. In initial experiments, homologous α and β globin cloned DNA were not available, and so the selection of globin clones was attempted using radioactive cDNA prepared from rabbit globin mRNA. Rabbit reticulocytes contain no H5 protein and so the probe will be free of H5 mRNA sequences, however, the conservation of globin sequences between species will ensure cross-reaction between the rabbit globin cDNA and the chicken globin recombinants. Unfortunately the results of colony hybridisation experiments using this probe were most unsatisfactory. The intensity of response varied greatly from colony to colony and a large number of colonies, later shown to contain

globin mRNA recombinants, were not detected at all. It seemed preferable to use homologous chicken globin probe and fortunately at this time it became possible to use clones containing chicken α and β-globin mRNA sequences, (due to the completion of the facilities required for recombinant DNA work).

Ribosomal sequences might also be expected to be present in a population of recombinants representing 10S RNA sequences and so filters were probed with ribosomal cDNA to identify those colonies containing ribosomal recombinants.

As a final step in the preliminary screening all colonies were probed with cDNA prepared from total 10S RNA to ensure that all recombinants actually contained sequences derived from the starting RNA.

Those recombinants that had been screened as negative to both globin and ribosomal probes, but positive to cDNA prepared from reticulocyte 10S RNA, were selected for more detailed examination. Small preparations of plasmid DNA were made and the recombinants examined on agarose gels. The plasmid DNA was transferred to nitrocellulose filters using the procedure of Southern, (1975), and the clones again screened for the presence of globin or ribosomal sequences.

This chapter describes the experiments undertaken to identify possible H5 recombinants. As described above, the negative screening procedure

involves the removal of those members of the population of ds cDNA recombinants that contain globin and ribosomal sequences. Any H5 mRNA sequences that have been cloned will be found amongst the recombinants remaining after this selection.

7.2 RESULTS

7.2.1 Preparation of Globin sequence Probes

Richards *et al.*, (1979), have inserted chicken α and β globin mRNA sequences into pBR 322 vector molecules using the oligonucleotide linker method of Ullrich *et al.*, (1977). DNA sequencing data, (Richards and Wells, 1980), has shown that the cloned α-globin sequence does not correspond to the nucleotide sequences of either α_A or α_D globin, predicted from the amino-acid sequences, (Matsuda *et al.*, 1971; Takei *et al.*, 1975). It does however, correspond to the 'stress' α-globin sequence, (α_S), reported by Cummings *et al.*, (1978), which is believed to arise in response to the phenylhydrazine treatment used to induce anaemia in chickens, (see chapter 2). While α_S-globin differs from the expected α_A and α_D sequences at a number of sites within the coding region alone, a large degree of cross-reaction can still be expected between the cloned sequence and the true α_A and α_D sequences. The non-coding regions should also exhibit a high degree of homology, as these sequences have been shown to be highly conserved in the chicken globin

mRNAs, (R. Richards - personal communication). Since the reticulocyte 10S RNA used in the preparation of the ds cDNA clones was also isolated from phenylhydrazine-treated chickens, it is likely that all the α-globin sequences in this population are also α_S-globin and so will react strongly with the cloned probe.

The DNA-sequence determined from the β-globin mRNA clone, (Richards *et al.*, 1979), agrees exactly with that predicted by the amino-acid sequence of the β-globin protein, (Matsuda *et al.*, 1973).

A problem arises in the use of these cloned globin mRNA sequences as probes because they are carried by the plasmid vector pBR 322. Since this is the same vector used in the recombinant DNA experiments in which reticulocyte 10S RNA sequences were cloned, all pBR 322 sequences will have to be removed before the α and β globin mRNA sequences can be successfully used as probes. Unless this is done, all colonies will yield a positive response to the cloned globin probe, regardless of the nature of the inserted sequence, due to the presence of pBR 322 sequences in the labelled probe and bound to the filter.

The α-globin plasmid used in these experiments carried a 600 base-pair insert of α-globin mRNA sequence and the β-globin plasmid carried two fragments of β-globin mRNA sequence, 630 and 550 base-pairs long respectively, incorporated into the vector molecule as a double insert. These inserts were

excisable by cleavage of the recombinant plasmids with Hind III restriction endonuclease. Large amounts of both the α and β-globin plasmid DNA were prepared by a modified heat-shock lysis procedure and separation of the DNA from low molecular-weight RNA by Sephadex column chromatography, (details are presented in chapter 2). This procedure was very quick and efficient and yields of 1.4 to 1.6 milligrams of plasmid DNA per litre of culture were routinely obtained. 100 micrograms of each plasmid preparation was digested with Hind III enzyme and the globin insert sequences separated from the vector DNA on sucrose gradients, (figure 7.1). An aliquot of the insert DNA was labelled by nick-translation, (Maniatis *et al.*, 1975), and the labelled DNA electrophoresed on 1 percent agarose gels, to achieve a further separation of insert sequences from contaminating pBR 322 DNA. The intact labelled insert DNA was eluted from the gel, boiled, and then used as pure globin probe in the screening experiments.

7.2.2 Preparation of cDNA from Ribosomal RNA and Reticulocyte 10S RNA sequences

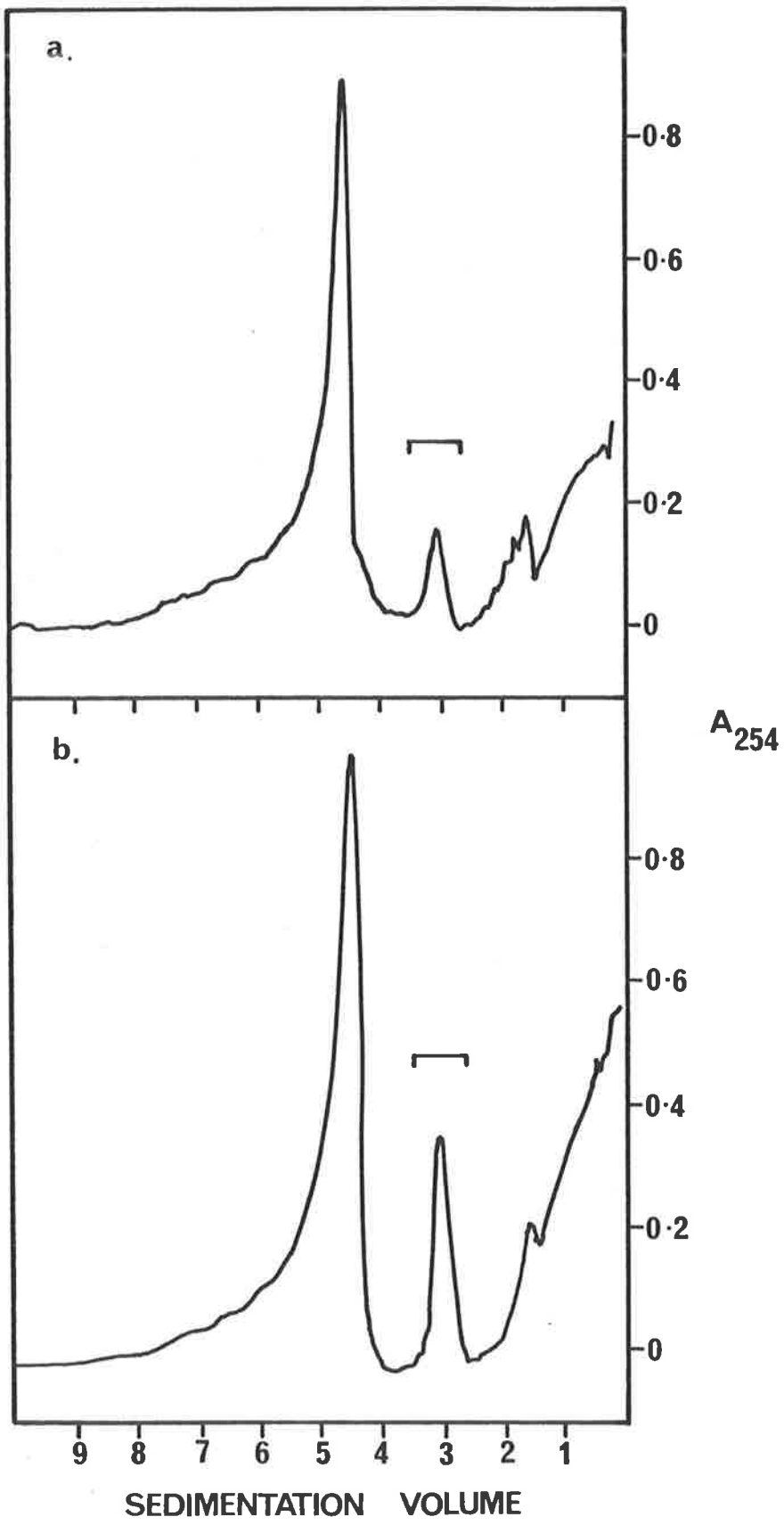
18S and 28S ribosomal RNA prepared from chicken scale tissue was used as template for the synthesis of ribosomal cDNA. Scale tissue ribosomal RNA was used, in preference to erythroid cell RNA, to ensure that no H5 mRNA sequences contaminated the ribosomal

Figure 7.1

Isolation of cloned α_S and β globin mRNA sequences. About 50 μg of each recombinant plasmid was digested with Hind III restriction enzyme to excise the insert. The digestion mix was adjusted to 0.5% SDS, loaded directly onto a 10-40% sucrose gradient and centrifuged for 16 hours at 160,000 g. The β-globin plasmid contained two insert sequences, (550 and 630 base pairs long), and so the β insert peak is about twice as large as the peak resulting from the single α insert.

a. Hind III digested α-globin plasmid.

b. Hind III digested β-globin plasmid.


a.
0·8
0·6
0·4
0·2
0
b.
0·8
0·6
0·4
0·2
0
A254
9 8 7 6 5 4 3 2 1
SEDIMENTATION VOLUME

probe. Labelled 10S cDNA was prepared by copying total unfractionated reticulocyte 10S RNA, (figure 3.1). Both the ribosomal and 10S cDNA probes were prepared using the random-primed cDNA synthesis method of Taylor *et al.*, (1976).

7.2.3 Preparation of Filters and Colony Screening

The preparation of filters for colony screening follows essentially the method of Grunstein and Hogness, (1975). 100 individual recombinant colonies were transferred from the master plates to a sheet of nitrocellulose lying in contact with a nutrient agar plate. The toothpick used to transfer the bacteria was touched to the surface of the nitrocellulose twice, to give rise to two identical colonies side-by-side on the filter. This precaution was taken to prevent confusion between a spot due to true hybridisation to a colony and the other spurious spots that sometimes appear on auto-radiographs. Included on the filter were a number of controls, including colonies containing parental plasmid pBR 322 and colonies containing keratin ds cDNA sequences inserted into pBR 322 by the dC-dG-tailing procedure, (keratin clones supplied by Rob Saint). The colonies containing plasmid alone will give an estimate of the background hybridisation of the system and the keratin sequence-containing colonies will provide an estimate of the extent of hybridisation, if any, due to the presence of dC-dG joining sequences.

The filters were prepared for colony screening, (Grunstein and Hogness, 1975), and the hybridisation and washing procedures were carried out as described by Wahl *et al.*, (1979). The results of the colony-screening experiment, using cloned α and β-globin mRNA probes, are presented in figure 7.2.

Two observations may be made immediately. Firstly, all colonies, even the parental pBR 322 and keratin recombinant colonies, can be detected on the autoradiograph. This indicates that, despite the precautions taken, the isolated α and β-globin sequences still contain some labelled pBR 322 sequences and these have hybridised to plasmid sequences bound to the filter. Despite the presence of this background exposure it is possible to identify globin recombinants with some certainty since many colonies exhibit a degree of hybridisation to the globin probe well above the background level. Examination of the autoradiographs suggested that of the 300 colonies screened, 208 contained globin sequences.

The second feature to note is that, ignoring the background response common to all colonies, many of the colonies have not hybridised to the globin probe. The 100-fold increase in the number of transformants obtained with the annealed ds cDNA-pBR 322 DNA, over that obtained for the vector DNA alone, (section 6.2.3), suggests that most of the colonies contain recombinant molecules. The initial

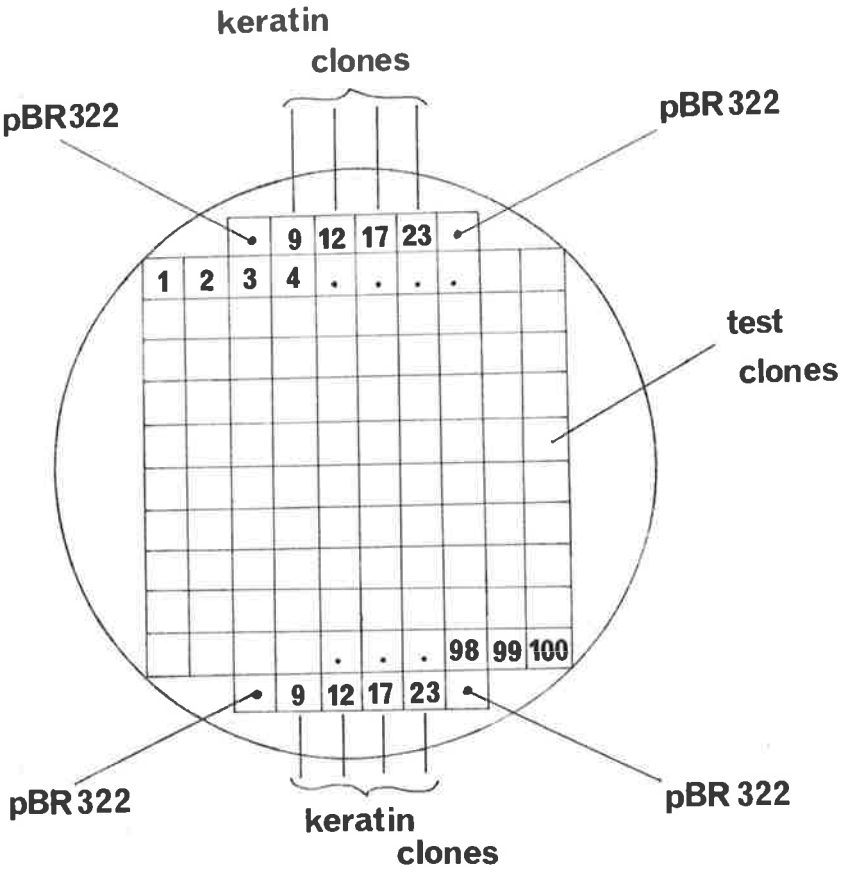
Figure 7.2

Detection of globin sequences, amongst recombinants formed from reticulocyte 10S RNA, using the colony screening procedure. Colonies containing recombinant plasmids were denatured *in situ*, immobilised onto nitro-cellulose filter paper, annealed to ^{32}P-labelled α and β globin sequences, washed in 0.2 x SSC, dried and autoradiographed. (Colonies were grown on the nitrocellulose in duplicate).

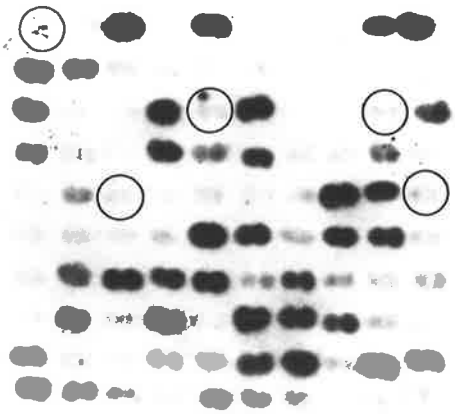
a. Distribution of control colonies and newly-formed recombinant colonies on the nitro-cellulose filter.

b. Autoradiograph resulting when filters were annealed with cloned α and β globin mRNA sequences.

a.


keratin
clones
pBR322
pBR322
9 12 17 23
1 2 3 4 . . . .
test
clones
. . . 98 99 100
9 12 17 23
pBR322
pBR322
keratin
clones

b.



miniscreen plasmid preparations also suggest that most colonies contain recombinants, (figure 6.5). Since the 10S RNA used as the initial template for the preparation of the ds cDNA had been bound to poly-U-Sepharose, it seemed unlikely that ribosomal sequences should comprise a major proportion of the recombinant population.

The nitrocellulose filters were boiled for 10 minutes in double-distilled water to elute most of the labelled globin sequences hybridised to the filters. The filters were then hybridised with ribosomal cDNA, washed, and placed in contact with X-ray film. The resulting autoradiograph is shown in figure 7.3.

Contrary to predictions a large proportion of the colonies exhibited a positive response to ribosomal probe. Of 300 recombinant colonies screened, 72 appeared to contain ribosomal RNA sequences. Assuming equal copy of all sequences during the cloning procedure, nearly a quarter of the original 10S RNA used as template consisted of ribosomal RNA despite the fact that this material had been bound to poly-U-Sepharose.

