KERATIN mRNA STRUCTURE AND GENE ORGANIZATION

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

TREVOR JOHN LOCKETT B.Sc.(Hons.)

Department of Biochemistry,
University of Adelaide,
South Australia.

February, 1979.

Awarded July 1979

To Angela

For her distractions during our courtship and her support during our marriage.

TABLE OF CONTENTS

SUMMARY

1. The work described in this thesis is divided into two sections:-

(a) A study of the organization of sequences within purified feather keratin mRNA.

(b) The investigation of a density gradient system, employing RNA:DNA hybrids, for analysing the arrangement of tandemly linked families of genes.

2. Two keratin complementary DNAs (cDNAs) of different lengths were prepared, characterized and used to study the organization of sequences within polysomal chick feather keratin mRNA by the analysis of their reassociation kinetics, when annealed to a vast excess of chick erythrocyte DNA, and by the thermal stability of the duplexes formed at different stages during the reassociation reaction. This work has produced the following conclusions.

(a) Each keratin mRNA molecule carries a 150 nucleotide sequence adjacent to the 3' poly(A) tract, which is represented only once in the chick genome.

(b) A reiterated sequence, which is either short (about 50 bases) and faithfully conserved or longer and mismatching, is covalently attached to the 3' unique sequence and located further toward the 5' end of the mRNA.

3. The possibility of enriching high molecular weight

chick erythrocyte DNA samples for keratin sequences by physico-chemical methods for future use in the analysis of keratin gene organization was investigated. It was found that thermal elution from hydroxylapatite (HAP) in aqueous buffers could produce a 2.8-fold enrichment of keratin sequences and a very large enrichment for ribosomal sequences at the expense of DNA integrity. Similar elutions in the presence of formamide led to decreased DNA degradation and decreased enrichment for keratin (about 1.8-fold) and ribosomal (5-fold) sequences. Caesium chloride density gradient fractionation permitted a quantitative separation of ribosomal and keratin sequences with the possibility of further enrichment of keratin sequences from total chick DNA by antibiotic (actinomycin D and netropsin sulphate) CsCl gradients.

4. A method was developed for the preparative isolation of relatively homogeneous single-stranded DNA size classes from randomly sheared chick erythrocyte DNA using alkaline agarose gels for the size fractionation. A number of different procedures for extracting DNA from agarose gels were investigated with simple centrifugation of the agarose slice and collection of the supernatant proving to be the most effective.

5. A caesium sulphate density gradient system, employing pre-hybridization of RNA to single-stranded DNA, for the analysis of tandemly repeated gene families, was evaluated using the chick ribosomal cistrons as a model system. By using single-stranded DNA size classes in the pre-hybridization reaction and examining the relationship between the

density of the generated hybrid (relative to the densities of single-stranded ribosomal sequences and 100% hybrid) and the proportion of the ribosomal sequences expected to be in hybrid form, it was concluded that the method was insufficiently accurate to be generally applicable to the analysis of the arrangement of multi-gene families.

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.

Signed

Trevor J. Lockett

ACKNOWLEDGMENTS

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

I am much indebted to Professor G.E. Rogers for his enthusiastic supervision, advice and encouragement throughout the course of this work.

Thanks are also due to my colleagues in the feather project for their friendship and stimulating discussions - Messrs. R.B. Saint, B.C. Powell and C.P. Morris, who were present throughout the course of this project and Drs. D.J. Kemp and P.E.M. Gibbs who were present during the formative years. In particular, my thanks go to Dr. Kemp for his collaboration with some experiments described in Chapter III, and for permission to quote some of his results.

I am grateful to Mrs. Lesley Crocker for her excellent technical assistance and invaluable help in preparing and photographing diagrams. Thanks are also due to Miss P.Y. Dyer for assistance with electron microscopy and useful discussions.

I also thank Mrs. J. Somerville for typing the thesis.

I was supported throughout this work by a Commonwealth Postgraduate Research Award. The work was supported by grants from the Australian Research Grants Committee and the Australian Wool Corporation (to G.E. Rogers).

ABBREVIATIONS

The following abbreviations have been used in this thesis.

B.S.A.:	bovine serum albumin.
C:	curie.
cDNA:	synthetic DNA complementary to messenger RNA.
CsCl:	caesium chloride.
Cs_2SO_4:	caesium sulphate.
DNA:	deoxyribonucleic acid.
EDTA:	ethylenediaminetetracetic acid.
HAP:	hydroxylapatite.
hnRNA:	heterogeneous nuclear RNA.
kb:	kilobase (RNA or single-stranded DNA) kilobase pairs (Native DNA).
mRNA:	messenger ribonucleic acid.
oligo dT:	oligodeoxythymidylic acid.
PolI:	E. coli DNA polymerase I.
Poly(A):	polyadenylic acid.
RNA:	ribonucleic acid.
RNA'ase:	ribonuclease.
mRNP:	messenger ribonucleoprotein complex.
rRNA:	ribosomal RNA.
SDS:	sodium dodecyl sulphate.
SSC:	standard saline citrate (0.15 M NaCl, 0.015 M Na citrate).
SSKI:	saturated solution of potassium iodide.
TCA:	trichloroacetic acid.
T.E.:	10 mM Tris-HCl pH 7.5, 1 mM EDTA.
T_m:	temperature at which half the molecules in

	the DNA duplex or RNA:DNA hybrid under study are denatured.
tRNA:	transfer RNA.

CHAPTER I

INTRODUCTION

CHAPTER I

INTRODUCTION

A. GENERAL INTRODUCTION

The embryonic chick down feather constitutes an example of a terminally differentiating tissue. After an initial period of growth, the cells become filled with the fibrous protein, keratin, and eventually die. The general aim of the work in this laboratory is to gain an understanding of the molecular events leading to a feather cell's commitment to keratin production. The work described in this thesis falls into two parts. The first has been to study the nature of different regions of the keratin mRNA molecule. The second has been directed towards establishing a density gradient method for analysing gene organization with the intention of applying it to a study of the arrangement of keratin genes in the chick genome.

The ensuing literature survey is intended to provide relevant background information on feather development in addition to information on current views concerning mRNA structure and eukaryote gene organization.

B. FEATHER DEVELOPMENT

1. Keratins

Keratins are fibrous, insoluble, intracellular proteins which have a high cystine content. They are found in the epidermis and tissues derived from it, e.g., the feathers, claws, beaks and scales in birds and the claws and hair in animals. Within the cells, the proteins are

aggregated into fibrillar masses, the structure of which has been extensively studied (see Fraser et al., 1972, for review). Owing to the stabilization of these protein aggregates by disulphide bonding (Goddard and Michaelis, 1934), the study of the proteins requires disulphide bond breakage by reducing agents, followed by stabilization by alkylation, e.g., carboxymethylation (Harrap and Woods, 1964a; Kemp and Rogers, 1972).

S-carboxymethylated chick feather keratins appear to be homogeneous in molecular weight, the estimates varying between 10,500 and 11,500 daltons (Harrap and Woods, 1964b; Walker and Rogers, 1976a), while reduced, non-alkylated keratin had a molecular weight of 10,000 daltons (Woodin, 1954). Despite this molecular weight homogeneity, the proteins appear to have different charge characteristics, as determined by ion exchange chromatography and electrophoresis, even when the material is obtained from one particular keratinized tissue (Harrap and Woods, 1964a; Kemp and Rogers, 1972; Walker and Rogers, 1976a). Some ten electrophoretically distinguishable protein species from the rachis component of adult feathers were resolved by Woods (1971), while Walker and Rogers (1976a) have found a minimum of twenty-two protein chains in the embryonic feather. Sequence analysis of the proteins has suggested that each chain is the product of a separate gene, but that all the proteins are closely related in primary structure (Walker and Rogers, 1976b).

Although embryonic feathers, and the specific parts of the adult feather (Rawles, 1965), contain an assembly of

closely related protein chains, each tissue has its own distinct population of protein species (Harrap and Woods, 1964a; Kemp and Rogers, 1972). Furthermore, the proteins of feather exhibit fundamental differences from those of other keratinized tissues, such as scale (Kemp and Rogers, 1972; Walker and Bridgen, 1976).

Despite these differences, avian keratins appear to have evolved from common ancestral proteins. The spectrum of proteins observed in different species of birds reveal similar degrees of heterogeneity and tissue specificity, although the patterns for a given tissue show some inter-species variation (Harrap and Woods, 1967; Woods, 1971; O'Donnell, 1973a). Primary structure analysis of chains from the adult feather calamus of emu (O'Donnell, 1973b) and gull (O'Donnell and Inglis, 1974), from chick embryonic feather (Walker and Rogers, 1976b) and from chick scale (Walker and Bridgen, 1976) illustrates the presence of major regions of sequence homology between the proteins from different tissues.

2. Embryonic Feather Differentiation

(a) Development and keratinization

The cellular events involved in embryonic feather morphogenesis have been extensively studied and detailed reviews can be found elsewhere (Watterson, 1942; Romanoff, 1960; Lillie, 1965; Voitkevich, 1966; Matulionis, 1970). Since keratinization is the manifestation of terminal differentiation, this survey will only deal with events occurring during and sub-

sequent to the onset of keratin synthesis.

Keratin synthesis begins at about the twelfth day of embryonic development, as judged by the appearance of keratin fibrils in the cells (Matulionis, 1970) and, as shown in Figure 1.1, by the appearance of keratin proteins on acrylamide gels of S-carboxymethylated proteins extracted from feathers of different ages. Figure 1.1 shows that bands corresponding to the major feather keratin proteins were only present in trace amounts in 11- and 12-day feather extracts. These bands rapidly increased in quantity after day 12 and had become the most abundant protein species in the feathers by day 15 (Kemp et al., 1974a). It was also observed, from labelling studies and gel electrophoresis of proteins, that the keratin proteins are co-ordinately synthesized in the tissue (Kemp et al., 1974a). Keratinization of the feather is completed by about the eighteenth day of embryonic development and the chick hatches at 21 days.

(b) Factors affecting keratinocyte development

A number of factors have been shown to affect the growth and development of keratinizing tissues and these are briefly discussed below. A more extensive review may be found in Fraser et al., (1972).

(i) The role of the dermis

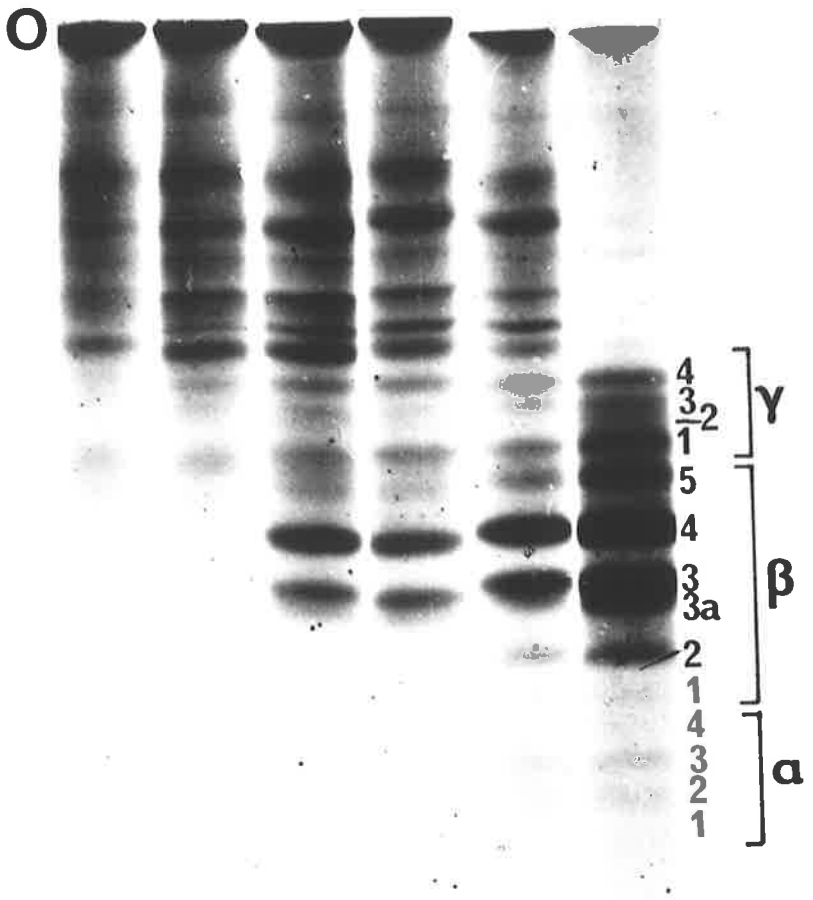
When the dermis and epidermis are taken from different regions of an embryo,

FIGURE 1.1.

TIME COURSE OF KERATIN SYNTHESIS

Samples (100 μg/gel) of the reduced and carboxymethylated protein preparations from feathers at 11-19 days of embryonic feather development were subjected to polyacrylamide gel electrophoresis at pH 9.5 and stained with Coomassie blue. O, origin; + anode; α, β and γ bands, keratin.

Reproduced from Kemp *et al.*, 1974a.



+

11 12 13 14 15 19

DAYS

recombined and allowed to develop in tissue culture, the epidermis develops structures characteristic of those normally associated with the dermis used (Sengel, 1957, 1971; Rawles, 1963, 1965; Wessells, 1962, 1965). Furthermore, the keratins produced in these recombinants are identical to those found in the same epidermal structures where no recombination has been performed (Dhouailly et al., 1978). Thus the recombination of presumptive scale dermis with presumptive feather epidermis causes the epidermis to form scales, and the keratins present in these scales are the same as those present in the normal tissue. These effects, exerted by the dermis, only occur at specific stages of development (Rawles, 1963, 1965) which are much earlier than the onset of keratinization in these tissues (Matulionis, 1970; Kemp et al., 1974a; Beckingham Smith, 1973a).

While the major role of the dermis appears to be to control the mitosis of epidermal cells, thereby controlling epidermal morphology (Wessells, 1962; Dodson, 1963), it is clear that it also specifies the genes which are being expressed in the differentiating tissue. It would appear that the control by the dermis over the epidermis is most likely mediated by some extracellular substance from the mesoderm influencing the environment of epidermal cells, since Wessells (1962) has shown

that isolated epidermis would grow even if a millipore filter was interposed between the dermis and epidermis prior to culturing the whole assembly *in vitro*. In studies using heterospecific dermal-edpidermal recombinants, Dhouailly (1967) has shown that duck dermis can direct chicken epidermis to produce feathers resembling those of the duck. In similar experiments (Dhouailly *et al.*, 1978), mouse footpad dermis was shown to direct the production of a footpad structure in presumptive feather producing chick epidermis. The keratins of these footpad structures were shown to be of the scale type. These results suggest that the effector of dermal control over the epidermis is not species specific, and even in heterospecific recombinants, it can dictate the type of gene to be expressed.

(ii) Vitamin A

It has been shown that if immature (Fell and Mellanby, 1953) or highly differentiated (Fell, 1957) embryonic chick skin is cultured in the presence of excess Vitamin A, keratinization is inhibited and the cells undergo a mucous metaplasia. The synthesis of keratin proteins is completely repressed under these conditions (Beckingham Smith, 1973b). For the older tissue at least, the effects are reversible on the removal of vitamin A (Fitton-

Jackson and Fell, 1963).

(iii) Hormones

Several hormones have been implicated in the development of avian keratinocytes. The role of pituitary and thyroid glands has been discussed at some length by Voitkevich (1966), but few studies of these phenomena have been made at the molecular level.

In vivo (Bartels, 1943) and in vitro experiments (Wessells, 1961; Kitano and Kuroda, 1967) have indicated that thyroxine accelerates epidermal keratinization. In the light of this evidence, it is interesting to note that the thyroid gland reaches maximum thyroxine secretion (Shain et al., 1972) at about the same time as the onset of keratin synthesis.

Hydrocortisone also hastens keratin synthesis in skin cultures (Fell, 1962; Sugimoto and Endo, 1969) and causes feather germs to abort. The presence of hydrocortisone in the medium was found to stimulate the synthesis of one class of scale epidermal proteins, although another class is not synthesised (Sugimoto et al., 1974).

(iv) Epidermal growth factor

This polypeptide growth factor was

first isolated from the submaxilliary gland of the male mouse by Cohen (1962) and produced precocious epidermal development when administered to *in vivo* or *in vitro* organ culture (Cohen, 1965). The growth factor stimulated RNA and protein synthesis (Hoober and Cohen, 1967) and an accompanying increase in the proportion of the ribosomes associated with polysomes (Cohen and Stastny, 1968). The material isolated from mouse is effective against skin tissue from a variety of animal sources, including the chicken, and similar material has recently been isolated from man (Cohen and Carpenter, 1975), suggesting that it might be widely distributed amongst animals. An *in vivo* requirement for epidermal growth factor has not, however, been demonstrated.

(c) DNA synthesis and mitosis

From the above discussion it is apparent that many factors affect keratinization and that there can be no simple mechanism whereby keratin synthesis is initiated in feather or skin cells. It is noteworthy that many of the factors discussed appear to affect mitosis, and that some effects on differentiation (or pseudo-differentiation in the case of vitamin A) are observed in post-mitotic cell populations. It has been shown that DNA synthesis in skin (Rothberg and Ekel, 1971) and DNA polymerase activity in feather (Kischer and Furlong, 1967) are at their maximum levels immediately prior to the onset of

keratinization in the tissues, and decrease significantly by the time keratin synthesis is established.

It is possible that the cells are committed to keratin synthesis early in development, as has been observed for myogenic cells (e.g., Holtzer et al., 1972, 1973). These cells, although not actively synthesizing muscle proteins, undergo several rounds of DNA synthesis and mitosis, until after a "quantal" division, the cells become fully differentiated. Depending on the environment in which they are located, epidermal cells show a certain degree of developmental variability and perturbations of that environment can cause the cells to follow a different developmental pathway. Nevertheless, the cells could be initially constrained to follow one of a limited number of fates, and the association with a particular mesenchyme could lead to the cells being finally committed to one path. Thus, with the appropriate dermal stimulus, the cells could undergo a programmed pattern of cell division, culminating in a "quantal" mitosis and keratinization. After the quantal division, the cells would be incapable of responding to external stimuli; however prior to it, major changes in environment could lead to a different set of genes being expressed, as, for example, in the case of the mucous metaplasia induced by vitamin A (Fell, 1957).

C. SOME CHARACTERISTICS OF FEATHER KERATIN mRNA

The isolation of feather keratin mRNA (Partington et al., 1973; Kemp et al., 1974b) has greatly aided studies on

feather keratin synthesis. The mRNA is isolated from 14-day embryonic feathers and, when purified, sediments at about 12S in sucrose gradients (Partington et al., 1973; Kemp et al., 1974b). By the comparison of the mobilities of feather keratin mRNA to those of 28S and 18S rRNA, 5S RNA and tRNA on acrylamide gels in the presence of 98% formamide, the molecular weight of the mRNA was calculated to be 2.5×10^5 daltons or 760 nucleotides (Kemp et al., 1974b). The mRNA stimulates keratin synthesis in both the rabbit reticulocyte (Partington et al., 1973) and wheat embryo cell free translation systems (Kemp et al., 1974c). In the wheat embryo system, the only proteins synthesized were keratins. From these translation studies and the observation that purified keratin mRNA electrophoresed as a single band on formamide gel electrophoresis (Kemp et al., 1974b), the mRNA preparations were judged to be about 95% pure.

Keratin mRNA was found to bind to cellulose (Partington et al., 1973; Kemp et al., 1974b) suggesting that it contains a 3' poly(A) tract (Kitos et al., 1972; Schultz et al., 1972; DeLarco and Guroff, 1973). This is a feature common to most other messenger RNAs studied (see Brawermann (1974) for review) with the notable exception of histone mRNA (Adesnik and Darnell, 1972). The average length of the poly(A) tract at the 3' end of feather keratin mRNA is 60 nucleotides (Morris and Rogers, in press).

Many animal (Adams and Cory, 1975; Furuichi et al., 1975a; Perry et al., 1975) and viral mRNAs (Furuichi et al., 1975b; Wei and Moss, 1975) have been shown to possess a structure having the sequence $m^7G^{5'}ppp^{5'}X^mpY^{(m)}pZp\text{---}$,

where m^7G is 7-methylguanosine, which is coupled via a triphosphate linkage to the next nucleotide in the RNA chain (X), which bears a 2'-O-methyl group. The third nucleotide (Y) is also occasionally methylated at the 2'-position. Evidence has recently been obtained suggesting the presence of 7-methylguanosine at the 5' terminus of keratin mRNA coupled by a triphosphate linkage to the penultimate residue (Morris and Rogers, in press). The sequence of the adjacent residues could not be determined, as it was necessary to label the mRNA in vitro in order to detect the 5'-terminal structure (c.f. Symons, 1975).

The hybridization kinetics of keratin mRNA with cDNA, prepared using the RNA dependent DNA polymerase of avian myeloblastosis virus, indicated that keratin mRNA is a heterogeneous mixture of species (Kemp, 1975). From this hybridization analysis it was estimated that there were 25 to 35 keratin mRNA species, an estimate consistent with the number of protein chains in the tissue (Walker and Rogers, 1976a). Each mRNA species appears to contain a sequence which is unique to it, and a sequence homologous to the other mRNAs in the mixture. Since the mRNA is about 760 bases long and feather keratin chains are only 100 amino acids in length, only 300 of the 760 bases can be attributed to the coding sequence, the rest presumably being untranslated sequences. Early reassociation kinetic analysis suggested that there are between 100 and 240 genes coding for keratin in the chick genome (Kemp, 1975).

Keratin mRNA, therefore, shows many features common to most eukaryotic mRNAs. However, the sequence complexity

indicates that keratin mRNA is a mixture of closely related sequences, presumably exhibiting equivalent heterogeneity to that observed in the keratin proteins.

D. STRUCTURE OF EUKARYOTIC mRNAs AND THEIR BIOGENESIS

The following discussion focuses attention on the relationship between the coding and untranslated sequences of a number of eukaryotic mRNA species and the mechanism by which the final mRNA species are generated.

1. mRNA Structure

All eukaryotic mRNAs which have been characterized in detail are known to contain more bases than is necessary for coding, in addition to poly(A) (Gould and Hamlyn, 1973; Berns et al., 1972; Brownlee et al., 1973; Milstein et al., 1974; McReynolds et al., 1978). With the advent of rapid DNA sequencing techniques (Maxam and Gilbert, 1977; Sanger and Coulson, 1975; Brownlee and Cartwright, 1977) and cDNA cloning (Maniatis et al., 1976), rapid progress has been made in the understanding of mRNA structures.

The most widely studied messengers have been those coding for globins. Rabbit β globin mRNA was the first mRNA to be completely sequenced (Baralle, 1977a; Proudfoot, 1976a, 1976b; Efstratiadis et al., 1977). Its length, not including poly(A), is 589 bases; the 3' non-coding sequence is 95 residues long (Proudfoot, 1976a) and the 5' non-coding sequence is 53 residues long (Baralle, 1977a). Human β globin has also been completely sequenced (Proudfoot, 1977; Marotta et al., 1977; Baralle, 1977c). Its length, minus poly(A), is 626 bases (Baralle, 1977c), 134

of these being attributable to the 3' non-coding sequence (Proudfoot, 1977), and 50 to the 5' non-coding sequence (Baralle, 1977c).

Comparison of the primary structure of rabbit and human β globin mRNA sequences reveals that, ignoring the 39 base insert which is present in the 3' non-coding region of the human, but not the rabbit, mRNA, the 3' non-coding sequences are 77% homologous (Proudfoot, 1977) and account for about $^{2}/_{3}$ of all the non-coding sequences in both of the mRNAs. The 5' non-coding region of rabbit β globin mRNA is 3 nucleotides longer than that for human but they show 80% homology (Baralle, 1977a; Baralle, 1977c). The homology of these non-coding sequences in these two mammals suggests that these sequences are under moderate selective pressure; less than those nucleotides specifying amino acid sequence but more than those whose function may be regarded as not being sequence specific, for example, satellite DNA (Southern, 1975a).

Large portions of human (Forget et al., 1974; Proudfoot and Longley, 1976; Proudfoot et al., 1977; Baralle, 1977c) and rabbit (Proudfoot, 1976a; Baralle, 1977b; Proudfoot et al., 1977) α globin mRNAs have also been sequenced. Their lengths, minus poly(A), are 575 and 551 nucleotides respectively. Human α globin mRNA has a 24 base insert in the 3' non-coding sequence which is not present in the rabbit messenger. Ignoring this, these sequences are 80% homologous (Proudfoot et al., 1977). The 5' non-coding sequences for human and rabbit α globin mRNA are 37 and 36 nucleotides respectively and exhibit 81%

homology (Baralle, 1977c). As for the β globins, about ¼ of the nucleotides of the α globin mRNAs are untranslated with $^{2}/_{3}$ of these being at the 3' end.

Chick ovalbumin mRNA has also been completely sequenced (McReynolds et al., 1978; Cheng et al., 1977; Brownlee and Cartwright, 1976) and shares some interesting structural similarities with the globin mRNAs. The messenger is 1859 residues long, not including poly(A) and 5' cap structures. This is made up of a 673 base 3' non-coding sequence, a 1158 nucleotide sequence coding for the ovalbumin protein, and a 5' untranslated sequence of 64 residues. It is interesting to note that the 5' non-coding sequences of human and rabbit α and β globin mRNAs and chick ovalbumin mRNA are all of similar length. There is, however, no sequence homology between the 5' untranslated sequences of ovalbumin and globin mRNAs (McReynolds et al., 1978). This lack of homology has led to the speculation that the 5' cap structure and the initiating AUG codon are the only important general protein synthesis initiating signals in higher organisms (Baralle and Brownlee, 1977). Many authors have also alluded to the possibility of important interactions between 18S rRNA and sequences in this 5' region similar to those proposed to occur between E. coli 16S rRNA and sequences at the 5' end of a number of procaryotic mRNAs (Shine and Dalgarno, 1975).

The consistent discovery of long untranslated 3' terminal sequences in many different mRNAs (McReynolds et al., 1978; Proudfoot et al., 1977; Proudfoot, 1977; Hamlyn et al., 1977) suggests that these sequences may have an

important role in regulating mRNA function (Proudfoot et al., 1977). Baralle (1977a) has proposed that the 5' and 3' non-coding sequences of rabbit β globin mRNA could interact via the 5' U-C-C-C-C 3' sequence between nucleotides 36 and 40 and the 5' G-G-G-G-A 3' sequence between residues 525-529. He suggests that, by such an interaction, the 3' non-coding sequence could be involved in the initiation of protein synthesis. Alternatively the 3' non-coding sequence might be important for the regulatory functions relating to mRNA processing from its presumptive precursor, transport across the nuclear membrane or stability (Perry, 1976). Certainly the conservation of the sequence 5' A-A-U-A-A-A 3' at a site about 20 nucleotides from the 3' poly(A) tract in rabbit (Proudfoot, 1976a) and human α and β globin mRNAs (Proudfoot et al., 1977; Proudfoot, 1977), mouse immunoglobulin mRNA (Milstein et al., 1974) and chick ovalbumin mRNA (Cheng et al., 1976) is suggestive of some important general function for at least this portion of the sequence. At present, however, the assignment of functions to different portions of the untranslated sequences in the different messengers studied can only be considered to be speculative. While it is often possible to draw secondary structures for the sequenced mRNAs on the basis of base pairing, the unknown effects of the proteins, associated with the mRNAs in vivo, on secondary structure, render it difficult to make any firm conclusions about the existence of such postulated structures.

2. RNA Processing

Several types of structural and kinetic evidence

indicate that heterogeneous nuclear RNA (hnRNA) contains mRNA precursors in a variety of forms (see Perry et al., 1976 for review). This section, however, will concentrate on recent results, obtained from restriction enzyme analysis and recombinant DNA technique, which demonstrate conclusively the existence of mRNA precursors and the intricacies of RNA processing.

The first strong evidence that mRNA biogenesis need not necessarily be a simple event, came from the adenovirus system, where transcripts from several parts of the adenovirus genome were found covalently linked together in a single mRNA molecule (Berget, et al., 1977; Chow et al., 1977). These results suggested that one or more of a number of mechanisms of mRNA biogenesis could be acting (Klessig, 1977).

(a) RNA could be transcribed from each site and subsequently ligated.

(b) RNA polymerase could skip the sequence between those regions which constitute the mRNA (intervening sequences) forming an intact mRNA from the one transcriptional event.

(c) RNA polymerase could copy the entire segment of DNA into an RNA product, with the portions not required in the mRNA being deleted, and the RNA re-ligated.

With the discovery of a 600 base intervening sequence in the rabbit β globin structural gene (Jeffereys and Flavell, 1977), it become apparent that a complex

mechanism for the synthesis of mature mRNA species in eukaryotes might be a common phenomenon. It is interesting to note that the mouse β globin mRNA precursor described by Curtis and Weissmann (1976) and Bastos and Aviv (1977), is only slightly longer than an hypothetical precursor mRNA containing the rabbit β globin coding sequences, intervening sequences and 3' poly(A) tract. This observation suggests that the mechanism of mRNA biogenesis most likely to be acting in the maturation of rabbit β globin mRNA, is that described in (c) above.

Intervening sequences have been found in the genes coding for mouse β globin, mouse immunoglobulin light chains (both κ and λ), chick ovalbumin, SV40 large T antigen, adenovirus mRNA and a subset of the genes coding for the 28S rRNA of Drosophila melanogaster. The details of these are outlined in Table 1.1. The arrangement of intervening sequences in the chick ovalbumin gene is particularly complex and is worthy of further discussion.

The presence of intervening sequences in the ovalbumin gene was first reported by Breathnach et al., (1977). Using restriction enzyme analysis of total chick DNA in conjunction with the Southern transfer procedure (Southern, 1975b), the natural ovalbumin gene was shown to contain at least two intervening sequences of lengths 1.0 and 1.5 kb (Breathnach et al., 1977; Lai et al., 1978; Doel et al., 1977). At least one of these intervening sequences was shown to interrupt the coding sequence and there was no difference in the sequence organization observed in ovalbumin producing and non-producing cells

TABLE 1.1

Gene	No. of Intervening Sequences	Size of Intervening Sequences	References
Rabbit β globin	1	600 bases	Jeffereys & Flavell, (1977).
Mouse β globin	2	550 bases <125 bases	Tilghman et al.(1978a) Kinniburgh et al.(1978).
Chick ovalbumin	7	200 to 1,600 bases	Mandel et al.(1978) Dugaiczyk et al.(1978).
Mouse immunoglobulin λ light chain (Myeloma DNA)	2	93 bases 1,250 bases	Tonegawa et al.(1978) Brack & Tonegawa (1977).
Mouse immunoglobulin κ light chain (Myeloma DNA)	⩾1	-	Rabbitts & Forster,(1978) Matthyssens & Tonegawa (1978).
SV40 late large T antigen	1	280 bases	Lavi & Groner (1977) Bratosin et al.(1978) Berk & Sharp (1978).
Adenovirus mRNAs	1	-	Chow et al. (1977).
Drosophila 28S rRNA	1	5,000 bases	Glover & Hogness,(1977) White & Hogness,(1977) Pellegrini et al.(1977).

Number and length of intervening sequences in a number of eukaryotic structural genes.

(Weinstock et al., 1978). These and later results suggested that intervening sequences are not used to switch transcription of the ovalbumin gene on or off in the various cell types (Mandel et al., 1978). Thus, while the arrangement of immunoglobulin variable and constant region genes have been shown to vary for DNA derived from germline and myeloma DNA (Hozumi and Tonegawa, 1976; Rabbitts and Forster, 1978), Weinstock's results suggest that such gene re-arrangements are not a general prerequisite for gene expression (Weinstock et al., 1978). When a cloned genomic ovalbumin gene became available, it was possible to demonstrate that the actual arrangement of intervening sequences was even more complex (Mandel et al., 1978; Dugaiczyk et al., 1978) with seven such sequences all being confined to the region corresponding to the 5' half of the ovalbumin mRNA. The inserts varied in length from 0.2 to 1.6 kb and the eight coding fragments were shown to be arranged in the same order and relative orientation as in the mRNA (Mandel et al., 1978; Dugaiczyk et al., 1978). In addition, studies on the ovalbumin gene intervening sequences from different chickens indicated the existence of at least two alleles of the gene, the variation responsible for this polymorphism lying in an intervening region (Weinstock et al., 1978; Mandel et al., 1978).

If the entire natural ovalbumin gene is transcribed in stimulated chicken oviduct cells and the mature mRNA obtained by deletion of those sequences complementary to the intervening regions, then the pre-mRNA would be expected to be approximately 6 kb long (Dugaiczyk et al., 1978). Very recently, Roop et al., (1978) have described

the extraction of multiple species of hnRNA from stimulated chicken oviduct cells, which hybridized to both structural and intervening ovalbumin sequences. These RNA species ranged from 1.3 to 4 times the size of the mature ovalbumin mRNA. Catteral et al. (1978) have shown the existence of short, partial sequence homology at all splice points in the chick ovalbumin gene. In the light of the existence of a precursor for ovalbumin mRNA, it is tempting to speculate that these homologous sequences are signals for a splicing enzyme (Catteral et al., 1978). The mechanism by which adjacent coding sequences would be brought into close proximity for splicing was not immediately obvious from this sequencing study since there was no evidence for the occurrence of strong Watson-Crick base pairing between adjacent junctions.

The existence of precursor mRNAs for chick ovalbumin (Roop et al., 1978), immunoglobulin κ light chain (Gilmore-Herbert and Wall, 1978) and mouse β globin (Bastos and Aviv, 1977; Curtis and Weissmann, 1976), along with the observation by Tilghman et al. (1978b) that the intervening sequences in the natural mouse β globin gene are transcribed within the β globin mRNA precursor, suggest that most genes with intervening sequences might be completely transcribed, with deletion of intervening sequences and ligation of coding sequences occurring subsequently to produce the mature mRNA. The origins and functions of intervening sequences are unknown. It has been postulated that these sequences may be descended from some putative eukaryotic insertion element (see Bukhari, Shapiro and Adhya, 1977) inserted at some stage during evolution to perform an

unknown function. Alternatively, they may represent sequences which happened to separate regions of the genome which nature then chose to link at the level of the messenger to yield a new protein (Mandel et al., 1978). Gilbert (1978) has suggested that intervening sequences may allow an increased rate of evolution for eukaryotic organisms. They could also have important roles in cascade type regulation of gene expression at transcriptional and post-transcriptional levels (Mandel et al., 1978). In hormone-induced systems like ovalbumin, hormone-protein complexes may be directly involved by binding to intervening sequences in the DNA or to their transcripts; alternatively, they may act indirectly by inducing the synthesis of a specific splicing enzyme (Mandel et al., 1978).

While there is good evidence that at least some of the mRNAs derived from genes containing intervening sequences have high molecular weight precursors generated by the transcription of coding plus intervening sequences, the presence of an intervening sequence is not a prerequisite for a precursor:mRNA relationship. The most notable exceptions to the above pattern of mRNA biogenesis are the histone genes and mRNAs of sea urchin. The arrangement of these genes have been extensively characterized (Schaffner et al., 1976; Gross et al., 1976b; Sures et al., 1976; Wu et al., 1976) and will be discussed in detail below. The feature of their sequence organization relevant to the present discussion, however, is that they contain no intervening sequences. Nevertheless, Levy et al. (1978) have demonstrated the existence of high molecular weight precursors to each of the histone mRNAs of the sea urchin,

S. purpuratus.

E. MULTIGENE FAMILIES

When evolutionarily related genes code for products with identical, similar or overlapping functions and when they are closely linked on the chromosome, they are said to be a multigene family (Hood et al., 1975). Until the advent of recombinant DNA technology, the study of eukaryote gene arrangement and control was limited to those genes which were reiterated and relatively easily purified, by physico-chemical means, from the remaining genomic DNA. These included the genes coding for rRNA, 5S RNA (Brown et al., 1971) and tRNA (Clarkson and Birnstiel, 1973) in Xenopus laevis, and those coding for histones in sea urchin (Birnstiel et al., 1974). The ribosomal genes were studied in the greatest detail (see Birnstiel, Chipchase and Spiers, 1971, for review of this physico-chemical work). The following survey is limited to work involving restriction enzyme analysis and cloning, and is divided into two sections:

1. Faithfully repeated multigene families.
2. Heterogeneous multigene families.

1. Faithfully Repeated Multigene Families

This definition refers to genes which exist in multiple identical copies in the genome. The three families that will be dealt with are the ribosomal genes of Xenopus laevis and Drosophila melanogaster, the 5S genes of Xenopus laevis and the histone genes of sea urchin and D. melanogaster.

(a) Ribosomal genes

The ribosomal DNA of Xenopus laevis consists of about 500 repeating units, each containing a region coding for the 40S rRNA precursor and a non-transcribed spacer region. All 500 repeating units are tandemly arranged in the nucleolus organizer region of one of the 18 X. laevis chromosomes. While the region transcribed into the 40S rRNA precursor is constant in all repeating units, the non-transcribed spacer regions vary in length causing repeat unit lengths to vary from 10.8 kb to 16.7 kb (Wellauer et al., 1976a). It appears that a large fraction of the spacer region of the rDNA is composed of short subrepeats which constitute internally repetitious segments. Long spacers are distinguished from short spacers by having more copies of these subrepeats. The observed length variation of rDNA exists within a single nucleolus organizer and 50% to 68% of the adjacent repeats differ in length. It was observed, however, that the relative abundance of the size classes contained in the chromosomal and amplified ribosomal cistrons from a given individual can vary (Wellauer et al., 1976b). Furthermore, the preference for the amplification of a particular size class is inherited and some animals selectively amplify repeat units which are rarely found in their chromosomal repertoire. Most tandem repeats in a single amplified rDNA are of equal length (Wellauer et al., 1976b) which supports the hypothesis that rDNA is amplified by a rolling circle mechanism (Hourcade et al.,

1973a, b; Rochaix et al., 1974).

The ribosomal genes of Drosophila melanogaster also exhibit length heterogeneity. Glover and Hogness (1977) discovered the existence of a 5 kb insert in some structural genes coding for 28S rRNA, and an accompanying increase in repeat unit length from 12 kb to 17 kb. "R-loop" mapping confirmed the existence of the intervening sequence and "in situ" hybridization of this sequence to D. melanogaster salivary gland chromosomes, revealed that it was located in the centromeric heterochromatin and many bands on the euchromatic arms (White and Hogness, 1977). Subsequent studies showed that the insertions into the gene for 28S rRNA vary from about 0.5 kb to 6.0 kb, and occur in distinct size classes which are multiples of 0.5 kb (Wellauer and Dawid, 1977). While rDNA is found on the X and Y chromosomes of D. melanogaster, only rDNA from the X chromosome carries 28S rRNA genes with intervening sequences (Tartof and Dawid, 1976). On the X chromosome, genes carrying intervening sequences appeared to be randomly assorted with those carrying no such sequence (Pellegrini et al., 1977). About 45% of all the D. melanogaster ribosomal repeats appear to carry intervening sequences. Pellegrini et al. (1977) observed a short inverted repeat (100 to 400 base pairs), at the extremities of the intervening sequence in their "R-loop" studies, suggesting that these sequences may be translocatable elements.