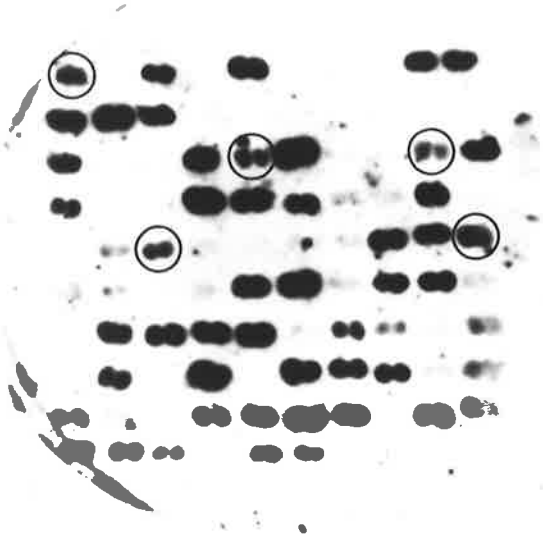
The nitrocellulose filters were again boiled, to remove most of the hybridised label, and then were hybridised with cDNA prepared from total 10S RNA. The autoradiograph resulting from this experiment is shown in figure 7.4. All the colonies originally detected with the purified globin probes have shown

Figure 7.3

Detection of ribosomal RNA sequences amongst the recombinants formed from reticulocyte 10S RNA. The filter annealed with globin mRNA sequences, (figure 7.2b), was boiled in distilled water to remove most of the label and then probed with ribosomal cDNA. Some residual label, however, is present at the position of the colonies originally detected by the globin probe. The colonies finally determined to contain non-globin, non-ribosomal sequences are circled on the diagram.

Figure 7.4

Detection of reticulocyte RNA sequences amongst the recombinants formed from reticulocyte 10S RNA. The nitro-cellulose filter previously annealed to globin and ribosomal RNA sequence probes, (figures 7.2 and 7.3), was boiled in distilled water and probed with cDNA prepared from reticulocyte 10S RNA. Comparison of this autoradiograph with the autoradiographs resulting from the previous two experiments allow the colonies hybridising to non-globin, non-ribosomal sequences to be detected. These colonies are circled on the diagram.



up once again with the 10S cDNA probe, many with considerably increased intensity. While in general the colonies containing globin sequences have been confirmed by the use of probe to reticulocyte 10S RNA sequences, five new colonies, (circled on the diagram), have appeared on the autoradiograph for the first time when probed with 10S cDNA. These colonies contain sequences present in the total 10S RNA that are not detected with either cloned globin or pure ribosomal probes. Any H5 mRNA sequences that have been cloned should be found amongst this group of recombinants, which represent the non-globin, non-ribosomal sequences present in reticulocyte 10S RNA.

7.2.4 Gel and Southern Blot Experiments

Small preparations of plasmid DNA were made from each of the non-globin, non-ribosomal colonies, (p1, p25, p29, p43 and p50), using the miniscreen procedure. In addition plasmid DNA was prepared from colonies containing a ribosomal sequence, (pR7), a globin sequence, (pG3) and one lacking an inserted sequence, (pO2), as judged from the autoradiographs. The plasmids were linearised with Eco RI restriction endonuclease and electrophoresed on a 1 percent agarose gel. A sample of linear pBR 322 DNA, (about 50 nanograms), was added to each track to act as a size marker. Plasmids containing inserted DNA sequences will migrate more slowly than the

parental pBR 322 DNA because of their increased molecular weight.

Samples of linearised plasmid DNA were electrophoresed on a gel and stained with ethidium bromide, as shown in figure 7.5. The first four tracks of the gel contain plasmid preparations which act as molecular weight markers and which will act as hybridisation controls in future experiments. The next eight tracks contain the miniscreen plasmid DNA preparations. The gel shows that all except one of the 10S RNA clones, (pO2), contains a visible insert, and that these range in size from about 150 up to nearly 500 base pairs in length. The plasmid preparation apparently lacking an inserted sequence was derived from the colony that gave no response to any of the labelled DNA probes, and is probably a parental pBR 322 plasmid.

The DNA was transferred from the gel to nitrocellulose using the method of Southern, (1975), as modified by Wahl *et al.*, (1979). After prehybridisation, the nitrocellulose filter was hybridised to different labelled DNA probes. The hybridisation and washing procedures followed the method of Wahl *et al.*, (1979). The washed filter was placed in contact with X-ray film and exposed at $-80^{o}C$.

The results of hybridisation of labelled globin, ribosomal and reticulocyte 10S RNA probes to the

Figure 7.5

Examination of recombinant plasmids on agarose gels. A number of recombinants were selected for closer examination on the basis of their response in the colony-screening experiments. A small amount of plasmid DNA was prepared using the miniscreen procedure, linearised with Eco RI and electrophoresed on a 1% agarose gel. Each track contained 50 ng of linear pBR 322 as a size marker. The tracks are as follows:

A : linear pBR 322

B : α-globin recombinant, (insert size 600 bp)

C : β-globin recombinant, (double insert, total 1200 bp)

D : keratin recombinant, pK 23, (insert size 740 bp)

E : plasmid p1 - non-globin, non-ribosomal

F : plasmid p02 - no response to probes

G : plasmid pG3 - globin positive

H : plasmid pR7 - ribosomal positive

I : plasmid p25 - non-globin, non-ribosomal

J : plasmid p29 - non-globin, non-ribosomal

K : plasmid p43 - non-globin, non-ribosomal

L : plasmid p50 - non-globin, non-ribosomal


A B C D E F G H I J K L
O

plasmid DNA immobilised on the nitrocellulose filter are shown in figure 7.6 (a, b and c respectively). The autoradiograph resulting from the hybridisation of globin probe to the nitrocellulose, (figure 7.6a), once again demonstrates the presence of contaminating pBR 322 DNA sequences in the probe. Every plasmid has shown up to some extent and this makes interpretation of the results somewhat difficult, however, it is quite clear that both the α and β-globin mRNA plasmids, from which the probe was derived, and the selected globin recombinant, pG3, have shown up strongly and certainly to a much greater extent than the control plasmids. The lack of hybridisation to the keratin sequence clone, containing a 740 base pair keratin sequence inserted by the dC-dG tailing procedure, shows that the presence of dC-dG sequences in a recombinant does not cause anomalous levels of hybridisation to be observed. While the tracks containing the recombinant plasmids pO2 and pR7 seem to show a response to the globin probe, this is probably due to contamination of these tracks with minute amounts of plasmid pG3 DNA. None of the other plasmids show hybridisation to the globin probe to any greater extent than that exhibited by the pBR 322 marker plasmid in the same track. While the results are not completely clear, it seems likely that of the 10S RNA recombinants tested, only the selected plasmid pG3 contains a globin sequence.

Figure 7.6

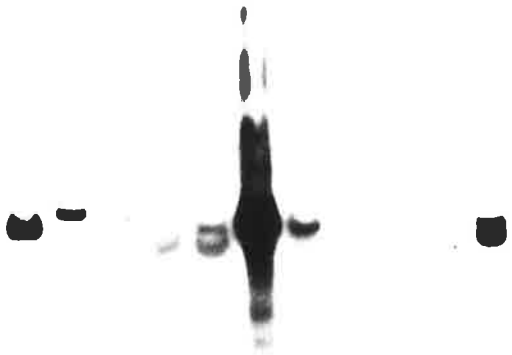
Characterisation of recombinant plasmids using the Southern blot procedure. Recombinant plasmid DNA was transferred from an agarose gel, (figure 7.5), to nitro-cellulose filter paper and then annealed to ^{32}P-labelled sequence probes. Between hybridisation experiments the filter was boiled in distilled water to remove annealed label.

a. Hybridisation to cloned α and β globin sequences
b. Hybridisation to 18S and 28S ribosomal cDNA
c. Hybridisation to cDNA prepared from reticulocyte 10S RNA

(see figure 7.5 for details of the tracks)

A B C D E F G H I J K L

a.



b.

c.

The experiment using ribosomal cDNA probe, (figure 7.6b), gives much clearer results, due to the low background hybridisation obtained when using cDNA probes. The recombinant plasmid, pR7, included as a control for ribosomal sequences, gives an extremely strong response to the ribosomal probe. Plasmids p1 and pO2 show an extremely low level of hybridisation to the ribosomal cDNA. This may be due to contamination of these two tracks with ribosomal sequences from plasmid pR7 or may indicate the presence of short ribosomal sequence inserts in both these plasmids. Notably, none of the other four plasmids prepared from colonies giving a non-globin, non-ribosomal response in the colony-screening experiments have shown a significant response to either the globin or ribosomal sequence probes.

Figure 7.6c shows the results of an experiment using total 10S cDNA probe. As expected, the known globin sequence recombinants show a very clear response, but, for the first time, the recombinants p25, p29, p43 and p50 show significant hybridisation to the labelled probe, indicating that they contain sequences present in 10S RNA. p1 has also shown up to a small extent, perhaps due to the presence of ribosomal sequences in the total 10S probe, however, the failure of the ribosomal control, pR7, to show a similar response leaves the exact nature of the p1 insert sequence in some doubt.

The overall result of these Southern blot experiments has been to confirm, at a much more sensitive level, the results of the colony hybridisation experiments. Recombinant plasmids p25, p29, p43 and p50 all contain sequences found in 10S RNA but which are not globin or ribosomal sequences. The clones chosen to represent globin and ribosomal sequence controls, on the basis of their response during the colony-screening experiments, have both hybridised extremely strongly to their respective probes. While p1 appeared to be non-globin, non-ribosomal from the colony hybridisations, the results of the Southern blot experiments have left the nature of the p1 insert sequence uncertain.

7.3 DISCUSSION

Making use of pure globin and pure ribosomal sequence probes, it has been possible to identify those members of the population of ds cDNA recombinants that contain globin and ribosomal RNA sequences, and to exclude these from further study. When these clones are removed, a small group of recombinants containing 10S RNA sequences remain. If H5 mRNA sequences have been successfully cloned they will be detected amongst this remaining population of recombinant molecules. Since this negative screening approach, involving the elimination of all identifiable non-H5 recombinants from further study, has been taken

as far as it can go, it is now essential to make a positive identification of H5 mRNA sequences.

The next chapter describes the detailed examination, including DNA-sequence determination, of four non-globin, non-ribosomal recombinants, p25, p29, p43 and p50.

CHAPTER 8

DETAILED EXAMINATION OF SELECTED ds cDNA RECOMBINANTS

8.1 INTRODUCTION

The complete amino acid sequence of goose H5 protein is known, (Yaguchi *et al.*, 1979), and 160 of the 189 residues of the chicken H5 protein have been sequenced (Sautiere *et al.*, 1975; 1976; Champagne - pers. commun). Of the remaining 29 unordered residues in the chicken H5 protein, 12 are lysine. Where the amino acid sequences of both goose and chicken H5 are known, a great deal of homology is observed, (differences occurring at 22 out of 160 positions), with most of the substitutions being conservative. Since the amino-acid sequence of H5 is known, the DNA sequence can largely be deduced and there is little doubt that a recombinant carrying an H5 coding sequence could readily be identified from its DNA sequence.

Since H5 mRNA is 10S in size, (Scott and Wells, 1976), corresponding to a length of approximately 650 bases, about 600 of which are required for amino acid coding, H5 mRNA is expected to have quite short non-coding regions. Any DNA sequence derived from an H5 mRNA clone is therefore very likely to include a portion of the amino acid coding sequence.

The four recombinant plasmids previously selected as non-globin, non-ribosomal, but containing 10S RNA sequences, were prepared in bulk, and a preliminary characterisation of the inserted DNA

was carried out. Fragments of the inserted DNA from each of the recombinants were sequenced using the chemical, chain-cleavage method, (Maxam and Gilbert, 1977), to determine whether any of these 10S mRNA clones carried an H5 mRNA sequence.

While in theory it would be practical merely to excise the inserted DNA with Pst I restriction enzyme, label the ends of the excised DNA, cleave the insert with another restriction enzyme and then sequence, this approach was not used for a number of reasons. Firstly, the ends of DNA molecules formed by Pst I cleavage are quite difficult to label. The polynucleotide kinase 5´ end-labelling reactions do not work well at Pst I sites, probably due to the 3´ overhanging sequence generated by the Pst I cut, but, even when the DNA strands are separated by heating, the labelling is not as efficient as that observed at other sites, (Richards - personal communication). While 'end-fill' labelling, using the Klenow fragment of *E. coli* DNA polymerase I (Klenow *et al.*, 1971; Englund, 1979), should work efficiently at the Pst I site, the lack of really high specific activity α-^{32}P-triphosphates has made it very difficult to achieve sufficient labelling using this method. Secondly, sequencing from the Pst I sites of the insert DNA is less desirable than sequencing from internal sites because of the

presence of the poly dc-dG tracts adjacent to the Pst I sites. A significant amount of 10S mRNA sequence information might be lost by sequencing long stretches of dC and dG residues.

For these reasons, sequencing was carried out from internal sites in the insert DNA. The procedure started with the cleavage of excised insert DNA with Hae III restriction endonuclease, which cuts at a four-base recognition sequence yielding blunt ends that label very efficiently. All the resulting restriction fragments had their 5´ terminal phosphate groups removed with bacterial alkaline phosphatase, and then were ^{32}P-5´-end labelled using polynucleotide kinase. Rather than search for restriction sites in the short lengths of labelled double-stranded DNA, the DNA was separated into single-strands and then electrophoresed on polyacrylamide gels to yield end-labelled single-stranded DNA molecules suitable for DNA-sequencing, (Maxam and Gilbert, 1979).

Fragments of the insert DNA from each of the four unidentified plasmids were subjected to the chemical degradation reactions of the Maxam and Gilbert procedure and the cleavage fragments separated on high-resolution acrylamide-urea gels so that the DNA-sequence could be determined.

This chapter describes the results of the DNA-sequencing experiments in detail.

8.2 RESULTS

8.2.1 Preparation of Plasmid DNA

Cultures of cells containing recombinant plasmids were treated with chloramphenicol to amplify the number of copies of plasmid per cell, (Clewell, 1972), and the cells were then lysed, using a heat-shock procedure. The plasmid DNA was separated from RNA by chromatography on a Sephadex G-150 column. Using this method, described in detail in chapter 2, 1.4-1.6 milligrams of clean plasmid DNA could be produced from one litre of amplified culture. A typical G-150 elution profile and a gel showing samples of four plasmids prepared by this procedure are shown in figure 8.1.

8.2.2 Excision and Sizing of Insert Sequences

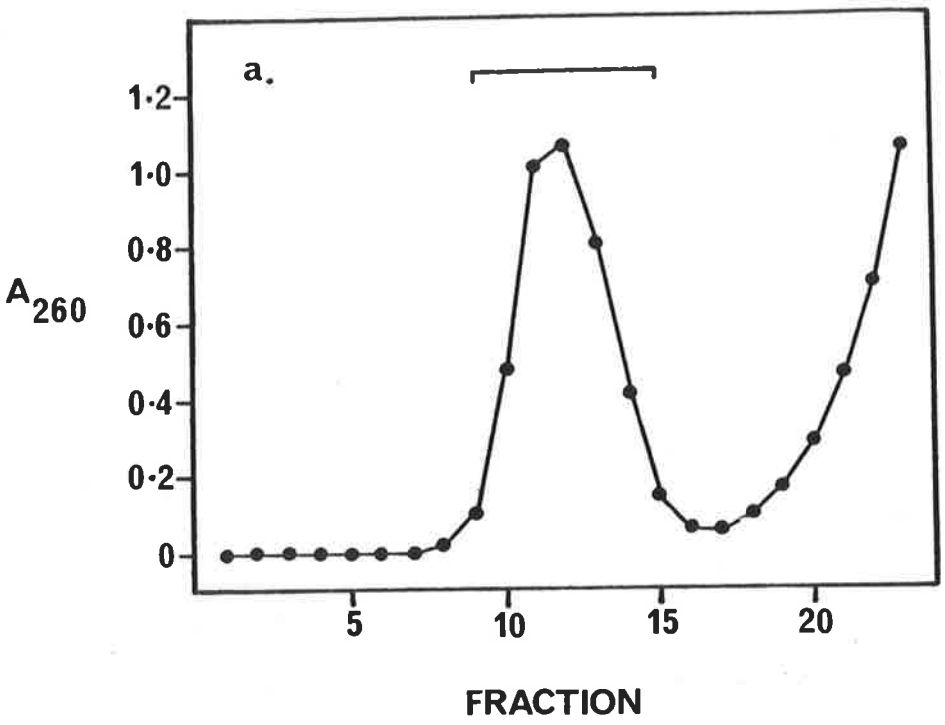
Samples of each plasmid DNA were digested with Pst I enzyme to excise the inserted DNA. Figure 8.2 shows the results of electrophoresing these digests on aqueous 4 percent polyacrylamide gels. Fragments of Hae III digested pBR322 DNA were used as the molecular weight markers on these gels, (Sutcliffe, 1978). The size of the inserts range from about 120 base pairs up to 450 base pairs long. While the G-C tailing procedure is designed to regenerate the Pst I site after annealing and ligation, it has been reported that as few as 40 percent of DNA sequences inserted in this way are finally excisable

Figure 8.1

Large-scale preparation of plasmid DNA, (section 2.2.14).

a. Elution profile of recombinant plasmid p43 chromatographed on Sephadex G-150. The material eluting at the exclusion volume was collected as plasmid DNA. The second, very large, peak of absorbance is due to degraded RNA.

b. Undigested first-peak material electrophoresed on a 1% agarose gel. Tracks A, B, C and D contain plasmid p25, p29, p43 and p50 DNA respectively.


a.
1·2
1·0
0·8
0·6
0·4
0·2
0
A260
5
10
15
20
FRACTION

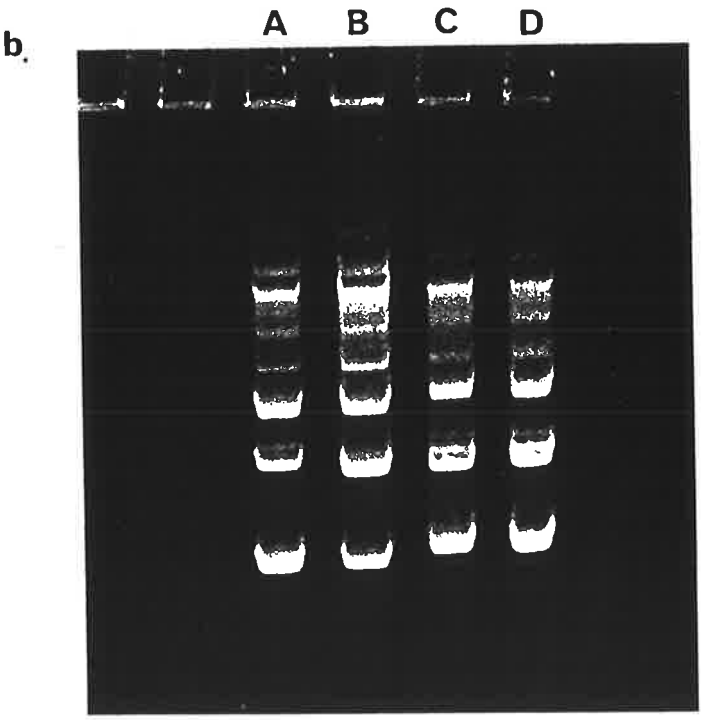

b.
A
B
C
D

Figure 8.2

Sizing of insert sequences on aqueous acrylamide gels. DNA from each of the recombinants prepared in bulk was digested with Pst I and electrophoresed on an aqueous 4% acrylamide gel. Hae III cleaved pBR 322 fragments were used as size markers, (Sutcliffe, 1978).

Tracks are as follows:

A and G : Hae III cleaved pBR 322.

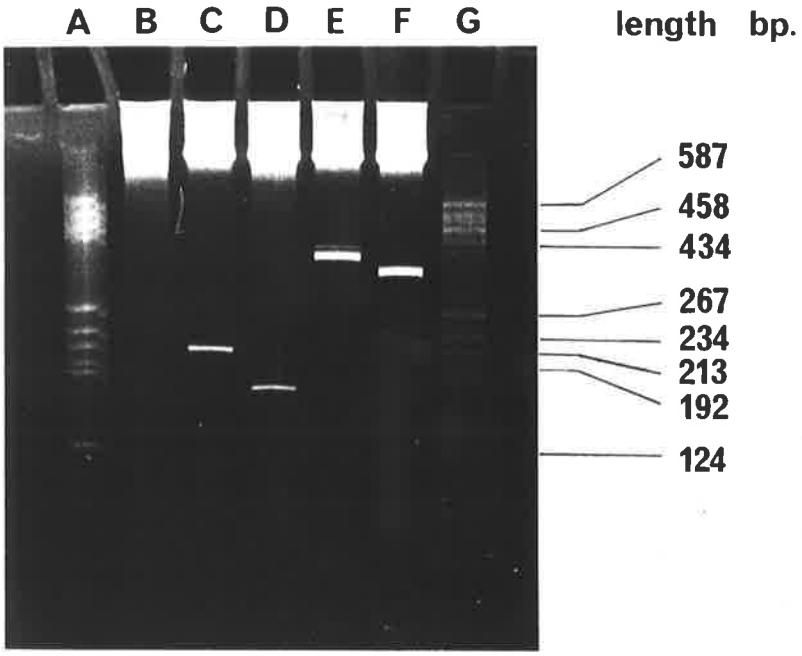
B : Pst I cleaved p1.

C : Pst I cleaved p25.

D : Pst I cleaved p29.

E : Pst I cleaved p43.

F : Pst I cleaved p50.


A
B
C
D
E
F
G
length bp.
587
458
434
267
234
213
192
124

by Pst I cleavage, (Villa-Komaroff *et al.*, 1978). As figure 8.2 shows, four of the five inserted sequences could be excised by Pst I. Plasmid p1 was linearised by Pst I cleavage, indicating the loss of the Pst I site at one of the insert-vector joins. In this small sample studied, 9 out of 10 Pst I restriction sites were regenerated using the G-C tailing insertion procedure, making about 80 percent of the inserts excisable by Pst I cleavage.

The accurate sizes of the four excised DNA fragments were determined by reference to the pBR322 marker tracks. The lengths of the inserts of plasmids p25, p29, p43 and p50 are 150, 120, 450 and 350 base pairs respectively. While the p43 and p50 inserts are quite large, it is possible that p25 and p29 carry only short lengths of 10S RNA sequence, since the total lengths of these inserts include poly dc-dG tracts at each end and possibly also a random primer DNA sequence of 10-20 bases length.

8.2.3 Hae III Cleavage of Recombinant Plasmid DNA

Due to the problems involved with the labelling of the Pst I sites at the ends of the excised insert sequences, it was necessary to label insert DNA at new ends generated by restriction cuts within the insert sequence. Samples of each plasmid DNA were digested with Hae III restriction endonuclease and

Figure 8.3

Hae III digestion of recombinant plasmids, p25, p29, p43 and p50. 3 µg of each of the recombinants was digested with Hae III and electrophoresed on an aqueous 6% acrylamide gel. Hae III digested pBR 322 DNA was also electrophoresed to show which of the fragments were derived from the vector molecule.

The tracks are as follows:

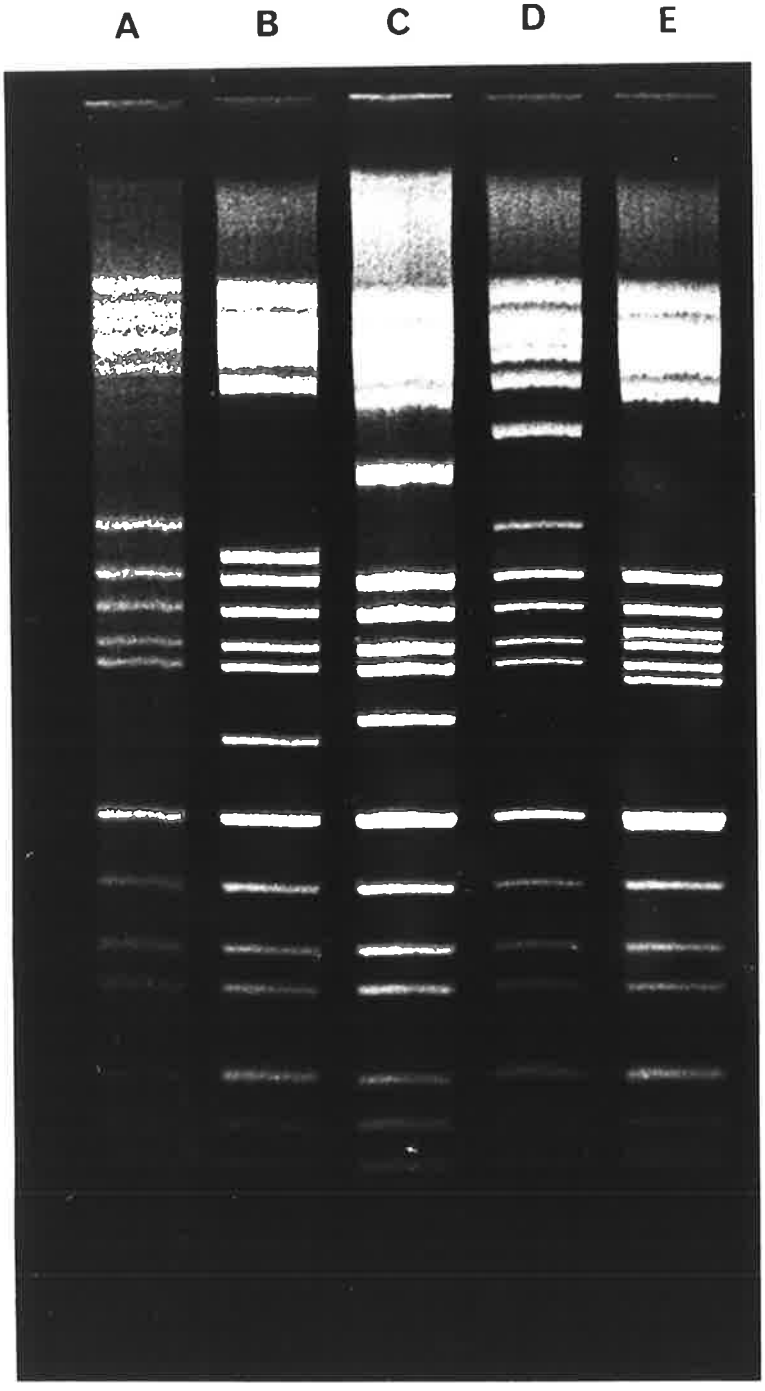
A : Hae III digested pBR 322.

B : Hae III digested p25.

C : Hae III digested p29.

D : Hae III digested p43.

E : Hae III digested p50.


A
B
C
D
E

the products electrophoresed on 6 percent polyacrylamide gels, (figure 8.3). Parental pBR 322 was digested with Hae III and the fragments run on the same gel to indicate the digestion pattern of the vector molecule. Any new bands appearing must be due to the presence of extra DNA inserted into the vector plasmid. As shown in figure 8.3, band number 6 of the parental plasmid, (containing the Pst I site), is not present in the digestion pattern of any of the recombinant plasmids, but is replaced, in each case, by at least three extra bands of varying sizes. This indicates the presence of at least two Hae III restriction sites within the insert sequence of each of the recombinants. It is interesting to note that in no case is an extra band common to more than one of the recombinants, and so there is no indication at this stage whether the plasmids share a common inserted sequence.

Similar digestion experiments were carried out in which plasmid DNA was digested with the four-base specificity restriction enzymes Hpa II and Alu I. In these cases, no restriction sites for these enzymes were detected within the inserted DNA sequences of any of the plasmids.

The presence of Hae III restriction sites within the inserted DNA will enable the fragments of the insert to be efficiently end-labelled, using polynucleotide kinase, in preparation for DNA-sequence determination.

8.2.4 Preparative Scale Isolation of Insert DNA

Results of previous experiments have shown that the insert DNA is readily resected by Pst I digestion of the recombinant plasmid and that each of the inserts contains the internal restriction endonuclease sites required for efficient end-labelling of the DNA. Since these requirements have been met, the first step in the DNA-sequencing procedure involves the isolation of working amounts of pure, insert DNA.

About 100 micrograms of each recombinant plasmid DNA was digested with Pst I restriction enzyme, and the resulting DNA fragments electrophoresed on preparative polyacrylamide gels. When ethidium bromide stained, the gels showed the presence of two bands; the excised insert DNA and the linear pBR 322 vector molecule. The preparative gels, used for the isolation of each of the insert DNAs, are shown in figure 8.4. The lower band, containing the insert DNA in each case, was cut from the gel and the DNA electrophoretically eluted from the gel slice.

Attempts were made to separate insert DNA from the parent vector molecule using sucrose-gradient centrifugation. While sufficient resolution of the insert and vector molecules was achieved, DNA from these preparations labelled very poorly after Hae III cleavage, possibly due to the presence of large molar amounts of RNA contaminating

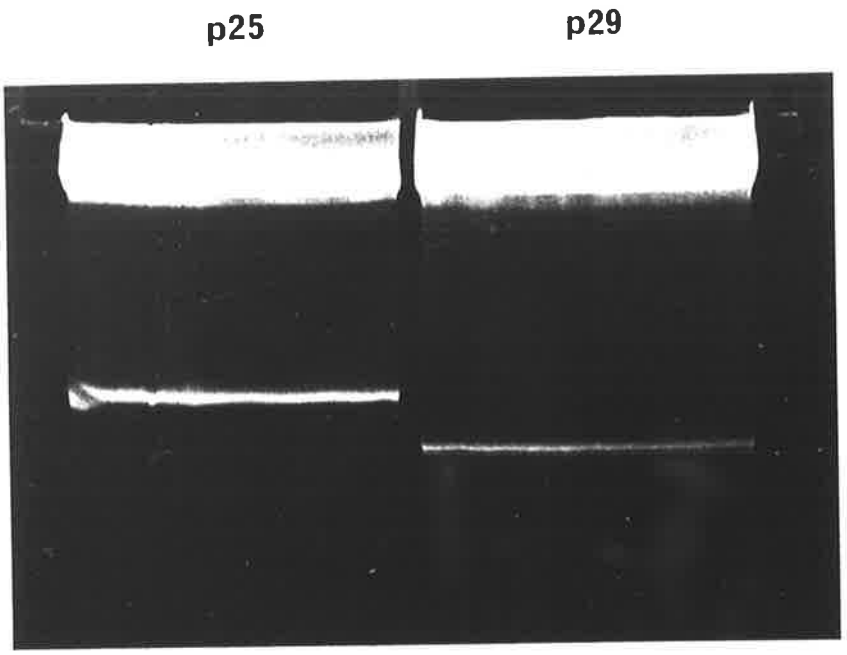
Figure 8.4

Large scale isolation of insert DNA. 50 µg of recombinant plasmid DNA was digested with Pst I to excise the insert sequence and electrophoresed on a 5% aqueous acrylamide gel. The insert band was excised from the gel and the DNA electrophoretically eluted from the gel slice.

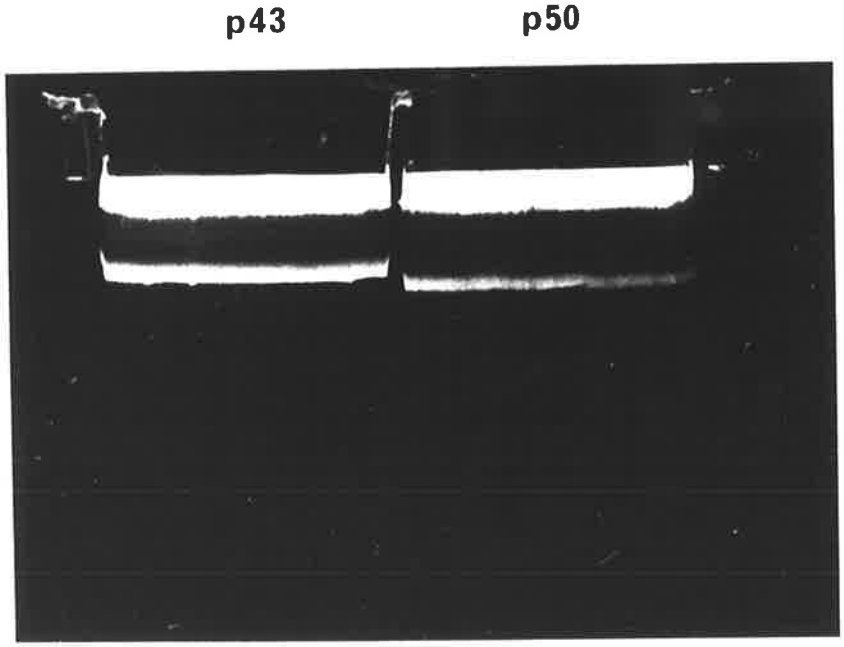
a. Pst I digested plasmids p25 and p29.

b. Pst I digested plasmids p43 and p50.

a.



b.



the insert DNA. Since the presence of RNA in the polynucleotide kinase ^{32}P-exchange reaction interferes with the efficient labelling of DNA, it is essential that the DNA preparation is free of contaminating RNA. The preparative gel procedure, however, allows excellent separation of the insert DNA from both the plasmid vector DNA and from residual RNA and, using electrophoretic elution, almost quantitative recovery of DNA from the gel can be achieved.