(b) 5S genes

The DNA coding for 5S RNA (5S DNA) can be isolated from total Xenopus laevis DNA by repeated cycles of density gradient centrifugation (Brown et al., 1971). Denaturation mapping of this 5S DNA indicated that each G + C rich 5S gene was associated with a longer A + T rich spacer to form a 0.7 kb unit in a tandemly repeated array of such units (Brown et al., 1971). Restriction enzyme analysis of this total 5S DNA revealed length heterogeneity in these A + T rich spacer sequences (Carroll and Brown, 1976a). Repeating units differ from each other by 15 base pair quanta (Carroll and Brown, 1976a) and Brownlee et al. (1974) have shown that these A + T rich sequences consisted of tandem repeats of just this size. Repeat length heterogeneity is therefore due to variations in the number of these subrepeats (Carroll and Brown, 1976a).

By cloning of 5S DNA in bacteria and restriction enzyme analysis, Carroll and Brown (1976b) showed that adjacent 5S repeats can differ in length. The 121 nucleotide G + C rich 5S gene, however, was not sufficiently large to account for all the G + C rich sequences in any given 5S repeat unit. Sequencing studies have revealed the presence of a "pseudo-gene" structure adjacent to the 5S structural gene (Jacq et al., 1977). The pseudo-gene appears to be a perfect copy of 101 residues of the structural gene. It is presumed that this "pseudo-gene" was

once a structural gene resulting from gene duplication and produced 5S RNA. Its sequence, however, diverged sufficiently for the gene to be no longer functional (Jacq et al., 1977). Thus it is thought that the "pseudo-gene" is a relic of evolution.

(c) Histone genes

The number of histone genes in the genome of sea urchin varies between individuals, but for those species studied to date, the histone genes appear to be reiterated 300 to 1,000 times (Weinberg et al., 1972; Grunstein et al., 1973). Furthermore, these genes appear to be tandemly linked as determined by the analysis of the buoyant density of histone genes with changing DNA molecular weight (Kedes and Birnstiel, 1971). These results suggested that histone genes were either linked in blocks, each block coding for one type of histone, or the genes for each of the histone proteins could be intermingled with each other. The actual arrangement of histone genes has been unequivocally elucidated by restriction endonuclease cleavage and cloning of the repetitive tandem histone DNA sequences. Restriction analysis of total sea urchin DNA (S. purpuratus, L. pictus and Ps. miliaris) revealed the existence of a 6 kb cluster containing sequences complementary to all the histone mRNAs (Kedes et al., 1975; Weinberg et al., 1975; Schaffner et al., 1976).

With the separation of the mRNAs coding for the different sea urchin (Ps. miliaris) histones

(Gross et al., 1976a), it became possible to obtain a restriction map of the 6 kb histone repeat. In this manner, Schaffner et al. (1976) were able to determine that the gene sequence within the repeat was H_4, H_2B, H_3, H_2A and H_1. By cloning the repeat unit into λ followed by thermal denaturation analysis of the amplified DNA, it was shown that about 50% of the histone repeat consisted of A + T rich DNA which would not be expected to code for histone proteins (Schaffner et al., 1976). It was suggested that this A + T rich DNA might constitute spacer sequences between the histone genes of the repeat unit, and this was confirmed by partial denaturation of amplified repeat unit DNA followed by electron microscopic analysis (Portmann et al., 1976). This electron microscopic analysis also indicated that the spacers between different genes varied in length from 420 base pairs (between H_1 and H_4) to 880 base pairs (between H_4 and H_2B) (Portmann et al., 1976). Using restriction enzymes, controlled λ exonuclease digestion of cloned DNA and hybridization of purified histone mRNAs, Gross et al. (1976b) demonstrated that the polarity of the DNA fragment carrying the genes for H_2B, H_4 and H_1 is 5' $H_2B \rightarrow H_4 \rightarrow H_1$ 3'. Since all highly purified mRNAs hybridize exclusively to one strand of the cloned DNA, the polarity of the histone gene cluster is 5' $H_1 \rightarrow H_2A \rightarrow H_3 \rightarrow H_2B \rightarrow H_4$ 3'. Similar restriction (Cohn et al., 1976) and electron microscopic (Wu et al., 1976) studies on S. purpuratus have demonstrated that both the polarity of the his-

tone repeat and the interdigitation of histone genes with A + T rich spacers of different lengths are conserved between these two species. Subsequent sequencing studies (Schaffner et al., 1978) have indicated that histones do not seem to be derived from longer precursor proteins, nor is there any evidence for intervening sequences within the coding regions. The A + T rich spacer segments between the genes differ from each other, are made up of relatively simple nucleotide arrangements, but are not repetitious, and apparently do not code for additional large proteins (Schaffner et al., 1978).

The histone genes of Drosophila melanogaster are repeated 100 times in the genome in two types of repeat unit of lengths 4.8 kb and 5.07 kb (Karp and Hogness, 1976; Lifton et al., 1977). Both repeat units code for all five histones but, in contrast to the sea urchin situation, H_3, H_2A and H_1 genes are transcribed from one strand, with H_4 and H_2B being transcribed from the other (Lifton et al., 1977). The difference between the two types of repeat derives from an additional block of 270 base pairs in the spacer separating H_1 from H_3. The 5.07 kb unit is three times as frequent as the 4.8 kb unit in the D. melanogaster genome (Lifton et al., 1977).

(d) General conclusions from these gene arrangement studies

It is interesting to note that there is size heterogeneity in the repeat units of both ribo-

somal and 5S genes. The fact that repeats of these genes are not identical and the nature of the heterogeneity, permit some conclusions to be made about mechanisms for the maintenance of tandemly linked genes. Sudden correction mechanisms such as the master-slave (Callan and Lloyd, 1960) and contraction-expansion (Brown et al., 1972) models, which require tandem homogeneity, cannot, therefore, explain these observations. Similarly, the hypothesis that the evolution of these tandem genes has occurred by gene duplication followed by mutation and genetic drift is not supported, since it is unlikely that the variation produced would be limited to the number of multiples of a subrepeat (Carroll and Brown, 1976b). The regularity of the observed heterogeneity, however, suggests that some correction mechanism does exist and the subrepeats of the spacer regions are strongly implicated in the correction process. A mechanism of unequal crossing over between members of family of sequences, similar to that proposed by Smith (1973), which leads to the elimination or fixation of variants while permitting heterogeneity, is a mechanism for tandem gene maintenance which is compatible with the results described for ribosomal and 5S gene repeat unit length variation. The comparative study of the spacer sequences of different histone repeat units from the same reiterated group of repeat units is, therefore, awaited with interest.

2. Heterogeneous Multigene Families

This definition refers to families of genes

which are similar but not identical, where the genes of a given family all code for a similar funciton. The immunoglobulin genes and the genes coding for silk moth chorion proteins are established examples of such multigene families.

(a) Immunoglobulin genes

Most studies have been carried out on myeloma cells producing immunoglobulin light chains. The immunoglobulin light chain molecule consists of two regions, the constant region and the variable region. It is the variable region of the chain which plays a part in antibody recognition. Dreyer and Bennett (1965) first suggested that the constant and variable regions might be separately encoded in the genome. This was supported by amino acid sequence studies (Hood et al., 1976) and nucleic acid hybridization (Tonegawa et al., 1976) studies which indicated that there were more germ line sequences coding for the variable region (V genes) than for the constant region (C genes) for a given type of immunoglobulin chain. The more recent evidence of somatic re-arrangement of V and C genes confirms this suggestion (Hozumi and Tonegawa, 1976; Rabbitts and Forster, 1978). Seidman et al. (1978a) have recently generated variable region probes against two different κ light chain subgroups and with them, identified two non-overlapping sets of EcoRI fragments of mouse DNA. Each set consisted of five to ten fragments and each fragment contained elements of variable region gene sequences. Since different sets of fragments

carrying V gene elements were detected by probes to different κ light chain subgroups, and since 25 to 30 different subgroups have already been identified from amino acid sequence studies, there are probably a minimum of 125-150 distinct V genes in the mouse genome. By comparing the nucleotide sequences of two V gene elements and portions of their flanking sequences from different fragments from the same subgroup set, extensive regions of homology were observed, with sequence variations being mainly limited to those regions coding for the antigen binding site (Seidman et al., 1978a). That this homology extended well beyond the structural V gene region, was demonstrated by heteroduplex analysis using two cloned EcoRI fragments from the same subgroup set (one 13 kb and the other 3 kb). A minimum of 2.5 kb from these two fragments were shown to be homologous, 0.29 kb of which can be attributed to the V region element (Seidman et al., 1978b). No such extensive stretch of homology was observed in a similar heteroduplex comparison of two mouse β globin genes (Tiemeier et al., 1978) (where the β globin gene is an example of an evolutionarily fixed gene). Seidman et al. (1978b) suggest that this homology creates a large target for intragenic recombination. Such recombination could occur in germ line or somatic cells expanding and contracting various subgroups while constantly testing and creating new diversity.

(b) Silk moth chorion proteins

Regier et al. (1978b) have recently

reported striking similarities among the amino-terminal sequences of a set of functionally related structural proteins from the chorion (eggshell) of the silk moth Antheraea polyphemus. The sequence similarities suggest that the proteins are related in an evolutionary sense, presumably being encoded by genes which have evolved by gene duplication followed by diversification via nucleotide substitutions and re-arrangements.

Hood et al. (1975) have proposed that certain groups of functionally related RNAs or proteins could be coded for by families of genes which are clustered in the genome and evolutionarily related (e.g., immunoglobulin genes (Hood et al., 1975)). Chorion genes are apparently linked and hence would appear to constitute a second example of an informational multigene family (Regier et al., 1978b). Studies on this multigene family are still in their early stages. The degree of evolutionary relatedness of the genes coding for chorion proteins is being examined by protein sequencing (Regier et al., 1978a, b) and sequencing of the mRNAs coding for these proteins. Ultimately the goal is to understand the organization of chorion genes and their pattern of expression during development.

(c) Keratin genes

From the currently available information about keratin genes, it seems that they most closely fit the heterogeneous multigene family classification.

The similarities between the chorion protein and keratin systems at the protein level are striking, and further comparison of these two systems at the gene organization and control levels are awaited with interest.

F. AIMS OF THE PROJECT

At the time of commencement of this work, studies on the structure of keratin mRNA were in their early stages and there was preliminary evidence that the keratin genes in the chick genome were clustered. The initial aims of the project, therefore, were two-fold.

1. To study the nature of the 3' end of keratin mRNA and the relationship between the unique and reiterated sequences which appeared to be present in the messenger.

2. To investigate methods for the partial purification of genomic keratin sequences and to examine the possibility of using a density gradient system, employing RNA:DNA hybrids, to study the arrangement of tandemly linked genes. Should this technique prove to be effective, it was intended to employ it in the analysis of keratin gene organization in the chick genome.

CHAPTER II

GENERAL MATERIALS AND METHODS

CHAPTER II

GENERAL MATERIALS AND METHODS

A. MATERIALS

1. Proteins and Enzymes

Bovine Serum Albumin : Fraction V, Sigma Chemical Co., St. Louis, Missouri.

DNA Polymerase I : Boehringer Mannheim, Mannheim, West Germany.

Pancreatic Ribonuclease : Type III, Sigma.

Proteinase K : E. Merck, Darmstadt, West Germany.

RNA-dependent DNA Polymerase : gift of Dr. J.R.E. Wells, prepared from avian myeloblastosis virus donated by Dr. J.W. Beard and the N.I.H. Cancer Program.

Restriction Endonuclease EcoRI : gift of R.B. Saint.

S_1 Nuclease : prepared by the method of Vogt (1973).

2. Radiochemicals

[^{3}H]-deoxycytidine triphosphate (16 - 26C / mmol)) : The Radiochemical Centre, Amersham, Buckinghamshire, England.

α-[^{32}P]-deoxyadenosine triphosphate (initial specific activity 90C/mmol) : prepared by Dr. R.H. Symons of this Department.

Na^{125}I (initial specific activity 500 mC/ml) : The Radiochemical Centre.

3. Chemicals for Specific Procedures

(a) Electrophoresis

Acrylamide : Merck, twice recrystallized from $CHCl_3$.

Agarose : Sigma.

N,N'-methylenebisacrylamide : BDH Chemicals Ltd., Poole, Dorset., England, recrystallized from $CHCl_3$.

N,N,N',N'-tetramethylethylenediamine : Eastman Organic Chemicals, Rochester, New York.

Formamide : BDH, deionized as described by Pinder et al. (1974).

Ethidium bromide : Sigma.

Toluidine Blue : George T. Gurr, London, England.

(b) Density gradient centrifugation

Actinomycin D : gift from Merck, Sharp and Dohme, Rahway, New Jersey.

Caesium Chloride Optical Density grade : Harshaw Chemical Co., Cleveland, Ohio.

Caesium Sulphate High Purity grade : Kawecki Berylco Industries, Inc., New York, New York, recrystallized from boiling water and ethanol.

(c) Complementary DNA synthesis

Deoxyribonucleoside triphosphates : Sigma.

Dithiothreitol : Sigma.

Oligodeoxythymidylic acid, free acid : P.L. Bio-

chemicals Inc., Milwaukee, Wisconsin.

2-Mercaptoethanol : Sigma.

(d) Radioactive counting

NCS Solubilizer : Amersham/Searle Corpn., Arlington Heights, Illinois.

POPOP (1,4-bis-(2,5-phenyloxazolyl)-benzene) : Sigma.

PPO (2,5-diphenyloxazole) : Sigma.

4. Miscellaneous Materials

Coalfish DNA : Sigma.

Diethylpyrocarbonate : Sigma.

Ficoll : Pharmacia Fine Chemicals, Uppsala, Sweden.

Nitrocellulose : Sartorius, Göttingen, West Germany.

Phenol : BDH, redistilled under N_2 and reduced pressure, stored at -15°C under N_2 prior to use.

Polyvinylpyrrolidone : Sigma.

Sarcosyl : Ciba-Geigy Ltd., Basle, Switzerland.

Sodium dodecyl sulphate (95% pure) : Sigma.

Sucrose, ultrapure, RNA'ase free : Schwarz-Mann, Orangeburg, New York.

All other chemicals used were of analytical reagent grade, or of the highest purity available.

5. Preparations of Solutions

All solutions were prepared in sterile glass distilled water followed by autoclaving or treatment with diethylpyrocarbonate to inactivate nucleases. Glassware was sterilzed by either autoclaving, incubation at 110°C

overnight, washing with 1 M KOH followed by rinsing with sterile glass distilled water, or by a combination of these procedures. Spatulas, etc., were washed in alkali and rinsed in sterile water, as described. Pipettes and micro-pipettes were washed in glass distilled water containing diethylpyrocarbonate and dried for 16 hours at 110°C.

B. GENERAL METHODS

1. Ethanol Precipitation

In all cases samples were made 0.1 M with respect to sodium acetate using a 3 M stock solution at pH 5.5. About 2.5 volumes of redistilled ethanol was added to each sample, samples were then shaken and stored at -20°C for a minimum of 2 hours. Precipitates were then collected by centrifugation at 15,000 r.p.m. for 15 to 30 min in a Beckman J-21B centrifuge at 2°C. Ethanol was poured off and any remaining ethanol was allowed to drain from the pellet by inversion of the tube for 20 min. The pellet was then redissolved in the appropriate solution.

2. Preparation of Keratin mRNA

Fertilized eggs of White Leghorn fowls (*Gallus domesticus*) strain Para. 3 were obtained from the Parafield Poultry Research Station of the Department of Agriculture, Parafield, South Australia. The eggs were stored at 10°C for no more than 10 days, and incubated at 37°C, 54% humidity in a forced draught incubator (Saunders Products Pty. Ltd., Adelaide) for 14 days. At 14 days, embryos were removed from the eggs and washed with Hanks balanced salt solution. Feathers were plucked into a solution containing 200 mM KCl,

5.3 mM $MgCl_2$, 10 mM Tris, pH 7.4, and keratin mRNA isolated from the mRNP particles produced by EDTA dissociation of polysomes, as described by Kemp et al. (1974b). The keratin mRNA was then purified by repeated cycles of sucrose gradient centrifugation.

3. Preparation of rRNA

The mRNA preparation procedure involves centrifugation of the EDTA treated polysomes on 10-40% sucrose gradients (Kemp et al., 1974b). Large and small ribosomal subunits were isolated from these gradients, the solution made 5% with respect to sarcosyl and the protein removed by three extractions with an equal volume of a mixture of phenol and chloroform (1:1). After the final extraction, the RNA was precipitated with 2½ volumes of ethanol at -20°C for a minimum of two hours. 28S rRNA from the large ribosomal subunit and 18S rRNA from the small ribosomal subunit were further purified by sucrose gradient centrifugation (10-40% gradients, Beckman SW41 rotor, 37,000 r.p.m., 16 hours). Purified RNA was ethanol precipitated twice and stored as an ethanol precipitate until required.

4. Preparation of Chick DNA

Chick blood was obtained by cardiac puncture (for large volumes) and from a wing vein (for small volumes). Blood was collected into tubes containing 0.15 M NaCl, 5 mM KCl, 2 mM $MgCl_2$ (NKM) made 1% w/v with respect to heparin. The blood was centrifuged at 2,000 r.p.m. for 5 min at 4°C and plasma and white cells removed. The erythrocytes were then washed a further two times with NKM. Approximately 5

volumes of 2 mM $MgCl_2$ was added to the packed erythrocytes and stirred at 0°C for 10 min to bring about lysis of the cell membrane. Nuclei were spun down at 4,000 r.p.m. for 10 min at 4°C and the lysis procedure repeated.

High molecular weight DNA was extracted from the erythrocyte nuclei by a modification of the procedure described by Gross-Bellard et al. (1973). A portion of the nuclear pellet was lysed in a solution containing 10 mM Tris-HCl pH 7.5, 10 mM EDTA, 10 mM NaCl, 0.5% SDS and Proteinase K at a concentration of 200 µg/ml. The solution was incubated at 37°C overnight then extracted once with water saturated phenol and twice with a mixture of phenol and chloroform (1:1). The aqueous phase was dialysed against 3 changes, each of 2 litres of 10 mM Tris-HCl, pH 7.5, 1 mM EDTA (TE). Pancreatic ribonuclease A, previously heated to 80°C for 10 min, was added to a concentration of 10 µg/ml and the solution incubated at 37°C for a minimum of 2 hours. The DNA-containing solution was then made up to 10 mM Tris-HCl pH 7.5, 10 mM EDTA, 10 mM NaCl, 0.5% SDS and Proteinase K added to a final concentration of 100 µg/ml. The solution was incubated for a minimum of 2 hours at 37°C then extracted 3 times with a mixture of phenol and chloroform (1:1). The DNA solution was dialysed against 3 changes, each of 2 litres, of TE and then stored at 4°C.

To obtain DNA of lower molecular weight, DNA was extracted from chick erythrocyte nuclei by a modification of the method of Marmur (Marushige et al., 1968).

5. Preparation of Hydroxylapatite (HAP)

500 ml Each of 0.5 M $CaCl_2$ and 0.5 M Na_2HPO_4 were slowly mixed with stirring over a period of about 2 hours. $CaPO_4$ crystals were allowed to settle and the supernatant removed. Crystals were washed 4 times with 750 ml of glass distilled water. After the final wash, 750 ml of hot water was added to the crystals along with 25 ml of a freshly prepared 40% (w/w) solution of NaOH. The mixture was boiled gently with stirring for 1 hour then removed from the heat source and the crystals allowed to settle for 5 min. The turbid supernatant was removed and the crystals washed 4 times with water. After the fourth wash, 1 litre of 0.01 M phosphate buffer, pH 7.0 was added to the crystals. The mixture was heated with stirring until boiling began. The supernatant was removed and the procedure repeated as before except that the boiling was allowed to continue for 5 min. The supernatant was again removed and the procedure repeated with boiling for 15 min. The supernatant was removed and the HAP crystals stored in 10 mM phosphate buffer, pH 7.0, at 4°C.

6. Preparation of cDNA

(a) cDNA to keratin mRNA

DNA complementary to keratin mRNA was synthesized in a system (25 µl) containing 50 mM Tris-HCl, pH 8.3, 8 mM dithiothreitol, 8 mM $MgCl_2$, 0.66 mM each of the three unlabelled deoxynucleoside triphosphates, 60 µM labelled deoxynucleoside triphosphate (^{32}P dATP or ^{3}H dCTP), 100 µg/ml actinomycin D, 0.5-1

µg mRNA, 0.1 µg oligo dT and AMV-DNA polymerase. Syntheses were carried out at 37°C for 1 hour in the early part of these studies, but subsequently it was found that in such preparations, anything up to 15% of the total cDNA could be complementary to rRNA. This contamination could be reduced to insignificant levels by allowing the synthesis at 37°C to proceed for 15 min. After incubation, the RNA template was removed by alkaline hydrolysis with 0.3 M NaOH for 1 to 2 hours at 37°C. The cDNA was separated from unincorporated triphosphates by chromatography on Sephadex G-50, in 0.2 M NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 7.5 followed by ethanol precipitation.

(b) cDNA to rRNA

Synthesis of cDNA to rRNA also employed the reverse transcriptase of avian myeloblastosis virus, but used random oligonucleotides of salmon sperm DNA, prepared as described by Taylor et al. (1976), to prime the synthesis at non-specific points along the 18S and 28S rRNAs.

Using saturating amounts of enzyme, cDNA was synthesized in a 25 µl reaction volume containing 100 mM KCl, 3 mM $MgCl_2$, 8 mM dithiothreitol, 0.67 mM each of dGTP, dTTP and dATP and 60 µM ^{3}H dCTP, 50 mM Tris-HCl, pH 8.3, 100 µg/ml actinomycin D, 1 µg of RNA, 50 µg of primer DNA and 1 µl (2.5 units) of enzyme. Reaction mixes were incubated at 37°C for 60 min and cDNA isolated as above.

7. Radioiodination of RNA

Radioiodinations were performed by a modification of the method of Prensky (1976). Radioiodinations were carried out in a 10 µl volume using 1 to 2 µg of RNA. A typical reaction mix is set out below with the important final concentrations in parenthesis.

1 mg/ml herparin	1 µl	(0.1 mg/ml)
RNA	2 µl	(1-2 µg)
*Acidification mix	1 µl	($[I^-]$ = 3 x 10^{-5} - 1 x 10^{-4} M
Na ^{125}I	1 µl	400 - 500 µC of ^{125}I)
0.1 M $TlCl_3$ in 1 M HNO_3, diluted $^1/_{20}$ in 1 M sodium acetate pH 4.7	1 µl	(Sodium acetate 0.1 M $[Tl^{+++}]$ = 5 x 10^{-4} M)
H_2O	4 µl	

*Acidification mix = 0.75 M HNO_3, 10^{-4} M KI, 0.4 M sodium acetate pH 4.7 (Prensky, 1976).

Heparin was used as a ribonuclease inhibitor. The acidification mix was used to neutralize the Na^{125}I since it is delivered in a solution of pH between 9 and 11. This mix also contains KI to allow the concentration of I^- to be raised to within the optimum range (Prensky, 1976).

The reaction mix was incubated at 60°C for 10 min then chilled on ice. 1 µl Of 2-mercaptoethanol was added and the iodinated RNA was separated from the unincorporated ^{125}I by chromatography on Sephadex G-50 in 0.2 M NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 7.5. Fractions containing iodinated RNA were pooled and ethanol precipitated.

The radioiodinated products generally had specific activities of between 2 x 10^7 and 8 x 10^7 c.p.m./µg.

8. Determination of Radioactivity

For liquid hybridization and reassociation studies, samples were precipitated with 10% trichloroacetic acid, in the presence of bovine serum albumin or coalfish DNA as carrier, at 0°C. After 30-60 min, the precipitates were collected on glass fibre filters (Whatman, GF/A). Filters were washed twice with 5 ml of 5% TCA then extensively with ether, placed into 2.5 ml glass vials and dried at 110°C for a minimum of 30 min. Toluene-based scintillation fluid (3.5 g PPO, 0.35 g POPOP per litre of toluene) was then added, and the samples counted in a Packard liquid scintillation spectrometer. Nitrocellulose filters from filter hybridizations, after washing and drying, were counted in a similar way.

In cases where small volumes (up to 10 µl) of aqueous samples were to be counted, these were dissolved in 2 ml of "toluene-triton" scintillation fluid (5 volumes of toluene scintillation fluid plus 3 volumes of triton X - 100).

^3H labelled radioactive products in gel slices were detected after solubilization of each slice with 0.2 ml NCS solubilizer plus 0.025 ml of 8 M NH_4OH, to which was added 2 ml of scintillation fluid. Gradient fractions and gel slices containing ^{32}P labelled products were determined by Cerenkov counting.

Samples were counted for as long as practicable in order to minimize counting errors.

9. Digestion of λ Phage DNA with $EcoR_I$ Restriction Endonuclease

$EcoR_I$ digestions of λ phage DNA were performed in a 16 µl incubation mix containing 1 µg of λ DNA, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 10 mM Tris-HCl, pH 7.5 and 2 units of $EcoR_I$ restriction endonuclease. (One unit of enzyme completely digests 1 µg of λ DNA in 60 min.) Incubations were carried out at 37°C for 60 min and the reaction was stopped by standing on ice.

CHAPTER III

STRUCTURE OF KERATIN mRNA

CHAPTER III

STRUCTURE OF KERATIN mRNA

A. INTRODUCTION

The work described in this chapter deals with studies on the 3' end 150 nucleotides of keratin mRNA and their relationship to the rest of the molecule. This work was begun by D.J. Kemp, formerly of this laboratory, and a complete appreciation of the study will require the presentation of a limited amount of data obtained by him and published jointly (Kemp, Lockett and Rogers, in preparation). Data obtained by Kemp will be indicated in the text.

Feather keratin has distinctive properties relevant to studies on the organization of different classes of sequence. Feather keratin consists of a large number of homologous polypeptide chains, all containing about 100 amino acids but differing in primary structure by multiple amino acid substitutions (O'Donnell, 1973b; O'Donnell and Inglis, 1974; Kemp et al., 1975). Highly purified keratin mRNA (Partington et al., 1973; Kemp et al., 1974b; Kemp et al., 1974c) was about 0.8 kb long, indicating the presence of 0.5 kb in addition to the keratin coding sequence of 0.3 kb. Keratin mRNA acted as a template for the synthesis of complementary DNA (cDNA) by the RNA dependent DNA polymerase of avian myeloblastosis virus (AMV) (Kemp et al., 1975; Kemp, 1975). Hybridization of keratin cDNA to keratin mRNA indicated that the sequence complexity of keratin mRNA was 25 to 35 times that expected for a single molecular species (Kemp, 1975). The apparent rate of hybridization

was about ten times greater when assayed on HAP than when assayed by S_1 nuclease and the hybrids increased in thermal stability with increasing $R_o t$. These results were explained by the hypothesis that each of the homologous keratin mRNA species contained a common (but non-identical) sequence (reiterated sequence) covalently attached to a sequence so different in base sequence from that of any other keratin mRNA species (unique sequence) that it appeared unique in the chick genome, under the experimental criteria used. It was estimated that there were about 100 to 240 copies of the reiterated sequence in the chick genome and suggested that this sequence is the keratin coding sequence (Kemp, 1975).

It was possible to obtain shorter keratin cDNA molecules which were representative of the 3' end of keratin mRNA, by reverse transcribing the mRNA with E. coli DNA polymerase I (PolI). Using this short cDNA as probe, it was possible to determine the nature of the 3' end sequences and whether they were representative of the unique or reiterated class of sequences.

Initial studies involved the comparison of the hybridization kinetics for PolI- and AMV-cDNAs to keratin mRNA using the equivalent hybridizations in the rabbit globin system for kinetic standards. While the observed kinetics of hybridization for AMV-cDNA to keratin mRNA were faster when assayed on HAP than when assayed by nuclease S_1 (Kemp, 1975), no such change in rate was observed when keratin PolI-cDNA was used as probe (Kemp et al., in preparation). Denaturation studies showed that, while both keratin AMV-cDNA

hybrids and PolI-cDNA hybrids changed in thermal stability as a function of $R_o t$, the change observed for the PolI-cDNA hybrids was nowhere near as marked as that for the AMV-cDNA hybrids (Kemp et al., in preparation). In addition, unlike the results with AMV-cDNA, the PolI- study showed no significant transfer of hybrids from low thermal stability to high thermal stability (Kemp et al., in preparation). The T_m of the PolI-cDNA hybrids was about 83°C after complete hybridization compared to 90°C for AMV-cDNA hybrids. At least a part of this T_m reduction was probably caused by the shortness of the hybrid, since globin PolI-cDNA hybrids showed a 4°C lower T_m than globin AMV-cDNA hybrids (Kemp et al., in preparation).

These results suggested that the PolI-cDNA had different characteristics from the AMV-cDNA which inferred that different regions of the mRNA molecule might have been subject to different divergence pressures. This regional variation was, therefore, studied in more detail.

B. METHODS

1. Preparation of mRNA, rRNA, cDNAs and DNA

Keratin mRNA was prepared from EDTA treated polysomes as described by Kemp et al. (1974b). Ribosomal RNA was prepared as described in Chapter II.B.3. AMV-cDNAs to keratin mRNA and rRNA were prepared as described in Chapter II.B.6 (a) and (b) respectively. PolI-cDNAs were synthesized in a system (50 µl) containing glycine buffer, (67 mM, pH 9.2) KCl (50 mM), dATP, dTTP, dGTP (50 µM each), ^{3}H-dCTP (20 µM) (Sp. Act. 26.2 mC/µmole) 2-mercaptoethanol (10

mM), $MnCl_2$ (0.5 mM), dT_{10} (2.5 µg/ml), mRNA (0.6 µg), actinomycin D (100 µg/ml) and E. coli DNA polymerase I (1 µg) incubated for 90 min at 37°C and purified as for AMV-cDNA (personal communication of O. Bernard). Chick erythrocyte chromosomal DNA was prepared essentially as described by Marushige, Brutlag and Bonner (1968) and sonicated to an average length of about 500 bases as determined by electron microscopy. Chick DNA was labelled by "nick translation" with E. coli DNA polymerase I in the presence of DNAse I as described by Schachat and Hogness (1973), using ^{3}H-dCTP (Sp. Act. 26 C/mmole) as the only labelled deoxynucleoside triphosphate. After labelling, the DNA was purified as for cDNA. PolI-cDNA was elongated as described by Crawford and Wells (1976).

2. RNA-cDNA Hybridizations

(a) Hybridization kinetics

18S rRNA, 28S rRNA and keratin mRNA were used in the hybridization kinetic analysis. Hybridizations were performed at 60°C for 4 hours in 0.18 M NaCl, 0.001 M EDTA, 0.01 M Tris-HCl, 0.05% SDS, pH 7.0. The reaction mixtures (25 µl) contained 2,000 c.p.m. of cDNA and varying amounts of RNA. For S_1 nuclease assays, 25 µl reaction mixes were diluted into 1.0 ml of 0.03 M Na-acetate, 0.05 M NaCl, 0.001 M $ZnSO_4$, 5% glycerol, pH 4.6 (low salt S_1 assay buffer) containing 12 µg of sonicated, denatured coalfish DNA. An aliquot of 0.5 ml was removed from each sample and identical samples were incubated with or without S_1 nuclease (4 units, Vogt (1973)) at 45°C for 30 min,

TCA precipitated (Chapter II.B.8) and counted.

(b) Estimation of ribosomal contaminants in cDNA preparations

Hybridizations to determine the percentage of PolI- and AMV-cDNAs complementary to rRNA were carried out by mixing 1.3 µg each of 18S and 28S rRNA with 4,000 c.p.m. of the appropriate ^{3}H labelled cDNA in a final volume of 20 µl of the same hybridization buffer as above. Samples were sealed in sterile 50 µl microcapillaries, placed in boiling water for 2 min, then incubated at 60°C for 18 hours giving a $R_o t$ value of about 14 mol.s.l^{-1} for each ribosomal species. Similar hybridizations were performed using 0.5 µg of keratin mRNA giving a $R_o t$ value of about 5 mol.s.l^{-1}. Under these conditions all hybridizations should have gone to completion. The hybrids were then assayed by S_1 nuclease in the above low salt S_1 nuclease digestion buffer as described by Kemp (1975), and the percentage of the counts in hybrid form estimated.

(c) Competition hybridizations

Competition hybridizations were carried out in the same buffer as described above. 25 µl Aliquots of a hybridization solution containing a molar ratio of cold AMV-cDNA:keratin mRNA:^{3}H labelled PolI-cDNA of 100:40:1, or of mRNA:PolI-cDNA of 40:1, were sealed in sterile 50 µl microcapillaries, placed in boiling water for 1 min and then transferred to a 60°C waterbath. Aliquots were removed at different

times, snap chilled in ice water and stored frozen at -20°C. When the last sample had been removed, earlier samples were thawed and assayed by nuclease S_1 as previously described (Kemp, 1975).

3. cDNA-DNA Reassociations and Thermal Stability Analyses

(a) Reassociation

Reassociation of keratin AMV-, PolI- and elongated PolI-cDNAs with sonicated chick erythrocyte DNA,10-15 mg/ml,was performed in the same buffer as described in Chapter III.B.2.(a), using a ratio of 2-3,000 c.p.m. cDNA to 1 mg of DNA. Reaction mixtures were heated to 100°C for 5 min to denature the DNA prior to incubation at 60°C to the appropriate $C_o t$ values. For S_1 nuclease assays, 50 µl time samples were diluted into 1 ml of 0.03 M Na-acetate, 0.3 M NaCl, 0.001 M $ZnSO_4$, 5% glycerol, pH 4.6 (high salt S_1 assay buffer), the samples divided into two and incubated at 37°C with or without nuclease S_1 (150 units). For HAP assays, 50 µl aliquots were diluted into 2 ml of 0.12 M phosphate buffer (0.18 M Na^+). Single and double-stranded DNA were then separated on HAP at 60°C as described by Britten and Kohne (1968). After S_1 nuclease digestion or HAP fractionation, all samples for any one curve were made up to the same final concentration of DNA before TCA precipitation to avoid errors resulting from differential quenching.

(b) Thermal stability

For the determination of the thermal stabilities of cDNA-DNA duplexes, 50 µl aliquots taken at different values of $C_o t$ were diluted into 2 ml of 0.12 M phosphate buffer and loaded on to HAP at 60°C. Duplexes which could not be eluted at 60°C were melted by raising the temperature in 5°C increments. Two 2.5 ml washes (0.12 M phosphate buffer) were collected at each temperature and TCA precipitated.

4. Preparation of Keratin AMV- and PolI-cDNA Duplexes for Density Gradient Centrifugation

AMV- or PolI-cDNA duplexes were prepared by incubation to a $C_o t$ of 2×10^4 mol.s.l^{-1} as above. Samples (100 µl) containing 1 mg DNA and 3,000 c.p.m. cDNA were added to 4.0 ml high salt nuclease S_1 assay buffer and incubated at 37°C for 15 min with 100 units of nuclease S_1. The reaction was stopped by addition of 0.25 ml of 0.4 M Na^+ phosphate, pH 7.0 and the mixture was dialysed against 3 changes of buffer (10 mM Tris-HCl, pH 8.4, 1 mM EDTA, 10 mM NaCl). Chick DNA (10 kb) was added to a total of 1 mg DNA and buffer and solid CsCl were added to a volume of 12 ml and a density of 1.6 gm/cc. The mixtures were chilled on ice and 500 µg actinomycin D was added. The refractive index was adjusted to 1.3903 at 20°C. The mixtures were centrifuged in a Beckman Ti50 rotor at 32,000 r.p.m. for 60 hours at 20°C. The gradients were fractionated by upward displacement and 0.5 ml fractions were collected. The A_{260} of each fraction was determined and the amount of DNA in each fraction was made constant by addition of sonicated

calf thymus DNA. Actinomycin D was extracted with CsCl-saturated isopropanol. Water (1.5 ml) and 20% TCA containing 1% sodium pyrophosphate (2.0 ml) were added to each fraction. The precipitates were collected on Whatman GF/A filters, rinsed twice with 5% TCA - 1% sodium pyrophosphate, twice with ether, dried and counted in toluene-based scintillation fluid.

C. RESULTS

1. mRNA Purity

The work described in this chapter depends on the PolI- and AMV-cDNAs being representative of particular regions of the keratin mRNA molecule. It was, therefore, necessary to prove either that the messenger preparations were completely pure or that any contaminating sequences were not copied by the enzymes. It has been published previously (Partington *et al.*, 1973; Kemp *et al.*, 1974b; Kemp *et al.*, 1974c) that keratin mRNA preparations were greater than 95% pure as assayed by formamide gel electrophoresis and *in vitro* translation. While the latter piece of information suggests the absence of contaminating messenger species, these methods of analysis would not have detected rRNA breakdown products of the same size as the mRNA (0.8 kb). Contaminating ribosomal sequences could be detected, however, by liquid hybridization using cDNA made to rRNA by the random priming method of Taylor *et al.* (1976) (see Chapter II.B.6.(b)). Complementary DNA was made to purified 18S and 28S rRNA (see Chapter II.B.3.) and the kinetics of hybridization of these cDNAs to keratin mRNA were compared to those for the homologous reactions (i.e., 18S rRNA with

cDNA made to 18S rRNA etc.). The results are shown in Figure 3.1. By dividing the $R_o t_{\frac{1}{2}}$ for the homologous hybridization by the $R_o t_{\frac{1}{2}}$ for the heterologous reaction, the proportion of the keratin mRNA preparation consisting of 18S or 28S rRNA breakdown products could be determined.

From Figure 3.1(a) the $R_o t_{\frac{1}{2}}$ for the hybridization of cDNA, made to 28S rRNA, to 28S rRNA was 1.2×10^{-2} mol.s.l^{-1} while the $R_o t_{\frac{1}{2}}$ for the hybridization of the same cDNA to keratin mRNA was 5.5×10^{-1} mol.s.l^{-1}. The proportion of the keratin mRNA preparation which could be attributed to 28S rRNA breakdown products was therefore $\frac{1.2 \times 10^{-2}}{5.5 \times 10^{-1}} =$ 0.022 or 2.2% of the preparation. From Figure 3.1(b) the $R_o t_{\frac{1}{2}}$ for the hybridization of cDNA made to 18S rRNA to 18S rRNA was 2.6×10^{-3} mol.s.l^{-1}. The $R_o t_{\frac{1}{2}}$ for the hybridization of this cDNA to keratin mRNA was 9×10^{-3} mol.s.l^{-1}. Thus the proportion of the keratin mRNA preparation attributable to 18S rRNA breakdown products was $\frac{2.6 \times 10^{-3}}{9 \times 10^{-3}} = 0.289$ or 28.9% of the keratin mRNA preparation.

It was apparent from these results that approximately 30% of this particular mRNA preparation was derived from ribosomal breakdown material with most of the contaminant coming from 18S rRNA. Although the actual proportion of any given mRNA preparation that could be attributed to ribosomal breakdown products would be expected to vary from preparation to preparation, at least 20% of the RNA in any particular preparation could be expected to be of ribosomal origin.

FIGURE 3.1.

THE DEGREE OF CONTAMINATION OF KERATIN mRNA PREPARATIONS WITH 18S AND 28S rRNA BREAKDOWN PRODUCTS

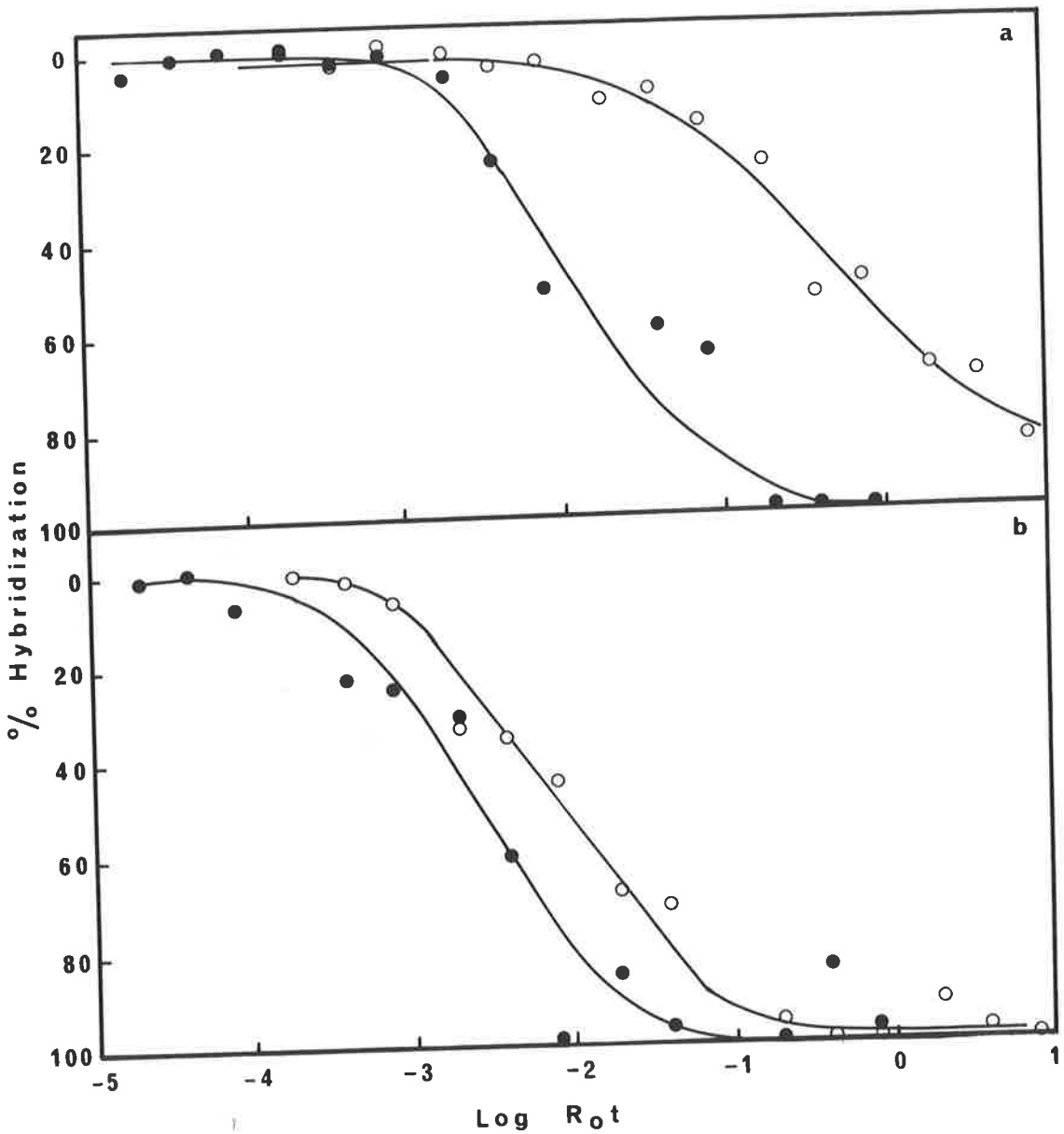
cDNA made to 18S or 28S rRNA (Chapter II.B.6.(b)) was hybridized to keratin mRNA (Chapter III.B.2.(a)). The rates of these heterologous hybridizations as assayed by S_1 nuclease, were compared with the rates of the homologous ribosomal hybridizations (e.g., 28S RNA to cDNA made to 28S rRNA) assayed in the same way. The reactions were carried out with a vast excess of RNA over cDNA. Samples were denatured at 100°C and hybridizations carried out at 60°C.

(a) cDNA to 28S rRNA hybridized to:

● 28S rRNA; o keratin mRNA.

(b) cDNA to 18S rRNA hybridized to:

● 18S rRNA; o keratin mRNA.


a
b
0
20
40
60
80
100
% Hybridization
-5
-4
-3
-2
-1
0
1
Log Rot

2. The Nature and Specificity of AMV- and PolI-cDNAs

The synthesis of AMV- and PolI-cDNAs both employ oligo-dT as a primer (see Chapter III.B.1 for PolI-cDNA synthesis conditions). Since keratin mRNA has a 60 base Poly(A) tract at its 3' end (Morris and Rogers, in press) while rRNA has no such structure, the AMV and PolI enzymes would be expected to copy only keratin mRNA, provided cDNA synthesis could be shown to be completely dependent on the presence of the primer. The primer dependence of PolI- and AMV-cDNA synthesis, using keratin mRNA as a template, was therefore examined. In addition, the efficiency with which these enzymes copied rRNA, in the presence and absence of oligo-dT, was also studied since rRNA breakdown products constituted a significant proportion of the keratin mRNA preparation. The results are shown in Table 3.1.

The PolI section of Table 3.1 shows that copying of keratin mRNA was almost completely dependent on the presence of oligo-dT primer. It also shows that PolI copied rRNA quite readily in the presence or absence of oligo-dT primer. This latter result suggests that the 7.7% background synthesis observed using keratin mRNA as template in the absence of oligo-dT, might be attributed to copying of the contaminating ribosomal sequences. Presumably, however, 92.3% of the cDNA copied from keratin mRNA in the presence of oligo-dT represents keratin sequences and so the low contamination with ribosomal cDNA would not significantly affect any of the conclusions made where PolI-cDNA has been used as a probe.

TABLE 3.1

DEPENDENCE OF POLI AND AMV COPYING OF KERATIN mRNA ON THE PRESENCE OF OLIGO dT PRIMER AND THE RELATIVE EFFICIENCIES OF THESE ENZYMES FOR COPYING rRNA

Experimental details	C.p.m. incorporated into TCA precipitable material as a percentage of those incorporated in the enzyme + primer + keratin mRNA synthesis
PolI + oligo dT + keratin mRNA	100 %
PolI - oligo dT + keratin mRNA	7.7%*
PolI + oligo dT + rRNA	38.8%
PolI - oligo dT + rRNA	52.9%*
AMV + oligo dT + keratin mRNA	100 %
AMV - oligo dT + keratin mRNA	3.5%
AMV + oligo dT + rRNA	2.0%*
AMV - oligo dT + rRNA	1.7%*

PolI-cDNA synthesis was as described in Chapter III.B. 1. except that mixes were made up ± oligo dT in 10 µl reaction volumes. 0.5 µg Of purified mRNA or 0.25 µg each of 18S and 28S rRNA was used as template in the reaction mix in the presence or absence of oligo dT primer.

AMV-cDNA syntheses were as described in Chapter II.B. 6.(a) except that mixes were made up ± oligo dT in 10 µl reaction volumes. Templates were as for PolI.

*Denotes that the result is the average of two independent experiments.

The AMV section of Table 3.1 shows that AMV-cDNA synthesis was almost completely dependent on the presence of oligo-dT primer and that no significant copying of rRNA occurred, regardless of the presence or absence of oligo-dT primer.

Keratin PolI- and AMV-cDNAs were also hybridized with excess rRNA to provide an independent estimate of the proportion of the cDNAs complementary to rRNA. Table 3.2 shows that only 3% of the PolI-cDNA and 2.15% of the AMV-cDNA hybridized to rRNA. In contrast 90.9% and 75.2% of PolI- and AMV-cDNAs respectively hybridized to keratin mRNA. Although these latter values might have been expected to be 100%, in practice keratin AMV-cDNA:mRNA and PolI-cDNA:mRNA hybridization reactions have never been observed to exceed about 80% and 90% hybridization respectively when assayed by S_1 nuclease. Rabbit globin hybridizations, however, always went to 100%.

Thus reverse transcription of keratin mRNA by both PolI and AMV reverse transcriptase was completely dependent on the presence of oligo-dT in the synthesis mix. In addition, the proportion of the cDNA complementary to rRNA made by either enzyme was insignificant.

3. Molecular Weights of Keratin AMV-cDNA and PolI-cDNA

Keratin cDNA was prepared using reverse transcriptase (AMV-cDNA) and also using *E. coli* DNA polymerase I (PolI-cDNA). The molecular weights of the keratin AMV-cDNA and PolI-cDNA were estimated by polyacrylamide gel electro-

TABLE 3.2

DETECTION OF REVERSE TRANSCRIPTS OF rRNA IN KERATIN POLI- AND AMV-cDNAs BY LIQUID HYBRIDIZATION

Hybridization	% Counts resistant to nuclease S_1
PolI-cDNA : rRNA	3 %*
PolI-cDNA : keratin mRNA	90.9%
AMV-cDNA : rRNA	2.15%*
AMV-cDNA : keratin mRNA	75.2%

Keratin AMV-cDNA or PolI-cDNA was hybridized to rRNA or keratin mRNA as described in Chapter III.B.2.(b) and assayed using S_1 nuclease in low salt assay buffer (Chapter III.B.2.(a)). The resistances of PolI- and AMV-cDNAs to S_1 nuclease were 5% and 7% respectively and the figures in the table have been normalized to account for this as in Kemp (1975).

*Denotes that the result is the average of two independent experiments.

phoresis in the presence of formamide (Pinder, Staynov and Gratzer, 1974) using RNA markers (Figure 3.2). Both cDNA preparations were heterogeneous with regard to size. However, the main peak of keratin AMV-cDNA was of apparent MW $\sim 1.7 \times 10^5$ (about 0.55 kb) while that of keratin PolI-cDNA was of apparent MW $\sim 4.5 \times 10^4$ (about 0.15 kb). Since transcription of mRNA into cDNA in the presence of oligo-dT has been shown to initiate at the poly(A) region of mRNA (Milstein et al., 1974; Proudfoot and Brownlee, 1974; Crawford et al., 1977) and since PolI-cDNAs are shorter than the respective AMV-cDNAs (Milstein et al., 1974; Proudfoot and Brownlee, 1974; Modak et al., 1973; Gulati et al., 1974; Rabbitts, 1974; Crawford et al., 1977), keratin PolI- and AMV-cDNAs were expected to be different length copies of keratin mRNA, both being initiated at the 3' end of the mRNA. The poly(dT) segments of chick globin PolI- and AMV-cDNAs, prepared under conditions identical to those used here, have been shown to average 18-20 bases in length (Crawford et al., 1977) so the keratin PolI- and AMV-cDNAs are presumably complementary to regions of keratin mRNA extending about 0.13 and 0.53 kb from the poly(A) tract respectively.

4. Competition Hybridization Between PolI- and AMV-cDNA

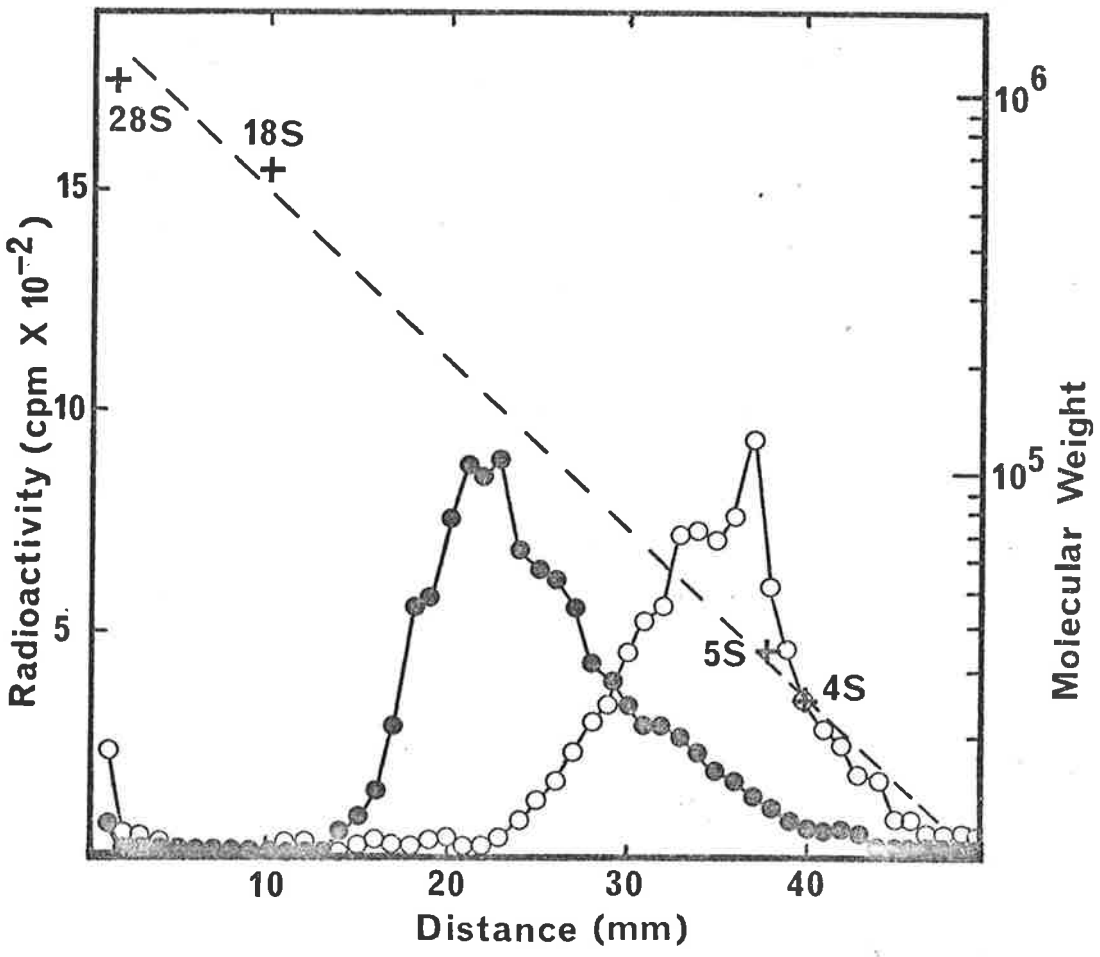
To be certain that the AMV reverse transcriptase and DNA polymerase I were copying the same set of mRNA molecules, the AMV- and PolI-cDNAs were tested for the presence of common sequences by studying the competition between the two cDNAs for sequences in the mRNA during hybridization. Figure 3.3 shows the hybridization kinetics of ^{3}H-labelled

FIGURE 3.2.

POLYACRYLAMIDE GEL ELECTROPHORESIS OF KERATIN AMV-cDNA AND POLI-cDNA IN THE PRESENCE OF FORMAMIDE

Aliquots of AMV-cDNA and PolI-cDNA were co-electrophoresed with the marker RNAs on 4% polyacrylamide gels in formamide containing 0.02 M Barbital, pH 9.0, by the procedure of Pinder et al. (1974). After electrophoresis for 1 hour at 5 mA/gel, the gels were stained with 0.05% Toluidine blue, photographed, and cut into 1 mm slices. The slices were counted in Toluene-NCS scintillation fluid. The c.p.m. scale for PolI-cDNA was half that shown.

● AMV-cDNA ○ PolI-cDNA. The positions of markers (rabbit reticulocyte 28S, 18S, 5S and 4S RNAs) are shown.


28S
18S
5S
4S
15
10
5
Radioactivity (cpm X 10^{-2})
10^6
10^5
Molecular Weight
10
20
30
40
Distance (mm)

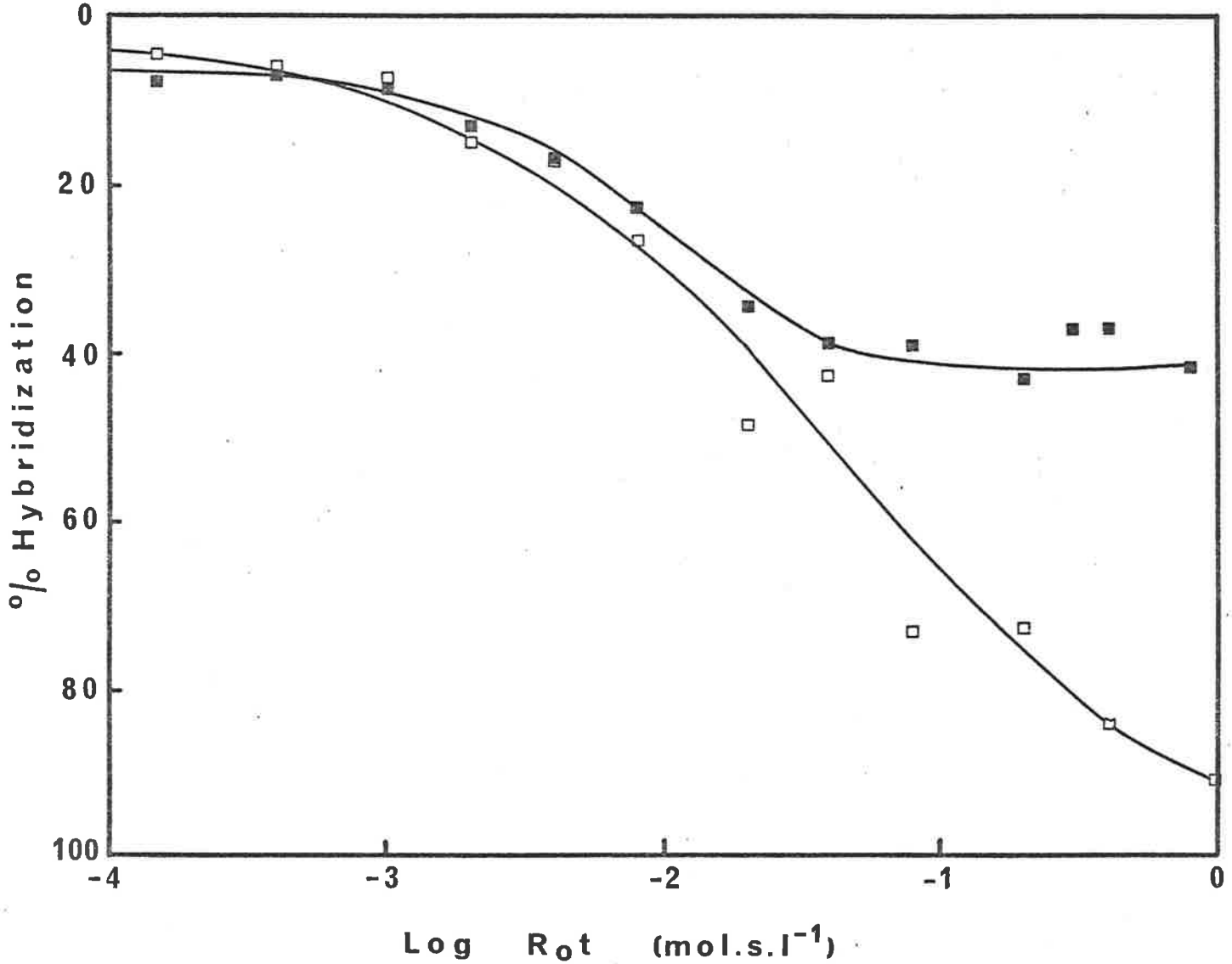
PolI-cDNA to mRNA in the presence and absence of cold competing AMV-cDNA. The molar ratio of cold AMV-cDNA:mRNA: labelled PolI-cDNA was 100:40:1. As described by Gambino et al. (1974), the percentage of a radioactive probe in duplex form at the end of a DNA reannealing reaction is dependent on the ratio of the reacting species. Consider the DNA strand with the same sequence as the cDNA as the + strand and the strand with the same sequence as the RNA, the - strand. In a reannealing reaction there would always be equal quantities of the cold reacting + and - species since DNA is double-stranded. If only a trace amount of cDNA probe was used relative to the amount of the cold + strand in the reaction, the extent of the reassociation, as measured by the radioactive probe, would be almost 100% at completion. If, however, the cDNA probe was present in the same amount as the cold + strand, there would be 2 times as many + strands as - strands. Thus, the chance of incorporating a labelled cDNA into a duplex is one in two and so the observed extent of the reaction will be 50%. If the amount of cDNA was further increased, then the excess of + strand over - strand would also increase and so a decreasing proportion of the input radioactively labelled probe would be incorporated into duplex form. In fact, under these conditions of + strand excess, the additional cDNA rather than the cold sequences alone, drives the reannealing reaction. In an analogous manner, when a trace amount of labelled cDNA is used in an RNA driven DNA:RNA hybridization reaction, the observed extent of hybridization would be expected to be 100%. This should remain true for ratios of cDNA:RNA up to 1:1. When the amount of cDNA exceeds the amount of

FIGURE 3.3.

HYBRIDIZATION KINETICS OF ^{3}H LABELLED POLI-cDNA TO KERATIN mRNA IN THE PRESENCE AND ABSENCE OF COLD COMPETING AMV-cDNA

Hybridization solutions containing molar ratios of cold AMV-cDNA : mRNA : PolI-cDNA of 100:40:1 and mRNA : ^{3}H labelled PolI-cDNA of 40:1 were divided into 25 µl aliquots and hybridizations performed as described in Chapter III.B. 2.(c). Hybridizations were assayed by nuclease S_1. Curves were normalized to remove background (Kemp, 1975). The normalization factor is shown in parenthesis for each curve.

- ■ Hybridization in the presence of cold AMV competitor (5%).
- □ Hybridization in the absence of cold AMV competitor (5%).


0
20
40
60
80
100
% Hybridization
-4
-3
-2
-1
0
Log Rot (mol.s.l-1)

driving RNA, the proportion of probe which can be in hybrid form is dependent on the ratio of cDNA to RNA. If twice as much cDNA as RNA is used in the hybridization, only half of the probe can be in hybrid form at completion. In the experiment shown in Figure 3.3, 2.5 times as much cold AMV-cDNA was present in the hybridization as mRNA. Thus it would be expected that only 40% of the AMV-cDNA could be in hybrid form at completion. Using a trace amount of PolI-cDNA as probe, the extent of PolI-cDNA hybridization at completion would also be expected to be 40%, if the AMV- and PolI-cDNAs were copies of the same mRNA population. (The same result would be expected if a trace amount of labelled AMV-cDNA was used instead of PolI-cDNA.) If AMV- and PolI-cDNAs represented copies of different mRNA populations, however, the extent of PolI-cDNA hybridization at completion would be expected to be 100%, regardless of the presence or absence of cold AMV-cDNA. The observation that the extent of PolI hybridization at completion in the presence of AMV competitor was about 40% of that in the absence of competitor (Figure 3.3) supports the view that PolI- and AMV-cDNAs truly represent different length copies of the same mRNA population.

5. Reassociation of Keratin AMV-cDNA and PolI-cDNA to Chick Erythrocyte DNA

Figure 3.4 shows the reassociation kinetics of keratin AMV-cDNA and PolI-cDNA in the presence of a vast excess of chick erythrocyte DNA. Keratin AMV-cDNA reassociated in a biphasic curve with major transitions at $C_o t_{1/2}$ values at about 9×10^1 and 2×10^3 mol.s.l^{-1} when reasso-

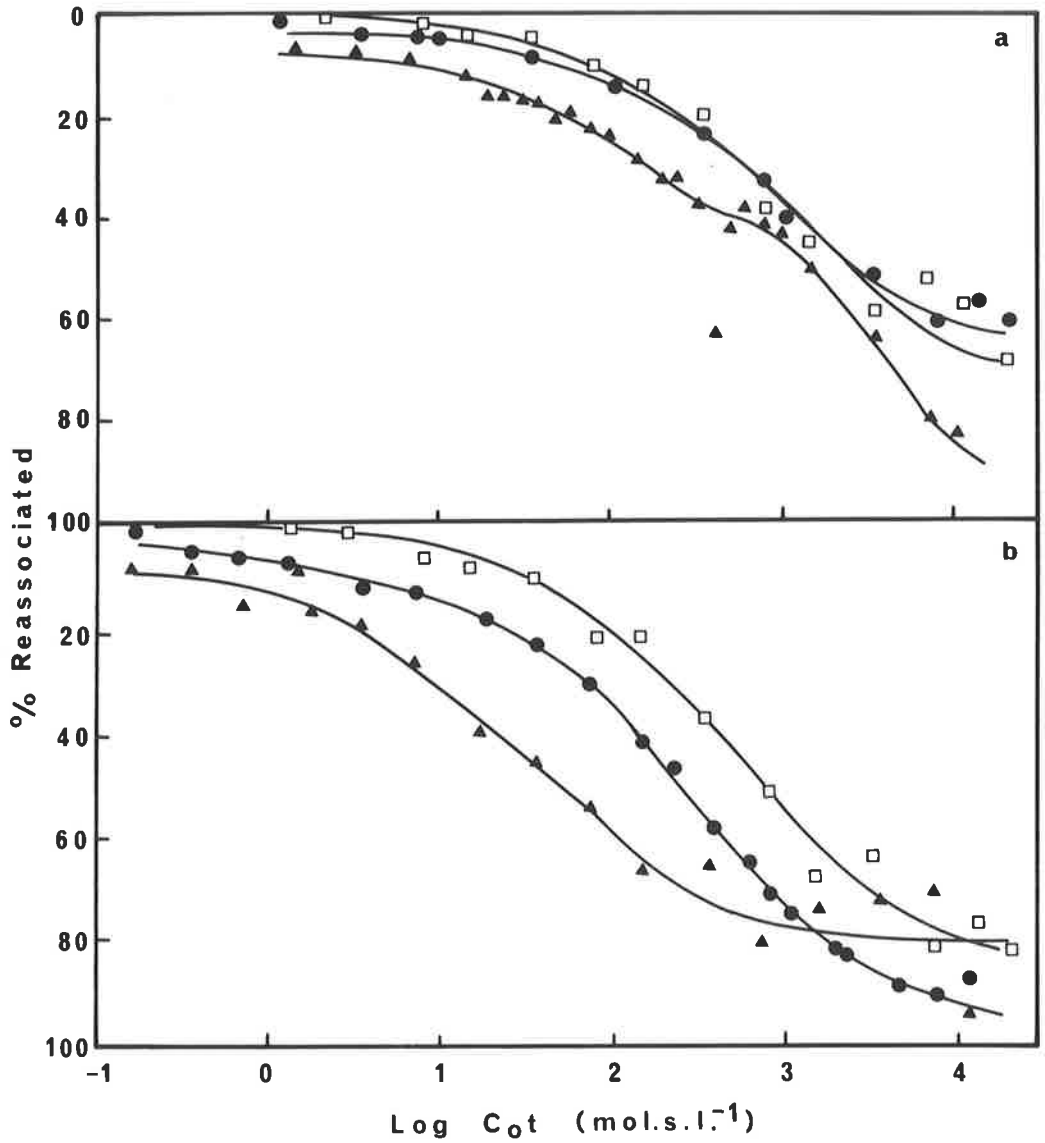
FIGURE 3.4.

REASSOCIATION OF KERATIN AMV- AND POLI-cDNAs WITH CHICK ERYTHROCYTE CHROMOSOMAL DNA

Chick DNA (average length 0.5 kb) was mixed with keratin AMV- or PolI-cDNA, denatured at 100°C and allowed to renature at 60°C. Aliquots were removed at appropriate intervals and the extent of renaturation estimated using S_1 nuclease (high salt buffer) and HAP. Normalization factors are shown.

(a) S_1 nuclease assays in high salt buffer; ▲ keratin AMV-cDNA (0%), □ keratin PolI-cDNA (10%), ● chick erythrocyte DNA, labelled by nick translation (14%).

(b) HAP assays; ▲ keratin AMV-cDNA (10%), □ keratin PolI-cDNA (11%), ● chick erythrocyte DNA (A_{260}) (18%).


a
b
0
20
40
60
80
100
% Reassociated
-1
0
1
2
3
4
Log Cot (mol.s.l.-1)

ciation was measured by resistance of the cDNAs to digestion with S_1 nuclease (Figure 3.4(a)). In contrast, less than 10% of keratin PolI-cDNA had reassociated by a $C_o t_{\frac{1}{2}}$ of 9×10^1 mol.s.l^{-1} and there was only one major transition with a $C_o t_{\frac{1}{2}}$ of 9.5×10^2 mol.s.l^{-1}, a value indistinguishable from that of the unique sequence fraction of chick erythrocyte DNA (Figure 3.4(a)).

When assayed on HAP, as shown in Figure 3.4(b), keratin AMV-cDNA reassociated in a broad curve, indicating the presence of both reiterated and unique sequences. In some experiments, the curve appeared to be biphasic, but insufficient points made it difficult to make accurate estimates of the $C_o t_{\frac{1}{2}}$ for the transitions. A biphasic curve was previously reported by Kemp (1975) for the AMV-cDNA reassociation reaction. The $C_o t_{\frac{1}{2}}$s for the transitions were estimated as 7×10^0 and 9×10^2 mol.s.l^{-1}. Keratin PolI-cDNA reassociated with a single transition at a $C_o t_{\frac{1}{2}}$ of about 4.7×10^2 mol.s.l^{-1}, a value again indistinguishable from that of the unique sequence fraction of chick erythrocyte DNA (Figure 3.4(b)). The differences in reassociation of AMV- and PolI-cDNAs were reproducible with different batches of each cDNA and also when the cDNAs were each annealed with the same batch of sheared chick erythrocyte DNA, ruling out the possibility that the differences were caused by variations in the fragment size of the chick erythrocyte DNA.

6. Reassociation of Elongated PolI-cDNA to Chick Erythrocyte DNA

If PolI- and AMV-cDNAs were different length

copies of the same mRNA species, then elongated PolI-cDNA should show similar reassociation kinetics to AMV-cDNA. PolI-cDNA elongation was performed as described by Crawford and Wells (1976) and the method is outlined in Figure 3.5. ^{3}H-labelled PolI-cDNA was hybridized to keratin mRNA and the hybridized cDNA was used as primer for AMV reverse transcriptase in a synthesis mix containing all 4 deoxyribonucleoside triphosphates unlabelled. After reverse transcription, the elongated probe with a labelled 3' end and an unlabelled 5' section was separated from the mRNA by alkali digestion. Radioactively labelled products ranging in length from 0.2 to 0.65 kb were obtained, as estimated from their velocity of sedimentation through linear 10% to 40% sucrose gradients relative to PolI- and AMV-cDNA markers (Figure 3.6). Three size classes of increasing length were selected from these gradients and the HAP assayed reassociation kinetics of each of these were compared with those of the starting PolI-cDNA (Figure 3.7). From Figure 3.7 it is clear that the apparent rate of reassociation increased with increasing length of the probe. In fact, the curve obtained with elongated PolI-cDNA of average length 0.55 kb was almost indistinguishable from that for the AMV-cDNA reassociation reaction (Figure 3.4(b)).

The radioactive label is confined to the 5' 150 nucleotides of these elongated molecules and since the PolI-cDNA before extension only showed unique reassociation kinetics (Figure 3.4(b)) the sequence of cold nucleotides added by the AMV polymerase, as directed by the mRNA template, must represent a reiterated sequence in the chick genome.

FIGURE 3.5.

SCHEMATIC REPRESENTATION OF THE ELONGATION OF POLI-cDNA WITH AMV POLYMERASE

See text for description. For experimental details see Crawford and Wells (1976).

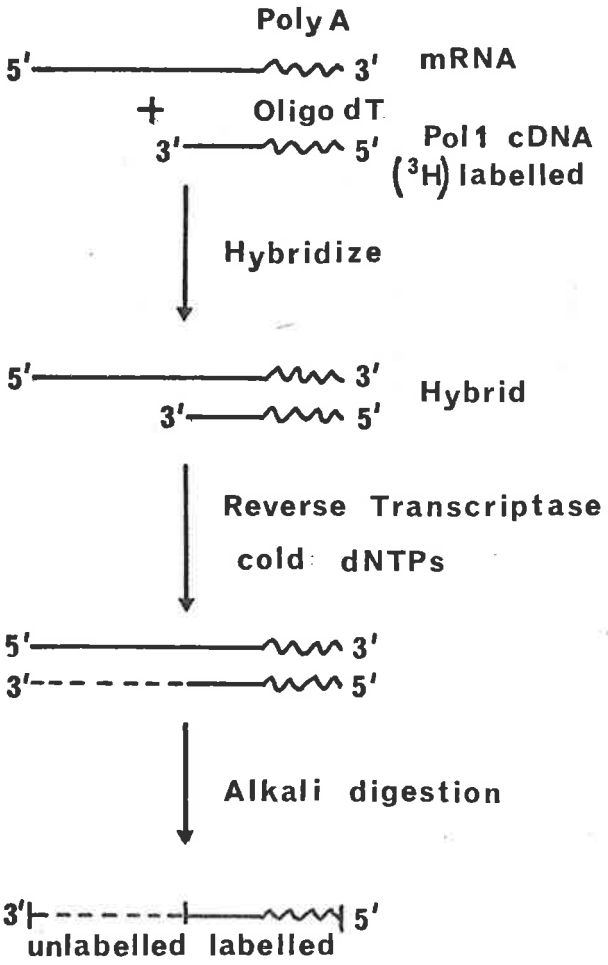

Poly A
5'
3'
mRNA
+
Oligo dT
3'
5'
Pol 1 cDNA
(3H) labelled
Hybridize
5'
3'
3'
5'
Hybrid
Reverse Transcriptase
cold dNTPs
5'
3'
3'
5'
Alkali digestion
3'
5'
unlabelled
labelled

FIGURE 3.6.

SELECTION OF SIZE CLASSES OF ELONGATED POLI-cDNA

Elongated keratin PolI-cDNA was centrifuged through 10-40% sucrose gradients in 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1 M NaCl at 37,000 r.p.m. for 16 hours at 3°C using a Beckman SW41 rotor in a Beckman L5-50 preparative ultra-centrifuge. Gradients were fractionated into 0.5 ml fractions using an ISCO density gradient fractionator, and 10 µl aliquots from each fraction were counted in Toluene Triton scintillation fluid to detect elongated molecules. AMV- and PolI-cDNAs were centrifuged on parallel gradients to provide molecular weight markers, and their banding positions were determined as described above. The bars indicate fractions which were pooled to give the different size classes; ■ PolI-cDNA; ● AMV-cDNA; O elongated PolI.

Sedimentation is from right to left.

1. 220 Bases average length.

2. 350 Bases average length.

3. 550 Bases average length.

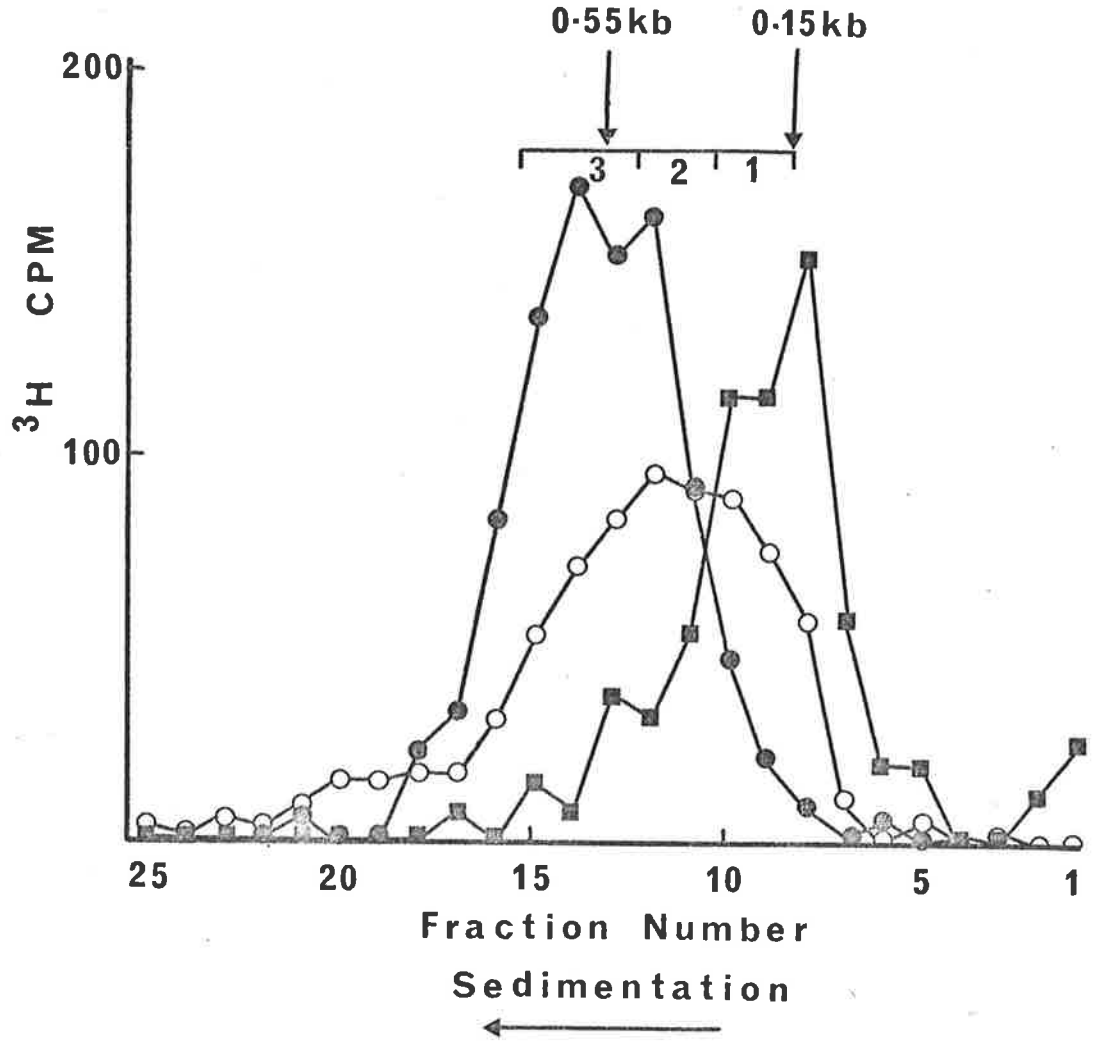
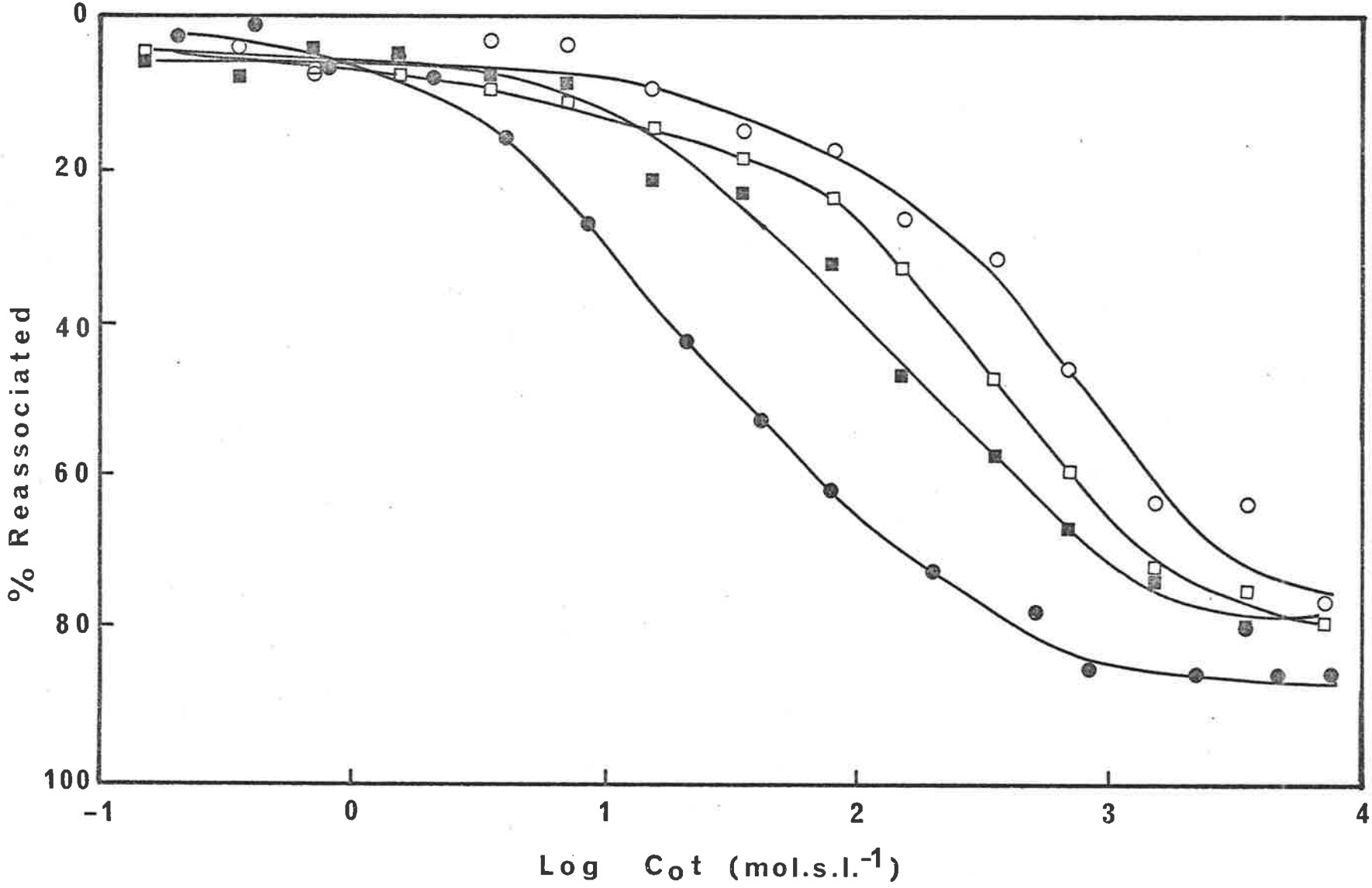

0·55kb
0·15kb
3
2
1
200
100
^{3}H CPM
25
20
15
10
5
1
Fraction Number
Sedimentation

FIGURE 3.7.