8.2.5 Restriction and end-labelling of insert DNA

The methods employed for the preparation of end-labelled DNA and for the chemical degradation of DNA in the sequencing reactions were essentially those described by Maxam and Gilbert, (1977; 1979). Any alterations to these procedures will be noted in the text or described in chapter 2.

One microgram of each insert DNA was cleaved with Hae III enzyme in an overnight incubation, to ensure complete digestion. High specific activity γ-labelled ^{32}P-ATP, (supplied by R. Richards), was used, with polynucleotide kinase, to end-label the Hae III digested DNA fragments by forward phosphorylation of the 5′-hydroxyl groups, (Lillehaug *et al.*, 1976). Forward phosphorylation of the ends of restriction fragments requires that the 5′-phosphates be removed beforehand, and so the Hae III restriction fragments of the insert DNA

were dephosphorylated with bacterial alkaline phosphatase. After phenol-extraction to remove the alkaline phosphatase the dephosphorylated insert DNA was heated in the kinase buffer, in the presence of spermidine, to separate the two DNA strands, and then the polynucleotide kinase end-labelling reaction was carried out.

8.2.6 Separation of ^{32}P end-labelled strands

The separation of dissociated DNA strands was first reported by Hayward, (1972), for whole bacteriophage DNA using agarose gels, and since then the technique has been applied to much smaller pieces of DNA using polyacrylamide gels, (Maxam and Gilbert, 1977). Making use of a DMSO denaturation and loading buffer and a modified gel composition, complementary DNA strands several hundreds of nucleotides long can be successfully resolved, (Maxam and Gilbert, 1979).

The Hae III digested, end-labelled insert DNA was DMSO dissociated and electrophoresed on a 5 percent strand-separation gel and the position of the labelled DNA detected by autoradiography. Figure 8.5 shows the position of single-stranded, end-labelled DNA fragments when insert DNA from recombinant plasmid p50 was electrophoresed on a strand separation gel. Clearly separation of the DNA strands has been achieved, because the pattern of bands is much more complex than would be observed

Figure 8.5

Strand separation of ^{32}P-labelled insert DNA fragments. Purified p50 insert DNA, (figure 8.4), was digested with Hae III and the fragments 5´ end-labelled. The DNA strands were separated by heating in 30% DMSO and then loaded directly onto a strand separation gel.

Track A : double-stranded ^{32}P-labelled DNA fragments.

Track B : single-stranded ^{32}P-labelled DNA fragments.

The 5´ ends of DNA contained within the Pst I cleavage sites label very poorly and so two single-stranded fragments of DNA present in track B are not visible on the autoradiograph.


A
B
–O

for the intact double-stranded restriction fragments of the insert DNA. The DNA fragments containing the Pst I site 5′ ends are poorly labelled and so are not visible on a short exposure. It has also been observed, (Harvey and Wells, 1979), that short pieces of DNA label more efficiently than long pieces, in the polynucleotide kinase reaction, and this differential labelling effect is visible in figure 8.5.

An autoradiograph was accurately aligned over the strand separation gel, by reference to the marker spots, and the regions of the gel containing labelled DNA were excised. The single-stranded DNA was electrophoretically eluted from the gel slices into dialysis bags and, in general, more than 80 percent of the labelled DNA was recovered from the gel slice. As a convenient rule, it was practical to proceed with the DNA-sequencing reactions if about 10^5 dpm of ^{32}P-label was present in each single-stranded fragment of DNA, at this stage.

8.2.7 DNA-sequencing reactions

Base-specific chemical cleavage reactions were carried out exactly as described by Maxam and Gilbert, (1979), except for the guanine plus adenine, (G + A), specific reaction, where pyridinium formate pH2 was replaced by 100 percent formic acid.

Samples of ^{32}P-labelled single-stranded DNA were separated into four aliquots and each aliquot

was taken through one of the base specific sequencing reactions. (Wherever possible, at least 6000 dpm of ^{32}P-DNA was used per sequencing reaction). After chemical stand-cleavage each reaction mix was dried down, resuspended in formamide loading buffer and loaded onto a separate track of a 0.5mm x 30 cm x 40 cm, 10 percent polyacrylamide sequencing gel. Electrophoresis was carried out at the maximum possible voltage, (1200 v) until the bromophenol-blue tracker dye, (which runs at a position equivalent to about a 20 nucleotide-long piece of single-stranded DNA), was 2 centimetres from the bottom of the gel. The gel was transferred from the glass plate onto a large sheet of exposed film which acted as a support for the gel during handling. After covering the gel with plastic-wrap, the gel was placed in contact with X-ray film and exposed at -80^{o}C in the presence of an intensifying screen. A 24 hour exposure was usually sufficient to give a clearly discernable banding pattern from which a DNA-sequence could be read.

Figure 8.6 shows the autoradiograph resulting when the base-specific cleavage products of a fragment of the p29 insert DNA were separated on a sequencing gel. From this autoradiograph it was possible to determine the nucleotide sequence of nearly 60 bases of the p29 insert. In the same manner DNA sequences were determined for fragments of the insert DNA of plasmids p25, p43 and p50.

Figure 8.6

DNA sequence determination. Aliquots of single-stranded end-labelled p29 insert DNA were degraded using the base-specific cleavage reactions of Maxam and Gilbert, (1977), and electrophoresed on a 10% acrylamide DNA-sequencing gel. On this gel the DNA-sequence starts at about 20 bases from the labelled end of the DNA fragment.


G A T C
C C A G T C C C G T C G T C C C C A C C C

Examination of the autoradiograph, (figure 8.6), shows the presence of a faint band in every track when the actual nucleotide in that position is a guanine. This is thought to be due to impurities in the hydrazine causing a low level of cleavage at all guanine residues, (R. Richards - personal communication). Fortunately the presence of these extra bands in the other tracks does not interfere with the reading of the correct nucleotide sequence.

In the course of these experiments several fragments that contained the dG-dC joining regions and the terminal Pst I site were sequenced. Figure 8.7 shows the sequencing gel of a terminal fragments of the p25 insert. The cytosine-specific track shows the presence of 26 consecutive dC residues making up the joining-region between the vector and ds cDNA sequences. The two other joining regions sequenced contained dC or dG tracts 43 and 23 bases in length. The 43 base length was considerably longer than expected from the kinetics of the tailing reaction, (figure 6.4), but no doubt indicates the heterogeneity in lengths of tails added to the ds cDNA and vector molecules.

8.2.8 Examination of the DNA-Sequences

The major DNA sequences determined, representing portions of the inserts from each of the recombinant plasmids, are presented in figure 8.8. While in each case the sequence of only one strand has been determined, the complementary nucleotide

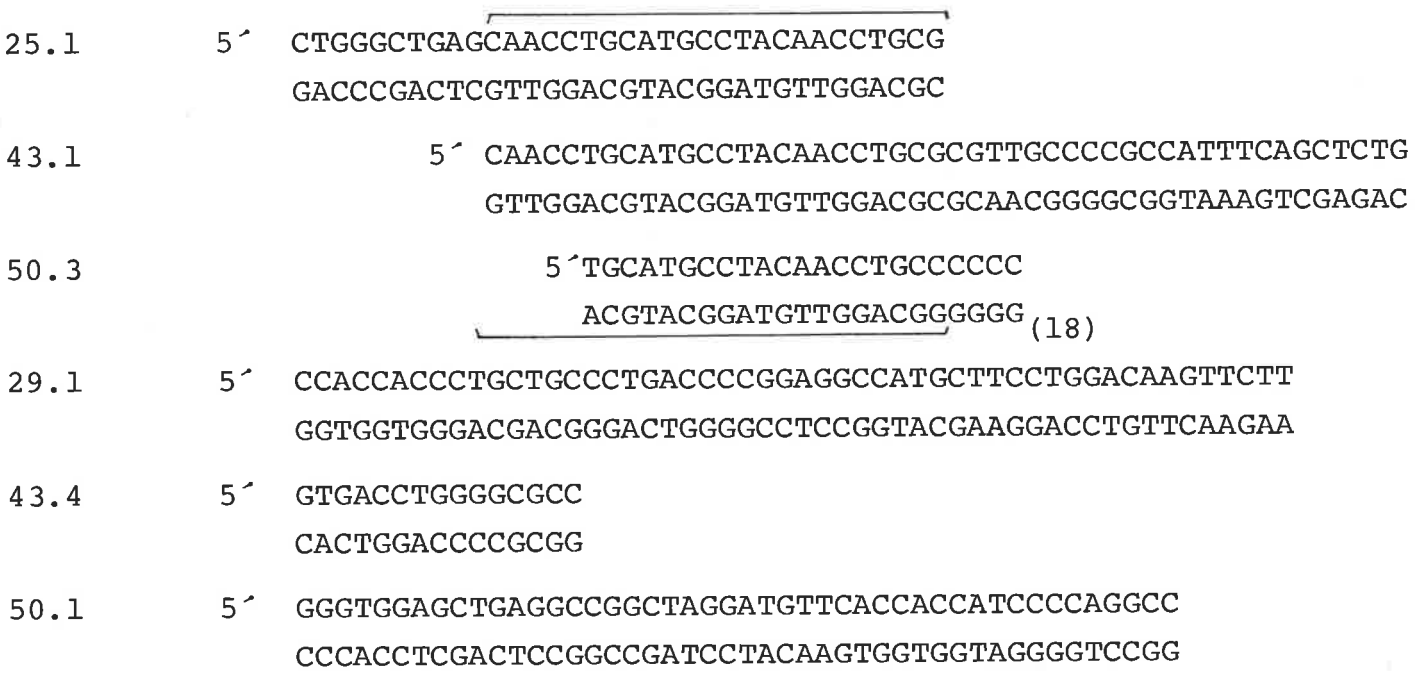
Figure 8.7

DNA sequence of a fragment of the p25 insert containing the dC-dG joining region. Details are the same as for the previous figure.


G A T C

Figure 8.8

DNA sequences determined from the recombinant plasmid inserts. The overlapping sequences detected in the p25, p43 and p50 inserts have been aligned.


25.1
5´ CTGGGCTGAGCAACCTGCATGCCTACAACCTGCG
GACCCGACTCGTTGGACGTACGGATGTTGGACGC
43.1
5´ CAACCTGCATGCCTACAACCTGCGCGTTGCCCCGCCATTTCAGCTCTG
GTTGGACGTACGGATGTTGGACGCGCAACGGGGCGGTAAAGTCGAGAC
50.3
5´TGCATGCCTACAACCTGCCCCCC
ACGTACGGATGTTGGACGGGGGG(18)
29.1
5´ CCACCACCCTGCTGCCCTGACCCCGGAGGCCATGCTTCCTGGACAAGTTCTT
GGTGGTGGGACGACGGGACTGGGGCCTCCGGTACGAAGGACCTGTTCAAGAA
43.4
5´ GTGACCTGGGGCGCC
CACTGGACCCCGCGG
50.1
5´ GGGTGGAGCTGAGGCCGGCTAGGATGTTCACCACCATCCCCAGGCC
CCCACCTCGACTCCGGCCGATCCTACAAGTGGTGGTAGGGGTCCGG

sequence of the other strand has been included in the diagram to make interpretation of the results easier.

Inspection of the sequences shows that three of the clones, p25, p43 and p50, share a common nucleotide sequence. This is a rather fortunate observation considering that only a small proportion of the total nucleotide sequence of the recombinants was determined. The detection of a common nucleotide sequence is important for two reasons, firstly because it demonstrates that the DNA-sequences obtained are accurate, (since 100 percent sequence agreement was found in the three independent sequencing determinations), and secondly, because it suggests that these three clones were derived from the same original RNA species. No overlap between the sequence determined for p29 and any part of the other plasmid DNA-sequences could be found. At this stage therefore, it is not possible to decide whether p29 was derived from the same original RNA species as the other clones, or represents a totally independent sequence.

The first question to ask is whether any of the nucleotide sequences determined represent sequences from the coding region of H5 mRNA. Figure 8.9 presents the available amino-acid sequence data for chicken histone H5. H5 is an extremely lysine-rich protein, lysine residues accounting for 44 out of the total of 189 amino-acids, (Sautiere *et al.*, 1976). While

Figure 8.9

The amino-acid sequence of chicken histone H5, (Sautiere *et al.*, 1975; 1977; Champagne - personal communication). The order of the last 29 residues has not been determined.

PRIMARY STRUCTURE OF CHICKEN HISTONE H5

Thr-Glu-Ser-Leu-Val-Leu-Ser-Pro-Ala-Pro-Ala-Lys-Pro-Lys-$\frac{\text{Gln}}{\text{Arg}}$-

Val-Lys-Ala-Ser-Arg-Arg-Ser-Ala-Ser-His-Pro-Thr-Tyr-Ser-Glu

Met-Ile-Ala-Ala-Ala-Ile-Arg-Ala-Glu-Lys-Ser-Arg-Gly-Gly-Ser-

Ser-Arg-Gln-Ser-Ile-Gln-Lys-Tyr-Ile-Lys-Ser-His-Tyr-Lys-Val-

Gly-His-Asn-Ala-Asp-Leu-Gln-Ile-Lys-Leu-Ser-Ile-Arg-Arg-Leu-

Leu-Ala-Ala-Gly-Val-Leu-Lys-Gln-Thr-Lys-Gly-Val-Gly-Ala-Ser-

Gly-Ser-Phe-Arg-Leu-Ala-Lys-Ser-Asp-Lys-Ala-Lys-Arg-Ser-Pro-

Gly-Lys-Lys-Lys-Lys-Ala-Val-Arg-Arg-Ser-Thr-Ser-Pro-Lys-Lys-

Ala-Ala-Arg-Pro-Arg-Lys-Ala-Arg-Ser-Pro-Ala-Lys-Lys-Pro-Lys-

Ala-Thr-Ala-Arg-Lys-Ala-Arg-Lys-Lys-Ser-Arg-Ala-Ser-Pro-Lys-

Lys-Ala-Lys-Lys-Pro-Lys-Thr-Val-Lys-Arg-

$(Ser_5\ Lys_{12}\ Pro_2\ Arg_3\ Ala_5\ Val_1\ Gly_1)$

the lysine residues are concentrated at the C-terminal end of the protein, (figure 8.9), there is no segment of protein more than 22 residues long that does not contain at least one lysine, and, in general, lysine residues occur much more frequently than this. The nucleotide-triplet codon for lysine is either AAA or AAG, with a strong preference for the AAG codon observed in eukaryotes. In the sea urchin histone genes, for example, about 25 in every 40 lysine residues are coded for by the AAG triplet, (Kedes, 1979).

Inspection of the DNA-sequences presented in figure 8.8 shows the presence of very few potential lysine codons. The number of possible lysine codons is, in fact, even fewer still, since several of the putative lysine codons appear in conflicting reading-frames and so could not all contribute to the final protein product. If the reading-frame of the sequence is aligned so that any particular AAG or AAA sequence is read as a lysine, the amino-acid sequence around the putative lysine residue can be determined from the DNA-sequence. When this is done, no correspondence is found between the potential amino-acid sequence, dictated by the nucleotide sequence, and the known amino-acid sequence of H5 protein, (Sautiere *et al.*, 1975; 1976; Champagne - personal communication). The remaining regions of the H5 protein that contain no lysine residues have also been examined, and, once again, no agreement

is found between the predicted and known amino-acid sequences. While it is possible that the DNA-sequences determined might belong to the untranslated regions of the H5 mRNA, this seems unlikely since sequences from all four independent clones have failed to detect any H5 coding sequences, and because the untranslated regions comprise only a very small part of the total length of the H5 mRNA molecule, (section 8.1).

The failure to detect a recombinant containing H5 mRNA sequences amongst the clones formed from 10S RNA was both disappointing and surprising. Certainly it was expected that, after globin and ribosomal RNA, the most common component of 10S reticulocyte RNA would be H5 mRNA, and so the failure to detect an H5 mRNA clone amongst the four non-globin, non-ribosomal candidates was difficult to understand.

The plasmid insert sequences were examined to determine whether they corresponded to any known nucleotide sequence that might appear in chicken reticulocyte RNA. Two RNA species that could have been present at a low level in the 10S RNA are 5.8S and 5S ribosomal RNA. The chicken 5.8S ribosomal RNA, (Khan and Maden, 1977) and the chicken 5S ribosomal RNA, (Pace *et al.*, 1974; Brownlee and Cartwright, 1975), have been sequenced, but neither molecule shows any homology with the nucleotide sequences determined for the plasmid inserts. In

addition, none of the known chicken globin mRNA sequences, (Richards *et al.*, 1979; Richards and Wells, 1980), (including the 5´ and 3´ untranslated regions), show any similarity to the insert DNA-sequences.