REASSOCIATION OF ELONGATED KERATIN POLI-cDNA TO CHICK CHROMOSOMAL DNA

All assays were carried out on HAP. The average size of the elongated PolI-cDNA size classes was estimated from their velocity of sedimentation through linear 10-40% sucrose gradients relative to AMV-cDNA and PolI-cDNA markers run on parallel gradients (Figure 3.6). Normalization factors are shown.

○ PolI-cDNA (10%), □ 220 base elongated PolI-cDNA (0%), ■ 350 base elongated PolI-cDNA (8%), ● 550 base elongated PolI-cDNA (21%).


0
20
40
60
80
100
% Reassociated
-1
0
1
2
3
4
Log Cot (mol.s.l.-1)

These results suggest that keratin mRNA has a 3' end of 150 nucleotides representing unique sequences in the chick genome, covalently linked to a reiterated sequence located further towards the 5' end of the messenger.

7. Thermal Stabilities of Reassociated Duplexes of Keratin AMV- and PolI-cDNA

In this section, the thermal denaturation studies described for AMV-cDNA reassociated duplexes were carried out by Kemp (Figure 3.8(a) and (b)).

As shown in Figure 3.8(a), the thermal stabilities of the keratin AMV-cDNA duplexes changed markedly with $C_o t$, two major components with T_ms of about 73°C and 90°C being evident. Replots of these curves (Figure 3.8(b)) demonstrate clearly that more than 60% of the duplexes with T_ms less than 85°C, which were present at $C_o t$ 1.9 x 10^1 mol. $s.1^{-1}$, had transferred into duplexes of high thermal stability by $C_o t$ 1.14 x 10^4 mol.$s.1^{-1}$. In contrast, there was a much smaller change in thermal stability of keratin PolI-cDNA duplexes with $C_o t$ (Figure 3.8(c)) and no significant transfer of duplexes from low thermal stability to high thermal stability (Figure 3.8(d)). One major component with a T_m of approximately 83°C was evident, although in addition there were some duplexes with T_ms lower than 70°C.

8. Buoyant Densities of Reassociated Duplexes of Keratin AMV- and PolI-cDNAs in the Presence of Actinomycin D

Studies on the buoyant density in actinomycin D CsCl of native DNA fragments containing sequences comple-

FIGURE 3.8.

THERMAL STABILITIES OF KERATIN AMV- AND POLI-cDNA REASSOCIATED DUPLEXES

Reaction mixes identical to those described in Chapter III.B.3.(a) were incubated to the various $C_o t$ values shown and the thermal stabilities determined (Chapter III.B.3.(b).

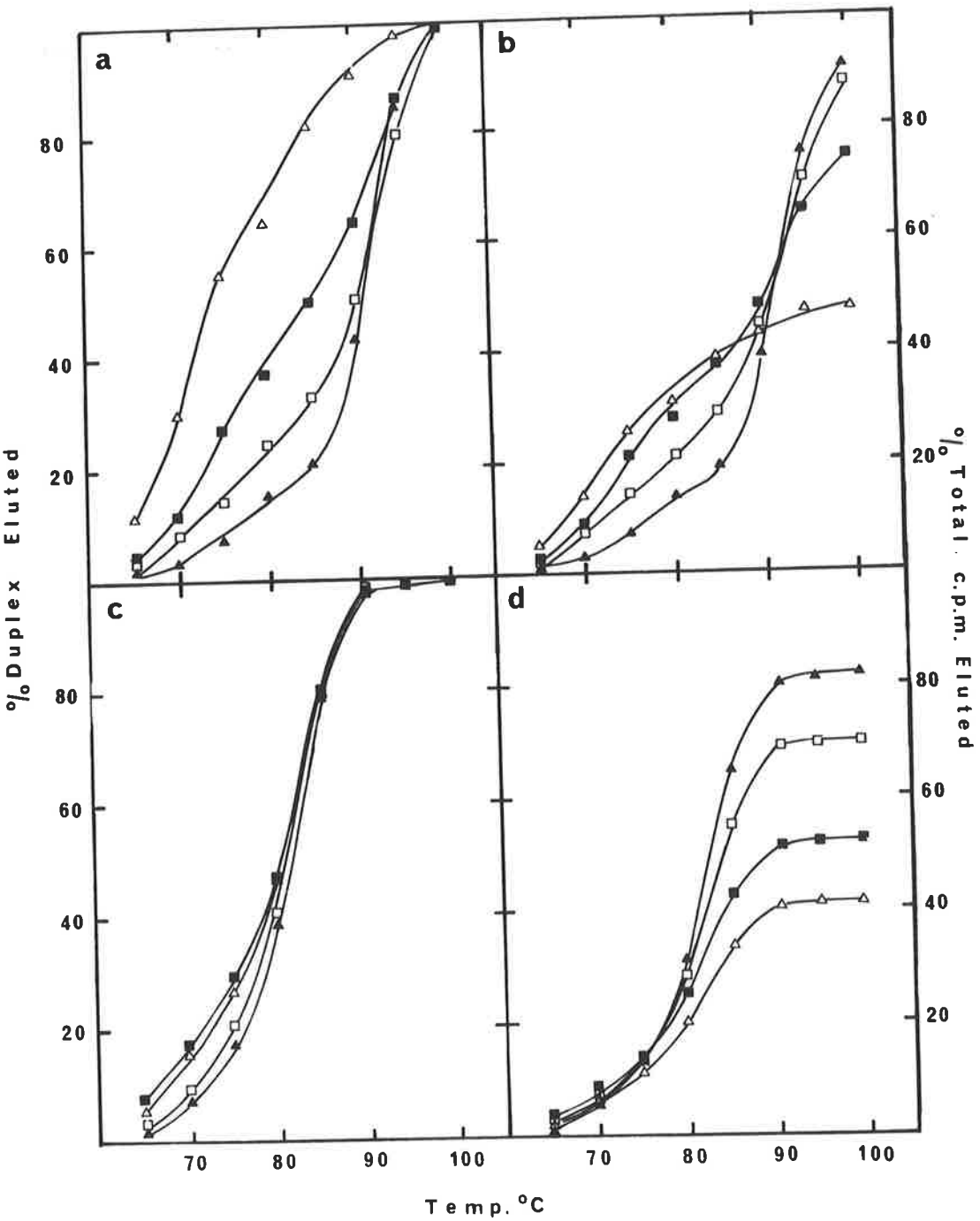
(a) Keratin AMV-cDNA duplexes after incubation to $C_o t$ values (mol.s.l^{-1}) of △ 1.9×10^{-1}, ■ 5.7×10^{2}, □ 3.75×10^{3}, ▲ 1.14×10^{4}.

(b) The same data as in (a) re-plotted after multiplying each value by the normalized % of total cDNA which bound to HAP at 60°C.

(c) Keratin PolI-cDNA duplexes after incubation to $C_o t$ values (mol.s.l^{-1}) of △ 4.9×10^{2}, ■ 9.3×10^{2}, □ 3.1×10^{3}, ▲ 1.6×10^{4}.

(d) The same data as in (c) re-plotted as in (b).

N.B.: Data in Figure 3.8(a) and (b) was obtained by Kemp.


a
b
c
d
% Duplex Eluted
% Total c.p.m. Eluted
Temp. °C
20
40
60
80
70
80
90
100

mentary to keratin AMV-cDNA (Lockett and Kemp, 1975) demonstrated that this DNA contained G + C rich sequences. The density of duplexes of keratin AMV-cDNA and PolI-cDNA, formed by reassociation of the cDNAs with chick erythrocyte DNA to high $C_o t$, was therefore investigated. The duplexes were treated with S_1 nuclease before actinomycin D-CsCl gradient centrifugation in order to remove single-stranded regions, mismatched regions and oligo-dT tails on the cDNA molecules.

The main band of reassociated chick erythrocyte DNA had a density of 1.590 gm/cc on actinomycin D-CsCl gradients (Figure 3.9). Keratin AMV-cDNA duplexes banded at an average density of about 1.566 gm/cc (Figure 3.9(a)). In contrast, keratin PolI-cDNA duplexes banded at an average density of 1.582 gm/cc (Figure 3.9(b)). Since actinomycin D binds preferentially to G + C rich sequences (Kersten, Kersten and Szybalski, 1966) resulting in a displacement of such sequences to the light side of CsCl gradients, keratin AMV-cDNA must contain a G + C rich sequence which is not present in keratin PolI-cDNA. Although there was slight variation in the apparent density of the main band DNA in independent experiments, the variation in density difference between main band DNA and cDNA duplexes was 0.002 gm/cc or less in 3 independent experiments. When the chick erythrocyte DNA was replaced with calf thymus DNA, no radioactivity above background was detectable on the gradients (results not shown). Since the duplexes had been treated with nuclease S_1 prior to centrifugation, we assume that the apparent densities of the cDNA duplexes would not have been significantly affected by the presence of single-stranded or mismatched regions.

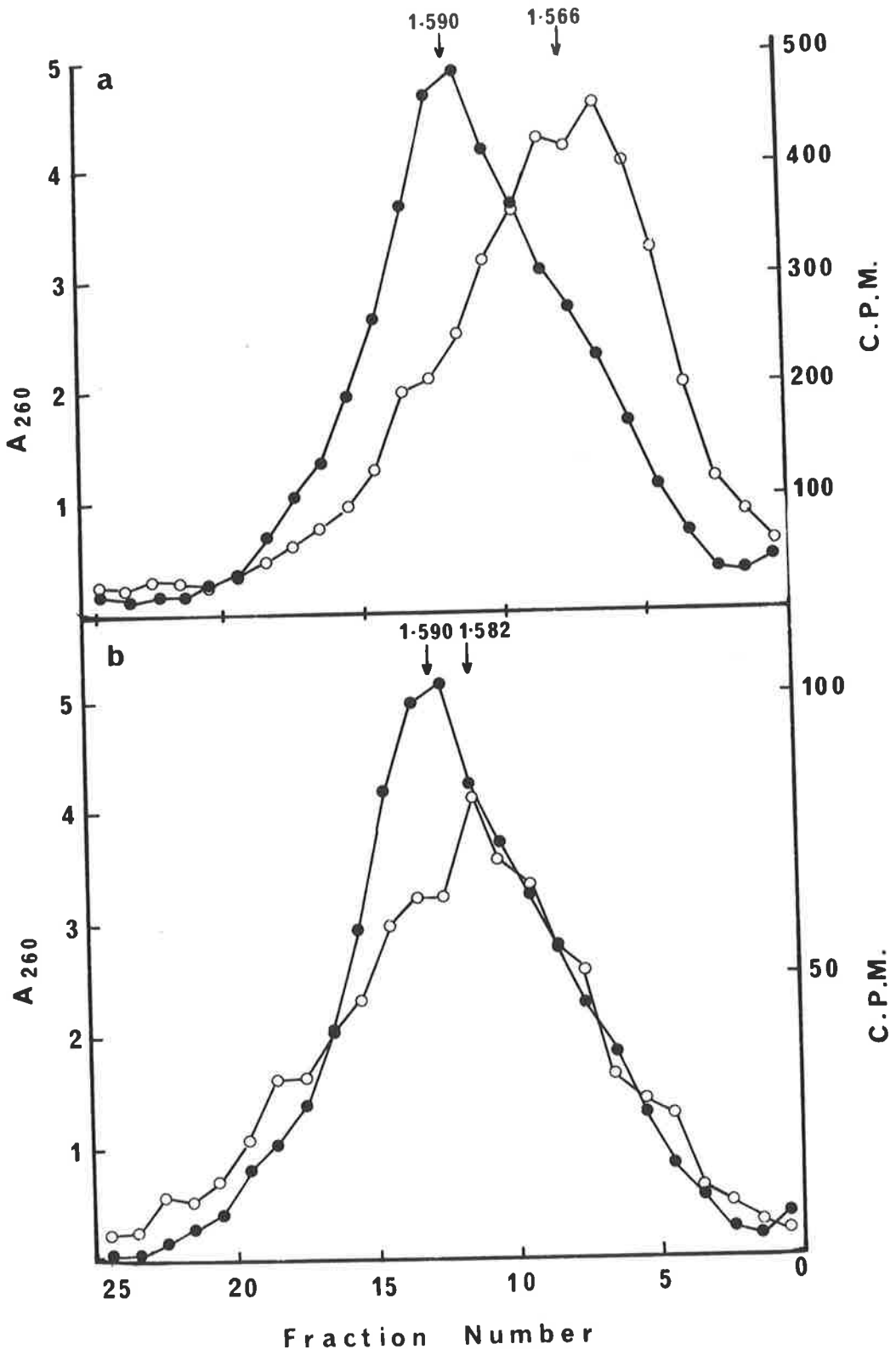
FIGURE 3.9.

CENTRIFUGATION OF KERATIN AMV- AND POLI-cDNA REASSOCIATED DUPLEXES ON ACTINOMYCIN D-CsCl GRADIENTS

Duplexes formed at $C_o t$ 2.0 x 10^4 mol.s.l^{-1} were treated with S_1 nuclease and then centrifuged on actinomycin D-CsCl gradients as described in Chapter III.B.4.

(a) ○ Keratin AMV-cDNA duplexes (c.p.m.),
● chick erythrocyte DNA (A_{260}).

(b) ○ Keratin PolI-cDNA duplexes (c.p.m.),
● chick erythrocyte DNA (A_{260}).


a
1·590
1·566
5
4
3
2
1
A 260
500
400
300
200
100
C.P.M.
b
1·590 1·582
5
4
3
2
1
A 260
100
50
C.P.M.
25
20
15
10
5
0
Fraction Number

D. DISCUSSION

1. Analysis of mRNA Preparations and cDNAs

Although keratin mRNA preparations have been previously reported to be greater than 95% pure by the criteria of cell-free translation and formamide gel electrophoresis (Partington et al., 1973; Kemp et al., 1974b; Kemp et al., 1974c) the data in Figure 3.1 indicate that as much as 30% of any given preparation might be ribosomal breakdown products. Most of the ribosomal contaminant appeared to be breakdown material from 18S rRNA. Tables 3.1 and 3.2, however, provide evidence that the cDNA transcribed from these messenger preparations contained insignificant amounts of ribosomal reverse transcripts, regardless of whether AMV reverse transcriptase or PolI was used for the synthesis. The facts that both of these enzymes required oligo-dT to prime cDNA synthesis (Table 3.1) and AMV- and PolI-cDNAs averaged 0.55 and 0.15 kb in length respectively (Figure 3.2), suggest that these two cDNAs represent different length copies of the keratin mRNA, both copies initiating at the 3' poly(A) tract. The competition hybridization experiments shown in Figure 3.3 clearly demonstrated that PolI- and AMV-cDNAs share common sequences and hence that the two enzymes did not copy different sets of the mRNA population. All of these results, taken together, suggest that AMV- and PolI-cDNAs truly represent different length copies of the keratin mRNA population. The AMV-cDNA represents a 0.55 kb copy of the messenger starting at the 3' poly(A) tract, while PolI-cDNA is representative of the 150 nucleotides at the 3' end of the keratin mRNA.

2. Reassociation and Thermal Properties of AMV-cDNA and PolI-cDNA Duplexes

The data in Figure 3.4 indicate that the kinetics of reassociation of AMV-cDNA and PolI-cDNA to a vast excess of chick erythrocyte DNA were quite different. The AMV-cDNA showed a broad transition, indicative of the presence of reiterated and unique sequences, while the PolI-cDNA reassociated at a rate indistinguishable from that of unique sequences.

The reassociation kinetics of AMV-cDNA have been previously explained (Kemp, 1975) by proposing that each keratin AMV-cDNA molecule contained a region of very similar sequence which could form poorly matched duplexes with the corresponding regions of other non-identical keratin genes. The reiterated kinetics were attributed to cross-hybridization since they were not apparent when the reassociation reaction was assayed under very stringent low salt S_1 assay conditions (Kemp, 1975). As the assay conditions became less stringent (high salt S_1 assay, HAP assay, Figure 3.4) the reiterated kinetics became more obvious. Another interpretation, however, which is indistinguishable from this mismatching hypothesis on the presently available data, would propose that each AMV-cDNA molecule contained a short faithfully conserved region which could form short perfectly matched duplexes with the corresponding regions of non-identical keratin genes. This sequence will be denoted the repetitive sequence in the sense that it must occur many times in the chick genome, while avoiding detailed description of the nature of this repetitive sequence (i.e., whether

it is a short common repeat or a longer mismatching repeat. Figure 3.8(a) and (b) indicate that the duplexes formed early in the AMV-cDNA reassociation were of low thermal stability, a result in complete agreement with the hypothesis that the reiterated sequence is either short or mismatched. The observation that the same cDNA molecules present in these low melting point duplexes early in the reactions were later transferred into duplexes of much greater thermal stability can be explained by proposing that each keratin AMV-cDNA molecule also contains a sequence which has diverged to the extent that it cannot cross-hybridize with non-identical keratin genes and, therefore, behaves as if it is unique in the genome. On the assumption that the multiple keratin genes have arisen by gene duplication, this model would predict that different regions of the keratin genes have diverged to varying extents, the reiterated regions showing little divergence while the unique regions have diverged to a much greater extent.

The behaviour of keratin PolI-cDNA observed here was entirely consistent with this model since they behaved as if they contained only unique sequences. It follows that unique sequences must be located adjacent to the poly(A) tract of keratin mRNA. The reassociation kinetic data obtained using elongated PolI-cDNA as probe would then suggest that the unique sequences at the 3' end of the mRNA molecule are covalently linked to a reiterated sequence further to the 5' end of the molecule.

An alternative model may also be considered in relation to the data. In this model it is assumed that

the degree of sequence divergence between different keratin mRNA species is distributed randomly along the molecules. This "random divergence" model can account for the observed behaviour of keratin AMV-cDNA:DNA duplexes since cDNA can be displaced from a duplex by other molecules, provided that more stable complexes can be formed (Beckmann and Daniel, 1974). In fact, the latter process can be invoked to explain the high thermal stability of the hybrids formed late in the reaction using either of the two models.

The two models lead to different predictions about the behaviour of PolI-cDNA however. Whereas the unique and repetitive sequence model predicts all the observed properties of the PolI-cDNA molecules, the random divergence model predicts, in principle, that the PolI-cDNA should be equally as repetitive in nature as the AMV-cDNA. While the latter was clearly not the case, the interpretation is complicated by the fact that short duplexes are intrinsically less stable than longer duplexes. It is, therefore, possible that mismatched duplexes of PolI-cDNA did not form because their thermal stability was too low relative to the incubation temperature. A quantitative evaluation of this possibility is necessary in order to distinguish between the two models.

Southern (1975a) demonstrated that restriction endonuclease fragments of mouse satellite DNA ranging in length from 120 to 600 bases could all reassociate to form mismatched duplexes. The depression of the T_ms of these duplexes from those of the native molecules due to mismatching was constant, irrespective of the length of the

fragments. The depression of T_ms of the shortest reassociated duplexes relative to long native duplexes could be accounted for as the sum of the depression due to chain length and the constant mismatching factor of 6°C. These data can be used to predict the thermal stability of the keratin PolI-cDNA duplexes expected for the random divergence model. The maximum T_m of keratin PolI-cDNA duplexes was lower by 7°C than that of the corresponding AMV-duplexes (Figure 3.8).

The thermal stability of globin PolI- and AMV-cDNA hybrids has been compared (Kemp et al., in preparation) and it was found that the PolI hybrids had a T_m 4°C lower than the AMV hybrid. This reduction in T_m was presumed to be due to the shorter length of the hybrid. Of the 7°C reduction in T_m for the keratin PolI-cDNA duplexes when compared to the AMV-cDNA duplexes, 4°C can presumably be accounted for by the reduced length of the duplex while the remaining 3°C can be attributed to the difference in G + C content of the PolI- and AMV-cDNA duplexes (Figure 3.9) The T_ms of mismatched PolI-cDNA duplexes should therefore be 7°C below those of mismatched AMV-cDNA duplexes (72-73°C; Figures 4 and 7) or about 65-66°C. Clearly, since the reaction was done at 60°C, the rate of formation of many of these duplexes would be negligible (Bonner et al., 1973). However, because of the broad range of T_ms exhibited by mismatched AMV-cDNA duplexes (Figure 3.8(a) and (b)) about half of the mismatched PolI-cDNA duplexes would be expected to have T_ms distributed between 65°C and 75°C. Such duplexes should form at a rate comparable to those of the mismatched AMV-cDNA duplexes (Bonner et al., 1973) after

correcting for the effect of length on rate.

The random divergence model, therefore, predicts that the assay of keratin PolI-cDNA reassociation kinetics on HAP should result in a complex reassociation curve showing both unique and reiterated characteristics. Since PolI-cDNA was observed to reassociate with unique kinetics, it was concluded that the random divergence model is incompatible with the data and hence that it is most probable that keratin mRNA contains distinct regions of unique and repetitive sequence. This conclusion depends on the validity of the assumption that the depression of the T_m of a duplex due to random mismatching is independent of length, an effect that has apparently only been shown for mouse satellite DNA.

3. Model for the Sequence Organization of Keratin mRNA

The data available suggest a model for the sequence organization in keratin mRNA (Figure 3.10). The experimental data, assumptions and deductions upon which this model is based, are summarized below.

(a) The length of keratin mRNA is about 0.8 kb (Kemp et al., 1974b), it contains a 3' poly(A) tract of average length 60 nucleotides and a 5' 7 methyl G cap structure (Morris and Rogers, in press).

(b) At least 150 bases adjacent to the poly(A) tract are unique sequences (present results). We assume that the actual length is 150 bases and designate this region as 3'

unique sequences.

(c) About 30-50% of the length of each keratin AMV-cDNA molecule, 550 bases long, is composed of repetitive sequences, since in separate experiments 30-50% of the radioactivity in these molecules appears as a fast transition resistant to nuclease S_1 at high salt concentration when reannealed with chick erythrocyte DNA (Figure 3.4(a)).

(d) The reassociation kinetic analysis of elongated PolI-cDNAs of varying lengths (Figure 3.7) also suggests that the repetitive sequence is fairly long and extends to within close proximity (100 nucleotides) of the 3' unique region of the mRNA. The presently available data do not allow any further conclusions to be made about the nature, size or position of this repetitive sequence. Most or all of the repetitive sequences are covalently attached to unique sequences since elongated PolI-cDNA shows reiterated reassociation kinetics when reannealed with chick erythrocyte DNA (Figure 3.7).

(e) Low melting point duplexes of the repetitive region were depressed in T_m by 18-20°C from well-matched duplexes (Figure 3.8(a) and (b)). Therefore, the sequences are mismatched at about 20% of the bases

(Bonner et al., 1973) and hence have diverged by about 10% in sequence (Southern, 1971). Alternatively the repetitive region is very short and highly conserved.

(f) Keratin chains are about 100 amino acids long (O'Donnell, 1973b; Walker and Rogers, 1976a). They do not appear to be synthesized as larger polypeptide precursors, since the molecular weight of keratin translated from keratin mRNA in both rabbit reticulocyte and wheat embryo cell-free systems was identical to that of native keratin and the N-terminal sequence was identical (Partington et al., 1973; Kemp et al., 1974c). Therefore the keratin coding sequence is about 300 bases long.

(g) Complete amino acid sequences have been determined for feather keratin chains from other species (O'Donnell, 1973b; O'Donnell and Inglis, 1974) and partial sequences of tryptic peptides from embryonic chick feather keratins have also been determined (Walker and Rogers, 1976b; Kemp et al., 1975). While regions of about 10 amino acids at the N- and C-termini vary considerably in sequence, the internal section of about 80 amino acids is highly conserved in the keratin chains which have been

sequenced. The limited sequence studies on embryonic chick feather keratins in this internal region demonstrated that the sequences of 4 different N-terminal tryptic peptides were identical in the region from amino acids 12-28 except for a single proline → serine substitution in one of the four sequences. In the absence of further sequence data we conclude that the amino acid sequence of the internal region of all embryonic chick feather keratins are very similar. If these regions were identical in amino acid sequence then they could differ by a maximum of about 17% in their mRNA sequence (Kohne, 1970). From (e), if the low melting point of early duplexes was due to mismatching then the repetitive sequences in the keratin mRNA would have diverged by about 10%, a value compatible with a suggestion that the repetitive sequence in keratin mRNA is in the coding sequence.

(h) From the amino acid composition of feather keratin (Walker and Rogers, 1976a) it can be predicted that the DNA coding for keratins should have a G + C content of about 54.7%, compared with 41% for total chick DNA. Such a G + C rich region is present at a distance greater than 150 bases from the poly(A) region, again

suggesting that the coding sequence is in this region.

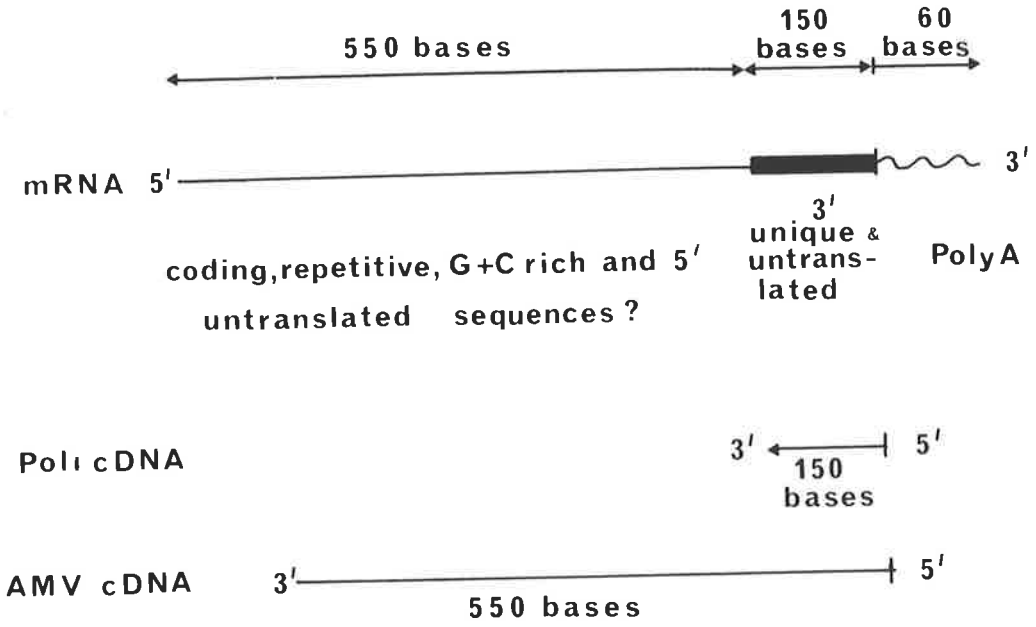
Thus it appears that keratin mRNA contains unique and repetitive sequences covalently linked. The 3' 0.15 kb of the mRNA constitutes at least a portion of the unique sequence, while the reiterated sequence is located further towards the 5' end of the mRNA molecule. This reiterated sequence could be relatively long and mismatching or short (about 50 bases) and faithfully conserved. The apparent increase in the rate of reassociation of elongated PolI-cDNA with increasing length, however, favours the mismatching interpretation. On the basis of the sequence homology of different feather keratin proteins, it is tempting to speculate that the reiterated sequence corresponds to the keratin coding sequence

FIGURE 3.10.

MODEL FOR THE ARRANGEMENT OF SEQUENCES IN KERATIN mRNA

The 150 nucleotides adjacent to the 3' poly(A) tract constitutes a unique untranslated sequence. Further to the 5' end of the mRNA and covalently linked to this 3' unique sequence is a repetitive sequence (short and faithfully conserved or longer and mismatching), a G + C rich sequence, the coding sequence and presumably a 5' untranslated sequence. The currently available data permit no further conclusions about the sequence arrangement within this 5' region of 550 bases.

The orientations and approximate sizes of AMV- and PolI-cDNAs are also shown.


550 bases
150 bases
60 bases
mRNA 5'
3'
3' unique & untrans-lated
PolyA
coding, repetitive, G+C rich and 5' untranslated sequences ?
Poli cDNA
3'
5'
150 bases
AMV cDNA
3'
5'
550 bases

CHAPTER IV

SEPARATION OF RIBOSOMAL AND KERATIN SEQUENCES BY PHYSICO-CHEMICAL METHODS

CHAPTER IV

SEPARATION OF RIBOSOMAL AND KERATIN SEQUENCES BY PHYSICO-CHEMICAL METHODS

A. INTRODUCTION

This chapter describes attempts to obtain a partial purification or quantitative separation of chick keratin and ribosomal sequences. The proposed hybrid-density gradient procedure for the investigation of gene arrangement relies on the density difference between RNA:DNA hybrids and single-stranded DNA in Cs_2SO_4 (see Chapter V.A). To generate hybrid without any DNA reannealing, hybrid formation is carried out in vast RNA excess. For the method to be able to contribute information about gene arrangement, there are two basic requirements:

1. That the sequences being studied constitute a high enough proportion of the total DNA being used in the experiment to allow their detection.

2. That the RNA being used to form hybrids is pure or that the DNA used in hybrid formation is devoid of those sequences which are complementary to the major contaminating RNA species.

From the $C_o t_{\frac{1}{2}}$ for total chick DNA (Kemp, 1975), it can be calculated from Laird (1971) that the genome size of chick is about 1.2×10^{12} daltons. This corresponds to 1.9×10^6 kb of DNA per genome. Reassociation kinetic analysis has revealed that there are about 100 ribosomal genes in the chick genome (P. Krieg, personal communication). Since the

amount of DNA coding for mature rRNA species per ribosomal cistron is about 7 kb, the total length of sequence coding for mature rRNA per genome is about 700 kb. From this it can be calculated that the sequences coding for mature rRNA species constitute about 0.037% of the chick genome, a figure consistent with that calculated by Sinclair and Brown (1971). This represents a sufficiently large proportion of the genome to permit ribosomal sequences to be readily detected without any pre-selection for these sequences. $C_o t$ analysis has also shown that there are about 100 genes coding for keratin in the chick genome (Kemp, 1975). The length of keratin mRNA is about 800 bases (Kemp *et al.*, 1974b), 60 of which make up the poly(A) tail at the 3' end of the molecule (Morris and Rogers, in press). Thus the total length of DNA complementary to keratin sequences in the chick genome is about 74 kb. It can therefore be calculated that keratin sequences constitute 0.0039% of the chick genome; about $^1/_{10}$th the amount attributable to ribosomal sequences.

From Chapter III.C.1 it is known that ribosomal sequences are the major contaminating RNA species in keratin mRNA preparations. While it is possible to remove some of these contaminating species by oligo-dT cellulose chromatography of the mRNA, some ribosomal sequences remain adsorbed to the column and elute with the messenger. Although the keratin mRNA has a fairly long poly(A) tract, for some unknown reason only 10% to 30% of the total mRNA loaded on to oligo-dT cellulose binds to it. The unbound material has been shown to contain about 70% of the loaded keratin messenger as determined by *in vitro* translation (B.C. Powell, personal communication). The hybrid-density gradient pro-

cedure, however, requires large amounts of highly purified mRNA and any attempts to remove contaminants from keratin mRNA preparations results in large losses of messenger. Added to this, the low yields of keratin mRNA from any given preparation (20-50 μg) and the low proportion of the genome coding for keratin, made it necessary to investigate the possibility of separating genomic ribosomal and keratin sequences with a view to enriching for keratin sequences.

It has been known for many years that the sequences coding for 18S and 28S rRNA in chick have a G + C content of 64% (Sinclair and Brown, 1971). While keratin genes are also G + C rich (54.5% G + C) compared with total chick DNA (41%), they appear to be associated with DNA of lower G + C content (Lockett and Kemp, 1975). The possibility of exploiting the different G + C contents of keratin and ribosomal sequences to obtain a quantitative separation of these sequences was, therefore, investigated. It has been reported (G. Partington, personal communication) that Xenopus globin sequences can be purified 10-fold by thermal elution from HAP using DNA of length 10 kb. Since DNA of higher molecular weight was required for the proposed hybrid-density gradient method for the analysis of gene organization, thermal elution studies were performed using DNA of single-stranded size $\geq$ 50 kb. The fractionation of ribosomal and keratin sequences on CsCl gradients was also studied using native DNA of high single-stranded molecular weight. It is known that when high molecular weight native DNA is fractionated on CsCl density gradients, the G + C rich keratin sequences band with the bulk of chick DNA as a result of their association with sequences of lower G + C content

(Lockett and Kemp, 1975). This effect may permit a greater separation of keratin and ribosomal sequences than would be expected on the basis of the G + C content of their coding sequences alone. After quantitative separations of ribosomal and keratin sequences by the CsCl density procedure, keratin sequences could be further purified by cycles of CsCl gradient centrifugation in the presence of actinomycin D and netropsin sulphate, as previously described (Lockett and Kemp, 1975).

B. METHODS

1. Thermal Elution from HAP

(a) In aqueous buffer

Using a glass column surrounded by a water jacket linked to a Haarke waterbath, HAP (prepared as described in Chapter II.B.5) was washed and equilibrated with 0.12 M phosphate, pH 7.0 at 60°C. High molecular weight DNA (prepared as described in Chapter II.B.4) ($\geqslant$ 50 kb) in 2 ml of the same buffer was loaded on to the column, allowed to equilibrate to 60°C for 2 min and then passed through the HAP under slight pressure. The column was washed a further 2 times with 2 ml of buffer equilibrated to 60°C. All 60°C eluates were pooled and stored. The equilibration temperature was raised in 5°C steps to 100°C and 3 washes, with 2 ml of 0.12 M phosphate (0.18 M Na^+), pH 7.0, were carried out at each temperature. After the 100°C wash, the column was eluted with 0.4 M (0.6 M Na^+) phosphate, pH 7.0, at 70°C to remove any

remaining DNA. Eluates from each equilibration temperature were pooled and stored separately.

Fractions were cleared of HAP fines by centrifugation in an MSE Superminor for 5 min at maximum speed. The DNA content of each eluate was determined from A_{260} measurement using a Zeiss spectrophotometer.

(b) In buffer containing 50% formamide

The procedure was exactly as described above except that DNA was loaded on the column at 40°C in 0.12 M phosphate, pH 7.0, containing 50% formamide (Ajax Chemicals Ltd., Laboratory Reagent grade). Elutions were carried out in 5°C steps up to 95°C using this same buffer as described above. Any remaining DNA was removed with 0.4 M phosphate, pH 7.0, containing 50% formamide at 60°C. The DNA content of each eluate was determined by measurement of A_{280} (since formamide absorbs at 260 nm but not at 280 nm).

2. CsCl Equilibrium Density Gradient Fractionation

8.53 g Of CsCl was dissolved in a minimum volume of TE (Chapter II.B.4) and the solution made up to a total weight of 15.3 g with high molecular weight DNA solution (100, 200 or 400 µg in TE). This gave a final gradient volume of 9 ml and a density of 1.700 g/cc. Gradients were covered with paraffin oil and centrifuged in a Beckman model L2-50 preparative ultracentrifuge using a fixed angle Ti50 rotor at 20°C for 60 to 70 hours at 32,000 r.p.m.

Gradients were fractionated from the bottom into 23 drop (0.3 ml) fractions using a Gilson Microfractionator model FC-80E. The density gradient achieved was determined by reading the refractive index of every third fraction. Where 100 µg of DNA was fractionated on a gradient the A_{260} of the fractions was determined using a Zeiss spectrophotometer. Where 200 µg and 400 µg of DNA was fractionated, fractions were diluted 2 and 4 times respectively with water before A_{260} determination.

3. Detection of Ribosomal and Keratin Sequences

All fractions from a given experiment were made up to the same DNA content with E. coli DNA. Fractions were made 0.1 M with respect to NaOH to denature the DNA. They were neutralized by the addition of 0.1 volumes each of 1 M Tris-HCl pH 7.5 and 1 M HCl, made up to 10 x SSC by the addition of an equal volume of 20 x SSC and the DNA immobilized by filtering through nitrocellulose discs using a Universal membrane filter apparatus. Filters were washed twice with 20 ml of 6 x SSC, cut in half, dried and baked at 80°C in vacuo for at least 2 hours. Half filters were incubated in 2 x SSC containing 1 x Denhardt solution (0.02% each of Ficoll, BSA and polyvinylpyrrolidone, see Denhardt, 1966) at 65°C for at least 6 hours, then blotted dry. Dry filters were stacked in a siliconized scintillation vial containing saturating amounts of ^{125}I labelled 18S and 28S ribosomal RNA (Chapter II.B.7) or ^{32}P labelled keratin cDNA in 2 x SSC, 0.5% SDS, pH 7.2, and hybridization carried out for 20 hours at 65°C. Half filters were given 3 x 8 hour washes in 2 x SSC, 0.5% SDS, at 65°C. Filters were then

rinsed in 70% ethanol, dried at 65°C and the radioactivity bound to each filter determined.

When 200 or 400 µg of DNA was fractionated on CsCl gradients, ½ or ¼ respectively of each gradient fraction was treated in the manner described above.