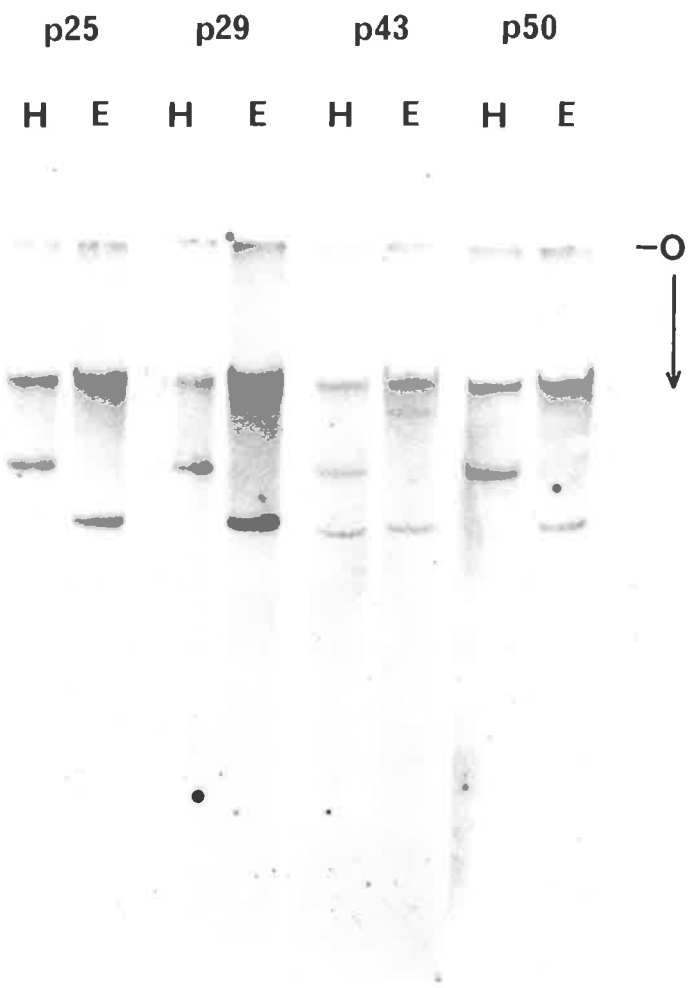
8.2.9 Southern-Blot Experiments

In an effort to determine something of the nature of the cloned sequences, DNA fragments homologous to the insert-DNA sequences were detected in restriction enzyme digests of chicken genomal DNA. Eco RI and Hind III restriction endonuclease-digested chicken DNA was electrophoresed on a low-percentage agarose gel, transferred to nitro-cellulose using the method of Southern, (1975), and the filter probed with ^{32}P-labelled nick-translated DNA representing each of the insert sequences. The resulting autoradiograph is shown in figure 8.10.

Despite small differences observed when Hind III-cleaved chicken DNA was probed with plasmid p43 insert, it is clear that the same genomal-DNA fragments are being detected by each of the cloned probes. Although limited nucleotide sequence data had failed to show homology between p29 and the other clones, this result shows that all four selected recombinants contain common sequences. Presumably therefore, each of these molecular recombinants arose from a separate copy of the same RNA species, originally present in the reticulocyte

Figure 8.10

Detection of fragments homologous to recombinant sequences in Eco RI and Hind III-digested chicken genomal DNA. Chicken DNA was digested to completion with either Eco RI or Hind III restriction enzyme, electrophoresed on a 0.8% agarose gel and transferred to nitrocellulose filter. One track of each digest was hybridised to ^{32}P-labelled p25, p29, p43 and p50 insert DNA, washed at high stringency, (0.1 X SSC) and auto-radiographed.


p25
p29
p43
p50
H E H E H E H E
O

10S RNA. The bands detected in the genomic DNA show no resemblance to the pattern obtained when chicken DNA is probed with either ribosomal sequences, (Saint, 1979), or with chicken globin sequences, (Engel and Dodgson, 1978; A. Robins, personal communication), and so the recombinants clearly contain non-globin, non-ribosomal sequences, as suggested by previous experiments. The weakness of the response to the labelled probe suggests that the genes located in the genomal DNA are present at only a very low copy number and are possibly unique.

Labelled p43 insert DNA detects an extra band in Hind III-digested chicken DNA. Since the p43 insert is the longest of the four, it is reasonable to assume that it contains sequences that are not found in any of the other plasmid inserts and so may detect an extra fragment in genomal-DNA digests.

The Southern transfer experiments have been useful for two main reasons, firstly, to demonstrate quite clearly that the selected clones contain neither globin nor ribosomal sequences, and secondly, to show that all the selected recombinants were derived from the same original RNA species. The experiments have yielded no clue as to the nature of the cloned sequences. While the original RNA molecule may be a non-coding species, it is also possible that it is an mRNA coding for a protein that is not readily detected in the translation assay system used. While extensive amino-acid sequence

data on proteins that are expressed in the red blood cell might allow identification of the cloned sequences, it seems unlikely that any positive identification will be made without further experiments being performed, (perhaps, for example, using the hybrid-arrested translation procedure, (Paterson *et al.*, 1977)). At present it is sufficient to know that the cloned sequences are not H5 mRNA sequences.

CHAPTER 9

RE-EXAMINATION OF RETICULOCYTE RNA TRANSLATION PRODUCTS

9.1 INTRODUCTION

The failure to detect H5 mRNA sequences in any of the non-ribosomal, non-globin recombinants suggested that a major error had been made in the estimation of the proportion of H5 mRNA sequences in the 10S RNA used as the original template for cDNA synthesis.

Only two translation products of 10S reticulocyte RNA are visible on polyacrylamide tube gels, with the minor non-globin protein comprising only a few percent of the total translation products. This peak has been identified as H5 protein because it co-electrophoreses with authentic histone H5 marker, (Scott, 1975, and section 3.2.2(i)). In the recombinant-DNA studies, four non-ribosomal, non-globin clones were isolated and analysed in detail. These clones comprised only a few percent of the total recombinants and while Southern-blot experiments, (section 8.2.9), demonstrated that each of these clones contained parts of the same gene sequence, DNA-sequencing studies showed that the recombinants did not contain H5 mRNA sequences.

Only one kind of non-globin, non-ribosomal sequence was detected amongst the recombinants formed from 10S RNA, and only one non-globin protein was observed in the translation products of 10S RNA, and so it seems possible that the cloned RNA-sequence may code for this protein. Since the ds cDNA clone does not contain H5 mRNA sequences,

it is possible that the non-globin protein peak observed in the translation products of 10S RNA is not histone H5. A series of experiments was undertaken to test this possibility.

For reasons explained previously, (section 3.1), all assays for H5 protein amongst the translation products of 10S RNA were performed by slicing and counting SDS-urea tube gels. An identical assay procedure was used by Scott, (1975), to analyse the translation products of RNA factions during the isolation of H5 mRNA by the immunoprecipitation technique. Using this assay system the minor, non-globin translation product co-electrophoresed with authentic H5 marker protein, (see figure 3.2).

In order to determine whether the minor translation product was in fact H5 protein it was necessary to apply more stringent assay conditions, and so the 10S RNA translation products were electrophoresed on high-resolution discontinuous SDS slab gels. The translation products were labelled with ^{35}S-methionine and the slab gels were fluorographed to detect the position of the protein products. Due to the very tight banding pattern obtained with discontinuous slab gels and the high resolution of the fluorographic technique, this system provides a much more rigorous assay of protein products than that previously obtained using tube gels.

Electrophoresis of the ^{35}S-labelled translation products of 10S reticulocyte polysomal RNA on the slab gels showed no observable translation product

at the position of the histone H5 marker. However, the minor protein peak observed on tube gels resolved into two closely-migrating bands on the slab gel, and both bands migrated a small but significant amount more slowly than the H5 marker protein. This experiment suggested that the double-band of protein had been mistakenly identified as H5 in all experiments using tube gels, and that H5 protein was not detectable amongst the translation products of 10S RNA.

H5 mRNA had previously been isolated from reticulocyte polysomes, (Scott and Wells, 1975; 1976), and so H5 protein should have been detected in the translation products of the 10S reticulocyte RNA used in these experiments, (since the method of Pemberton *et al.*, (1972), also involves the isolation of polysomes as the initial step in the RNA preparation procedure). To ensure, however, that no selection had been made against H5 mRNA sequences by the particular RNA preparation procedure used, total cellular RNA was isolated from reticulocytes, chromatographed through oligo-dT-cellulose and the bound and unbound RNA samples fractionated on sucrose gradients. Samples of the RNA ranging in size from about 7S to 16S were translated in the wheat-germ cell-free system. The labelled translation products from all experiments were electrophoresed on slab gels in the presence of histone H5 markers. Once again, no translation product was observed to

co-electrophorese with the H5 marker under these high-resolution conditions.

The experiments described in this chapter show that, contrary to the interpretation of previous experiments, the non-globin translation product of reticulocyte 10S RNA observed on tube gels does not correspond to the mature H5 protein isolated from erythrocytes. Further, since H5 protein is undetectable in the translation products of any fraction of reticulocyte RNA, the results of these studies suggest that H5 mRNA is only present at extremely low levels in reticulocytes.

9.2 RESULTS

9.2.1 Analysis of 10S polysomal RNA translation products on discontinuous SDS gels

10S RNA prepared from reticulocyte polysomes, (section 3.2.1), was translated in the wheat-germ cell-free translation system using ^{35}S-methionine as the labelled amino-acid. While most proteins contain few methionine residues, the use of ^{35}S in fluorography procedures allows high sensitivity with reasonably short exposure times, (Bonner and Laskey, 1974). Reference to the amino-acid compositions of histone H5 and the α and β chicken globins, (Table 3.1), shows that each of the poly-peptides contains a single methionine residue and so the amount of ^{35}S-methionine label in each protein will reflect the amount of each protein translated.

The ^{35}S-labelled translation products were precipitated from the translation mix solution and electrophoresed on discontinuous SDS slab gels, (Laemmli, 1970), using a 3 percent acrylamide stacking gel and a 12.5 percent acrylamide separating gel. The gels were prepared for fluorography, (Bonner and Laskey, 1974), placed in contact with X-ray film and exposed at -80^{o}C in the presence of an intensifying screen. Unfortunately, although the use of SDS-urea slab gels, (Swank and Munkres, 1971), would have provided a direct comparison with the previous tube gel results, this gel system could not be used in these experiments because, in repeated attempts, the gels cracked during the drying-down process, permitting no useful interpretation of results. ^{3}H-labelled H5 marker was electrophoresed in tracks adjacent to the ^{35}S-labelled translation products. In a control experiment, ^{3}H-labelled H5 was added to a translation mix and incubated at 25^{o}C for the normal 60 minute translation period, to determine whether any degradation of H5 protein occurred during the incubation. The fluorograph showing the results of the translation and control experiments is shown in figure 9.1.

The shorter exposure, (figure 9.1a), shows a very intense band of globin protein and a less-intense double band of protein electrophoresing just above the position of the H5 marker. Together,

Figure 9.1

Analysis of reticulocyte RNA translation products on discontinuous-SDS slab gels, (Laemmli, 1970). The 10S RNA samples were prepared from polysomes using the RNP-procedure, (Pemberton *et al.*, 1972). The ^{35}S-labelled translation products were visualised using fluorography. The tracks are as follows:

A : translation products of polyadenylated chicken liver RNA, (control).

B : ^{3}H-labelled H5 marker.

C : translation products of 10S reticulocyte RNA, preparation 1.

D : translation products of 10S reticulocyte RNA, preparation 2.

E : translation products of 10S reticulocyte RNA, preparation 3.

F : translation products of 10S reticulocyte RNA, preparation 4.

G : ^{3}H-labelled H5 marker, taken through the complete translation procedure.

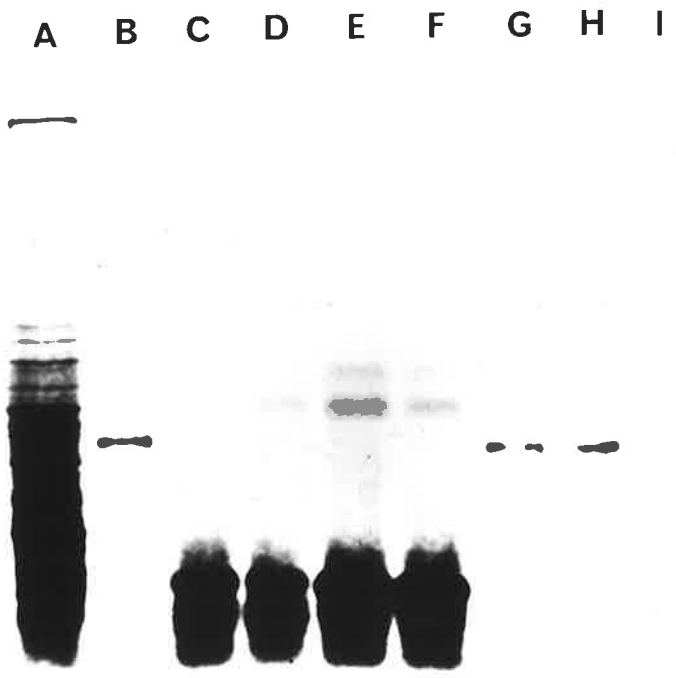
H : ^{3}H-labelled H5 marker.

I : wheat-germ background.

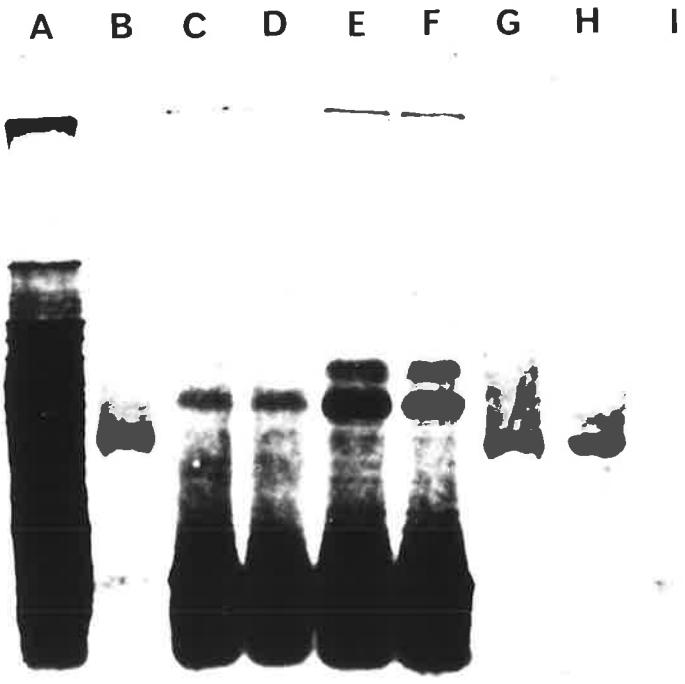
a. 11 hour exposure.

b. 58 hour exposure.

a.



b.



the double bands account for several percent of the labelled translation products and almost certainly represent the material that co-electrophoresed, (as a single peak on the low-resolution tube gels), with the H5 marker on the SDS-urea tube gels, (figure 3.2). Under the high resolution conditions used in these experiments, the H5 marker is clearly resolved from this double band of protein. Apart from the globin and the double-band of protein, no other translation products at all are visible on the fluorograph, either at the position of the H5 marker or elsewhere on the gel.

The longer exposure, (figure 9.1b), shows the presence of protein bands that were undetected in the short initial exposure. None of these proteins represent more than a fraction of a percent of the total labelled products. Despite the sensitivity of the assay, it is not possible to identify any translation product that co-electrophoreses with the H5 marker, indicating that H5 protein must represent only a very small fraction of the total translation products.

Since both Knöchel, (1975) and Scott and Wells, (1975; 1976), have shown the presence of H5 protein in the translation products of reticulocyte polysomal RNA, it is not clear why H5 protein was not detected in the ^{35}S-labelled translation products analysed on the slab gels. As described previously, (section 3.2.2(i)),

H5 mRNA was expected to account for a few percent of the total 10S RNA, and so the failure to detect any H5 protein at all in the translation products of this RNA suggested that some step in the 10S reticulocyte RNA preparation procedure may have selected against H5 mRNA sequences. To check this possibility, total cellular RNA was isolated from reticulocytes and assayed for the presence of H5 mRNA.