C. RESULTS

1. Thermal Elution from HAP

Figure 4.1 shows the results obtained from analytical thermal elution studies. The bulk of the high molecular weight DNA appeared to denature with a T_m of 93°C in aqueous buffer (Figure 4.1(a)). In this same aqueous buffer, keratin sequences denatured with a T_m of 97°C while at 100°C only 25% of the ribosomal sequences could be eluted from the column. It was possible, however, to elute the remaining ribosomal sequences with 0.4 M phosphate, pH 7.0, at 70°C. These results suggested that, by carrying out preparative thermal elution at 95°C in aqueous buffer, 70% of the keratin sequences would remain bound to the HAP along with only 23% of the total chick DNA. If the bound material, after elution at 95°C with 0.12 M phosphate, pH 7.0, was eluted with 0.4 M phosphate, pH 7.0, at 70°C, this material would be enriched 3-fold for keratin sequences, but 4-fold for ribosomal sequences as well. Thus, to obtain enrichment of keratin sequences without enrichment for ribosomal sequences by this procedure, thermal elution should be carried out at 95°C to remove most of the chick DNA and again at 100°C to elute 70% of the keratin sequences, but only 25% of the ribosomal sequences. This would decrease

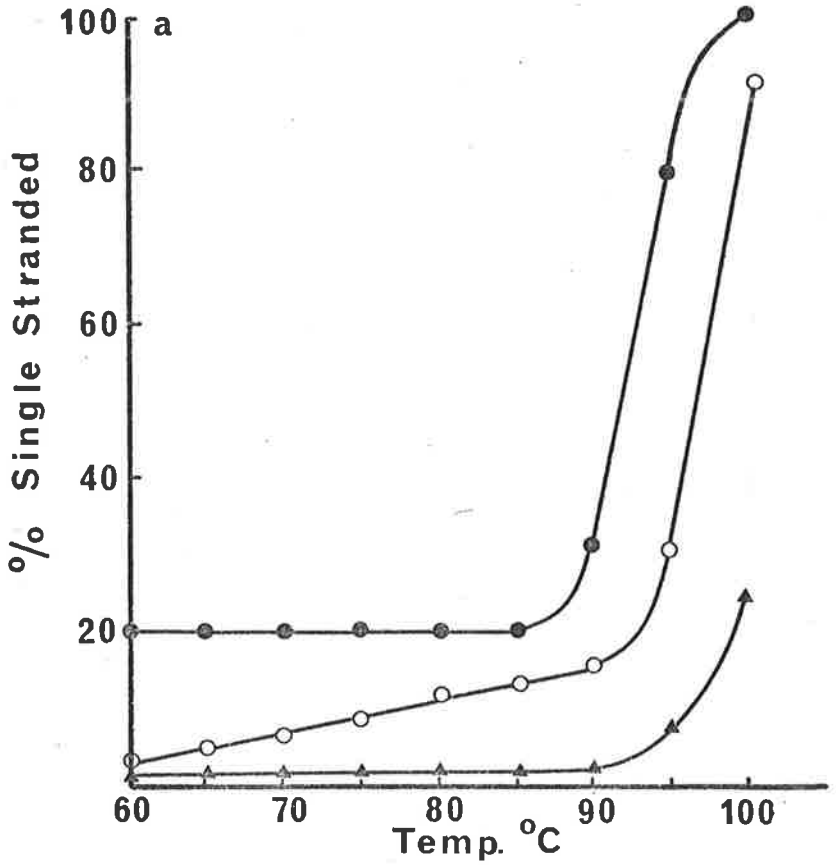
FIGURE 4.1.

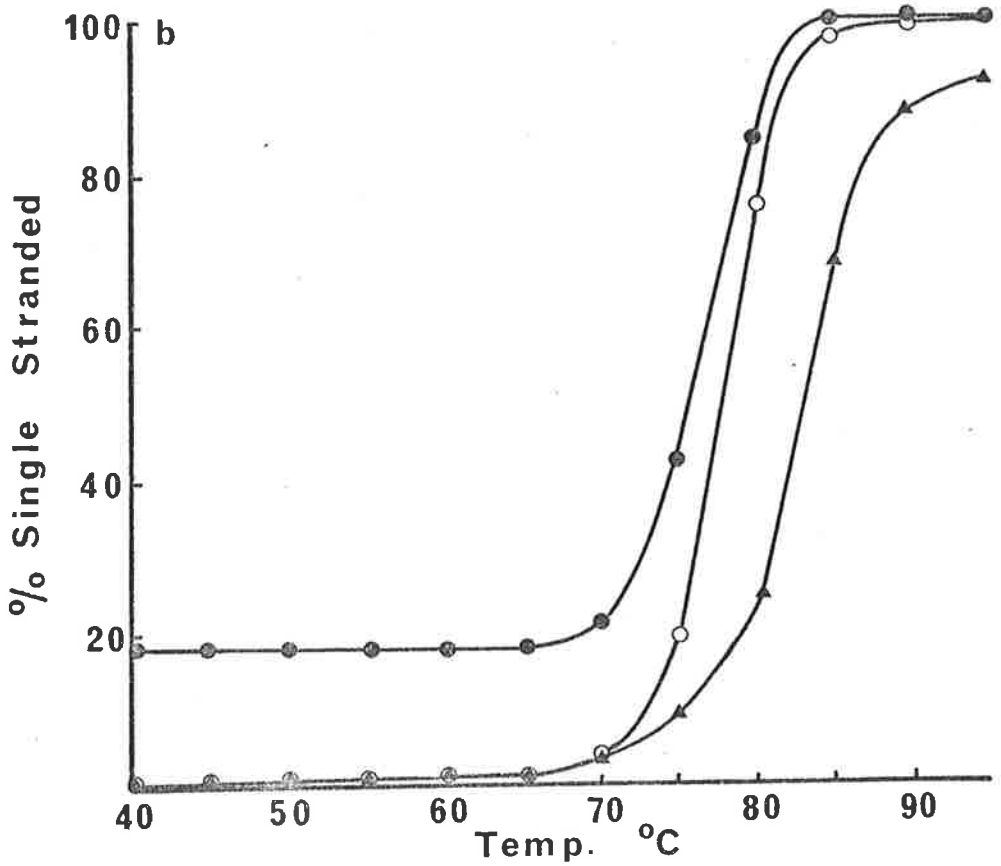
FRACTIONATION OF KERATIN AND RIBOSOMAL SEQUENCES BY THERMAL ELUTION FROM HAP

100 µg Of DNA was bound to HAP and eluted with phosphate buffer (0.18 M Na^+) at increasing temperatures using aqueous buffers and buffers containing 50% formamide as described in Chapter IV.B.1.(a) and (b). The DNA samples from each elution were denatured, neutralized and bound to nitrocellulose discs as described in Chapter IV.B.3. Filters were cut in half and challenged with either [^{3}H]-keratin AMV-cDNA or ^{125}I labelled rRNA. The points plotted for any given elution temperature represent the counts hybridized at that temperature and all lower elution temperatures, expressed as a percentage of the total counts hybridized after all elutions had been performed.

(a) In aqueous buffers; ● total chick DNA (A_{260}), ○ keratin sequences, ▲ ribosomal sequences.

(b) In buffers containing 50% formamide; ● total chick DNA (A_{280}), ○ keratin sequences, ▲ ribosomal sequences.


a
100
80
60
40
20
% Single Stranded
60
70
80
90
100
Temp. °C


b
100
80
60
40
20
% Single Stranded
40
50
60
70
80
90
Temp. °C

the enrichment for keratin sequences to 2.8-fold, but this fraction would show no significant enrichment for ribosomal sequences over unfractionated DNA. By elution with 0.4 M phosphate, pH 7.0, at 70°C after the 100°C thermal elution, a vast purification of ribosomal sequences could be obtained. From Figure 4.2(a), however, it is apparent that, at the high temperatures necessary for the partial purification process in aqueous buffers, gross nicking of the DNA occurs. Since this breakdown was undesirable for the proposed hybrid-density gradient system, analytical studies on thermal elution were repeated using buffers containing formamide.

Figure 4.1(b) shows the thermal elution data using 0.12 M phosphate, pH 7.0 in 50% formamide. Under these conditions total chick DNA exhibited a T_m of 77°C, keratin sequences showed a T_m of 78.5°C while ribosomal sequences denatured with a T_m of 83.5°C. These data suggested that, under these conditions, it was not possible to obtain any significant partial purification of keratin sequences from total chick DNA. It should be possible, however, to obtain a significant purification of ribosomal sequences. If the double-stranded fraction was eluted with 0.4 M phosphate, pH 7.0 in 50% formamide at 60°C after thermal elution at 80°C with 0.12 M phosphate, pH 7.0 in 50% formamide, it would contain 75% of the ribosomal sequences and only 15% and 25% of the total chick DNA and keratin sequences respectively. This would correspond to a 5-fold enrichment of ribosomal sequences over total chick DNA.

The purification of keratin sequences from total chick DNA using thermal elution in formamide contain-

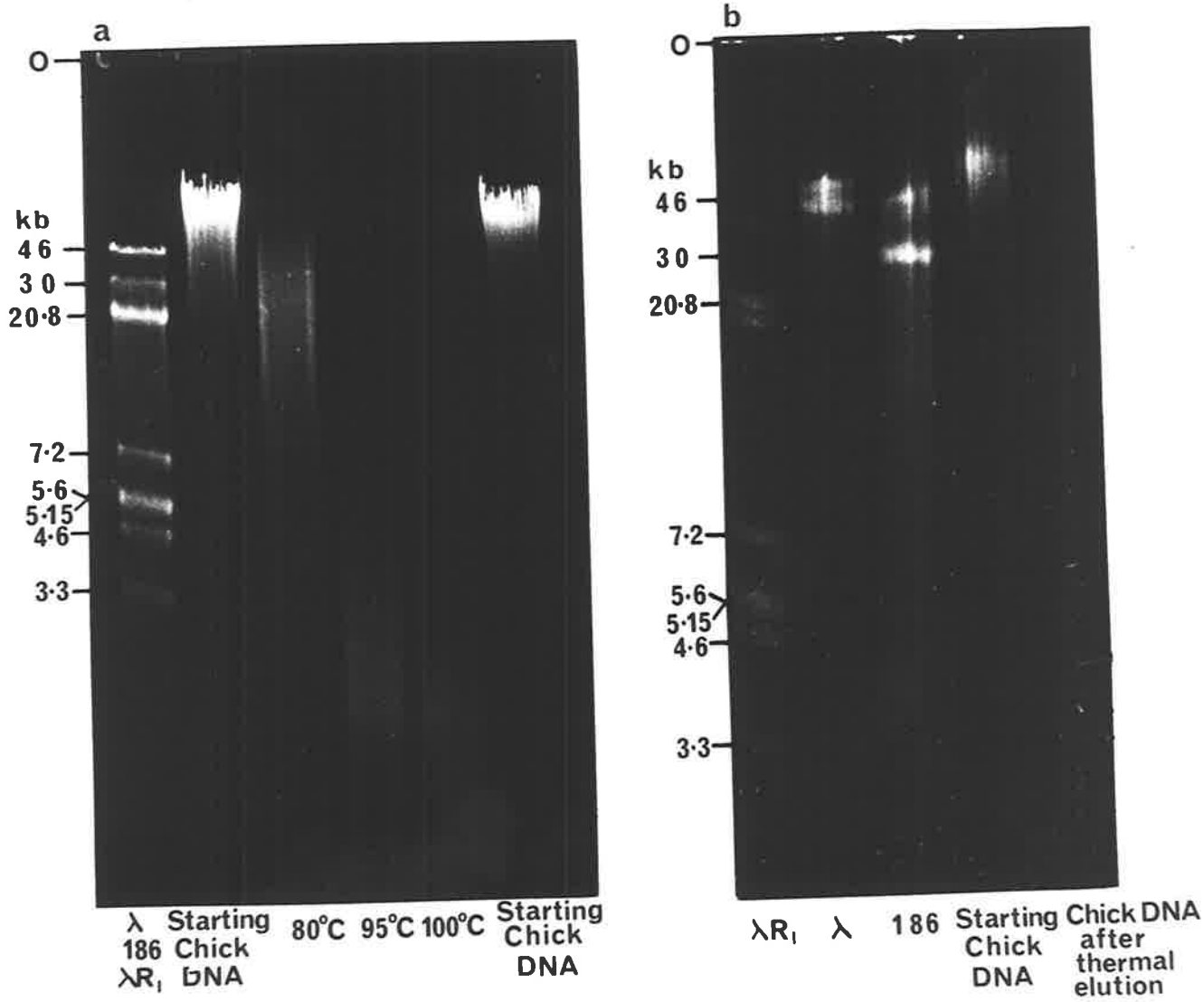
FIGURE 4.2.

EFFECT OF TEMPERATURE AND HAP CHROMATOGRAPHY ON SINGLE-STRANDED DNA MOLECULAR WEIGHT

(a) DNA (1 μg) of single-stranded size ≥ 50 kb in 0.5 ml of T.E. buffer was heated at the temperature indicated in the figure for 5 min. Samples were chilled on ice, ethanol precipitated, redissolved in 2.5 μl of 0.1 M NaOH, 10 mM EDTA and electrophoresed in a 0.4% alkaline agarose gel as described in Chapter V.B.1.(b)(ii). λ Phage DNA (46 kb), 186 phage DNA (30 kb) and λ DNA digested with $EcoR_I$ (Chapter II.B.9) were loaded in 0.1 M NaOH, 10 mM EDTA and used as molecular weight markers. The size of the starting chick DNA is also shown. The gel was stained with ethidium bromide and photographed as described in Chapter V.B.1.(b)(ii). O, origin.

(b) Chick DNA (1 μg), before and after thermal elution at 80°C in buffer containing 50% formamide, was electrophoresed in a 4% alkaline agarose gel, the gel was stained with ethidium bromide and photographed. Due to the lower voltage at which this gel was electrophoresed, the strands of λ, 186 and the 20.8 kb $EcoR_I$ fragment of λ DNA have separated. O, origin.

N.B.: Owing to the inadequacies of photographic reproduction and the broad distribution of DNA, it is difficult to see the DNA in the track showing chick DNA after thermal elution. The DNA was in fact distributed from about 20 kb to less than 1 kb.


a
O
kb
46
30
20·8
7·2
5·6
5·15
4·6
3·3
λ 186 λR1
Starting Chick DNA
80°C
95°C
100°C
Starting Chick DNA
b
O
kb
46
30
20·8
7·2
5·6
5·15
4·6
3·3
λR1
λ 186
Starting Chick DNA
Chick DNA after thermal elution

ing buffers was nowhere near as great as that obtained using aqueous buffers at higher temperatures. However, the fact that 75% of the ribosomal sequences could be removed from the bulk of the DNA, which harbours the keratin sequences, and the observation that lower temperatures lead to less thermal degradation of the DNA (Figure 4.2(a)) suggested that thermal elution on a preparative scale under these conditions should be investigated.

Table 4.1 shows the yields from a preparative thermal elution. Figure 4.2(b) shows the electrophoretic mobility of the DNA in alkaline agarose gels before and after thermal elution relative to markers consisting of phage λ DNA, phage 186 DNA and λ DNA digested with $EcoR_I$. It is apparent from this figure that chick DNA of initial length $\geqslant$ 50 kb is broken down to a range of fragments varying from less than 1 kb to 20 kb. Comparison of the electrophoretic pattern of chick DNA fragments after treatment at 80°C (Figure 4.2(a)) with the pattern after HAP chromatography at 80°C in the presence of formamide (Figure 4.2(b)) indicates that additional breakdown, over and above that caused by temperature, occurs during the chromatography step. It should be pointed out that the gels in Figure 4.2(a) and (b) were electrophoresed at different voltages. At the lower voltage used in Figure 4.2(b) the strands of λ, 186 and the 20.8 kb $EcoR_I$ fragment of λ DNA have separated while no such separation was observed at the higher voltage (Figure 4.2(a)).

The hybrid-density gradient procedure requires single-stranded DNA size classes ranging from 5 kb to 50 kb

TABLE 4.1

RECOVERY OF DNA FROM PREPARATIVE THERMAL ELUTION IN THE PRESENCE OF FORMAMIDE

Treatment	Amount	Recovery
DNA loaded	900 µg	-
0.12 M phosphate, pH 7.0, 50% formamide, 80°C	7½µg	0.83%
0.4 M phosphate, pH 7.0, 50% formamide, 60°C	200 µg	22.2 %
Total recovery 23%		

300 µg Of high molecular weight DNA (≥ 50 kb) was loaded on to each of 3 HAP columns, packed volume 2 mls. 5 ml Of 0.12 M phosphate (0.18 M Na^+), pH 7.0 in 50% formamide was equilibrated to 80°C and passed through the HAP. This was repeated twice with 2.5 ml volumes of the same buffer. 5 mls Of 0.4 M phosphate (0.6 M Na^+), pH 7.0 in 50% formamide was equilibrated to 60°C and passed through the HAP. This was repeated twice with 2.5 ml volumes of the same buffer. Total low and high salt eluates were dialysed against 2 changes of T.E. and ethanol precipitated. Precipitates were redissolved in T.E. and DNA content estimated from A_{260} readings.

as discussed in the next chapter. From Figure 4.2(b) it is apparent that the higher molecular weight classes could not be obtained from DNA which has undergone thermal elution. Since the recovery of DNA from this procedure was low and further losses were expected to occur on selection of the single-stranded size classes (Chapter V), the technique was of no practical use for the quantitative separation of keratin and ribosomal sequences.

2. CsCl Gradient Centrifugation

It has been known for many years that the ribosomal genes of chick are G + C rich and band to the heavy side of the main band of chick DNA in a CsCl equilibrium density gradient (Sinclair and Brown, 1971). While keratin genes are also G + C rich, they have been shown to band directly under the main peak of chick DNA in CsCl gradients when the DNA being studied is of high molecular weight (Lockett and Kemp, 1975). It therefore seemed that CsCl gradient centrifugation might allow the partial purification of ribosomal sequences and the separation of ribosomal and keratin sequences on a preparative basis.

Increasing amounts of DNA ($\geq$ 50 kb) were centrifuged in CsCl and the banding position of ribosomal and keratin sequences determined in each case by filter hybridization (Figure 4.3). When 100 µg of DNA was centrifuged to equilibrium in CsCl (Figure 4.3(a)), 80% of the ribosomal counts appeared as a single sharp band on the heavy side of the gradient around fraction 9. The rest of the ribosomal counts appeared as a lower density shoulder trailing toward the chick DNA main band. The rRNA, iodinated for use as

FIGURE 4.3.

CsCl GRADIENT FRACTIONATION OF KERATIN AND RIBOSOMAL SEQUENCES

Different amounts of native chick DNA of single-stranded size ≥ 50 kb were centrifuged to equilibrium in CsCl gradients as described in Chapter IV.B.2. Gradients were fractionated, the DNA from each fraction immobilized on nitrocellulose discs, the discs split in half and one set of half discs challenged with ^{125}I rRNA and the other with ^{32}P cDNA to keratin mRNA.

(a) 100 µg DNA; ● A_{260} (total chick DNA),
▲ ^{125}I c.p.m. (ribosomal sequences),
○ ^{32}P c.p.m. (keratin sequences).

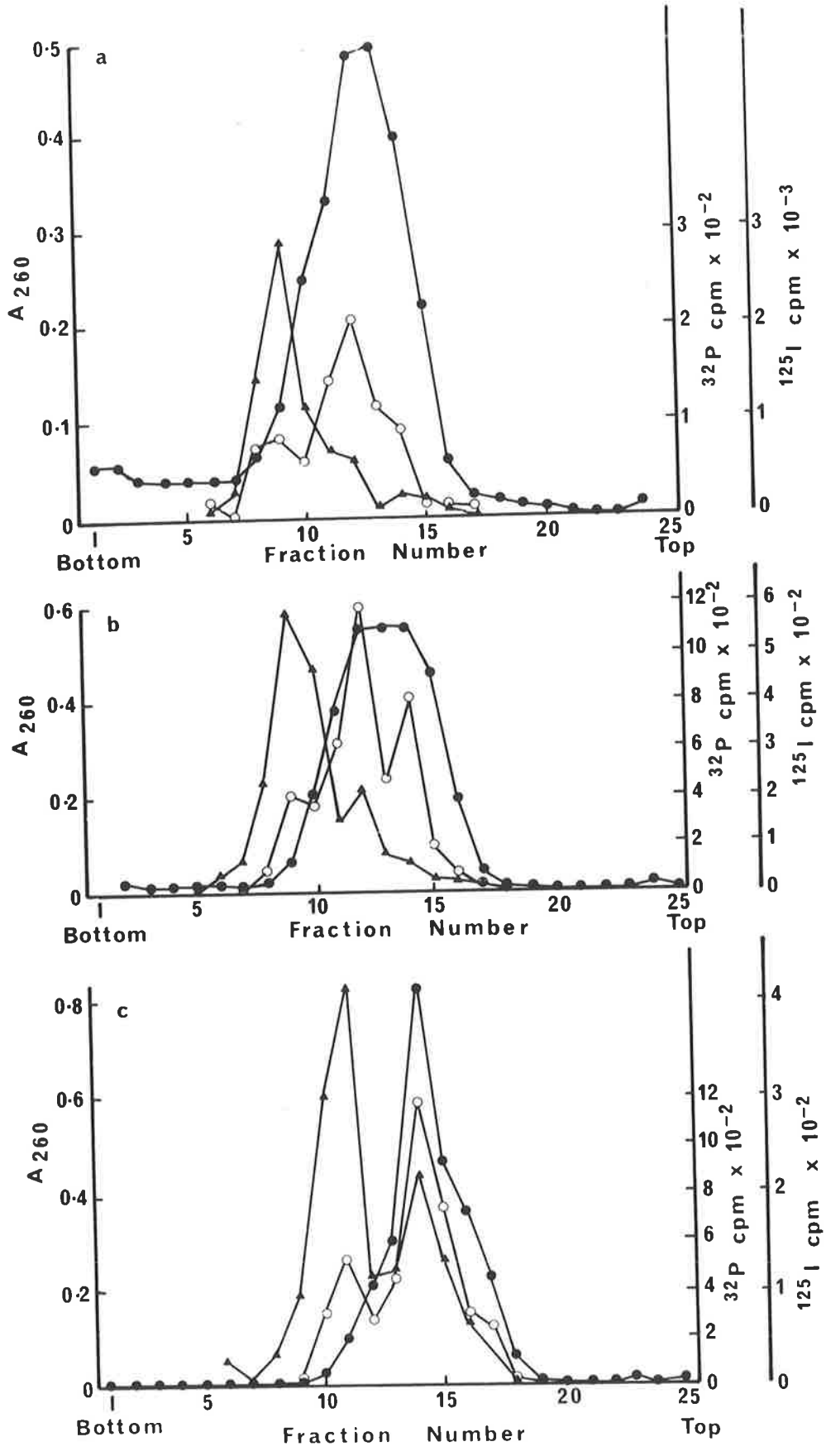
(b) 200 µg DNA; ● A_{260} (total chick DNA),
▲ ^{125}I c.p.m. (ribosomal sequences),
○ ^{32}P c.p.m. (keratin sequences).

(c) 400 µg DNA; ● A_{260} (total chick DNA),
▲ ^{125}I c.p.m. (ribosomal sequences),
○ ^{32}P c.p.m. (keratin sequences).

Density increases from right to left with the main band of chick DNA having a density of 1.700 g/cc. Densities were determined by refractometry.

N.B.: 1. ^{32}P cDNA used in this experiment was prepared using a 60 min incubation at 37°C (see Chapter II.B.6.(a)).

2. CsCl gradient centrifugation is not expected to lead to any degradation of the DNA (Tomizawa and Anraku, 1965).


a
b
c
A 260
32P cpm x 10-2
125I cpm x 10-3
125I cpm x 10-2
Bottom
Fraction Number
Top

the probe, was obtained from small and large ribosomal subunits and the RNA species were subsequently purified as described in Chapter II.B.3. Nevertheless, the counts in the trailing shoulder could still be due to contaminating non-ribosomal sequences in the radioactive probe. Alternatively, the trailing edge could be due to some non-specific aggregation artefact brought about by the high viscosity of the high molecular weight DNA in the main band of chick DNA (Brown and Weber, 1968a). Keratin sequences appear to band at two positions. One peak, carrying 80% of the counts, appears directly under the main band of chick DNA as expected, while the smaller peak bands along with the ribosomal sequences. While it is possible that keratin genes exist in two different density classes, a more likely explanation for this bimodal distribution is that 20% of the cDNA, in this particular preparation, was copied from contaminating rRNA in the keratin mRNA preparation since the cDNA synthesis mix was incubated at 37°C for 1 hour (See Chapter II.B.6.(b)). These data indicated that keratin and ribosomal sequences could be separated by CsCl density gradient centrifugation. The next problem was to determine how much DNA could be loaded on a single gradient without adversely affecting the resolution.

When 200 µg of DNA was used in the density gradient analysis (Figure 4.3(b)), a secondary peak of ribosomal hybridization appeared underneath the main band of chick DNA. Under these conditions, the secondary ribosomal peak (fractions 12 to 15) constituted about 12% of the total ribosomal counts, while the distribution of keratin counts between main band and ribosomal peaks remained the same as

for Figure 4.3(a).

Figure 4.3(c) shows the profiles obtained when 400 µg of DNA was centrifuged in a CsCl gradient. The secondary peak of ribosomal hybridization, first observed when 200 µg of DNA was subjected to CsCl density analysis, has become more prominent accounting for 37% of all the ribosomal counts, suggesting that it is probably a viscosity artefact. Again the distribution of keratin counts between main band and ribosomal peaks remained the same as before.

From Figure 4.3(b) and 4.3(c) it appears that the maximum amount of DNA which can be fractionated on a single CsCl gradient, without inviting artefacts caused by viscosity effects, is 200 µg. Under these conditions (Figure 4.3(b)) fractions 5 to 10 contained 72% of the ribosomal counts and only 9.4% of the total chick DNA. By pooling these fractions, therefore, it should be possible to obtain a 7.7-fold enrichment of ribosomal sequences over total chick DNA. The rest of the DNA would contain 80% of the keratin sequences and only 18% of the ribosomal sequences.

D. DISCUSSION

The results shown in Figures 4.1 and 4.3 indicate that keratin sequences cannot be effectively separated from the bulk of chick DNA by either thermal elution from HAP (Figure 4.1) or CsCl density gradient fractionation (Figure 4.3). Ribosomal sequences, however, can be purified 5-fold and 7.7-fold using these methods. More important than the partial purification of ribosomal sequences, however, is

the fact that these procedures can provide a quantitative separation of keratin and ribosomal sequences.

1. Thermal Elution

From Figure 4.1(a) it can be calculated that thermal elution carried out in aqueous buffers, employing elution steps at 95°C and 100°C, would allow a 2.8-fold purification of keratin sequences. In addition, the DNA of this keratin-containing fraction would only contain 25% of the total chick ribosomal sequences. Figure 4.2(a), however, shows that gross breakdown of the DNA occurs during these thermal elution steps when carried out at such high temperatures.

When DNA melting experiments are performed in buffers containing formamide, the T_m of the DNA is, theoretically decreased by 0.72°C for each 1% of formamide in the buffer (McConaughy *et al.*, 1969). To reduce the amount of thermal degradation occurring during the thermal elution study, buffers containing 50% formamide were used. Figure 4.1(b) shows that while no significant separation of keratin sequences from total chick DNA could be achieved under these conditions, it was possible to obtain a 5-fold purification of ribosomal sequences. After thermal elution at 80°C, the single-stranded fraction contained 75% of the keratin sequences and only 25% of the ribosomal sequences. Thus no enrichment for keratin genes was possible using thermal elution in the presence of formamide, but a quantitative separation of keratin and ribosomal sequences could be obtained. Less thermal degradation of the DNA occurred under these formamide conditions than in aqueous buffers at higher

temperatures (Figure 4.2).

From Figure 4.1 it is apparent that there is a greater difference in T_m between keratin sequences and total chick DNA in aqueous buffers than in buffers containing formamide. This is presumably an artefact caused by the difference in single-stranded DNA molecular weight at the different temperatures (Figure 4.2(a)) since, if all other variables were constant, the presence of formamide would be expected merely to displace both the total chick and keratin sequence melting profiles to a lower temperature range without changing the difference between their T_ms. Poorly matched DNA duplexes are known to adsorb to HAP, but different explanations for the mechanism by which single-stranded DNA molecular weight could affect the relative T_ms of keratin sequences and total chick DNA must be invoked depending on the nature of the DNA:HAP interaction.

(a) <u>If DNA with any vestige of double-stranded structure adheres to HAP</u>

Actinomycin D-CsCl gradient analysis has revealed that keratin genes have a higher G + C content than their surrounding DNA (Lockett and Kemp, 1975). Thus, regardless of the length of DNA containing keratin sequences (within reason), the T_m of these sequences should be dictated by the T_m of the structural genes. If the total chick DNA has regions of G + C rich sequences, then the probability of any particular fragment containing such sequences increases with increasing DNA size. Thus, with increasing single-stranded DNA size, the apparent T_m of total

chick DNA would be expected to increase. If this is the mode of action of HAP, then the observed thermal elution results would suggest that G + C rich sequences are present on all chick DNA fragments of lengths 15 kb to 40 kb. (This is the size range of fragments after incubation at 80°C (Figure 4.2(a).)

(b) If the proportion of the molecule in double-stranded form is the important criterion for DNA binding to HAP

Since DNA is badly degraded at the high temperatures required for thermal elution in aqueous buffers, keratin genes would constitute a large proportion of any fragment containing keratin sequences. Thus, the T_m of such fragments would be dictated by the T_m of the keratin gene, and the resolution between this value and the T_m for total chick DNA would reflect the difference in their G + C content. At the lower temperatures required for thermal elution in the presence of formamide, however, the DNA is subjected to less thermal degradation and hence keratin genes might constitute too small a proportion of the duplex to permit it to adhere to HAP when the rest of the molecule is in single-stranded form. Thus, where the DNA is of relatively high single-stranded size, the melting characteristics of keratin sequences would be dictated by those of the surrounding sequences which are of lower G + C content (Lockett and Kemp, 1975). If this is the mode of action of HAP and the above explanation of the T_m anomaly is correct, then the T_m

data would imply that each keratin gene is associated with a large length of DNA of lower G + C content. This would suggest that, if the keratin genes are linked, they are either associated with large intervening sequences of lower G + C content or large spacers of lower G + C content.

Comparison of the T_m for total chick DNA in the presence and absence of formamide (Figure 4.1) indicates that the apparent T_m depression per 1% formamide in the elution buffer, is considerably less than 0.72°C (McConaughy et al., 1969). There are two possible explanations for this.

(1) The arrangement of G + C rich sequences in chick genome may cause high molecular weight DNA to exhibit an abnormally high T_m when assayed on HAP, as described in (a) above. This interpretation assumes that these G + C rich sequences constitute a sufficiently high proportion of the DNA fragment to allow the fragment to remain adsorbed to the HAP. If this combined effect of G + C rich sequence distribution and single-stranded DNA molecular weight is the reason for the unexpectedly small effect of formamide on T_m in these studies, the melting curve in the presence of formamide would be expected to be steeper than that observed using aqueous buffers. This effect is not obvious in Figure 4.1.

(2) There may be some interaction between single-stranded DNA and HAP in 0.12 M phosphate, pH 7.0; it must be independent of the formamide content of the buffer and the magnitude of the interaction would have to increase with increasing single-stranded DNA size. As discussed earlier, the DNA obtained from thermal elution studies in the presence of formamide is of higher single-stranded molecular weight than that obtained in aqueous buffers. In this situation the interaction of the DNA with HAP would be greater when T_m analysis is carried out using buffers containing formamide. Higher temperatures would therefore be required to elute these structures from the column and this would lead to an anomalous increase in the apparent T_m of the DNA.

Neither of these explanations can, by itself, account satisfactorily for the abnormal melting characteristics of total chick DNA observed during denaturation analysis on HAP using buffers containing formamide. It is quite likely that all the factors mentioned above play some part.

When preparative thermal elution was attempted using buffers containing formamide, the yields of DNA were very low (Table 4.1). This observation, along with the fact that the DNA obtained from the thermal fractionation

procedure did not contain molecules of the largest size class necessary for the hybrid-density gradient procedure (Figure 4.2(b)), indicated that the method was inadequate for the separation or partial purification of keratin and ribosomal sequences.

2. CsCl Gradient Fractionation

CsCl density gradient centrifugation allowed a 7.7-fold purification of ribosomal sequences from total chick DNA but no enrichment for keratin sequences (Figure 4.3). The bulk of the DNA left after removal of the ribosomal peak contained 80% of the keratin sequences and only 18% of the ribosomal sequences. For preparative separation of ribosomal and keratin sequences using DNA of length $\geq$ 50 kb, however, a maximum of 200 µg of DNA could be loaded on to each gradient. The relative abundance of ribosomal sequences in the chick genome rendered it unnecessary to partially purify these sequences prior to their use in the characterisation of the proposed hybrid-density gradient procedure. On the other hand, keratin sequences are present at a 10-fold lower level than ribosomal sequences and so enrichment for these sequences might be necessary before keratin gene arrangement could be analysed by the hybrid-density gradient technique. The CsCl fractionation described was to be used as the first step of this partial purification, providing a quantitative separation of keratin and ribosomal sequences. Keratin sequences could than be further purified by cycles of CsCl density gradient centrifugation in the presence of antibiotics actinomycin D and netropsin sulphate (Lockett and Kemp, 1975).

E. APPENDIX

The conclusions made in this chapter regarding the proportions of keratin and ribosomal sequences present in different fractions are only valid if the radioactive probe used in the filter hybridization was present in saturating amounts. In order to investigate this, 100 µg of chick DNA was immobilized on filters as described in Chapter IV.B.3 and the filters challenged with increasing concentrations of ^{125}I labelled 28S rRNA and ^{3}H labelled keratin cDNA respectively. Typical results for these saturation hybridization analyses are shown in Figure 4.4.

From these results it appeared that the levels of the different radioactive probes necessary for the saturation of 100 µg of chick DNA on a filter were 300 ng/ml for 28S rRNA and 60 ng/ml for keratin cDNA. In accordance with these findings then, these were the concentrations of the radioactive probes used in the experiments described in Chapter IV. Saturation filter hybridizations were not performed for ^{125}I labelled 18S rRNA, but in all experiments, equal concentrations of ^{125}I labelled 18S and 28S rRNA were used. Since 28S rRNA and 18S rRNA were isolated from large and small ribosomal subunits respectively and re-purified by sucrose gradient centrifugation, the two rRNA species should show little cross contamination. Thus the bulk of the ^{125}I labelled 18S rRNA should be 18S sequences rather than 28S rRNA breakdown products. In this situation, given that 18S rRNA is slightly less than half the length of 28S rRNA, the hybridization conditions used for ^{125}I labelled 18S rRNA would be expected to be saturating.

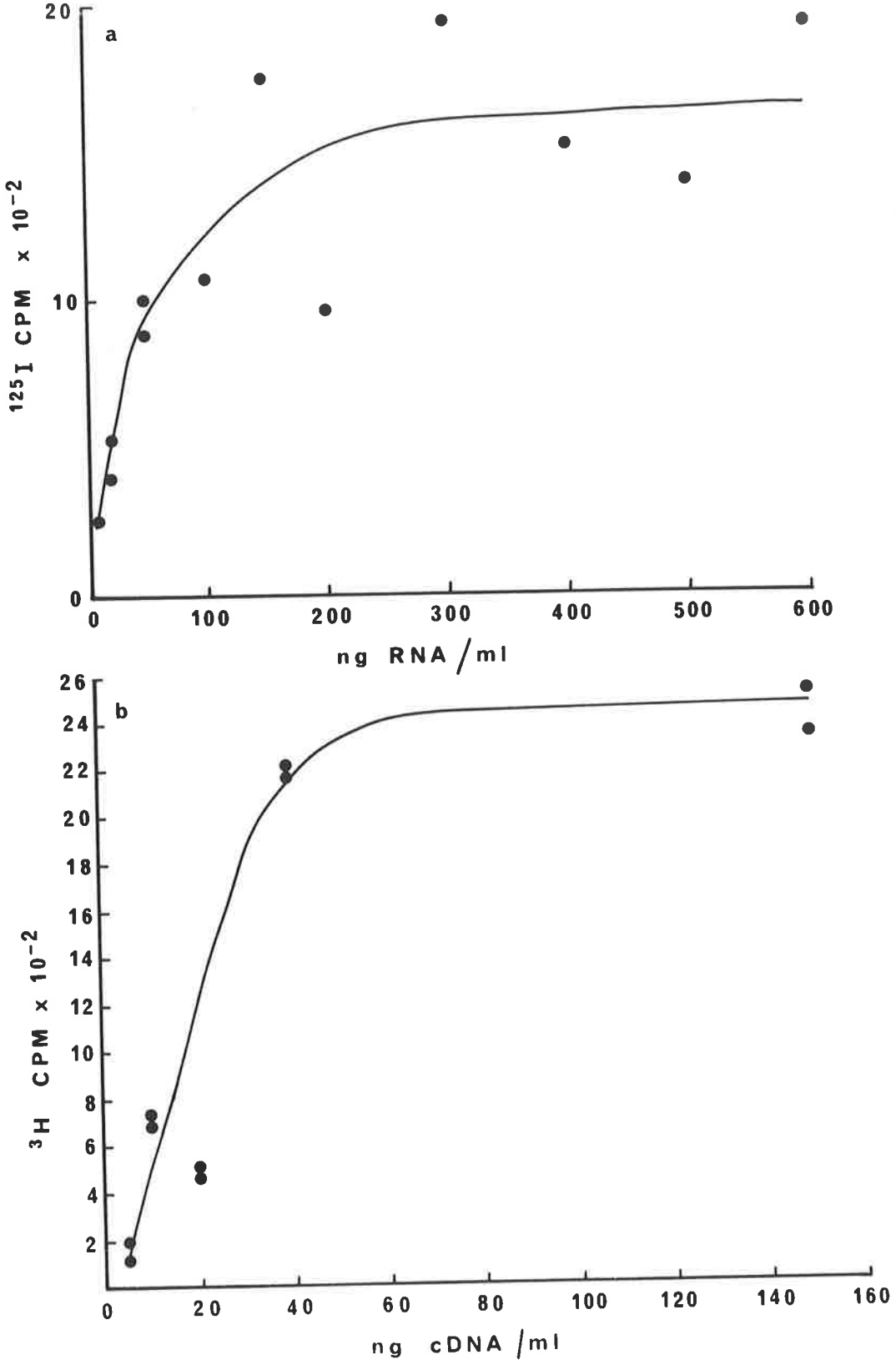
FIGURE 4.4.

CONDITIONS FOR SATURATION OF FILTER HYBRIDIZATIONS

Chick DNA (100 µg) was immobilized on nitrocellulose discs. Filters were challenged with increasing amounts of ^{125}I labelled 28S rRNA or ^{3}H labelled keratin cDNA. Filters were washed extensively in 2 x SSC, 0.5% SDS, dried and counted.

(a) Filter hybridizations using increasing amounts of ^{125}I labelled 28S rRNA.