9.2.2 Preparation of total cellular RNA from reticulocytes and analysis of translation products

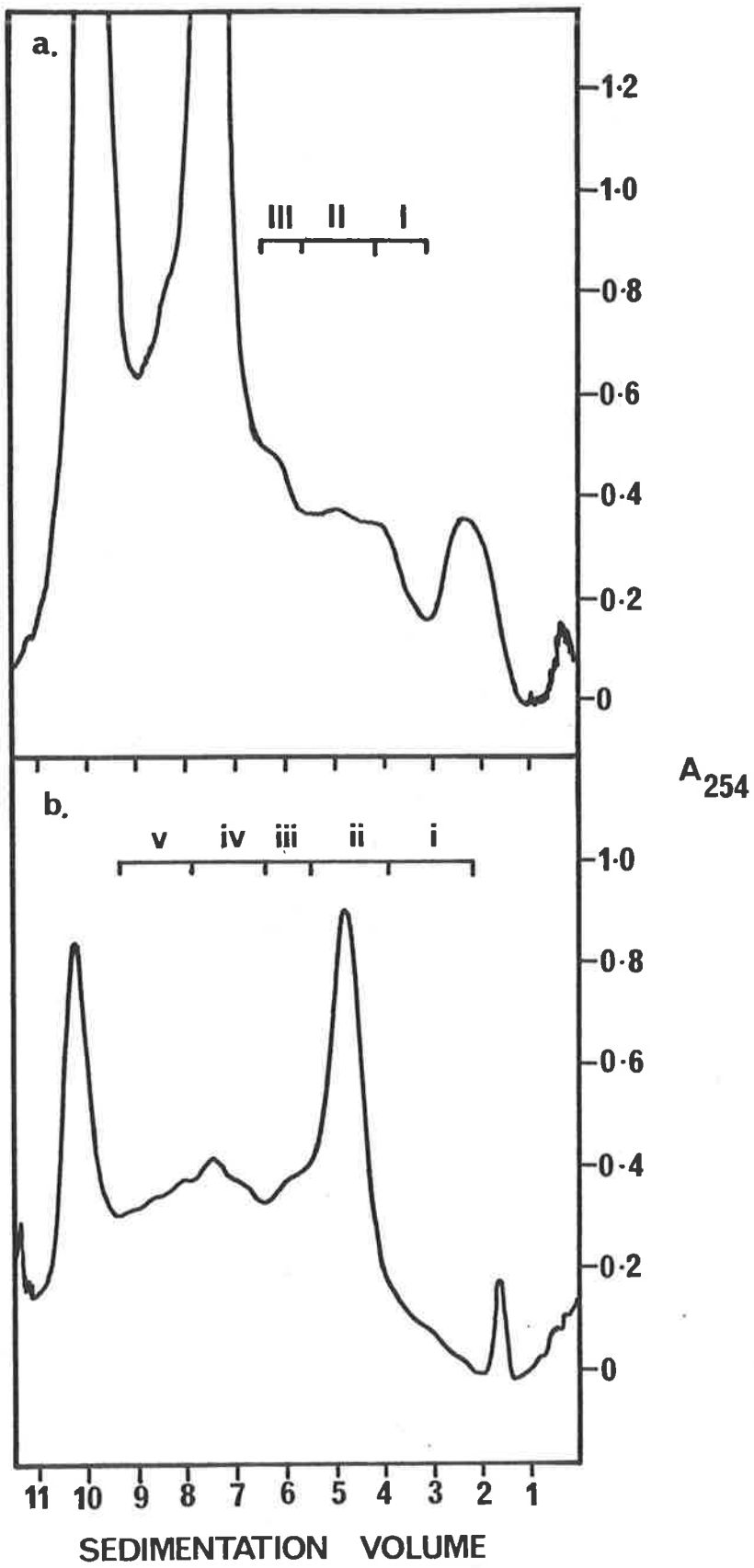
Total cellular RNA was extracted from reticulocytes, using a guanidinium-HCl denaturation procedure, and then the RNA was separated into polyadenylated and non-polyadenylated fractions by oligo-dT-cellulose chromatography. (Complete details of the extraction procedure are contained in chapter 2). The absorbance profiles obtained for oligo-dT-cellulose bound and unbound RNA samples, fractionated on sucrose gradients, are shown in figure 9.2. Fractions of RNA from about 7S to 16S in size were selected from the gradient, as indicated, and aliquots translated in the wheat-germ cell-free system. Since H5 mRNA is not polyadenylated, (Scott, 1975; Scott and Wells, 1976), the 10S RNA fraction unbound to oligo-dT-cellulose should be considerably enriched for H5 mRNA sequences.

Figure 9.2

Sucrose gradient fractionation of total chicken reticulocyte RNA, (prepared as described in section 2.2.2). Samples of RNA were collected as indicated.

a. reticulocyte RNA unbound to oligo-dT-cellulose.

b. reticulocyte RNA bound to oligo-dT-cellulose.


a.
III II I
1·2
1·0
0·8
0·6
0·4
0·2
0
b.
v iv iii ii i
1·0
0·8
0·6
0·4
0·2
0
A254
11 10 9 8 7 6 5 4 3 2 1
SEDIMENTATION VOLUME

The ^{35}S-labelled translation products of all RNA fractions were electrophoresed on discontinuous SDS slab gels, (Laemmli, 1970), together with ^{3}H-labelled H5 marker protein, and the position of the labelled proteins detected by fluorography, (figure 9.3).

The translation products of these RNA fractions contain a number of proteins that were either absent from, or present in only very small amounts in, the products of 10S polysomal RNA, (figure 9.1). These proteins may have been coded for by mRNAs that were not present in the polysomal fraction of reticulocyte RNA or, alternatively, the polysome/ribonucleoprotein-particle preparation technique, (Pemberton *et al.*, 1972), may have selected against these mRNA species specifically. Despite the increase in the number of proteins observed in the translation products of the total cellular RNA preparations, no observable band, in any fraction, co-electrophoreses with the H5 marker protein.

The translation products of the RNA fractions bound and unbound to oligo-dT-cellulose show differences in the intensity of specific bands, but no protein seems to be localised entirely to either the bound or unbound fraction. Despite the prediction that H5 mRNA will be considerably enriched in the non-polyadenylated RNA, no H5 protein is visible in the translation products of the unbound RNA.

Figure 9.3

Analysis of reticulocyte RNA translation products on discontinuous-SDS slab gels. Tracks A, J, K and T contain ^{3}H-labelled H5 marker. The other tracks contain the ^{35}S-labelled translation products produced by the following RNA fractions, (fraction numbers refer to figure 9.2):

B : reticulocyte 10S RNA prepared from polysomes, (preparation 1).

C : reticulocyte 10S RNA prepared from polysomes, (preparation 2).

D : fraction I RNA.

E : fraction II RNA.

F : fraction III RNA.

G : fraction i RNA.

H : fraction ii RNA.

I : fraction iii RNA.

L : fraction iv RNA.

M : fraction v RNA.

N,O,P,Q and R : reticulocyte RNA fractions from an unrelated experiment.

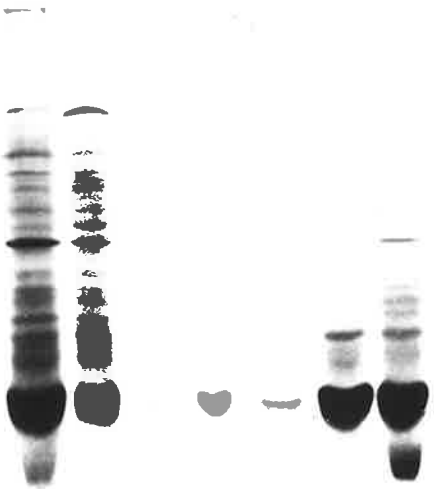
S : wheat-germ background.

A B C D E F G H I J

a.

K L M N O P Q R S T

b.



9.3 DISCUSSION

Control experiments have shown that authentic H5 protein may be recovered, quantitatively, from the translation mix in undegraded form, (figure 9.1), and amino-acid composition data indicates that H5 contains a methionine residue, (Sautiere *et al.*, 1975), and so histone H5 should have been detected amongst the translation products of the reticulocyte RNA preparations.

Several explanations are possible for the failure to identify H5 protein amongst the reticulocyte RNA translation products. Firstly, H5 protein may be synthesised in a precursor or unmodified form in the wheat germ cell-free translation system, and so may not electrophorese at exactly the same position on the gel as the mature H5 marker protein. For example H5 isolated from erythroid cells may be phosphorylated at two specific sites, (Tobin and Seligy, 1975), while the *in vitro* synthesised protein will probably not be phosphorylated. Such a difference in phosphorylation, however, is unlikely to significantly effect the migration of the protein on SDS gels. In the only other reports of detection of H5 protein as an *in vitro* translation product, (Knöchel, 1975; Scott and Wells, 1975;1976), the translated protein co-electrophoresed with the authentic H5 marker. Knöchel, (1975), translated reticulocyte RNA in the Ehrlich ascites cell-free

system, (Housman *et al.*, 1971), and analysed the products on low pH-urea tube gels, (Panyim and Chalkley, 1969). Scott and Wells, (1975; 1976), translated mRNA in the wheat-embryo cell-free translation system, (Shih and Kaesberg, 1973), and detected translation products which co-electrophoresed with H5 on low pH-urea tube gels, (Panyim and Chalkley, 1969), and SDS-urea tube gels, (Swank and Munkres, 1971). These *in vitro* translation results, analysed using low-resolution gel systems, were taken as evidence that H5 mRNA had been positively identified and purified, and also suggest that the wheat-germ cell-free system (Roberts and Paterson, 1973), will synthesise an H5 protein that co-electrophoreses with the authentic H5 marker.

Secondly, H5 mRNA may translate very badly, or not at all, in the wheat germ translation system, and so H5 protein will not be observed in the translation products. The results of Scott, (1975), however, show that when added to the wheat-embryo cell-free system, equal amounts of purified H5 mRNA and globin mRNA stimulated the incorporation of equal amounts of ^{3}H-leucine into the respective protein products. Since H5 contains only about half as many leucine residues as the globin protein chains, (see Table 3.1), this result indicates that H5 mRNA translates about twice as efficiently as globin mRNA in the wheat embryo cell-free system,

(in the absence of competing mRNA). Since globin mRNA itself has been shown to be an extremely efficient mRNA, these results demonstrate that H5 mRNA translates remarkably efficiently in cell-free systems, and suggest that H5 protein will be synthesised even when only small amounts of H5 mRNA are used to programme the wheat-germ cell-free system.

The third possible reason for the failure to observe H5 protein in the *in vitro* translation studies is that H5 mRNA is present at such low concentrations in the added RNA that, even with efficient translation, detectable amounts of H5 protein are not produced. This seems to be the most likely explanation of the results.

Knöchel, (1975), identified H5 protein as a component of the translation products of globin mRNA isolated from reticulocytes. The total translation products were extracted at high ionic strength and an acid pH to specifically select for histone H5, (and similar basic proteins), and the extracted proteins were identified on low pH-urea gels. While this study made no estimate of the amount of H5 mRNA present in the reticulocyte mRNA preparation, the data presented indicate that H5 protein comprises only a minute proportion of the total labelled translation products, (certainly much less than one percent). The amount of labelled H5 detected was extremely small even

though the labelled amino-acid used was ^{3}H-lysine, which will tend to overestimate the amount of H5 present relative to other proteins, (due to the extremely high lysine content of H5).

Scott and Wells, (1976), isolated microgram amounts of H5 mRNA from reticulocyte polysomes. If these yields had been quantitative, H5 mRNA would represent about one percent of the reticulocyte polysomal mRNA, but claims that the yields were extremely low, (Scott, 1975), suggest that H5 mRNA initially represented a much higher proportion of the total reticulocyte mRNA. The results of Knöchel, (1975), and the results presented in this chapter are not consistent with a concentration of H5 mRNA of more than a fraction of a percent of the total reticulocyte mRNA and therefore disagree with the estimate of about 4 percent of the polysomal mRNA made by Scott and Wells, (1975).

When reticulocyte 10S RNA was passaged through oligo-dT-cellulose, Scott, (1975), observed H5 protein only in the translation products of the RNA unbound to oligo-dT-cellulose, where it comprised about 10 percent of the labelled products. This result was verified by Scott on both the SDS-urea and the low pH-urea gel systems, however, it seems most unlikely that H5 mRNA could be present at anything approaching this level in the 10S RNA unbound to oligo-dT-cellulose, and remain undetected in the translation studies described in

this chapter and in chapter 3.

The results presented in chapter 3, (figure 3.2), show that the low pH-urea gel system is unreliable for the identification of H5 protein due to the presence of other labelled material electrophoresing at the H5 position. In addition, the results presented in this chapter show that the protein peak that co-electrophoreses with H5 marker on the SDS-urea tube gels does not co-electrophorese with marker H5 on higher resolution gels. It seems possible that in the oligo-dT-cellulose chromatography experiments, and perhaps in other experiments, Scott over-estimated the amounts of H5 mRNA present and the purity of the H5 mRNA preparations due to inadequacies in the gel assay systems used.

If H5 mRNA is in fact present at levels much lower than those estimated by Scott and Wells, (1975), it is doubtful whether any physico-chemical approach could achieve a useful purification of H5 mRNA. Even with 100 percent recovery at each step, very large amounts of reticulocyte RNA would have to be handled to isolate even the smallest practical quantities of H5 mRNA. If H5 mRNA is present at the very low levels suggested by the slab gel experiments, it is not surprising that no H5 mRNA sequences were detected in the ds cDNA recombinants prepared from

10S RNA. Despite this failure, however, recombinant DNA techniques probably provide the only practical means to the isolation of H5 mRNA sequences.

CHAPTER 10

GENERAL DISCUSSION

10. GENERAL DISCUSSION

The discovery that the original estimate of the concentration of H5 mRNA in reticulocyte RNA, (Scott and Wells, 1975), might be error was delayed by the wait for appropriate containment facilities to perform recombinant-DNA experiments. Once a population of recombinant molecules had been produced from reticulocyte 10S RNA, the failure to isolate a clone containing H5 mRNA sequences immediately suggested that at least one of the basic pieces of information about H5 mRNA, (Scott and Wells, 1976), was wrong. Examination of the translation products of reticulocyte RNA on high-resolution gels indicated that the peak of protein that had been identified as H5 protein in experiments using low resolution gels, (Scott, 1975), no longer electrophoresed at the position of the H5 marker. No translation product was detectable at the position of authentic H5 marker on the high-resolution gels. This chapter, therefore, consists of a retrospective discussion of the results of the previous chapters with particular reference to the information that the minor protein product observed when 10S RNA translation products are assayed on SDS-urea tube gels is not histone H5, and that H5 mRNA is probably present at only very low levels in reticulocyte RNA. (For the purposes of this discussion the proteins that together comprise the major, non-globin translation products of

10S reticulocyte RNA have been called, collectively, protein U, (for unknown)).

The results of the initial characterisation experiments, (chapter 3), showed that undegraded 10S RNA could be reproducibly isolated from chicken reticulocytes and translated efficiently in the wheat germ cell-free translation system. When assayed on SDS-urea tube gels, the translation products of 10S RNA showed the presence of just two protein peaks, one corresponding to globin protein and the other co-electrophoresing with authentic histone H5 marker. The low pH-urea tube gels gave a complex radioactivity profile that did not permit the ready identification of minor translation products. Scott, (1975), assigned the larger and smaller peaks of translated protein as globin and H5 respectively, because they co-electrophoresed with the appropriate marker proteins on both SDS-urea and low pH-urea tube gels and because the minor peak of protein represented about that proportion of the total translation products expected for H5, (based on cell-labelling studies, (Scott and Wells, 1975)). This identification of translation products was followed through all the early experiments described in this thesis.

In experiments designed to separate H5 mRNA from globin mRNA on the basis of poly-A content, protein-U was always found to be slightly enriched in the translation products of the non-polyadenylated

RNA, (as expected for genuine H5), but always to be present, in significant amounts, in the poly-adenylated RNA translation products. While this result apparently contradicted the observation, (Scott and Wells, 1976), that H5 mRNA lacked a poly-A tract, the failure of Scott and Wells to detect H5 mRNA in the fraction of RNA bound to oligo-dT-cellulose could be explained by the lack of sensitivity of the wheat-embryo translation assay used. Since at no stage in the affinity-chromatography experiments, (described in chapter 3), was authentic histone H5 detected in the translation products of the various RNA fractions, no conclusion can be drawn as to the localisation of H5 mRNA to either the polyadenylated or non-polyadenylated RNA fractions, or as to the degree of enrichment of H5 mRNA relative to globin mRNA in any of the fractions. The translation assay results do suggest, however, that the combination of affinity-chromatography techniques did achieve some enrich-ment for protein-U mRNA sequences.

When applied to RNA preparations from 5-day old chick embryos, (chapter 4), the translation assay system permitted the positive identification of histone mRNA sequences and the affinity-chromatography techniques achieved a clean fractionation of histone and globin mRNA sequences. cDNA prepared from the RNA fraction enriched for histone mRNA has since been used to select clones

containing the genomic histone genes, (Harvey and Wells, 1979).