(b) Filter hybridizations using increasing amounts of ^{3}H keratin cDNA.


a
20
10
0
^{125}I CPM x 10^{-2}
0
100
200
300
400
500
600
ng RNA /ml
b
26
24
22
20
18
16
14
12
10
8
6
4
2
0
^{3}H CPM x 10^{-2}
20
40
60
80
100
120
140
160
ng cDNA /ml

CHAPTER V

THE HYBRID-DENSITY GRADIENT PROCEDURE FOR THE ANALYSIS OF GENE ARRANGEMENT

CHAPTER V

THE HYBRID-DENSITY GRADIENT PROCEDURE FOR THE ANALYSIS OF GENE ARRANGEMENT

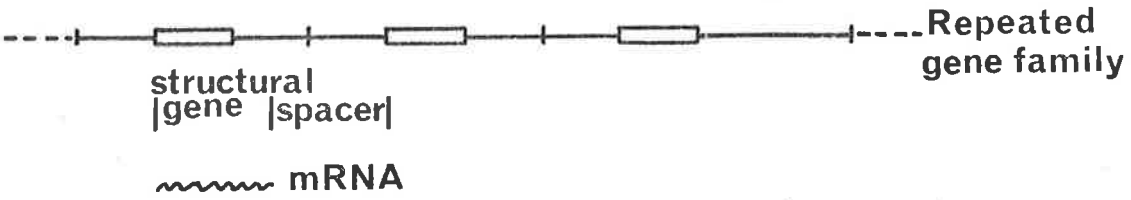
A. INTRODUCTION

This chapter deals with studies on the isolation of relatively homogeneous size classes of single-stranded DNA and their use in the preliminary investigation of a density gradient method for analysing the arrangement of eukaryotic genes.

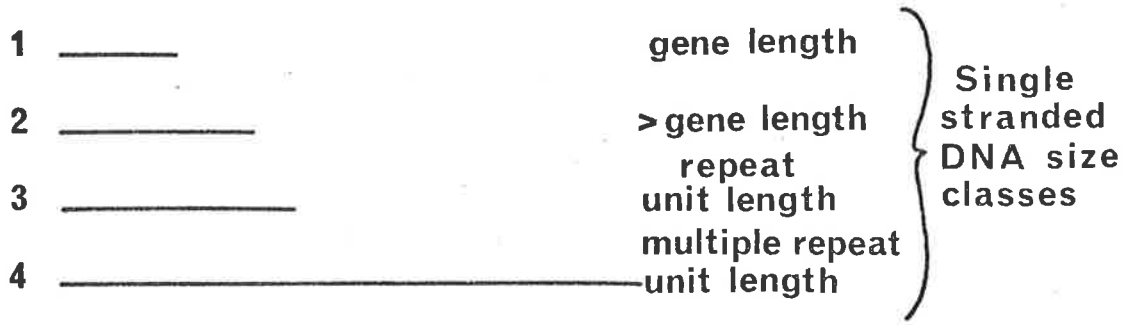
The technique is a slight modification of the pre-hybridization technique used to study the arrangement of the sequences coding for 18S and 28S rRNA of Xenopus laevis (Wallace and Birnstiel, 1966; Birnstiel et al., 1968; Brown and Weber, 1968b) and the tRNA genes of this toad (Clarkson and Birnstiel, 1973). It relies on the density difference between single-stranded DNA and RNA:DNA hybrids. Figure 5.1 outlines the principle of the technique. Consider the tandemly repeated gene family shown. The structural gene gives rise in vivo to a mRNA species which can then be isolated and hybridized to single-stranded DNA of specific size classes. This pre-hybridization step must be carried out with a vast excess of RNA so as to drive RNA:DNA hybridization ahead of DNA reassociation. When the resultant structures are centrifuged in Cs_2SO_4 gradients, the position at which the hybrid bands, relative to the two extremes of 100% hybrid and single-stranded DNA, should be dependent on the proportion of the DNA molecule in hybrid form. Hybrids made using DNA of single-stranded size much larger than that

FIGURE 5.1.

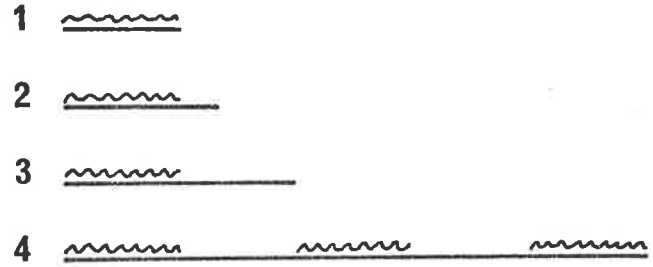
SCHEMATIC REPRESENTATION OF THE PRINCIPLE BEHIND THE HYBRID-DENSITY GRADIENT PROCEDURE FOR THE ANALYSIS OF GENE ARRANGEMENT



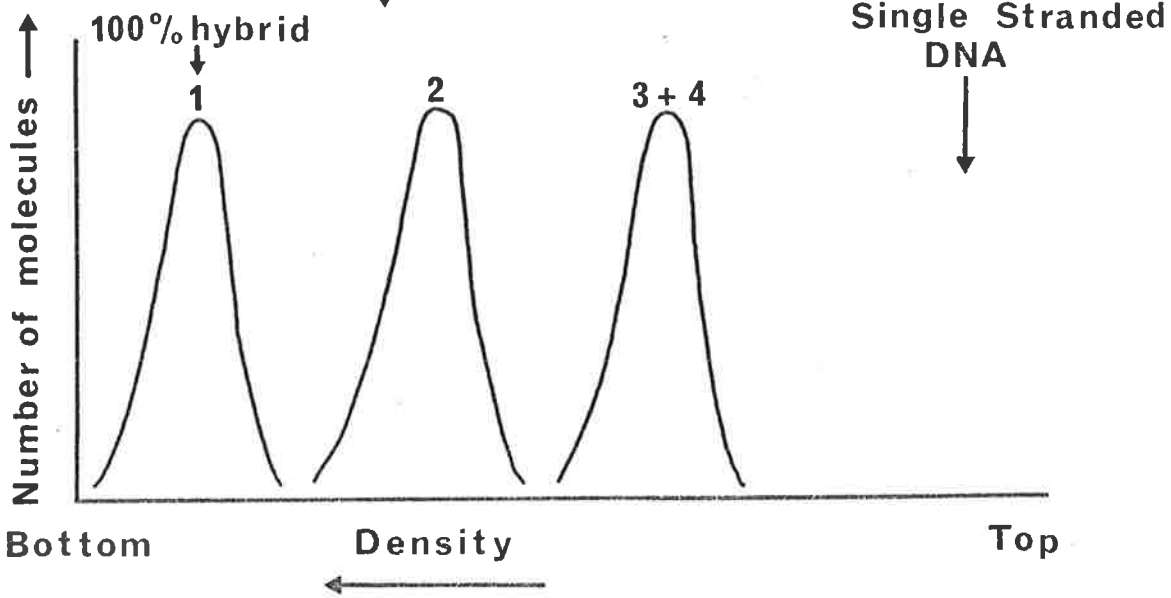
mRNA



Hybridize with vast excess of RNA



Cs_2SO_4 gradient



of a single repeat unit should have a density identical to that of hybrids generated using DNA of repeat unit length if the genes are tandemly linked. If the genes being studied are not tandemly linked, the density of the hybrid generated should approach that of single-stranded DNA as the size of the DNA used in hybrid formation increases. For tandemly linked repeated gene families, it should be possible, by this technique, to determine the average length of a repeat unit by knowing the length of the RNA gene product and using single-stranded DNA, which is expected to be much larger than the repeat unit in the pre-hybridization reaction. This would, of course, require a knowledge of the relationship between the density at which a hybrid bands and the proportion of the DNA in hybrid form relative to the two extremes mentioned above. The repeat unit length calculated in this way could be confirmed by carrying out similar hybrid-density analyses using single-stranded DNA of size classes increasing up to and beyond the repeat unit length calculated from the initial investigation. Analyses of these data in a manner similar to that used by Lizardi and Brown (1975) to determine the length of the silk fibroin gene using CsCl gradient centrifugation in the presence of actinomycin D, should allow the accurate determination of the repeat unit length for tandemly linked repeated gene families.

The technique was to be characterized using the ribosomal cistrons of chick as a model system. When perfected, it was to be used for the analysis of keratin gene arrangement since it is believed that the keratin genes of chick constitute a tandemly repeated gene family (Lockett and

Kemp, 1975). The requirements for the characterization of the system were:

1. The isolation of homogeneous single-stranded DNA size classes.

2. The ability to detect the generated hybrids in the density gradient.

3. To obtain complete RNA:DNA hybridization without DNA reassociation during pre-hybridization.

4. To empirically determine the relationship between the density of the hybrid formed and the proportion of the DNA molecule in hybrid form, relative to the two extremes of single-stranded DNA and 100% hybrid.

It should be stressed that this project was conceived and had reached an advanced state of development before restriction enzyme analysis and recombinant DNA technology became readily available to Australian molecular biologists. Indeed, Australia's first recombinant DNA facility was only built in late 1977. These studies were continued in the belief that the results obtained by this approach would provide independent information about average gene arrangement which could then be critically compared with that obtained from restriction or recombinant DNA analysis. During the course of this study, the restriction enzyme map of the chick ribosomal genes was published (McClements and Skalka, 1977) permitting a critical evaluation of the hybrid-density technique.

B. METHODS

1. Isolation of Specific Size Classes of Single-stranded Chick DNA

(a) Alkaline sucrose gradient centrifugation

Ethanol precipitated, high molecular weight chick DNA was dissolved in 0.1 M NaOH, 10 mM EDTA at a concentration of about 1 mg per ml. 5-20% Sucrose gradients made up in 0.1 M NaOH, 10 mM EDTA, 0.9 M NaCl, were overlayed with 500 µl of the DNA solution and centrifuged in an SW41 rotor using a Beckman L5-50 preparative ultracentrifuge for 4 hours at 37,000 r.p.m. at 15°C. Gradients were fractionated by upward displacement using an ISCO density gradient fractionator, model 640.

Fractions were neutralized by the addition of 0.1 volumes each of 1 M Tris-HCl, pH 7.5, and 1 M HCl, diluted 2-fold with sterile bidistilled water and precipitated with 2.5 volumes of ethanol at -20°C for a minimum of 2 hours.

(b) Alkaline agarose gel electrophoresis

Alkaline agarose gel electrophoresis was essentially as described by McDonell *et al*. (1977).

(i) Preparative

20 cm x 20 cm x0.5 cm, 0.5% Agarose gels in 30 mM NaOH, 2 mM EDTA, were poured either vertically between glass plates or hori-

zontally inside 0.5 cm high borders. The circuit between the electrophoresis tanks containing 30 mM NaOH, 2 mM EDTA and the gel was completed by wicks made from Whatmann 3 MM paper soaked in electrophoresis buffer. 0.2-2 mg Of randomly sheared chick DNA in 2 ml of 0.1 M NaOH, 10 mM EDTA, 5% glycerol, containing bromophenol blue, was loaded on top of the gel (or in the loading slot of a horizontal gel) and electrophoresis performed at 100 volts overnight or at 300 volts for 6 hours.

(ii) Analytical

20 cm x 20 cm x 0.2 cm Horizontal 0.4% agarose gels in the buffer described above were used for this purpose. 1 µg Samples were loaded in the loading slots in 40 µl of 0.1 M NaOH, 10 mM EDTA, 5% glycerol, containing bromophenol blue, and electrophoresis carried out at 250 volts for 4 to 6 hours.

After electrophoresis, the gel was neutralized by soaking in 0.5 M Tris-HCl, pH 6.8 for 10 min, stained with a 0.01% solution of ethidium bromide in bidistilled water for 15 min and destained in mono-distilled water for 30 min. The destained gel, illuminated with ultra-violet light in an Oliphant ultra-violet viewing box, was photographed with a Nikormat E.L. camera fitted with a red filter using

Kodak recording film 2475. Films were developed with Kodak HC-110 developer.

Whole λ phage DNA (gift of Dr. R.C. Crawford), whole phage 186 DNA (gift of R.B. Saint) and λ DNA digested with $EcoR_I$ restriction endonuclease were used as molecular weight markers in these analytical gels allowing molecular weight estimation of fragments varying in size between 3.3 and 45 kb. New digestions of λ DNA with $EcoR_I$ were carried out for each analytical gel so as to minimize any nicking of the DNA on storage.

(c) Extraction of single-stranded DNA from preparative alkaline agarose gels

After electrophoresis of the DNA in a preparative alkaline agarose gel, the gel was sliced into 1 cm thick strips at right angles to the direction of electrophoresis. Three methods for the extraction of the DNA from these strips were investigated.

(i) Solubilization of agarose with KI followed by HAP chromatography

Agarose strips were broken into small pieces by passage through a syringe and were dissolved by the addition of 1.25 volumes of a saturated solution of KI (SSKI) followed by incubation at 37°C for 30 min. Solubilized agarose strips were then loaded on to an HAP

column of 2 ml packed volume, previously equilibrated at 37°C with SSKI, equilibrated to 37°C and passed through the HAP under slight pressure. The columns were washed twice with 5 ml of SSKI and twice with 5 ml of 10 mM phosphate, pH 7.0, to remove any remaining agarose. The DNA from the solubilized strips was then eluted with 0.4 M phosphate, pH 7.0. The 0.4 M phosphate eluate was centrifuged to remove any HAP as described in Chapter IV.B.1.(a), dialysed against 2 changes of T.E., each of volume 4 litres. The amount of DNA remaining after this procedure was determined by estimation of the A_{260}. The DNA was concentrated by ethanol precipitation (Chapter II.B.1).

(ii) KI gradient centrifugation

Agarose slices were solubilized as described above and neutralized by the addition of 0.1 gel volumes of 1 M Tris-HCl, pH 7.5, and 0.03 gel volumes of 1 M HCl. Ethidium bromide was added to a final concentration of 50 µg/ml. The solution was adjusted to a refractive index of 1.420 (density about 1.5 g/cc), and centrifuged at 40,000 r.p.m. for 40 to 60 hours at 15°C using a Ti50 rotor and Beckman L2-50 preparative ultracentrifuge (Wolf, 1975). After centrifugation, the position of the DNA band was determined from the fluorescence of the bound ethidium bromide when viewed under ultra-

violet light of wavelength 2573 Å. The DNA was removed by side puncture of the centrifuge tube using an 18 gauge needle attached to a 2 ml syringe. Ethidium bromide was removed from the DNA by 3 extractions with isoamyl alcohol while KI was removed by 2 dialyses against 4 litres of T.E. The amount of DNA remaining after this treatment was determined by reading the A_{260} and DNA was concentrated by precipitation with 2.5 volumes of ethanol at -20°C overnight.

(iii) Centrifugation of the agarose slice

Whole agarose slices were placed in screwcap 50 ml centrifuge tubes and centrifuged at 19,000 r.p.m. for 20 min at 4°C in a Beckman JA20 rotor using a Beckman J-21B preparative centrifuge. The supernatants were decanted, neutralized by the addition of 0.1 volumes of 1 M Tris-HCl, pH 7.5, and 0.03 volumes of 1 M HCl and concentrated by precipitation with 2.5 volumes of ethanol at -20°C for a minimum of 2 hours.

2. Pre-hybridization Conditions

20 to 100 µg Of DNA from a given single-stranded size class in 1 ml of T.E. was made up to 10 mM EDTA, 0.1 M NaOH. The DNA solution was neutralized by the addition of 0.05 volumes of 1 M Tris-HCl, pH 7.5, and 0.1 volumes of 1 M HCl and placed on ice. RNA, which had been denatured by

boiling, was added to the DNA and the solutions made up to a final volume of 2 ml by the addition of 0.2 ml of 20 x SSC and 0.65 ml of water. In the early experiments, 20 µg of ^{125}I labelled 28S rRNA or 20 µg each of ^{125}I labelled 18S and 28S rRNA of specific activity about 2 x 10^5 c.p.m./µg was added to the pre-hybridization mix which was then incubated at 60°C for 60 min. Under these conditions, the RNA driven hybridization reached a $R_o t$ of 0.125 mol.s.l^{-1} with DNA re-association reaching a $C_o t$ of 0.125 mol.s.l^{-1} to 0.625 mol.s.l^{-1}. These values should be compared with the S_1 assayed $R_o t_{½}$ for the hybridization of 28S cRNA to its cDNA of 1 x 10^{-2} mol.s.l^{-1} (Figure 3.1(a)) and the S_1 assayed $C_o t_{½}$ for the reassociation of chick ribosomal sequences of 10 mol.s.l^{-1} (P. Krieg, personal communication). Pre-hybridization was also carried out using unlabelled rRNA under these same conditions. In later experiments, 60 µg of 28S rRNA alone or 60 µg each of 18S and 28S rRNA was hybridized to 20 µg of DNA at 60°C for 30 min in a total volume of 2 ml. Under these conditions a $R_o t$ of about 0.2 mol.s.l^{-1} and a $C_o t$ of about 0.06 mol.s.l^{-1} was achieved. Hybridization was stopped by chilling the reaction mix on ice.

3. Generation of 100% Hybrid

A 50 µl solution containing keratin mRNA (2 µg) and ^{32}P labelled cDNA made to keratin mRNA (5 x 10^5 c.p.m.) in hybridization buffer (10 mM Tris-HCl, pH 7.0, 1 mM EDTA, 0.18 M NaCl, 0.05% SDS), overlayed with paraffin oil, was boiled for 2 min then incubated at 60°C for 1 hour. The hybridization solution was then added to medium salt S_1 assay buffer (0.03 M Na-acetate, 0.15 M NaCl, 0.001 M $ZnSO_4$, 5% glycerol, pH 4.6) containing 12 µg of denatured coalfish

DNA, 4 units of S_1 nuclease was added and S_1 digestion was allowed to proceed at 37°C for 30 min. The S_1 digestion was stopped by the addition of 0.02 volumes of 25% SDS. 5 µg Of E. coli tRNA was added as carrier and the S_1 nuclease resistant hybrid precipitated with 2.5 volumes of ethanol at -20°C for at least 2 hours. Hybrids were redissolved in T.E. About 100,000 c.p.m. of hybrid was added to 2 ml of a solution identical to that used for pre-hybridization prior to Cs_2SO_4 gradient centrifugation. 100% Hybrids using rRNA and cDNA made to rRNA (Chapter II.B.6.(b)) was prepared in the same way.

4. Cs_2SO_4 Gradient Centrifugation

High purity grade Cs_2SO_4 was obtained from Kawecki Berylco Industries Inc. and recrystallized from boiling water. 5.63 g Of recrystallized Cs_2SO_4 was dissolved in the minimum volume of water. The pre-hybridization solution was mixed with the Cs_2SO_4 solution and made up to a final volume of 9 ml (density 1.5 g/cc). The solution was transferred to siliconized centrifuge tubes and the gradients centrifuged at 28,000 r.p.m. for 60 to 70 hours at 20°C in a Ti50 rotor using a Beckman L2-50 preparative ultracentrifuge. Gradients were fractionated from the bottom into 0.3 ml fractions using a Gilson micro-fractionator and densities determined by refractometry. The amount of DNA in each fraction was determined by measurement of the A_{260} of the gradient fractions.

5. Detection of Hybrids

Three methods for detection of hybrid were

employed.

(a) RNA'ase resistance of iodinated hybrid

Hybrids were formed using an excess of ^{125}I labelled rRNA to drive the reaction. After fractionation of the Cs_2SO_4 gradients, 1.7 ml of 10 mM Tris-HCl, pH 7.5, 4 mM EDTA containing pancreatic ribonuclease A (Sigma) at a concentration of 7 µg/ml was added to each fraction. RNA'ase digestion was allowed to proceed at 37°C for 30 min and the amount of radioactivity resistant to RNA'ase in each fraction was determined by TCA precipitation.

(b) Binding of iodinated hybrid to nitrocellulose

This was performed essentially as described by Wallace and Birnstiel (1966). The method relies on the observation that RNA:DNA hybrids bind to nitrocellulose filters while RNA does not.

Hybrids were formed using an excess of ^{125}I labelled rRNA. After fractionation of the gradient, 150 µl of each fraction was added to 6 ml of 2 x SSC and heated at 60°C for 1 hour. Fractions were chilled on ice, filtered through nitrocellulose discs and both sides of each disc were washed with 50 ml of 2 x SSC. Filters were dried at 65°C and the radioactivity bound to each filter determined.

(c) Immobilization of DNA on nitrocellulose followed by filter hybridization

The third method was essentially as described by Birnstiel et al. (1968). Unlabelled RNA was used in the pre-hybridization step and, after Cs_2SO_4 density gradient fractionation, the fractions were diluted to 1 ml by the addition of water and made 0.1 M with respect to NaOH. Fractions were left in 0.1 M NaOH at room temperature for at least 1 hour to allow denaturation of DNA:RNA hybrids and extensive alkaline hydrolysis of the RNA. Neutralization, immobilization of DNA on nitrocellulose filters, pre-treatment of filters and conditions of hybridization were as described in Chapter IV.B.3 with the exception that one set of half filters was challenged with ^{125}I labelled rRNA to detect the hybrid, the other set being challenged with ^{3}H labelled cDNA made to rRNA to allow detection of the banding position of single-stranded ribosomal DNA sequences.

6. Determination of Single-stranded Sequence Density

In method (a) and (b) for hybrid detection, the A_{260} peak was used as the marker for single-stranded ribosomal sequences. Since the fractionation of DNA on the basis of G + C content is much lower in Cs_2SO_4 than in CsCl (Erikson and Szybalski, 1964) it was assumed that ribosomal sequences would band close to the main peak of single-stranded DNA. In method (c), however, detection of single-stranded ribosomal sequences was by filter hybridization as described above using cDNA made to rRNA as probe.

C. RESULTS

1. Isolation of Specific Size Classes of Single-stranded Chick DNA

(a) Alkaline sucrose gradient centrifugation

6 Gradients, each loaded with 200 μg of high molecular weight DNA in 0.1 M NaOH, 10 mM EDTA, were centrifuged as described in Chapter V.B.1.(a) and fractionated on an ISCO density gradient fractionator, model 640. The absorbance at 254 nm was monitored by the optical unit of the ISCO which was coupled to a W + W recorder. The profile obtained from all 6 gradients were identical and a representative is shown in Figure 5.2. This broad DNA distribution was subdivided into 5 fractions as shown in Figure 5.2. Corresponding fractions from each gradient were pooled, neutralized, diluted, ethanol precipitated, re-centrifuged on 5-20% alkaline sucrose gradients under the same conditions as before, and fractionated. The profiles obtained after re-centrifugation are shown in Figure 5.3. While these profiles demonstrate that successively higher molecular weight fractions contained DNA of successively larger single-stranded size, the peaks obtained were very broad indicating the presence of a wide range of molecular weight species within each fraction. Accordingly, narrower size ranges were selected from these gradients as shown in Figure 5.3 and the DNA re-centrifuged on alkaline sucrose gradients under the same conditions as before. The results from this third alkaline

FIGURE 5.2.

ISOLATION OF SINGLE-STRANDED DNA SIZE CLASSES BY ALKALINE SUCROSE GRADIENT CENTRIFUGATION 1

High molecular weight DNA (500 µg single-stranded size ≥ 50 kb) was loaded on to 5-20% alkaline sucrose gradients and centrifuged as described in Chapter V.B.1.(a). Gradients were fractionated using an ISCO density gradient fractionator coupled to a W and W recorder. The A_{254} profile is shown. The numbered sections at the top of the figure represent the molecular weight cuts selected for re-centrifugation.

Sedimentation is from right to left.

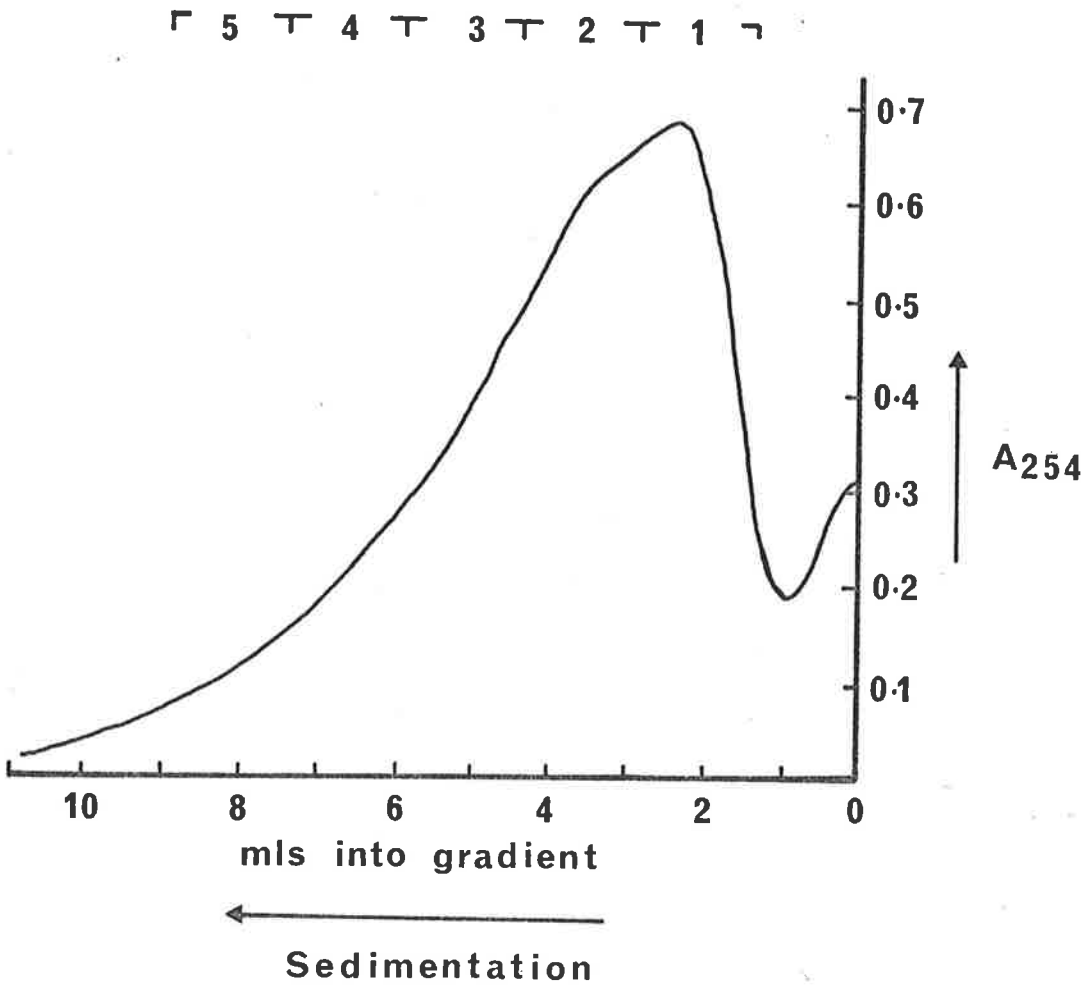

5
4
3
2
1
0·7
0·6
0·5
0·4
0·3
0·2
0·1
A_{254}
10
8
6
4
2
0
mls into gradient
Sedimentation

FIGURE 5.3.

ISOLATION OF SINGLE-STRANDED DNA SIZE CLASSES BY ALKALINE SUCROSE GRADIENT CENTRIFUGATION 2

Molecular weight cuts from Figure 5.2 were ethanol precipitated and recentrifuged on 5-20% alkaline sucrose gradients as described in Chapter V.B.1.(a). Sections marked at the top of each profile represent molecular weight cuts selected for recentrifugation.

Sedimentation is from right to left. Numbers 1 to 5 correspond to the fraction numbers shown in Figure 5.2.

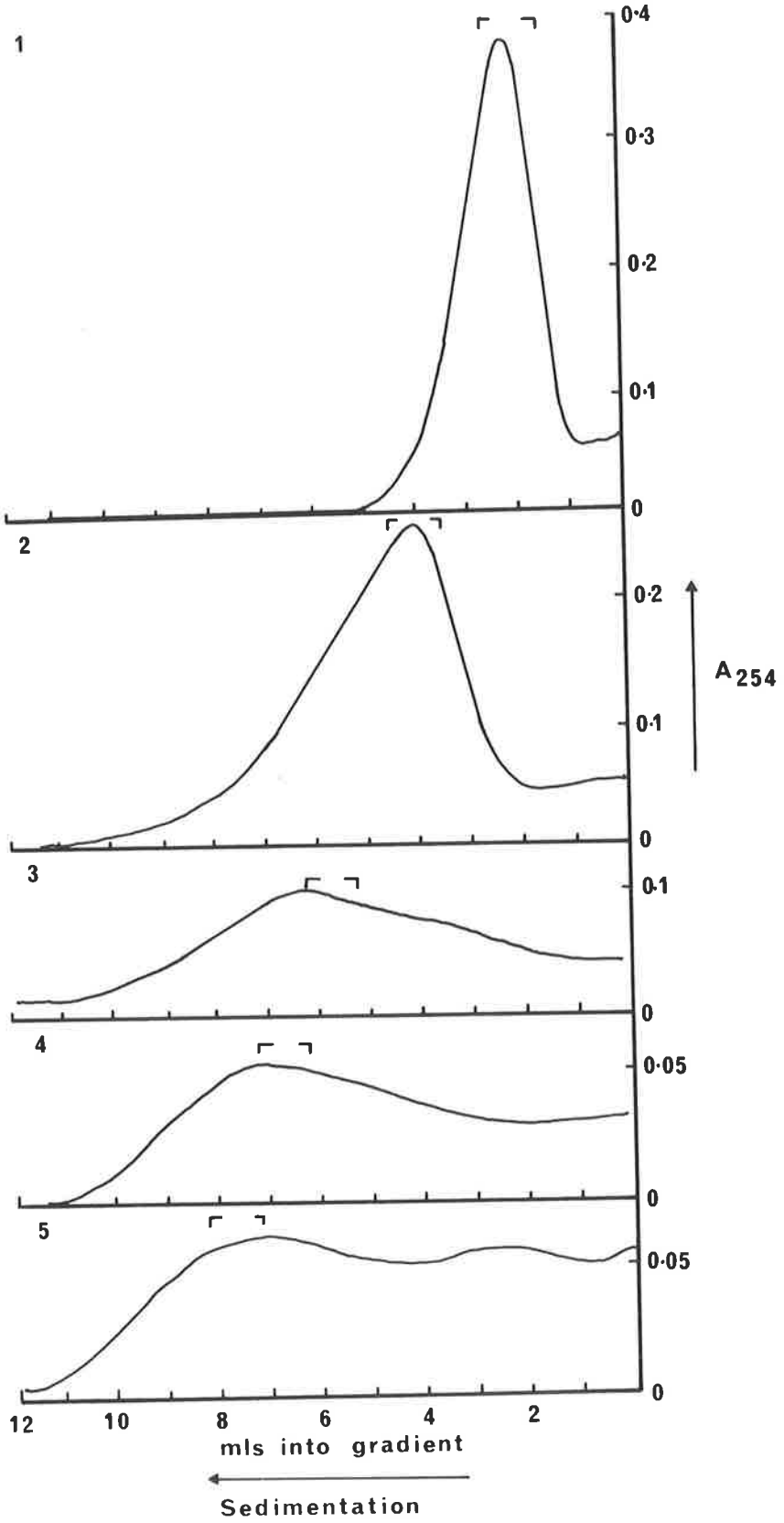

1
0·4
0·3
0·2
0·1
0
2
0·2
0·1
0
A254
3
0·1
0
4
0·05
0
5
0·05
0
12
10
8
6
4
2
mls into gradient
Sedimentation

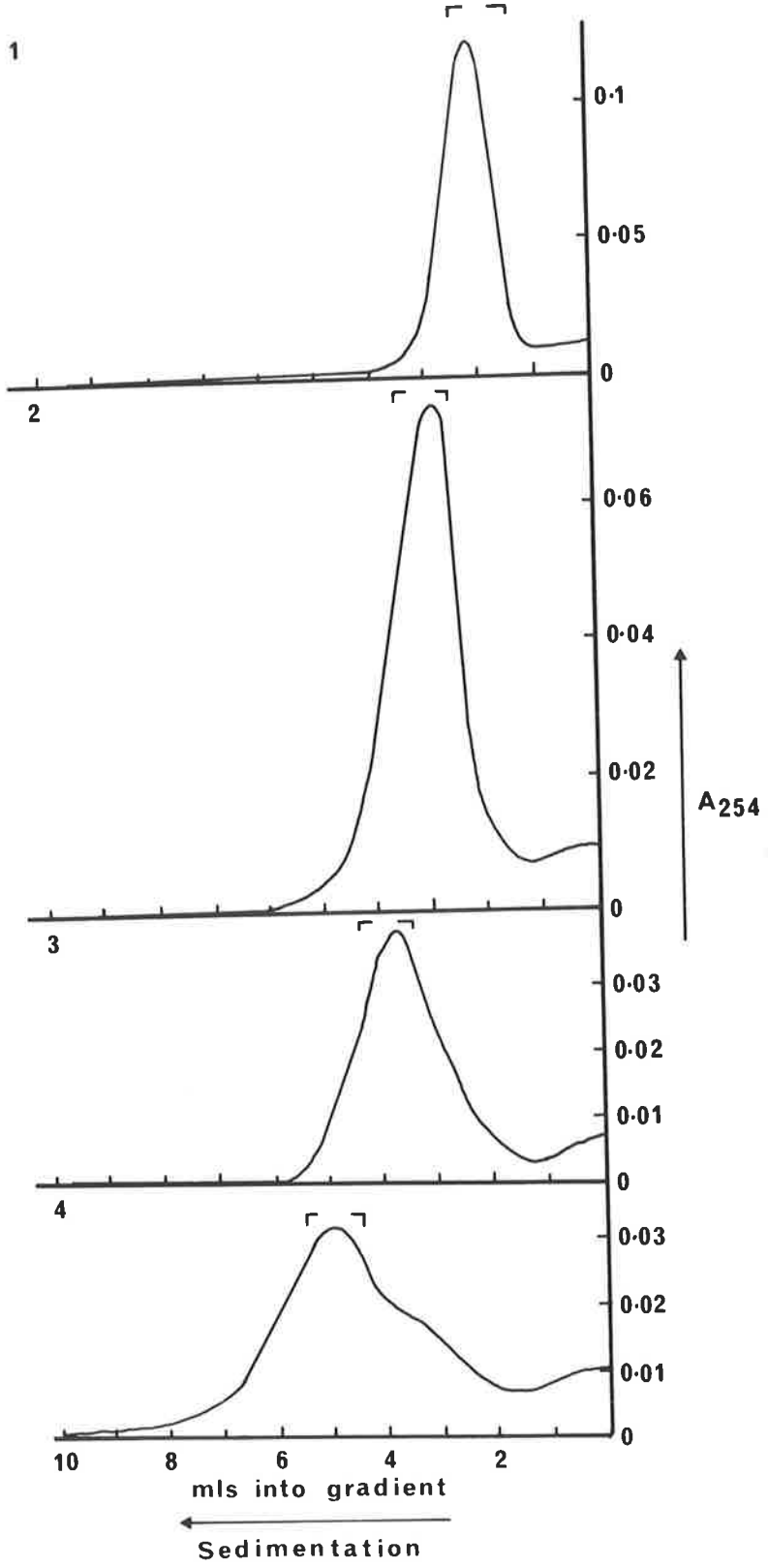
FIGURE 5.4.

ISOLATION OF SINGLE-STRANDED DNA SIZE CLASSES BY ALKALINE SUCROSE GRADIENT CENTRIFUGATION 3

Molecular weight cuts from Figure 5.3 were ethanol precipitated and recentrifuged on 5-20% alkaline sucrose gradients as described in Chapter V.B.1.(a). Sections marked at the top of each profile represent the final molecular weight cuts selected.

Sedimentation is from right to left. Numbers 1 to 4 correspond to fraction numbers shown in Figure 5.2.

N.B.: Fraction 5 from Figure 5.3 contained insufficient DNA to register as an absorbance peak on the 0.2 absorbance setting on the ISCO.


1
0·1
0·05
0
2
0·06
0·04
0·02
0
A254
3
0·03
0·02
0·01
0
4
0·03
0·02
0·01
0
10
8
6
4
2
mls into gradient
Sedimentation

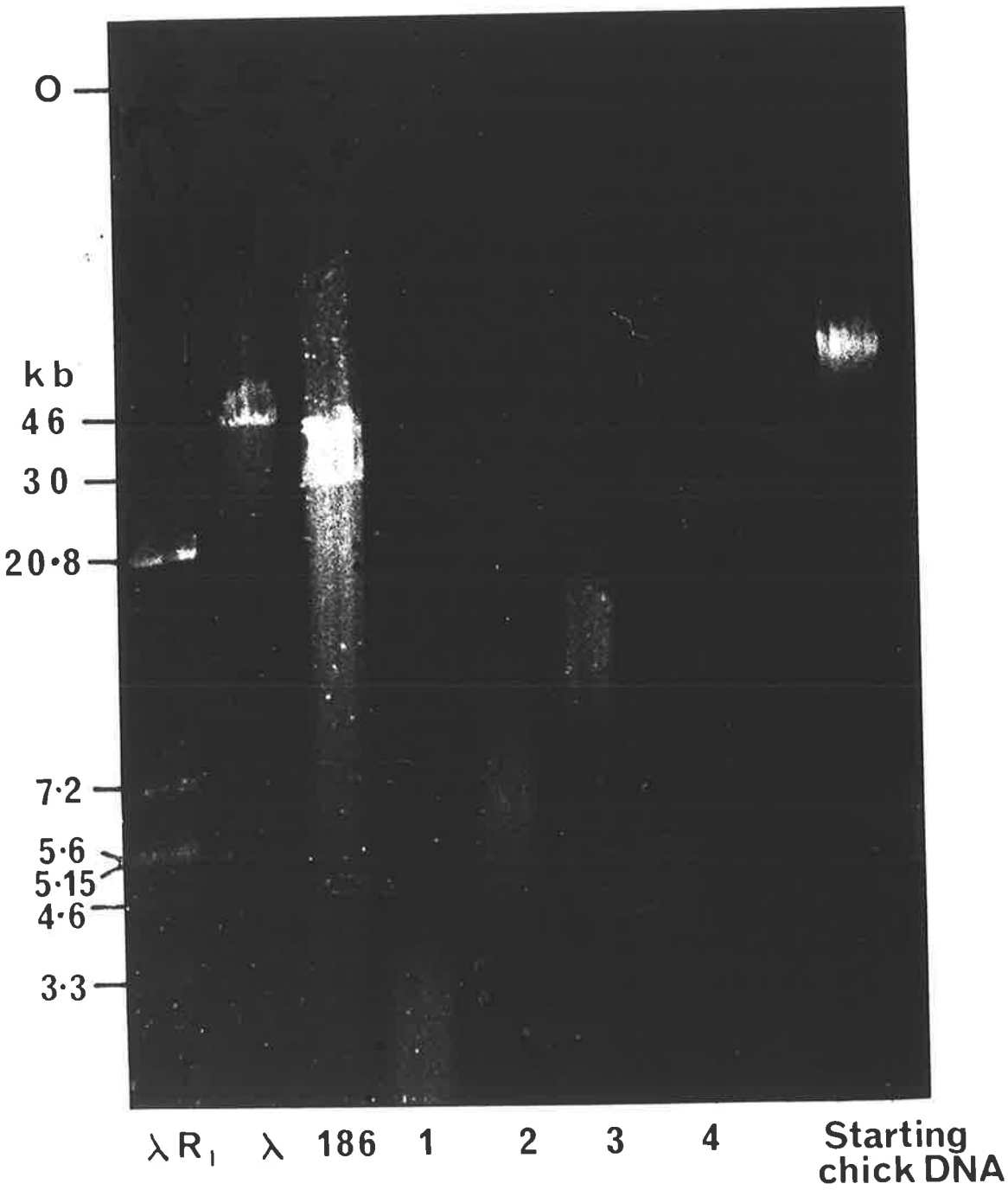
sucrose gradient step are shown in Figure 5.4. Fraction 5 did not contain enough DNA to register as an absorbance peak on the 0.2 absorbance setting of the ISCO. Fractions 1 to 4 showed considerably narrower peaks than in Figure 5.3 indicating greater DNA molecular weight homogeneity within each fraction. Despite this, these peaks were still too broad for use as single-stranded DNA molecular weight size classes in the density gradient procedure and so a narrower size range was again selected from these gradients (Figure 5.4). After concentration by ethanol precipitation, the single-strand molecular weight of each of these size fractions, along with that of the starting material, was determined by alkaline agarose gel electrophoresis relative to molecular weight markers λ phage DNA, 186 DNA and λ DNA digested with restriction endonuclease $EcoR_I$ (Figure 5.5).