In retrospect also, no conclusions concerning the distribution of H5 mRNA sequences can be drawn from the results of the experiments designed to separate globin mRNA and H5 mRNA on polyacrylamide gels, (chapter 5). Certainly the results show that it was not possible to fractionate the mRNA coding for protein-U from globin mRNA on these gels, even under conditions where rabbit α and β-globin mRNA were clearly resolved. It is no longer surprising that the ribonuclease H treatment of the mRNA did not enhance the resolution of the two mRNA species, and the result merely indicates that the protein-U mRNA and globin mRNA are polyadenylated to a similar extent. Because H5 was not detected in the translation assay it is not possible to show whether H5 mRNA was resolved from globin mRNA but, if H5 mRNA is indeed 10S in size, (Scott and Wells, 1976), the results obtained suggest that it would have been extremely difficult to separate globin and H5 mRNA sequences. Chicken globin mRNA has proved to be difficult material to work with, and, despite a number of attempts, (Knöchel and Grundmann, 1977; Longacre and Rutter, 1978), no successful fractionation of chicken globin mRNA into its component mRNA species has been reported.

In a series of experiments using recombinant-DNA techniques, a large number of ds cDNA clones were prepared from 10S reticulocyte RNA and screened

for H5 mRNA sequences, (chapters 6,7 and 8). The recombinants were characterised using hybridisation and DNA-sequence criteria, and so the identification of clones was independent of errors made using the translation assay. However, the RNA used as template for the preparation of the ds cDNA had been selected because it was enriched for H5 mRNA sequences, on the basis of the translation assay, and so, unfortunately, the recombinant-DNA experiments were not totally independent of the translation assay errors. The RNA fraction used as template was enriched for protein-U coding sequences and had formerly been unbound to oligo-dT-cellulose and bound to poly-U-Sepharose, (see section 3.2.5). If authentic H5 mRNA is non-polyadenylated, (Scott and Wells, 1976), then H5 mRNA would have been largely eliminated when the RNA was bound to poly-U-Sepharose. It is interesting to note however, that despite the affinity-chromatography procedures employed, a number of ribosomal RNA sequences were detected amongst the recombinants, demonstrating that non-polyadenylated sequences were cloned. If H5 mRNA sequences had been present as a sizeable percentage of the poly-A-minus RNA they may still have been detected amongst the non-globin, non-ribosomal clones.

Technically, the recombinant-DNA experiments were very successful. Large numbers of ds cDNA clones were prepared at high efficiency and the

selection procedure employed rapidly eliminated those clones containing ribosomal RNA and globin mRNA sequences from further consideration. The four clones examined by DNA-sequencing were found to contain sequences other than H5 mRNA sequences. While these cloned sequences have not been associated with a specific protein product, and may not even be coding sequences, it is possible that they contain mRNA sequences that code for one of the proteins that comprise the minor translation peak observed on tube gels, (protein-U). This proposition could be tested using hybrid-arrested translation procedures, (Paterson *et al.*, 1977). If the recombinants could be shown to contain sequences coding for an erythroid-cell specific product they could be of considerable interest in gene-control studies.

While the cloning experiments did not achieve the goal of isolating a specific H5 mRNA sequence, it was the results of these experiments that suggested that an error had been made in the interpretation of previous results. The use of the high resolution slab gel assay system to analyse the reticulocyte RNA translation products showed that the minor translation product of 10S RNA, previously interpreted as H5 protein, did not coelectrophorese with authentic H5 marker. Further, these experiments suggested that H5 mRNA was present at <u>very low</u> concentrations in 10S RNA. Despite their incomplete

nature these experiments suggest that Scott and Wells, (1976), may have over-estimated the amount and purity of H5 mRNA isolated from chicken reticulocyte polysomes by the indirect immuno-precipitation procedure.

Even if H5 mRNA is present in reticulocyte RNA at only a very low level it may still be possible to use recombinant DNA techniques to prepare clones containing H5 mRNA sequences.

Seeburg *et al.*, (1977b), and Shine *et al.*, (1977), have shown that it is possible to select specific fragments from restriction digests of complex mixtures of ds cDNA and to clone these fragments directly. Provided that a suitable restriction fragment can be found, this method should be applicable to mRNA sequences, such as H5 mRNA, which comprise as little as one percent of the original mRNA population.

Alternatively, deoxyoligonucleotide primers have been used to selectively prime the synthesis of cDNA from specific mRNA molecules, (for example pig gastrin mRNA, (Noyes *et al.*, 1979), and rat insulin mRNAs, (Chan *et al.*, 1979)), where the mRNA comprises only a small percentage of the total mRNA population. (Gastrin mRNA represents only 0.6 percent of the polyadenylated mRNA fraction). If the amino-acid sequence of a protein is known, oligonucleotide primers specific to a portion of the corresponding mRNA nucleotide sequence, may be

prepared. cDNA can be specifically synthesised from these mRNA molecules and the sequences subsequently inserted into recombinant DNA vectors. Since most of the amino-acid sequence of chicken H5 protein is known, it may be possible to synthesise a primer specific for H5 mRNA. Using this primer double-stranded cDNA containing H5 mRNA sequences could be prepared and cloned.

The fascinating question of the evolutionary relationship of H5 and Hl histones, and of the physical association of the H5 gene and the other histone genes justifies further attempts to isolate a clone containing H5 mRNA sequences.

CHAPTER 11

POSTSCRIPT

11. POSTSCRIPT

Recent results obtained by Molgaard *et al.*, (1980), are very relevant to the discussion of the work presented in this thesis. Using a procedure similar to that used by Scott and Wells, (1976), involving the indirect immunoprecipitation of polysomes, Molgaard *et al.*, have been able to isolate a fraction of chicken reticulocyte RNA enriched for H5 mRNA. Further studies of this RNA fraction indicate that H5 mRNA is polyadenylated, that H5 mRNA, (including poly-A tail), is about 1000 bases long and that H5 mRNA accounts for no more than 0.6 percent of reticulocyte mRNA activity. Similar to the results presented in chapter 9 of this thesis, Molgaard *et al.*, were unable to detect H5 protein in the translation products of unenriched reticulocyte mRNA.

The findings of Molgaard *et al.*, concerning H5 mRNA, therefore contradict those of Scott and Wells, who found H5 mRNA to be non-polyadenylated, 10S in size, (about 650 bases long), and to be present at about 3 to 4 percent of the total mRNA activity.

Since the results presented in this thesis do not confirm any of the data of Scott and Wells concerning H5 mRNA, it seems that the results of Molgaard *et al.*, are more likely to be correct.

REFERENCES

Abelson, J. (1979) *Ann. Rev. Biochem.* 48, 1035-69

Adesnick, M. and Darnell, J.E. (1972) *J. Mol. Biol.* 67, 397-406

Adler, A.J., Schaffhauser, R., Langan, T.A. and Fasman, G.D. (1971) *Biochemistry* 10, 909

Adler, K., Beyreutler, K., Fanning, E., Geisler, N., Gronenborn, B., Klamm, A., Muller-Hill, B., Pfahl, M. and Schmidt, A. (1972) *Nature* 237, 322

Alt, F.W., Kellems, R.E., Bertino, J.R. and Schimke, R.T. (1978) *J. Biol. Chem.* 253, 1357-1370

Appels, R. and Wells, J.R.E. (1972) *J. Mol. Biol.* 70, 425

Attardi, G., Parnas, H. and Attardi, B. (1970) *Exptl. Cell Res.* 62, 11

Aviles, F.J., Chapman, G.E., Kneale, G.G., Crane-Robinson, C. and Bradbury, E.M. (1978) *Eur. J. Biochem.* 88, 363-71

Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 1408

Bantle, J.A., Maxwell, I.H. and Hahn, W.E. (1976) *Anal. Biochem.* 72, 413-27

Bard, E., Efron, D., Marcus, A. and Perry, R.P. (1974) *Cell* 1, 101

Berkower, I., Leis, J. and Hurwitz, J. (1973) *J. Biol. Chem.* 248, 5914-5921

Bernard, O., Hozumi, N. and Tonegawa, S. (1978) *Cell* 15, 1133

Bishop, J.O., Pemberton, R. and Baglioni, C. (1972) *Nature New Biol.* 235, 231

Blake, C.C.F. (1979) *Nature* 277, 598

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L. Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) *Science* 196, 161-169

Bodenstein, D. (1943) *Biol. Bull.* 84, 34-58

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heymeker, H.L. and Boyer, H.W. (1977) *Gene* 2, 95-113

Bolund, L., Ringertz, N.R., Harris, H. (1969) *J. Cell Sci.* 4, 71

Bonner, W.M. and Laskey, R.A. (1974) *Eur. J. Biochem.* 46, 83-88

Borun, T.W., Ajiro, K., Zweidler, A., Dolby, T.W., Stephens, R.E. (1977) *J. Biol. Chem.* 252, 173-80

Bradbury, E.M., Carpenter, B.G. and Rattle, H.W.E. (1973) *Nature* 241, 123

Bradbury, E.M., Cary, P.D., Crane-Robinson, C. and Rattle, H.W.E. (1974a) *Ann. New York Acad. Sci.* 222, 266

Bradbury, E.M., Inglis, R.J., Matthews, H.R. and Langan, T.A. (1974b) *Nature* 249, 553

Brack, C. and Tonegawa, S. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5652-5656

Brasch, K., Seligy, V.L., Setterfield, G. (1971) *Exptl. Cell Res.* 65, 61

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

Breathnach, R., Mandel, J.L. and Chambon, P. (1977) *Nature* 270, 314

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

Britten, R.J. and Davidson, E.H. (1969) *Science* 165, 349-357

Brownlee, G.G. and Cartwright, E.M. (1975) *Nucleic Acids Res.* 2, 2279-2288

Bruschi, S. (1979) *Honours Thesis: University of Adelaide*

Burckhardt, J. and Birnstiel, M.L. (1978) *J. Mol. Biol.* 18 61-79

Busslinger, M., Portmann, R. and Birnstiel, M. (1979) *Nucleic Acids Research* 6, 2997-3008

Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) *Cell* 16, 739-751

Catterall, J.F., O'Malley, B.W., Robertson, M.A., Staden, R., Tanaka, Y. and Brownlee, G.G. (1978) *Nature* 275, 510

Chambon, P. (1975) *Ann. Rev. Biochem.* 44, 613-638

Champagne, M., Mazen, A. and Wilhelm, C. (1968) *Bull. Soc. Chim. Biol.* 50, 1261-72

Chan, S.J., Noyes, B.E., Agarwal, K.L. and Steiner, D.F. (1979) *Proc. Natl. Acad. Sci. USA* 76, 5036-5040

Chandler, M., Kedes, L.H. and Yunis, J. (1978) *Am. Soc. Hum. Genet. P.* 76A (Abstract)

Childs, G., Levy, S. and Kedes, L.H. (1979) *Biochemistry* 18, 208-213

Clewell, D.B. (1972) *J. Bacteriol.* 110, 667-676

Cohn, R.H., Lowry, J.C. and Kedes, L.H. (1976) *Cell* 9, 147-61

Cole, R.D. (1977) in *The Molecular Biology of the Mammalian Genetic Apparatus*, Vol. 1, ed. P.O.P. Ts'O pp 93-104, Amsterdam; North-Holland

Crawford, R. (1977) *PhD Thesis; University of Adelaide*

Crawford, R. and Wells, J.R.E. (1978) *Biochemistry* 17, 1591-1596

Crawford, R.J., Krieg, P., Harvey, R.P., Hewish, D.A. and Wells, J.R.E. (1979) *Nature* 279, 132-136

Cummings, I., Liu, A. and Salser, W. (1978) *Nature* 276, 418-419

Daneholt, B. and Hosick, H. (1974) *Cold Spring Harbor Symp Quant. Biol.* 38, 629

Darnell, J.E., Wall, R. and Tushinski, R.J. (1971) *Proc. Natl. Acad. Sci. USA* 68, 1321

Davidson, E.H. and Hough, B.R. (1971) *J. Mol. Biol.* 56, 491

Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973) *J. Mol. Biol.* 77, 1

De Robertis, E.M. and Olson, M.V. (1979) *Nature* 278, 137-143

Destree, O.H.J., Haenni, A. and Birnstiel, M.L. (1977) *Nucleic Acids Res.* 4, 801-11

Dugaiczyk, A., Woo, S.L.C., Lai, E.C., Mace, M.L., McReynolds, L.A. and O'Malley, B.W. (1978) *Nature* 274, 328

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

Efron, D. and Marcus, A. (1973) *FEBS Letts* 33, 23

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

Efstratiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T. (1976) *Cell* 7, 279-288

Engel, J.D. and Dodgson, J.B. (1978) *J. Biol. Chem.* 253, 8239-8246

Englund, P.T. (1979) *Methods in Enzymology, (in press).*

Finnegan, D.J., Rubin, G.M., Young, M.W. and Hogness, D.S. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1053-1063

Freigan, N., Marchionni, M. and Kedes, L.H. (1976) *Bio. Bull.* 151, 407 (Abstract)

Gall, J.G. (1968) *Proc. Natl. Acad. Sci. USA* 60, 553

Gallwitz, D. (1975) *Nature* 259, 242

Gallwitz, D. and Mueller, G.C. (1969) *J. Biol. Chem.* 244 5947-52

Gallwitz, D., Bos, E. and Stahl, H. (1978) *Methods in Cell Biol.* 19, 197-213

Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) *Nature* 278, 428-434

Garcia-Bellido, A. (1966) *Devl. Biol.* 14, 278-306

Glover, C.L.C. and Gorovsky, M.A. (1978) *Biochemistry* 17, 5705-5713

Glover, D.M. and Hogness, D.S. (1977) *Cell* 10, 167-176

Goodman, H.M., Olson, M.V., Hall, B.D. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5453-7

Greenaway, P.J. and Murray, K. (1971) *Nature New. Biol.* 229, 233-238

Gross, K., Probst, E., Schaffner, W. and Birnstiel, M.L. (1976a) *Cell* 8, 455-469

Gross, K., Schaffner, W., Telford, J. and Birnstiel, M.L. (1976b) *Cell* 8, 479-484

Grossbach, U. (1973) *Cold Spring Harbor Symp. Quant. Biol.* 38, 619

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

Grunstein, M.L. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3961-3965

Grunstein, M., Schedl, P. and Kedes, L.H. (1976) *J. Mol. Biol.* 104, 351-69

Grunstein, M. and Grunstein, J. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1083-92

Gurdon, J.B. (1962) *Dev. Biol.* 4, 256

Hackett, P.B., Traub, P. and Gallwitz, D. (1978) *J. Mol. Biol.* 126, 649

Hadorn, E. (1963) *Develop. Biol.* 7, 617-629

Hagenbüchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) *Cell* 13, 551-563

Hall, J., Gelbart, W. and Kankel, D. (1973) *Biology of Drosophila:* ed. Noritski, E., Ashburner, M. New York: Academic

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

Harris, H. (1967) *J. Cell Sci.* 2, 23

Harris, S.E., Rosen, J.M., Means, A.R. and O'Malley, B.W. (1975) *Biochemistry* 14, 2072

Harrison, P.R., Birnie, G.D., Hell, A., Humphries, S., Young, B.D. and Paul, J. (1974) *J. Mol. Biol.* 84, 539

Harvey, R.P. and Wells, J.R.E. (1979) *Nuc. Acids Res.* 7, 1787-98

Hayward, G.S. (1972) *Virology* 49, 342

Hillman, N., Sherman, M.I. and Graham, C. (1972) *J. Embryol. Exp. Morphol.* 28, 263

Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1929-1933

Hirsch, D. and Vanderslice, R. (1976) *Dev. Biol.* 49, 220-235

Hourcade, D., Dressler, D. and Wolfson, J. (1974) *Cold Spring Harbor Symp. Quant. Biol.* 38, 537

Housman, D., Pemberton, R. and Taber, R. (1971) *Proc. Natl. Acad. Sci. USA* 68, 2716-2719

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

Humphries, P., Cochet, M., Krust, A., Gerlinger, P., Kourilsky, P. and Chambon, P. (1977) *Nucleic Acids Res.* 4, 2389-2406

Humphries, P., Old, R., Coggins, L.W., McShane, T., Watson, C. and Paul, J. (1978) *Nucleic Acids Res.* 5, 905-924

Hunt, J.A. (1974) *Biochem. J.* 138, 499

Illmensee, K. (1972) *Wilhelm Roux' Archiv.* 170, 267

Illmensee, K. and Mahowald, A.P. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1016

Illmensee, K. and Mahowald, A.P. (1976) *Exp. Cell Res.* 97, 127

Ingram, V.M. (1974) *Annals of New York Academy of Science* 241, 93-98

Jackson, D.A., Symons, R.H. and Berg, P. (1972) *Proc. Natl. Acad. Sci. USA* 69, 2904-2909

Jacob, E. (1976) *Eur. J. Biochem.* 65, 275-84

Jacob, E., Malacinski, G., Birnstiel, M.L. (1976) *Eur. J. Biochem.* 69, 45-54

Jacob, F. and Monod, J. (1961) *J. Mol. Biol.* 3, 318-356

Jeffreys, A.J. and Flavell, R.A. (1977) *Cell* 12, 1097-1108

Johns, E.W. (1971) in *Histones and Nucleohistones* (Ed. D.M.P. Phillips), Plenum Press, London

Kates, J. (1970) *Cold Spring Harbor Symp. Quant. Biol.* 35, 743

Kates, J. (1973) *Methods in Cell Biology* 7, 53

Kedes, L.H. (1976) *Cell* 8, 321-31

Kedes, L.H. (1979) *Ann. Rev. Biochem.* 48, 837-70

Kedes, L.H. and Birnstiel, M.L. (1971) *Nature New Biol.* 230, 165-69

Kemp, D.J. (1975) *Nature* 254, 573-577

Khan, M.S.N. and Maden, B.E.H. (1977) *Nucleic Acid Res.* 4, 2495-2505

Klenow, H., Overgaard-Hansen, K. and Patkar, S.A. (1971) *Eur. J. Biochem.* 22, 371

Knöchel, W. (1975) *Biochim. Biophys, Acta.* 395, 501-508

Knöchel, W. and Grundmann, U. (1977) *Biochim. Biophys. Acta.* 476, 47

Kohne, D.E. and Byers, M.J. (1973) *Biochemistry* 12, 2373

Konkel, D.A., Tilghman, S.M. and Leder, P. (1978) *Cell* 15, 1125-1132

Kornberg, R.D. (1974) *Science* 184, 868-871

Kunkel, N.S., Hemminiki, K. and Weinberg, E.S. (1978) *Biochemistry* 17, 2591-98

Laemmli, U.K. (1970) *Nature* 227, 680-685

Laskey, R.A. and Gurdon, J.B. (1970) *Nature* 228, 1332

Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) *Cell* 15, 1157-1174

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

Lehrach, H., Frischauf, A., Hanahan, D., Wozney, J., Fuller, F., Crkvenjakov, R., Boedtker, H. and Doty, P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 5417-5421

Levenson, R. and Marcu, K. (1976) *Cell* 9, 311-22

Levy, C.C., Schmukler, M., Frank, J., Karpetsky, T., Jewett, B., Hieter, P., LeGendre, S. and Dorr, R. (1975) *Nature* 256, 340

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

Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1047-51

Lillehaug, J.R., Kleppe, R.K. and Kleppe, K. (1976) *Biochemistry* 15, 1858

Lindberg, U. and Persson, T. (1972) *Eur. J. Biochem.* 31, 246

Lobban, P.E. and Kaiser, A.D. (1973) *J. Mol. Biol.* 78, 453-471

Longacre, S. and Rutter, W.J. (1977) *J. Biol. Chem.* 252, 2742-2752

Macleod, A.R., Wong, N.C.W. and Dixon, G.H. (1977) *Eur. J. Biochem.* 78, 281-91

Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) *Proc. Natl. Acad. Sci. USA* 72, 1184-1188

Maniatis, T., Sim, G.K., Efstratiadis, A. and Kafatos, F.C. (1976) *Cell* 8, 163-182

Marcu, K. and Dudock, B. (1974) *Nucleic Acids Res.* 1, 1385-1397

Matsuda, G., Takei, H., Wu, K.C. and Shiozawa, T. (1971) *Protein Research* 3, 173-174

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

Max, E., Seidman, J. and Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3450

Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564

Maxam, A.M. and Gilbert, W. (1979) *Methods in Enzymology*, (in press)

McClintock, B. (1956) *Cold Spring Harbor Symp. Quant. Biol.* 21, 197-216

McKenzie, S.L., Henikoff, S. and Meselson, M. (1975) *Proc. Natl. Acad. Sci. USA* 72, 1117

Melli, M., Spinelli, G. and Arnold, E. (1977a) *Cell* 12, 167-74

Melli, M., Spinelli, G., Wyssling, H. and Arnold, E. (1977b) *Cell* 11, 651-61

Miki, B.L.A. and Neelin, J.M. (1975) *Can. J. Biochem.* 53, 1158-69

Milcarek, C., Price, R. and Penman, S. (1974) *Cell* 3, 1

Miller, H.I., Konkel, D.A. and Leder, P. (1978) *Nature* 275, 772-774

Mintz, B. and Illmensee, K. (1975) *Proc. Natl. Acad. Sci. USA* 72, 3585-3589

Moav, B. and Nemer, M. (1971) *Biochemistry* 10, 881-88

Molgaard, H.V., Perucho, M. and Ruiz-Carrillo, A. (1980) *Nature* 283, 502-504

Moll, R. and Wintersberger, E. (1976) *Proc. Natl. Acad. Sci. USA* 73, 1863-67

Morata, G. and Lawrence, P.A. (1975) *Nature* 225, 614-617

Morata, G. and Garcia-Bellido, A. (1976) *Wilhelm Roux' Archiv.* 179, 125-143

Morrison, M.R., Brinkley, S.R., Garski, J. and Lingrel, J.B. (1974) *J. Biol. Chem.* 249, 5290-5295

Neelin, J.M., Callahan, P.X., Lamb, D.C. and Murray, K. (1964) *Can. J. Biochem.* 42, 1743-52

Nemer, M., Graham, M. and Dubroff, L.M. (1974) *J. Mol. Biol.* 89, 435

Nigan, V., Guerrier, P. and Monin, H. (1960) *Bull. Biol. Fr. Belg.* 94, 131-202

Noll, M. and Kornberg, R.D. (1977) *J. Mol. Biol.* 109, 393-404

Nöthinger, R. (1964) *Wilhelm Roux' Archiv.* 155, 269-301

Noyes, B.E. and Stark, G.R. (1975) *Cell* 5, 301-310

Noyes, B.E., Mevarech, M., Stein, R. and Agarwal, K. (1979) *Proc. Natl. Acad. Sci. USA* 76, 1770-1774

Nudel, U., Ramirez, F., Marks, P.A. and Banks, A. (1977) *J. Biol. Chem.* 252, 2182-2186

Ohlenbusch, H.H., Olivera, B.M., Tuan, D. and Davidson, N. (1967) *J. Mol. Biol.* 25, 299

Okada, M., Kleinman, I.A., Schneiderman, H.A. (1974) *Dev. Biol.* 39, 286

O'Malley, B.W., Roop, D.R., Lai, E.C., Nordstrom, J.L., Catterall, J.F., Swaneck, G.E., Colbert, D.A., Tsai, M.J., Dugaiczyk, A. and Woo, S.L.C. (1978) Presented at: *Laurentian Hormone Conference* August, 1978

Pace, N.R., Walker, T.A. and Pace, B. (1974) *J. Mol. Evol.* 3, 151-159

Palmiter, R.D. (1973) *J. Biol. Chem.* 248, 8260

Palmiter, R.D. (1974) *J. Biol. Chem.* 249, 6779-6787

Panyim, S. and Chalkley, R. (1969) *Biochemistry* 8, 3972

Paterson, B.M., Roberts, B.E. and Kuff, E.L. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4370-4374

Pelling, C. (1964) *Chromosoma* 15, 71

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

Pinder, J.C., Staynov, D.Z. and Gratzer, W.B. (1974) *Biochemistry* 13, 5373-5378

Popa, L.M. and Bosch, L. (1969) *Fed. Eur. Biochem. Soc. Lett.* 4, 143

Portmann, R., Schaffner, W. and Birnstiel, M.L. (1976) *Nature* 264, 31-34

Postlethwait, J.H. and Schneiderman, H.A. (1973) *Ann. Review of Genetics* 381-433

Powell, B.C., Kemp, D.J., Partington, G.A., Gibbs, P.E.M. and Rogers, G.E. (1976) *Biochem. Biophys. Res. Commun.* 68, 1263-71

Prescott, D.M. (1966) *J. Cell Biol.* 31, 1-9

Proudfoot, N.J. (1979) *Nature* 279, 376

Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214

Ptashne, K. and Cohen, S.N. (1975) *J. Bacteriol.* 122, 776-781

Rabbitts, T.H. (1976) *Nature* 260, 221-225

Richards, R., Shine, J., Ullrich, A., Wells, J.R.E. and Goodman, H.M. (1979) *Nucleic Acids Res.* 7, 1137

Richards, R. and Wells, J.R.E. (1980) *J. Biol. Chem.* (accepted for publication)

Ritossa, F.M., Atwood, K.C., Lindsley, D.L. and Spiegelman, S. (1966) *Natl. Cancer Inst., Monogr.* 23, 449

Robbins, E. and Borun, T.W. (1967) *Proc. Natl. Acad. Sci. USA* 57, 409-416

Roberts, B.E. and Paterson, B.M. (1973) *Proc. Natl. Acad. Sci. USA* 70, 2330-2334

Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. (1978) *Cell* 15, 671

Rougen, F. and Mach, B. (1976) *Proc. Natl. Acad. Sci. USA* 73, 3418-3422

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J.L., LeMeur, M., Bregegegre, F., Gannon, F., LePennee, J.P., Chambon, P. and Kourilsky, P. *Nature* 279, 125-132

Roychoudhury, R., Jay, E. and Wu, R. (1976) *Nucleic Acids Res.* 3, 863-877

Ruderman, J.V. and Pardue, M.L. (1978) *J. Biol. Chem.* 253, 2018-25

Sadgopal, A. and Kabat, D. (1969) *Biochem. Biophys. Acta* 190, 486

Saint, R.B. (1979) *PhD Thesis; University of Adelaide*

Sakano, H., Rogers, J.H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S. (1979) *Nature* 277, 627

Sanger, F. and Coulson, A.R. (1975) *J. Mol. Biol.* 94, 441

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-67

Sargent, T.D., Wu, J.R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A. and Bonner, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3256-3260

Sautiere, P., Kmiecik, D., Loy, O., Briand, G., Biserte, G., Garel, A. and Champagne, M. (1975) *FEBS Lett* 50, 200-203

Sautiere, P., Briand, G., Kmiecik, D., Loy, O., Biserte, G., Garel, A. and Champagne, M. (1976) *FEBS Lett* 63, 164-166

Schaffner, W., Gross, K., Telford, J. and Birnstiel, M. (1976) *Cell* 8, 471-78

Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) *Cell* 14, 655-71

Schmeckpeper, B.J., Cory, S. and Adams, J.M. (1974) *Mol. Biol. Rep.* 1, 355

Schmid, C.W. and Deininger, P.L. (1975) *Cell* 6, 345

Schmidt, D.A., Mazaitis, A.J., Kasai, R. and Bautz, E.K.F. (1970) *Nature* 225, 1012

Schuerch, A., Mitchell, W. and Joklik, W. (1975) *Anal. Biochem.* 65, 331-345

Schultz, G.A., Manes, C. and Hahn, W.E. (1973) *Biochem. Genet.* 9, 247

Schutz, G., Beato, M. and Feigelson, P. (1972) *Biochem. Biophys. Res. Commun.* 49, 680

Scott, A.C. (1975) *PhD Thesis; University of Adelaide*

Scott, A.C. and Wells, J.R.E. (1975) *Biochem. Biophys. Res. Commun.* 64, 448-55

Scott, A.C. and Wells, J.R.E. (1976) *Nature* 259, 635-638

Scott, A.C. and Wells, J.R.E. (1978) *Methods in Cell Biol.* 19, 257-272

Seeburg, P.H., Shine, J., Martial, J., Baxter, J.D. and Goodman, H.M. (1977a) *Cell* 12, 157-165

Seeburg, P.H., Shine, J., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977b) *Nature* 270, 486-494

Seidman, J.G., Edgell, M.H. and Leder, P. (1978) *Nature* 271, 582-585

Seligy, V., Roy, C., Dove, M. and Yaguchi, M. (1976) *Biochem. Biophys. Res. Commun.* 71, 196-202

Shapiro, D.J., Taylor, J.M., McKnight, G.S., Palacios, R., Gonzalez, C., Kiely, M.L. and Schimke, R.T. (1974) *J. Biol. Chem.* 249, 3665

Shih, D.S. and Kaesberg, P. (1973) *Proc. Natl. Acad. Sci. USA* 70, 1799

Shine, J., Seeburg, P.H., Martial, J.A., Baxter, J.D. and Goodman, H.M. (1977) *Nature* 270, 494-499

Smith, L.D. and Ecker, R.E. (1965) *Science* 150, 777

Smith, L.D. and Ecker, R.E. (1970) *Current Topics in Develop. Biol.* 5, 1

Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517

Stein, G., Stein, J., Shepard, E., Park, W. and Phillips, I. (1977) *Biochem. Biophys. Res. Commun.* 77, 245-52

Steiner, D.F., Kemmler, W., Tager, H.S. and Peterson, J.D. (1974) *Federation Proceedings* 33, 2105

Stephens, R.E., Pan, C., Ajiro, K., Dolby, T.W. and Borun, T.W. (1977) *J. Biol. Chem.* 252, 166-72

Stern, M.S. (1973) *Differentiation* 1, 407

Sullivan, D., Palacios, R., Stavnezer, J., Taylor, J.M., Faras, A.J., Kiely, M.L., Summers, N.M., Bishop, J.M. and Schimke, R.T. (1973) *J. Biol. Chem.* 248, 7530

Sullivan, N. and Roberts, W.K. (1973) *Biochemistry* 12, 2395

Sures, I., Lowry, J. and Kedes, L. (1978) *Cell* 15, 1033-44

Sutcliffe, J.G. (1978) *Nucleic Acids Res.* 5, 2721-2728

Suzuki, Y., Gage, L.P. and Brown, D.D. (1972) *J. Mol. Biol.* 70, 637-49

Swank, R.T. and Munkres, K.D. (1971) *Anal. Biochem.* 39, 462

Takei, H., Ota, Y., Wu, K.C., Kiyohara, T. and Matsuda, G. (1975) *J. Biochem.* 77, 1345-47

Taylor, J.M., Illmensee, R. and Summers, J. (1976) *Biochim. Biophys. Acta.* 442, 324-330

Thrall, C.L., Park, W.D., Rashba, H.W., Stein, J.L., Mans, R.J. and Stein, G.S. (1974) *Biochem. Biophys. Res. Commun.* 61, 1443-1449

Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V. and Leder, P. (1978a) *Proc. Natl. Acad. Sci. USA* 75, 725-729

Tilghman, S.M., Curtis, P.J., Tiemeier, D.C., Leder, P. and Weissman, C. (1978b) *Proc. Natl. Acad. Sci. USA* 75, 1309

Tobin, R.S. and Seligy, V.L. (1975) *J. Biol. Chem.* 250, 358

Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) *Proc. Natl. Acad. Sci.* 75, 1485

Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. and Goodman, H.M. (1977) *Science* 196, 1313-1319

Ursprung, H. (1967) *Methods in Developmental Biology* ed. F. Wilt, N. Wessels, 485-492 New York: Crowell

van den Berg, J., van Ooyen, A., Mantei, N., Shambock, A., Grosveld, G., Flavell, R.A., Weissmann, C. (1978) *Nature* 276, 37-44

Vedeckis, W.V., Schrader, W.T. and O'Malley, B.W. (1978) In: *Biochemical Action of Hormones* ed: J. Litwack, Academic Press NY 5, 321

Villa-Komaroff, L., Efstratiadis, S., Broome, S., Lomedico, P., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3727-3731

Vogt, V.M. (1973) *Eur. J. Biochem.* 33, 192

Vournakis, J.N., Gelinas, R.E. and Kafatos, F.C. (1974) *Cell* 3, 265-273

Vournakis, J.N., Efstratiadis, A. and Kafatos, F.C. (1975) *Proc. Natl. Acad. Sci. USA* 72, 2959

Wahl, G.M., Stern, M. and Stark, G.R. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3683-3687

White, R.L. and Hogness, D.S. (1977) *Cell* 10, 177-192

Williamson, R., Crossley, J. and Humphries, S. (1974) *Biochemistry* 13, 703

Wilson, E.B. (1925) in *"The Cell in Development and Heredity"*, 3rd edition, Macmillan, New York

Wilson, M.C. and Melli, M. (1977) *J. Mol. Biol.* 110, 511-35

Wolfe, J. (1979) *Nature* 282, 15

Wu, M., Holmes, D.S., Davidson, N., Cohn, R.H. and Kedes, L.H. (1976) *Cell* 9, 163-169

Yaguchi, M., Roy, C., Dove, M. and Seligy, V. (1977) *Biochem. Biophys. Res. Commun.* 76, 100

Yaguchi, M., Roy, C. and Seligy, V. (1979) *XI Congres International De Biochimie, Canada* (Abstract)

Yoshimori, R.N. (1971) *PhD Thesis; University of California, San Francisco*

Yu, L.C., Szabo, P., Borun, T.W. and Prensky, W. (1978) *Cold Spring Harbor Symp. Quant. Biol.* 42, 1101-5

Zweidler, A. and Cohen, T.H. (1972) *Fed. Proc.* 31, 926A