From Figure 5.5 it is immediately obvious that there was a large amount of DNA breakdown during the size class isolation procedure. Much of this breakdown must have occurred during treatment of the DNA prior to or during loading on the first of the alkaline sucrose gradients, since the broad DNA distribution shown in Figure 5.2 contains those regions from which even the lowest molecular weight size class was derived. Consideration of Figures 5.2 and 5.4 shows that the position of the cut for a given size fraction from the original gradient (Figure 5.2) is slightly further toward the bottom of the gradient than the

FIGURE 5.5.

ANALYTICAL ALKALINE AGAROSE GEL ANALYSIS OF SINGLE-STRANDED DNA SIZE CLASSES OBTAINED FROM ALKALINE SUCROSE GRADIENT CENTRIFUGATION

1 μg Of DNA from each of the size cuts shown in Figure 5.4 was electrophoresed on a 0.4% alkaline agarose gel as described in Chapter V.B.1.(b)(ii). λ DNA, 186 DNA and $EcoR_I$ digested λ DNA (Chapter II.B.9) were used as molecular weight markers. O, origin.


O
kb
46
30
20·8
7·2
5·6
5·15
4·6
3·3
λR1
λ
186
1
2
3
4
Starting chick DNA

final cuts for the same size fraction (Figure 5.4) indicating that although most of the breakdown occurred on the first gradient, breakdown continued to occur on subsequent gradients. Some degree of breakdown of the DNA could have been tolerated since the density gradient system required size classes varying in length between 5 kb and 50 kb. By this procedure, however, most of the DNA appeared in the lowest molecular weight fraction (Figure 5.4) which was too small to be of practical use (Figure 5.5). Increasing the concentration of EDTA in the gradient and loading solutions failed to reduce the breakdown problem (data not shown). Figure 5.5 also shows that the single-stranded DNA size classes obtained by this procedure were still heterodisperse with regard to the molecular weight of their component molecules. These results, taken together, indicated that the alkaline sucrose gradient centrifugation procedure was inadequate for the isolation of relatively homogeneous single-stranded DNA size classes over the range of sizes required for density gradient analysis of eukaryote gene organization employing RNA:DNA hybrids.

(b) Preparative alkaline agarose gel electrophoresis

Agarose gels provide greater resolution of DNA on the basis of molecular weight than sucrose gradients. Recently McDonell et al. (1977) have used alkali as a denaturing solvent for agarose gel

electrophoresis. The possibility of using alkaline agarose gels for the preparative isolation of single-stranded DNA size classes was, therefore, investigated. Since the size classes required for the proposed hybrid-density gradient technique ranged from 5 kb to 50 kb, low percentage gels (0.5% agarose) which give an almost linear relationship between log MW and mobility over this range, were used.

(i) Quality of single-stranded DNA size classes

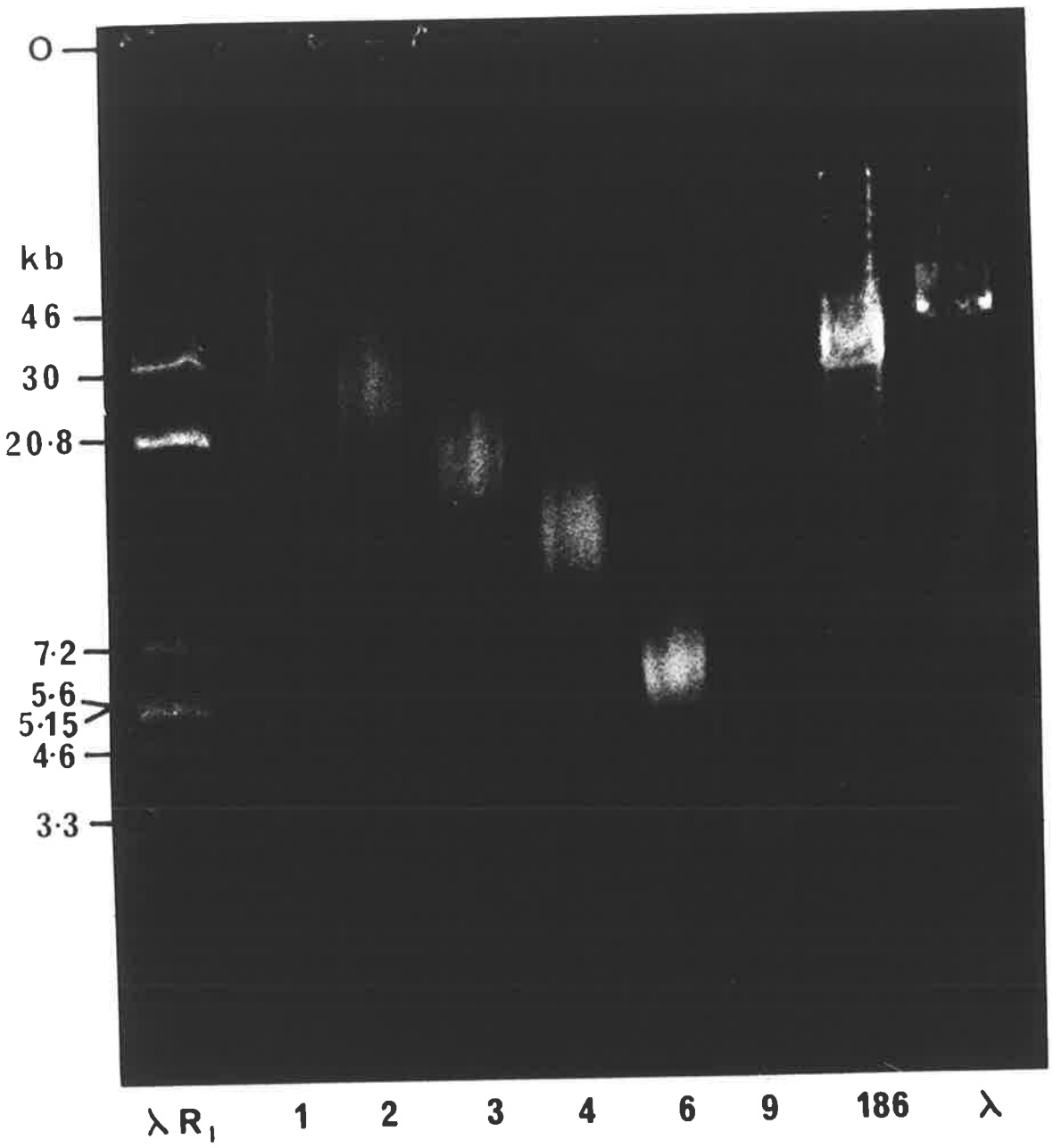
To determine whether preparative alkaline agarose gels could provide relatively homogeneous size classes of single-stranded DNA, 200 µg of randomly sheared chick DNA in 0.1 M NaOH, 10 mM EDTA was loaded in the loading slot of a 20 cm x 20 cm x 0.5 cm, horizontal 0.5% alkaline agarose gel and electrophoresed at 100 volts overnight. A slice was taken from the side of the gel and stained with ethidium bromide to determine the position of the DNA. The DNA containing section of the gel was divided into 9 slices, each slice was dissolved in SSKI and the DNA extracted by HAP chromatography as described in Chapter V.B.1.(c)(i). The DNA from 6 of these fractions was concentrated by ethanol precipitation and examined by analytical alkaline agarose gel electrophoresis. The result of this analysis is shown in Figure 5.6. The fractions examined showed excellent molecular

FIGURE 5.6.

QUALITY OF SINGLE-STRANDED DNA SIZE CLASSES OBTAINED FROM PREPARATIVE ALKALINE AGAROSE GELS

0.2 mg Of randomly sheared chick DNA in 0.1 M NaOH, 10 mM EDTA was electrophoresed in a preparative, horizontal 0.5% alkaline agarose gel at 100 volts overnight. The gel was sliced into 1 cm thick slices taken at right angles to the direction of electrophoresis. DNA was isolated from the agarose slices by solubilization of the agarose in KI followed by HAP chromatography (Chapter V.B.1.(c)(i)). DNA was dialysed against T.E. and concentrated by ethanol precipitation. 1 μg Of DNA from selected slices was electrophoresed on analytical 0.4% alkaline agarose gels using λ DNA, 186 DNA and λ DNA digested with $EcoR_I$ as molecular weight markers.

Numbers represent the distance, in cm, of the slice from which the DNA was extracted from the origin of the preparative gel. O, origin.


O
kb
46
30
20·8
7·2
5·6
5·15
4·6
3·3
λR1
1
2
3
4
6
9
186
λ

weight homogeneity and ranged in size from about 3 kb to 30 kb.

(ii) Conditions of electrophoresis

To investigate whether alkaline agarose gels could be used on a preparative basis for the isolation of single-stranded DNA size classes, 0.5 cm thick 0.5% alkaline agarose gels were loaded with 600 µg and 2 mg of randomly sheared DNA in 0.1 M NaOH, 10 mM EDTA and electrophoresed overnight at 100 volts. The gel was divided into 1 cm wide strips taken at right-angles to the direction of electrophoresis and the DNA extracted by centrifugation of the agarose slice (Chapter V.B.1.(c)(iii)). The DNA from the slices was examined on analytical alkaline agarose gels and the results are shown in Figure 5.7. If the preparative gels were overloaded with DNA, the DNA extracted from any given strip would be heterogeneous in molecular weight. In this situation the DNA from any given slice, when examined on analytical alkaline agarose gels, would appear as a broad distribution and there would be a high degree of overlap between those DNA distributions derived from different slices. Figure 5.7(a) shows the results for single-stranded DNA size classes taken from a preparative gel loaded with 0.6 mg of DNA, Figure 5.7(b) shows size classes taken from a gel loaded with 2 mg of DNA. The mole-

FIGURE 5.7.

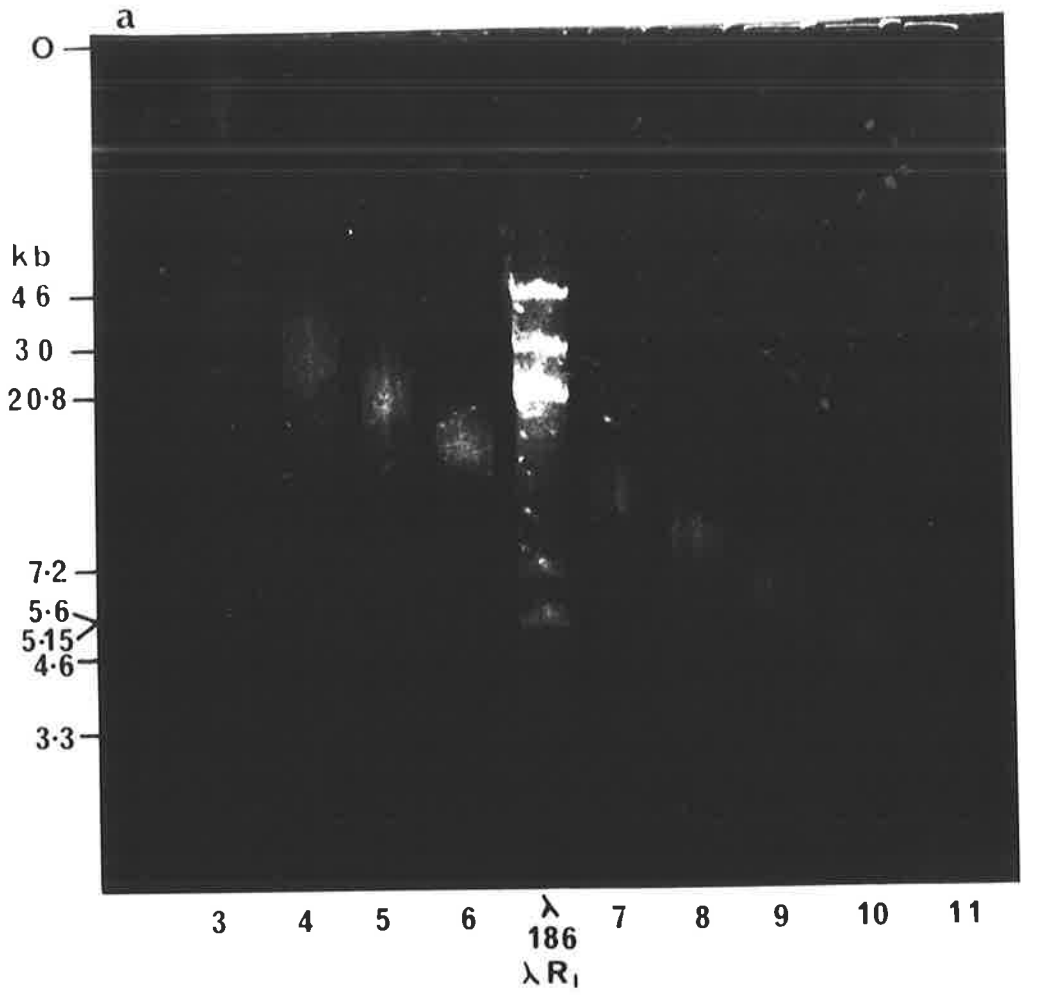
CAPACITY OF PREPARATIVE ALKALINE AGAROSE GELS

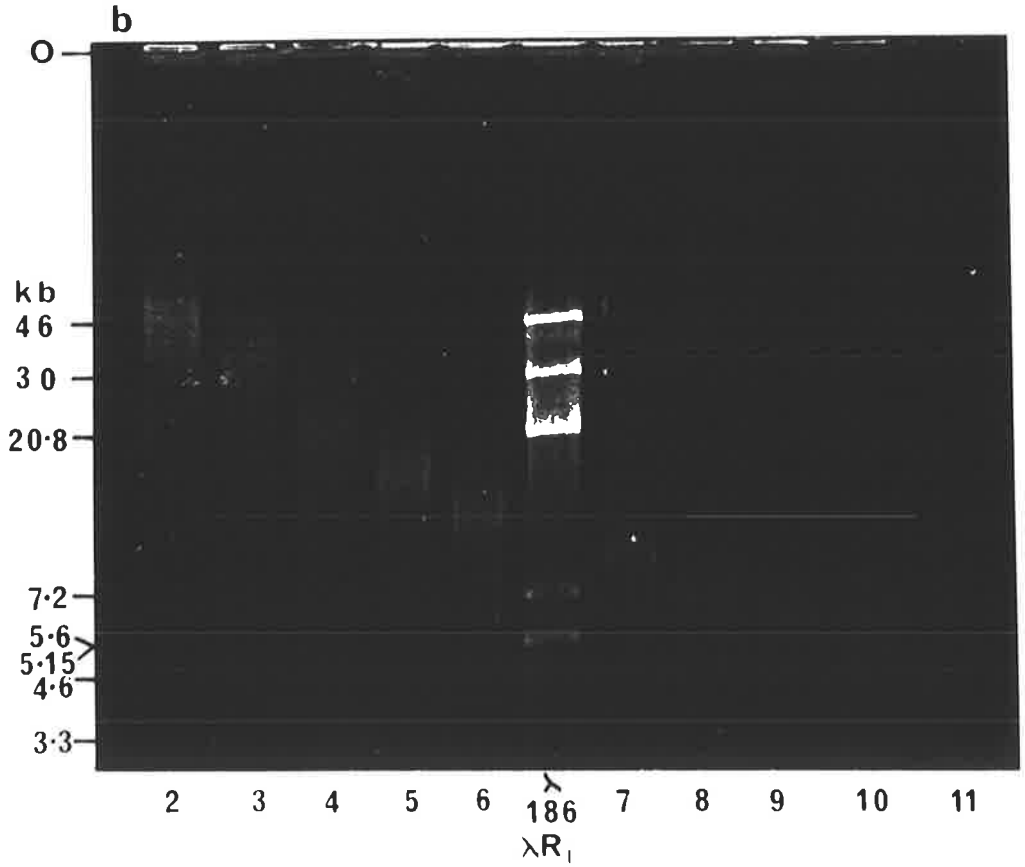
Horizontal preparative 0.5% alkaline agarose gels were loaded with 0.6 mg and 2 mg of randomly sheared chick DNA in 0.1 M NaOH, 10 mM EDTA and electrophoresed at 100 volts overnight. Gels were sliced into 1 cm thick slices taken at right angles to the direction of electrophoresis. DNA was extracted by centrifugation of the agarose slice (Chapter V.B.1.(c)(iii)) and 1 μg from each slice electrophoresed on an analytical 0.4% alkaline agarose gel using λ DNA, 186 DNA and λ DNA digested with $EcoR_I$ as molecular weight markers.

(a) DNA from a sliced preparative gel loaded with 0.6 mg DNA.

(b) DNA from a sliced preparative gel loaded with 2.0 mg DNA.

Numbers represent the distance, in cm, of the slice from which the DNA was extracted from the origin of the preparative gel. O, origin.


a
O
kb
46
30
20·8
7·2
5·6
5·15
4·6
3·3
3
4
5
6
λ 186 λR1
7
8
9
10
11


b
O
kb
46
30
20·8
7·2
5·6
5·15
4·6
3·3
2
3
4
5
6
λ 186 λR1
7
8
9
10
11

cular weight distributions for the size classes obtained with each of these loadings indicated that the preparative gels were not overloaded in either case. Accordingly, for all subsequent experiments, the initial alkaline agarose gel size fractionation was performed on 2 mg of randomly sheared chick DNA.

To determine whether the rate of the preparative size fractionation could be increased 2 mg of randomly sheared DNA was electrophoresed in a preparative alkaline agarose gel at 300 volts for 6 hours. The gel was sliced, DNA extracted from each slice and examined on analytical alkaline agarose gels (Figure 5.8). The DNA distributions obtained from the different slices were broad and distributions from different slices showed a high degree of overlap. This effect was most obvious for the slices which should have contained the highest single-stranded DNA molecular weight classes. The effect was the same regardless of whether vertical or horizontal gels were used for the initial size fractionation. For all subsequent experiments preparative electrophoresis was carried out at 100 volts overnight.

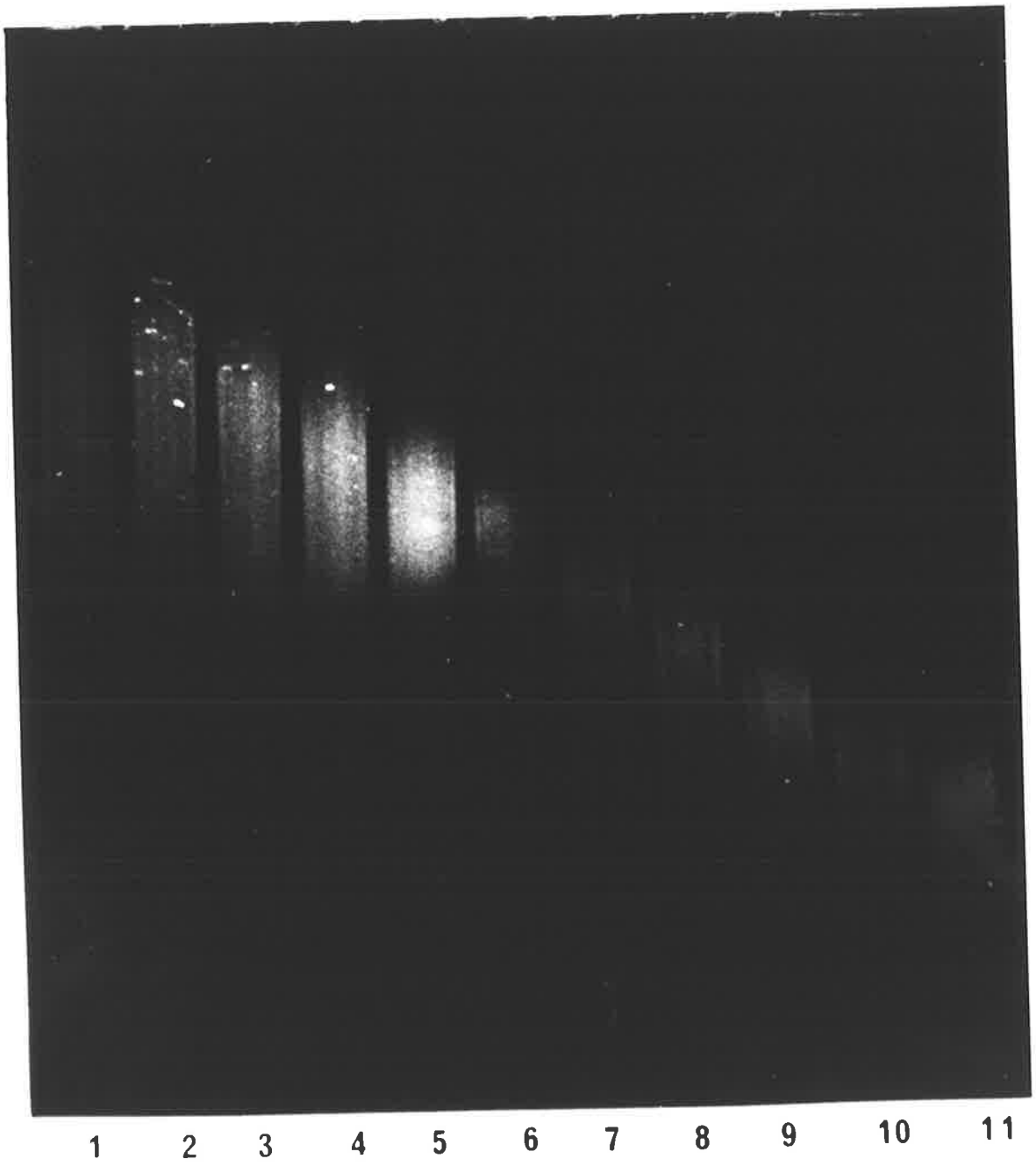
(iii) Extraction of DNA from agarose slices

2 mg Of randomly sheared DNA was electrophoresed in preparative alkaline agarose gels at 100 volts overnight. The DNA was

FIGURE 5.8.

EFFECT OF VOLTAGE ON THE QUALITY OF SINGLE-STRANDED DNA SIZE CLASSES OBTAINED FROM PREPARATIVE AGAROSE GEL ELECTROPHORESIS

Randomly sheared DNA (2 mg) was electrophoresed on a horizontal preparative 0.5% alkaline agarose gel at 300 volts for 6 hours. The gel was sliced into strips 1 cm wide and DNA extracted from each slice as for Figure 3.6. DNA (1 µg) from each slice was electrophoresed on an analytical 0.4% alkaline agarose gel. Numbers correspond to the distance, in cm, of the slice from which the DNA was extracted from the origin of the preparative gel. O, origin.


1
2
3
4
5
6
7
8
9
10
11

extracted from the slices of the gels by three methods and the recoveries obtained using these procedures were compared. The amount of DNA extracted from each slice was determined by measurement of the A_{260} of the fractions after all the manipulations necessary for the appropriate isolation procedure had been completed (Chapter V.B.2.(c)). The total recovery from each gel was then determined by the summation of the quantities of DNA obtained from each of the constituent slices. As can be seen from Table 5.1, centrifugation of the agarose slice was the most efficient of the extraction procedures returning 24% to 33% of the DNA initially loaded on to the gel. Gel solubilization with KI followed by HAP chromatography was next, reproducibly yielding about 12% of the DNA initially loaded, while KI gradient centrifugation yielded 9.9% of the DNA loaded. Under the conditions of preparative gel electrophoresis used, none of the DNA initially loaded was ever observed to electrophorese off the bottom of the gel, as determined by ethidium bromide staining. The recoveries shown in Table 5.1, therefore, represent accurate determinations of the efficiency of recovery of single-stranded DNA from alkaline agarose gels by each of the three procedures.

In all subsequent work, 2 mg of DNA was used for the preparative alkaline agarose

TABLE 5.1.

COMPARISON OF THE EFFICIENCES OF 3 METHODS FOR EXTRACTION OF SINGLE-STRANDED DNA FROM AGAROSE SLICES

Method	Percentage recovery
Agarose solubilization with KI followed by HAP chromatography.	10-12.5
Agarose solubilization with KI followed by KI density gradient centrifugation.	9.9
Centrifugation of the agarose slice.	24-33

2 mg Of DNA was electrophoresed on a preparative horizontal alkaline agarose gel overnight at 100 volts. The gel was sliced into 1 cm strips at right angles to the direction of electrophoresis and all strips from any given gel were subjected to the same extraction procedure (Chapter V.B.1.(c)). The amount of DNA extracted from each slice was estimated by reading the A_{260} of each resultant fraction after the fraction had been subjected to all the steps necessary for the particular extraction procedure. The amount of DNA extracted from each slice of a given gel was then summed giving the total amount of DNA extracted from the gel. This value was divided by the total amount of DNA loaded on the gel and multiplied by 100 to give percentage recovery.

N.B.: Under the conditions of preparative electrophoresis employed, no DNA was ever observed to run off the end of the gel as determined from ethidium bromide staining.

gel. Electrophoresis was performed at 100 volts overnight and DNA was extracted from the gel slices by centrifugation of the slice.

2. Studies on the Hybrid-density Gradient Procedure for the Analysis of Gene Arrangement

(a) Detection of 100% hybrid and single-stranded ribosomal sequence densities

The first requirement was to determine the densities of 100% hybrid and single-stranded DNA. 100% Hybrid was generated as described in Chapter V.B.3 and loaded on to Cs_2SO_4 gradients in pre-hybridization buffer. In a similar manner, 100 µg of DNA was alkali denatured, neutralized and loaded on to Cs_2SO_4 gradients. After centrifugation, the gradients were fractionated from the bottom. The banding position for pure hybrid (Figure 5.9(a)) was determined by TCA precipitation of each fraction followed by the determination of the amount of radioactivity in the precipitates. The banding position of single-stranded DNA was determined by reading the A_{260} of each fraction (Figure 5.9(b)). The banding position of 100% hybrid varied in different experiments between 1.548 g/cc and 1.552 g/cc. This variation was the same regardless of whether the hybrid was generated using keratin mRNA:cDNA hybrids or rRNA:cDNA hybrids. The density of 100% hybrid was therefore taken to be 1.55 g/cc under these conditions. The density of single-stranded DNA ranged from 1.452 to 1.456 g/cc in independent experiments and so the actual density for

FIGURE 5.9.

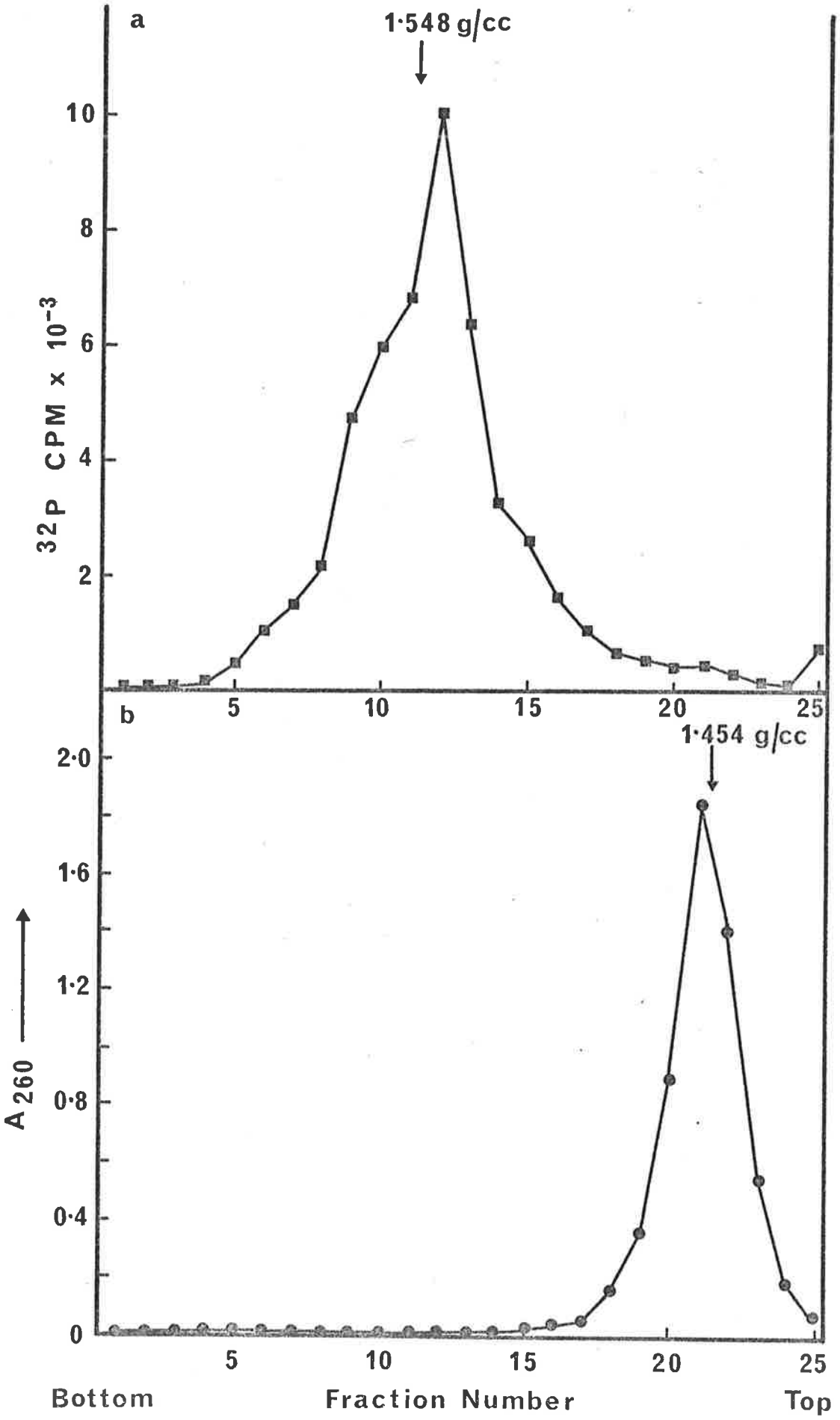
DENSITY OF SINGLE-STRANDED DNA AND 100% HYBRID IN $CsSO_4$ GRADIENTS

100% Hybrid was prepared from ^{32}P labelled keratin cDNA and keratin mRNA as described in Chapter V.B.3. 100,000 c.p.m. Of this hybrid was added to 2 ml of pre-hybridization buffer then added to a solution of Cs_2SO_4 to give a final volume of 9 ml and a density of 1.5 g/cc (Chapter V.B.4). 100 µg Of high molecular weight chick DNA was denatured with alkali, neutralized, added to 2 ml of pre-hybridization buffer and the gradient prepared as described for 100% hybrid. Conditions of centrifugation were as described in Chapter V.B.4. Gradients were fractionated from the bottom and densities determined by refractometry. Hybrid was detected by TCA precipitation of the fraction contents followed by radioactive counting. Single-stranded DNA was detected by reading the A_{260} of the fractions.

(a) 100% hybrid.

(b) Single-stranded DNA.

Density increases from right to left. Densities of 100% hybrid and single-stranded chick DNA are indicated.


a
1·548 g/cc
10
8
6
4
2
32P CPM x 10-3
5
10
15
20
25
b
1·454 g/cc
2·0
1·6
1·2
0·8
0·4
0
A260
5
10
15
20
25
Bottom
Fraction Number
Top

single-stranded DNA under these conditions was taken to be 1.454 g/cc.

(b) Pre-hybridization and detection of hybrid in Cs_2SO_4 gradients

Having determined the limiting density values it was necessary to be able to detect the hybrid generated in the pre-hybridization step after fractionation in the Cs_2SO_4 gradients. Pre-hybridization was performed using ^{125}I labelled 18S and 28S rRNA and 100 µg of DNA of single-stranded length $\geq$ 50 kb. Hybrids were centrifuged in Cs_2SO_4 gradients and the gradients fractionated from the bottom. Gradient fractions were divided into two. The fractions of one of these sets were treated with pancreatic RNA'ase followed by TCA precipitation (Chapter V.B.5.(a)) while the fractions of the other set were filtered through nitrocellulose discs to selectively immobilize the iodinated hybrid (Chapter V.B.5.(b)). The results from this experiment are shown in Figure 5.10(a). The figure shows that, after RNA'ase treatment, much of the RNA was still in TCA precipitable form. If the ribosomal sequences represent 0.035% of the chick genome (see Chapter IV.A.) and the specific activity of the ^{125}I labelled RNA used in the pre-hybridization step was 2×10^5 c.p.m./µg for each species, then the maximum number of counts expected in the hybrid generated under these conditions, is about 1.75×10^3 c.p.m., taking asymmetric transcription into account. The large amount of undigested

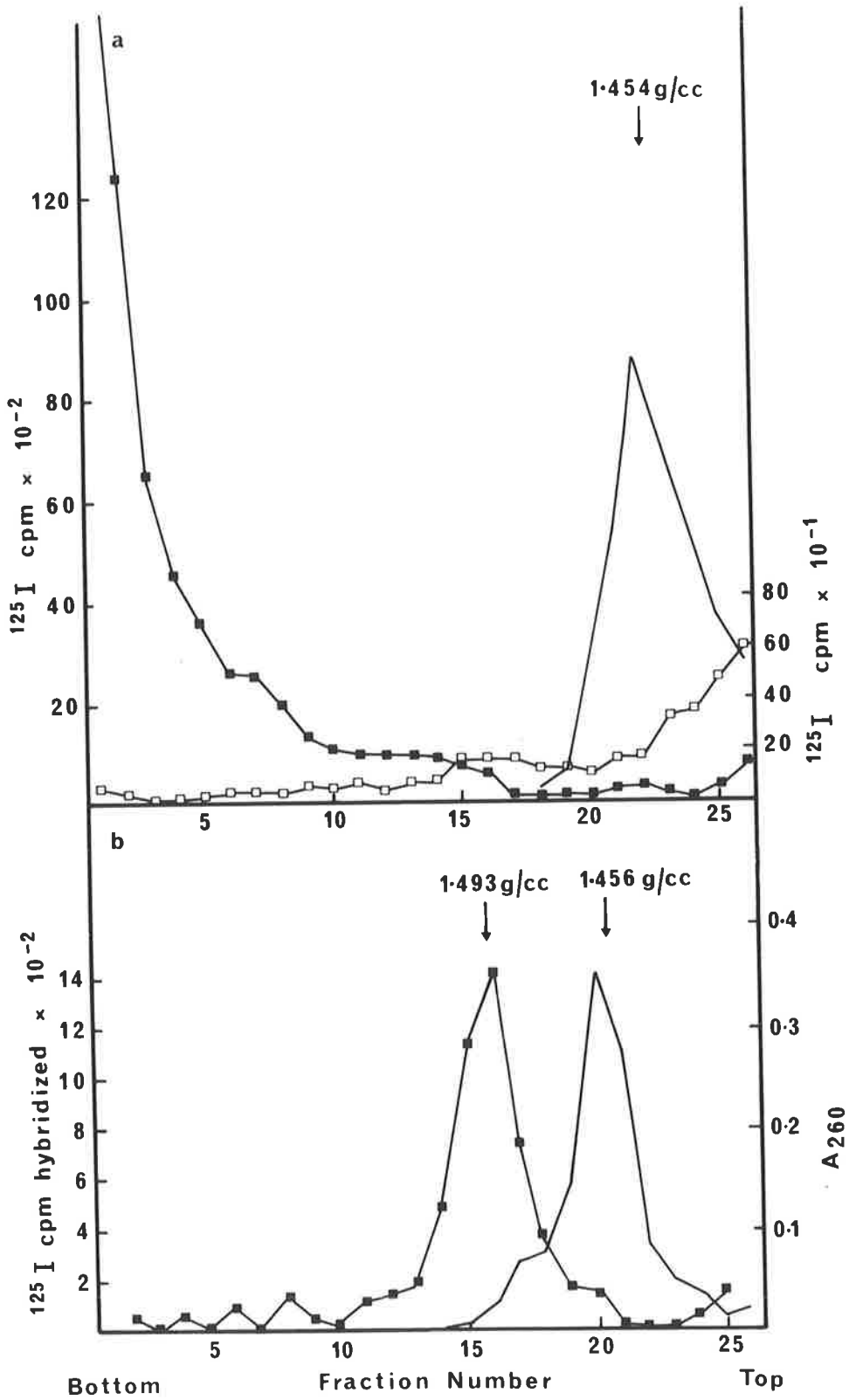
FIGURE 5.10.

HYBRID DETECTION TECHNIQUES

(a) Hybrids were prepared using 100 µg of high molecular weight DNA (≥ 50 kb) and 20 µg each of ^{125}I labelled 18S and 28S rRNA (Sp.Act. 5 x 10^5 c.p.m./µg) as described in Chapter V.B.2. The pre-hybridization solution was loaded on to Cs_2SO_4 gradients and centrifuged as described in Chapter V.B.4. Gradients were fractionated from the bottom, densities determined by refractometry and the fractions divided in two. One set of fractions was treated with pancreatic RNA'ase as described in Chapter V.B.5.(a), the other set being filtered through nitrocellulose discs as described in Chapter V.B.5.(b). —— A_{260} (single-stranded chick DNA), ■ RNA'ase treatment, □ filter binding. The density of single-stranded chick DNA is shown.

(b) Hybrids were prepared using 50 µg of high molecular weight DNA and 20 µg each of unlabelled 18S and 28S rRNA. Gradients were prepared and centrifuged as above. After fractionation of the gradients from the bottom, densities were determined by refractometry, DNA from individual fractions was immobilized on nitrocellulose discs and the filters challenged with ^{125}I labelled 18S and 28S rRNA (Sp.Act. 2.5 x 10^7 c.p.m./µg) as described in Chapter IV.B.3. —— A_{260} (single-stranded chick DNA), ■ ^{125}I c.p.m. hybridized (Hybrid).

Densities for single-stranded chick DNA and hybrid are indicated.


a
1·454g/cc
120
100
80
60
40
20
^{125}I cpm × 10^{-2}
80
60
40
20
^{125}I cpm × 10^{-1}
5
10
15
20
25
b
1·493g/cc
1·456 g/cc
14
12
10
8
6
4
2
^{125}I cpm hybridized × 10^{-2}
0·4
0·3
0·2
0·1
A_{260}
5
10
15
20
25
Bottom
Fraction Number
Top

iodinated RNA present in the bottom half of the gradient would, therefore, make detection of a hybrid peak, which could span a number of fractions, quite difficult. Backgrounds were much lower when hybrid retention on nitrocellulose discs was employed (Figure 5.10(a)). There was, however, no obvious hybrid peak. In addition, the number of filter retained counts was higher at the top of the gradient than elsewhere. This may be due to non-specific entrapment of ^{125}I labelled rRNA with single-stranded DNA since single-stranded DNA will bind to nitrocellulose. This cannot be the whole explanation, however, since the filter retained counts profile does not follow the A_{260} profile. These results were shown to be reproducible even when whole fractions were subjected to either of these treatments.

Owing to the failure of these simpler techniques to allow the detection of hybrids in Cs_2SO_4 gradients, a modification of the filter hybridization procedure, used by Birnstiel et al. (1968) to study the arrangement of ribosomal sequences in Xenopus laevis, was employed (Chapter V.B.5.(c)). 20 µg Each of unlabelled 28S and 18S rRNA was used in the pre-hybridization reaction with 50 µg of DNA of size $\geq$ 50 kb. Hybrids were loaded on to Cs_2SO_4 gradients and, after centrifugation, the gradients were fractionated from the bottom. The DNA content of the fractions was determined by measurement of their A_{260}, then the DNA was immobilized on nitrocellulose filters. Filters were challenged with ^{125}I labelled rRNA

(specific activity 25 x 10^6 c.p.m./μg), the filters washed and the radioactivity bound to the filters determined. The results are shown in Figure 5.10(b). Single-stranded DNA exhibited a density of 1.456 g/cc while the hybrid appeared as a sharp peak at a density of 1.50 g/cc. Thus, by using filter hybridization it was possible to detect the position of the hybrid in Cs_2SO_4 gradients.

To ensure that the peak ascribed to hybrid in Figure 5.10(b) was in fact due to hybrid and not some odd banding behaviour of single-stranded ribosomal sequences in Cs_2SO_4, single-stranded DNA of length $\geq$ 50 kb was subjected to pre-hybridization conditions (without RNA) followed by Cs_2SO_4 centrifugation. After centrifugation, the gradient was fractionated, the DNA distribution determined by measurement of the A_{260}, the DNA from each fraction immobilized on nitrocellulose filters, filters challenged with ^{125}I labelled rRNA and the distribution of radioactivity compared with that in Figure 5.10(b). Figure 5.11 shows that single-stranded ribosomal sequences banded to the heavy side of the A_{260} profile for single-stranded chick DNA, a result not initially expected. It was not possible, however, to completely explain the resolution between the hybrid and single-stranded DNA peaks of Figure 5.10(b) on the basis of the G + C content of the ribosomal sequences. This suggested that the filter hybridization technique was in fact detecting the generated hybrid.

From the results shown in Figure 5.10, it was apparent that the best method for detection of hybrids in the Cs_2SO_4 gradients was filter hybridization. Since the density of single-stranded ribosomal sequences did not coincide with that of total DNA (Figure 5.11) and since the average G + C content of these sequences may vary with DNA fragment size, it was necessary to determine the density of these sequences for each size class of DNA used. These two requirements were most easily satisfied by immobilizing the DNA from each fraction on nitrocellulose discs, splitting the discs in two and challenging one set of half filters with ^{125}I labelled rRNA and the other set with cDNA made to rRNA. Clearly the use of cDNA hybridization to indicate the density of the single-stranded sequence under study assumes that all these genes are transcribed from the same strand of the DNA. Thus the importance of control experiments subjecting single-stranded DNA to pre-hybridization conditions in the absence of any RNA, followed by Cs_2SO_4 gradient analysis, cannot be overstressed (see DISCUSSION).

(c) Comments on pre-hybridization conditions

The hybrid-density gradient method for the analysis of gene arrangement relies mainly on two conditions being met during the pre-hybridization reaction:-

(i) RNA:DNA hybridization must be driven

FIGURE 5.11.

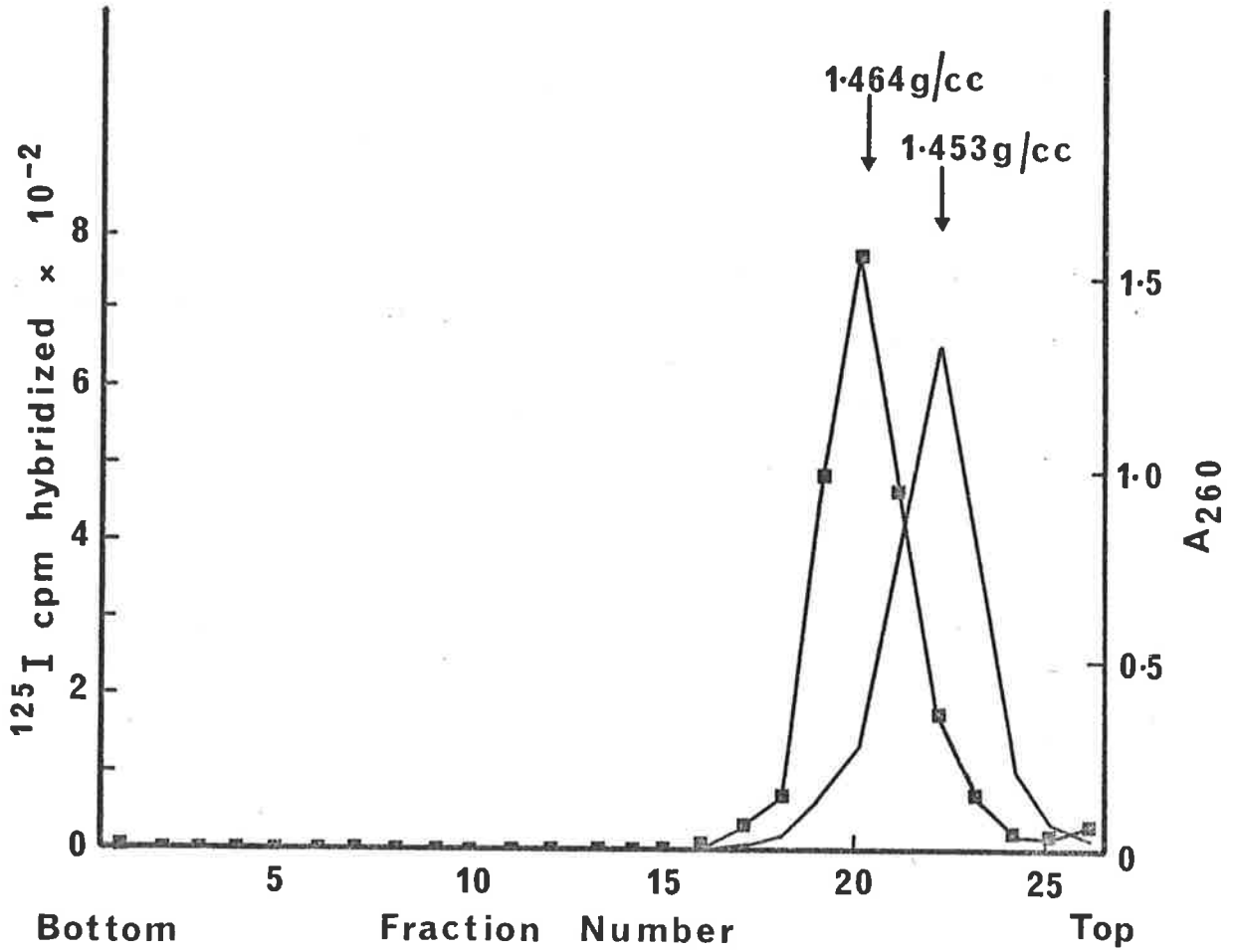
DENSITY OF SINGLE-STRANDED RIBOSOMAL SEQUENCES

High molecular weight chick DNA (50μg) was denatured with alkali, neutralized, added to pre-hybridization buffer and the Cs_2SO_4 gradient prepared and centrifuged as described in Chapter V.B.4. The gradient was fractionated from the bottom into 0.3 ml fractions, densities determined by refractometry and the A_{260} of each fraction determined. DNA from each fraction was immobilized on nitrocellulose discs and the filters challenged with ^{125}I labelled 18S and 28S rRNA (Sp. Act. 2.5×10^7 c.p.m./μg) to detect ribosomal sequences (Chapter IV.B.3).

—— A_{260} (single-stranded chick DNA).

■ ^{125}I c.p.m. hybridized (single-stranded ribosomal sequences).

Densities of single-stranded total chick DNA and ribosomal sequences are indicated.


1·464g/cc
1·453g/cc
^{125}I cpm hybridized × 10^{-2}
8
6
4
2
0
1·5
1·0
0·5
0
A_{260}
5
10
15
20
25
Bottom
Fraction Number
Top

ahead of DNA reassociation.

(ii) The conditions of pre-hybridization must not lead to the breakdown of single-stranded DNA.

It is known that the rate of DNA reassociation increases with increasing length of the reacting molecules and decreases with increasing viscosity of the solution. Taking these factors into account, it was possible that the apparent extent of reassociation attained, under the conditions of pre-hybridization employed, might have been greater than that expected for molecules of average length 500 bases at an equivalent value of $C_o t$, If this was so, then reassociation of the sequences under study would not permit saturation of the DNA with RNA. In addition, given that RNA saturation of the DNA occurred, reassociation of the DNA surrounding the sequences of interest would lead to the production of double-stranded complexes which could carry additional single-stranded DNA into the hybrid structure. On density gradient analysis, both of these effects would be expected to produce a broad hybrid peak of reduced density with a shoulder tailing towards the light side. In a similar way, the shape of the single-stranded peak, as determined by cDNA hybridization, would be expected to be of higher density and broad with a shoulder tailing towards the heavy side.

To determine whether DNA reassociation

was affecting the density gradient analysis of the ribosomal cistrons, pre-hybridizations were performed using 50 µg of high molecular weight DNA and two different levels of 18S and 28S rRNA. The resultant hybrid structures were analysed on Cs_2SO_4 gradients and a qualitative assessment of reassociation effects was made on the basis of the respective shapes of the peaks corresponding to the hybrid and single-stranded ribosomal sequences.

Figure 5.12(a) shows the profiles obtained using 20 µg each of 18S and 28S rRNA in the pre-hybridization reaction. The hybrid peak was broad and bimodal with the smaller peak coinciding with the main peak of cDNA hybridization. The cDNA hybridization pattern consisted of a sharp peak at a density similar to that expected for single-stranded ribosomal sequences (Figure 5.11) with a shoulder tailing off to the heavy side. This pattern was compatible with that expected if some reannealing of sequences within the ribosomal cistrons had occurred during the pre-hybridization reaction.

Figure 5.12(b) shows the profiles obtained using 60 µg each of 18S and 28S rRNA in the pre-hybridization reaction. The ^{125}I labelled rRNA hybridization pattern indicated that the hybrid peak was still broad but was no longer bimodal. The cDNA hybridization showed a main peak of hybridization similar to that for single-stranded ribosomal sequences and a trailing heavy shoulder. It should be noted, however,

FIGURE 5.12.

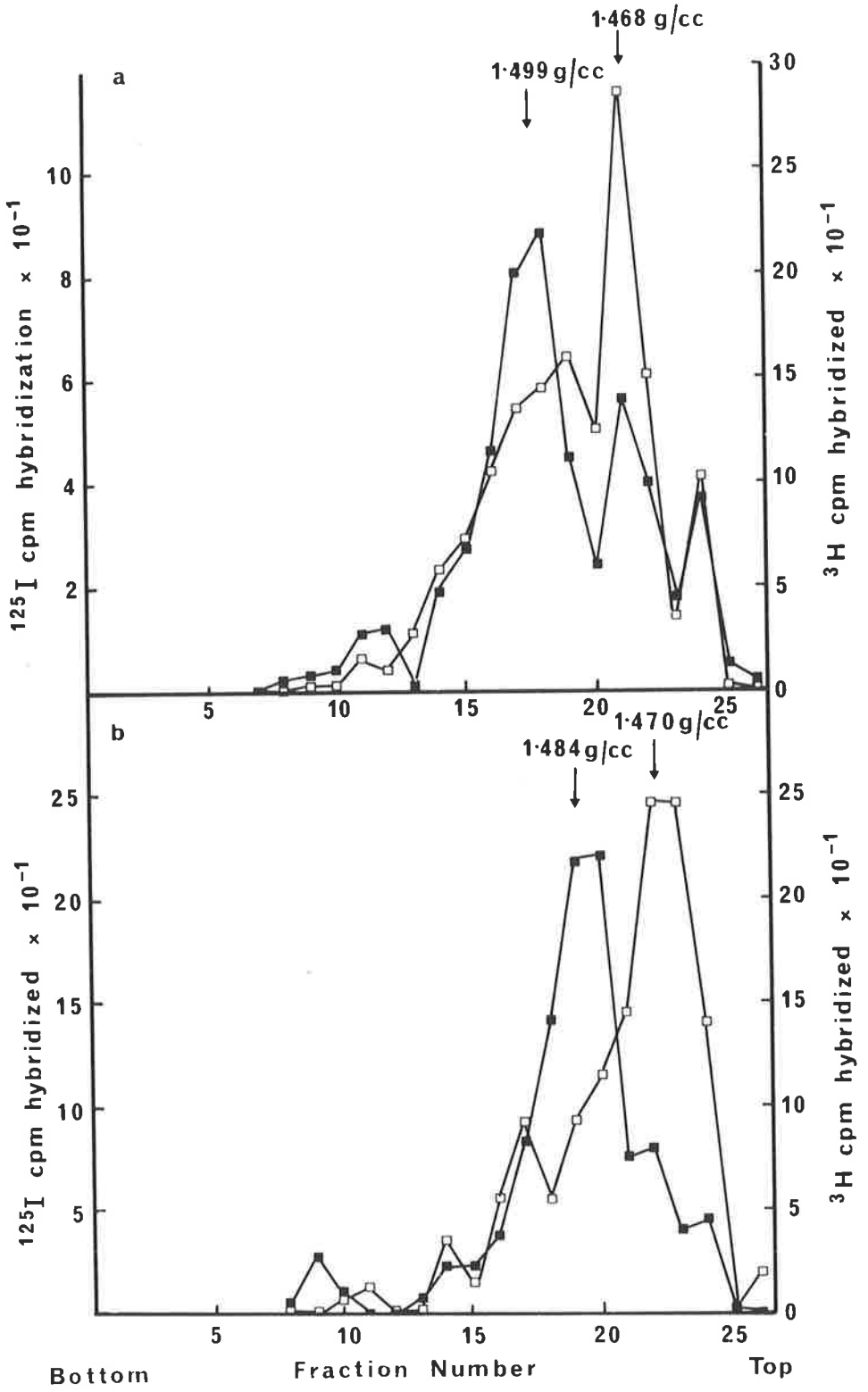
AMOUNT OF RNA REQUIRED FOR RNA:DNA HYBRIDIZATION IN THE ABSENCE OF DNA REANNEALING

Hybrids were generated using 50 µg of high molecular weight chick DNA and 20 µg or 60 µg each of 18S and 28S rRNA as described in Chapter V.B.2. The pre-hybridization solution was added to Cs_2SO_4 solution and the gradients prepared and centrifuged as described in Chapter V.B.4. Gradients were fractionated from the bottom, densities determined by refractometry, the DNA from each fraction immobilized on nitrocellulose filter discs and these filters split in half. One set of half filters was challenged with ^{125}I labelled rRNA (Sp.Act. 3×10^7 c.p.m./µg) to detect ribosomal sequences present in the hybrid and the other set was challenged with cDNA made to rRNA to determine the density of single-stranded ribosomal sequences.

(a) Hybrids generated using 20 µg each of 18S and 28S rRNA; ■ ^{125}I c.p.m. hybridized (hybrid), □ ^{3}H c.p.m. hybridized (single-stranded ribosomal sequences).

(b) Hybrids generated using 60 µg each of 18S and 28S rRNA; ■ ^{125}I c.p.m. hybridized (hybrid), □ ^{3}H c.p.m. hybridized (single-stranded ribosomal sequences).

Densities of hybrid and single-stranded ribosomal sequences are indicated.


a
1·468 g/cc
1·499 g/cc
^{125}I cpm hybridization × 10^{-1}
^{3}H cpm hybridized × 10^{-1}
b
1·470 g/cc
1·484 g/cc
^{125}I cpm hybridized × 10^{-1}
^{3}H cpm hybridized × 10^{-1}
Bottom
Fraction Number
Top

that this heavy shoulder was not as marked as that shown in Figure 5.12(a). Under these latter conditions, therefore, it appeared that DNA reassociation within the ribosomal cistrons was less marked.

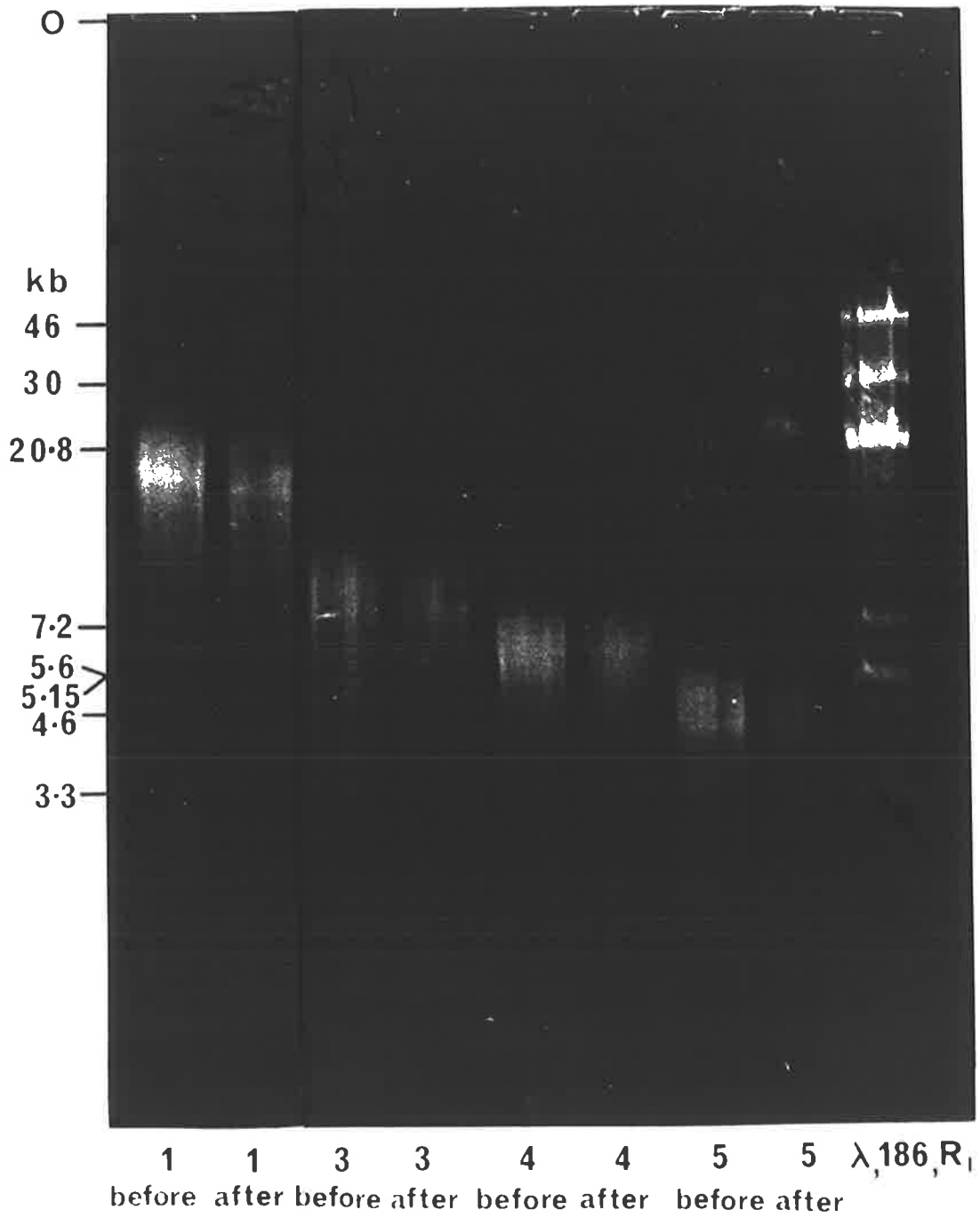
It can be seen from Figure 5.12 that the density of the peak of cDNA hybridization in these experiments did not vary significantly from that for single-stranded ribosomal sequences shown in Figure 5.11. From this observation and the fact that the difference between hybrid peaks and shoulders were clearly discernible, it was concluded that DNA reassociation had only affected the density of a proportion of the molecules present (about 25% in Figure 5.12(b)) and that the peak of ^{125}I labelled rRNA hybridization and cDNA hybridization probably represented the true densities of saturated hybrid and single-stranded ribosomal sequences respectively. It was also apparent that increasing the amount of RNA used in the pre-hybridization step, accompanied by a shorter incubation time, or decreasing the amount of DNA used to form the hybrid, should further decrease the effect of DNA reassociation on density analysis.

To determine whether single-stranded DNA size classes were degraded during the pre-hybridization process, samples from various size classes of single-stranded DNA were divided in two with one half being subjected to denaturation, neutralization and pre-hybridization conditions while the other half was used as a control. Figure 5.13 shows the comparison

FIGURE 5.13.

THE EFFECT ON THE CONDITIONS OF PRE-HYBRIDIZATION ON SINGLE-STRANDED DNA SIZE

Four DNA size classes were isolated from slices of a preparative alkaline agarose gel as described in Chapter V. B.1.(c)(iii). 20 µg Of each size class was subjected to pre-hybridization conditions in the absence of RNA (alkali denaturation, neutralization, made up to 2 ml of 2 x SSC, incubated at 60°C for 30 min). 1 µg Of each of these was ethanol precipitated, redissolved in 40 µl of 0.1 M NaOH, 1 mM EDTA and electrophoresed on an analytical 0.4% alkaline agarose gel as described in Chapter V.B.1.(b)(ii). 1 µg Of DNA from the same initial size class was electrophoresed in the adjacent track. λ DNA, 186 DNA and λ DNA digested with $EcoR_I$ were used as molecular weight markers. Numbers represent the distance, in cm, of the slice from which the DNA was extracted from the origin of the preparative gel. O, origin.


O
kb
46
30
20·8
7·2
5·6
5·15
4·6
3·3
1 before
1 after
3 before
3 after
4 before
4 after
5 before
5 after
λ,186,R1

of the size of these single-stranded size classes before and after pre-hybridization conditions as determined by alkaline agarose gel electrophoresis. These results indicated that no appreciable DNA breakdown occurred during pre-hybridization.

(d) The relationship between density and the proportion of the DNA molecule in hybrid form

To investigate the relationship between the proportion of the DNA molecule in hybrid form and the density of this hybrid, relative to the densities of the two extremes of single-stranded DNA and 100% hybrid, the ribosomal cistrons of chick were used as the model system. Restriction analysis of total chick DNA and cloned ribosomal sequences has shown that the ribosomal repeat unit is about 27 kb long and contains 1 copy each of the structural genes for 18S and 28S rRNA (McClements and Skalka, 1977). 60 µg Of 28S rRNA was hybridized to 20 µg of single-stranded DNA from different size classes under the conditions outlined in Chapter V.B.2. The resultant mixture was centrifuged in Cs_2SO_4 gradients, the gradients fractionated and fractions treated as in Chapter V.B.5.(c). Since only 28S rRNA was used to form the hybrid, only ^{125}I labelled 28S rRNA was used as probe to detect it. The profiles obtained using single-stranded DNA size classes ranging in size from 5.9 kb to 14.6 kb are shown in Figure 5.14. The sharp symmetrical peaks of cDNA hybridization,

FIGURE 5.14.

HYBRID-DENSITY ANALYSIS OF CHICK RIBOSOMAL CISTRONS USING SINGLE-STRANDED DNA OF DIFFERENT SIZES

Single-stranded DNA size classes were isolated from preparative alkaline agarose gels as described in Chapter V.B.1.(c)(iii). 20 µg Of DNA from each size class and 60 µg of 28S rRNA was used in the pre-hybridization reactions. Pre-hybridization mixes were centrifuged in Cs_2SO_4 gradients (Chapter V.B.4) and the gradients fractionated from the bottom (0.3 ml fractions). Densities were determined by refractometry. DNA from each fraction was immobilized on nitrocellulose filter discs, filters split in half, one set of half filters being challenged with ^{125}I labelled 28S rRNA (Sp.Act. 3×10^7 c.p.m./µg) to detect the hybrid and the other with ^3H cDNA to rRNA to detect the density of single-stranded ribosomal sequences.

(a) DNA 14.6 kb; ■ ^{125}I c.p.m. (hybrid), □ ^3H c.p.m. (single-stranded ribosomal sequences).

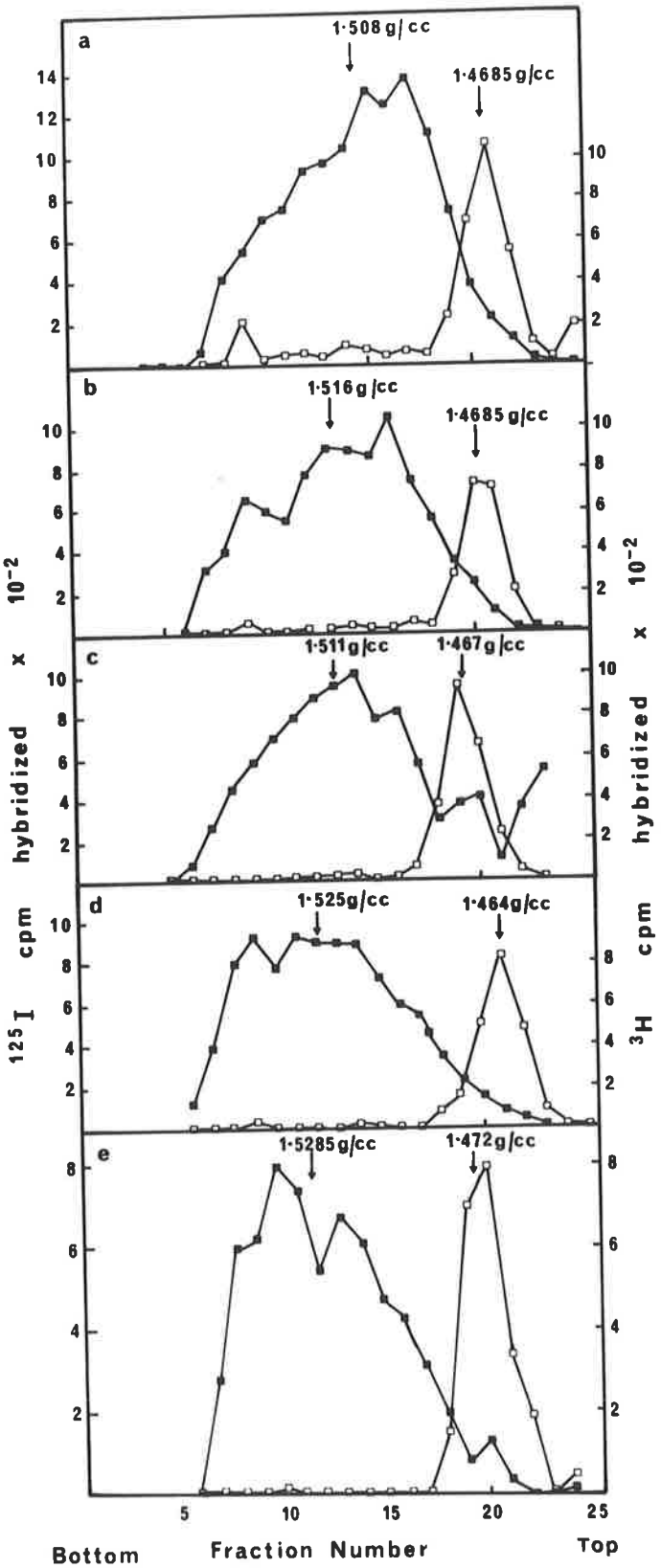
(b) DNA 12 kb; ■ ^{125}I c.p.m. (hybrid), □ ^3H c.p.m. (single-stranded ribosomal sequences).

(c) DNA 9 kb; ■ ^{125}I c.p.m. (hybrid), □ ^3H c.p.m. (single-stranded ribosomal sequences).

(d) DNA 7.2 kb; ■ ^{125}I c.p.m. (hybrid), □ ^3H c.p.m. (single-stranded ribosomal sequences).

(e) DNA 5.9 kb; ■ ^{125}I c.p.m. (hybrid), □ ^3H c.p.m. (single-stranded ribosomal sequences).

Densities of hybrid and single-stranded ribosomal sequences are indicated.


a
1·508 g/cc
1·4685 g/cc
b
1·516 g/cc
1·4685 g/cc
c
1·511 g/cc
1·467 g/cc
d
1·525 g/cc
1·464 g/cc
e
1·5285 g/cc
1·472 g/cc
^{125}I cpm hybridized x 10^{-2}
^{3}H cpm hybridized x 10^{-2}
Bottom
Fraction Number
Top

denoting the density of single-stranded ribosomal sequences, suggested that little or no DNA reassociation affected these sequences under the conditions of pre-hybridization used. The hybrid peaks, however, were very broad and it was difficult to obtain accurate estimates of the density of the hybrid. The hybrid densities shown in Figure 5.14 were determined by reading off the density corresponding to the midpoint of a line drawn horizontally across the hybrid peak at half the peak height.

Knowing the size of the single-stranded DNA used in hybrid formation, and knowing that 28S rRNA is 5 kb long, it was possible to calculate the average proportion of any given hybrid molecule expected to be in hybrid form by dividing the length of 28S rRNA by the length of the single-stranded DNA size class. This will only hold for single-stranded DNA up to the length of one repeat unit. The experimental value for the proportion of the DNA molecule in hybrid form was calculated by dividing the density difference between the generated hybrid and single-stranded ribosomal sequence by that between 100% hybrid and single-stranded sequences. Multiplication of this value by 100 gives the percent saturation. A comparison of theoretical and experimental estimates of the percent of the DNA saturated with RNA for the DNA size classes used in Figure 5.14 is shown in Table 5.2. The data in Table 5.2 indicate that there was very little similarity between the figures obtained by these two independent estimation procedures.

TABLE 5.2

THEORETICAL AND EXPERIMENTAL PERCENT SATURATION VALUES FOR THE DNA SIZE CLASSES OF FIGURE 5.14

Profile from Fig.5.14	Size (kb)	Theoretical Percent saturation	Experimental Percent saturation
a	14.6	34.2	48.5
b	12	41.7	58.3
c	9	55.6	53.0
d	7.2	69.4	70.9
e	5.9	84.7	72.4

Single-stranded DNA size classes were prepared by preparative alkaline agarose gel electrophoresis. Single-stranded DNA sizes were calculated by analytical alkaline agarose gel electrophoresis using λ phage DNA, 186 phage DNA and λ DNA digested with restriction endonuclease $EcoR_I$ as molecular weight markers. Pre-hybridization and density gradient centrifugation conditions were as described in the legend to Figure 5.14. Theoretical percent saturation values were calculated by dividing the length of 28S rRNA by the length of the DNA size class being used and multiplying by 100.

Experimental percent saturation was calculated by dividing the density difference between generated hybrid and single-stranded ribosomal sequences by that between 100% hybrid and single-stranded sequence and multiplying by 100.

This effect was expected for the hybrids produced using the lower molecular weight DNA fragments (5 kb to 8 kb) since many of the hybrid molecules produced under these conditions would be expected to have single-stranded RNA sections producing abnormally high hybrid densities and hence high percent saturation estimates. Where higher molecular weight DNA (9 kb - 50 kb) was used in the pre-hybridization, however, a much closer approximation of experimental results to theoretical estimates was expected since strucutres containing single-stranded RNA would be expected to be far less frequent. An improved agreement, however, was only observed when high molecular weight DNA was used in the pre-hybridization reaction.

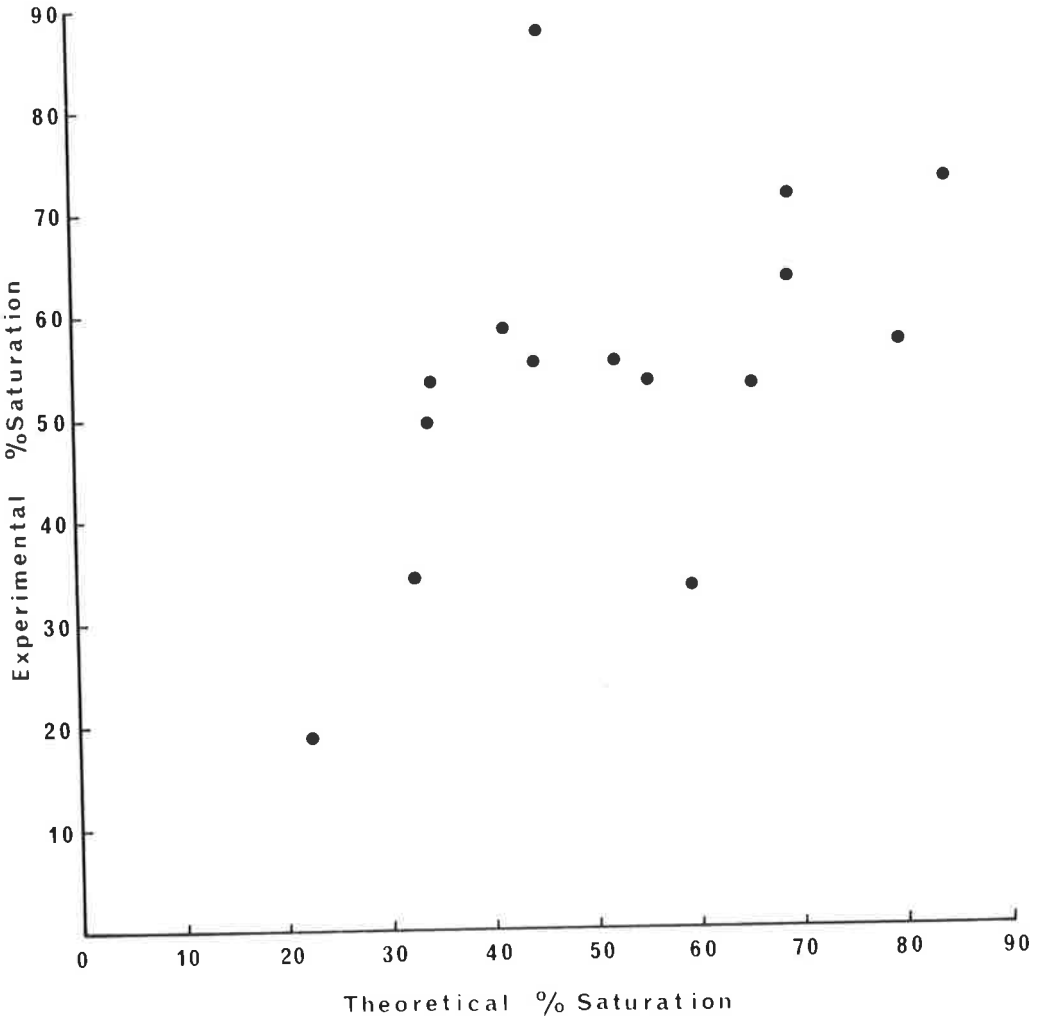
These results suggested that there was not a simple relationship between the theoretically estimated and the experimentally calculated percentage of the DNA saturated with RNA. Accordingly, theoretical and experimental percent saturations were determined for more DNA size classes, the data pooled and the experimentally determined percent saturation plotted against that expected for 28S rRNA hybridized to the different lengths of single-stranded DNA. These results are shown in Figure 5.15. The scatter of points is sufficient to discourage any attempt to draw a line of best fit through them. This result suggests that while there was a trend for the hybrid density to decrease with increasing DNA length, the basic information required for the calculation of the

FIGURE 5.15.

RELATIONSHIP BETWEEN THEORETICALLY AND EXPERIMENTALLY DERIVED ESTIMATES FOR THE AVERAGE PERCENTAGE OF A DNA MOLECULE CARRYING RIBOSOMAL CISTRONS IN HYBRID FORM

Hybrid-density analyses were performed for a number of single-stranded DNA size classes as described for Figure 5.14. From this experimental data an experimentally derived estimate for the average percentage of a DNA molecule, containing ribosomal cistrons, in hybrid form (percent saturation) could be made as described in the legend to Table 5.2. Knowing the length of the single-stranded DNA and the length of 28S rRNA, the theoretical value for the percent saturation could be determined as described in the legend to Table 5.2. The figure shows the relationship between these two values for a number of different single-stranded DNA size classes.

N.B.: All DNA size classes used were less than 1 chick ribosomal repeat unit in length.


90
80
70
60
50
40
30
20
10
Experimental %Saturation
0
10
20
30
40
50
60
70
80
90
Theoretical % Saturation

experimentally derived percentage saturation was insufficiently precise to permit any meaningful comparison with those values calculated by dividing the length of the saturating RNA by the molecular weight of the DNA size class being used. There are many possible sources of error in the experimental data obtained, and these will be dealt with in detail in the Discussion.

For the hybrid-density gradient procedure to be of any value in the elucidation of gene arrangement it was essential that some discernable relationship exist between the proportion of the DNA molecule in hybrid form and the density of that hybrid relative to single-stranded DNA and 100% hybrid. The inaccuracies of the density gradient procedure, however, render the technique inadequate for this purpose. For this reason no attempt was made to apply this system to the question of keratin gene organization.

D. DISCUSSION

1. Isolation of Single-stranded DNA Size Classes

Figures 5.2 to 5.4 show the successive stages of purification of single-stranded size classes of DNA obtained by repeated cycles of alkaline sucrose gradient centrifugation. It is obvious from these profiles that the procedure does not allow the selection of particularly homogeneous size classes of single-stranded DNA. This is reinforced by the analytical alkaline agarose gel patterns obtained from the purified size classes (Figure 5.5). The data in

Figure 5.5 also suggest that a large amount of DNA breakdown occurs during the alkaline sucrose gradient procedure. After the final alkaline sucrose selection step, the fraction containing the most DNA is that containing the smallest molecular weight class (< 3 kb). This DNA is too small to be of any practical use in the proposed hybrid-density gradient procedure for investigating gene arrangement. Since the size classes were heterogeneous and most of the DNA was too degraded, the alkaline sucrose gradient method for selection of single-stranded DNA size classes was not considered any further.

From Figure 5.6 it is apparent that preparative alkaline agarose gel electrophoresis can provide single-stranded DNA size classes of sufficient homogeneity for the proposed hybrid-density gradient analysis. Figures 5.7 and 5.8 show that as much as 2 mg of randomly sheared DNA can be fractionated per run on a preparative 0.5% alkaline agarose gel of dimensions 20 cm x 20 cm x 0.5 cm, if electrophoresis is carried out at 100 volts overnight. The major limitation of the preparative alkaline agarose gel method lies in the extraction of the DNA from agarose (Table 5.1).

Three methods for extracting DNA from agarose were tried; solubilization of agarose with KI followed by HAP chromatography, KI gradient centrifugation and centrifugation of the agarose slice followed by collection of the supernatant. The last of these was the most successful, yielding up to 30% of the DNA initially loaded on the gel. The success of this centrifugal squeezing procedure in extracting DNA from agarose, would be expected to decrease

as the percentage of agarose in the gel increased. Logically, the methods involving solubilization of agarose, followed by separation of DNA and agarose by physical procedures, should have been the most successful extraction techniques. The most likely reason for their lack of success in the present instance is that the DNA was being handled in its single-stranded form. Since both solubilization DNA extraction procedures involved extensive dialysis at some point (Chapter V.B.1.(c)) most of the losses using these methods were probably incurred by the single-stranded DNA adsorbing to the dialysis tubing.

To summarize, the results from the first part of this chapter suggest that single-stranded DNA size classes, suitable for use in the proposed hybrid-density gradient method for analysing gene organization, can be obtained by preparative alkaline agarose gel electrophoresis of randomly sheared chick DNA. After slicing the gel at right angles to the direction of electrophoresis, the DNA from each slice can be extracted by centrifugation of the slice and collection of the supernatant. All single-stranded DNA size classes used in the experiments described in the second part of this chapter were obtained using this procedure.

2. Studies on the Hybrid-density Gradient Procedure for the Analysis of Gene Arrangement

When these studies were initiated, the techniques of restriction enzyme analysis, Southern transfers (Southern, 1975(b)) and DNA cloning (Wensink *et al.*, 1974) were not in general use. There had been, however, a number of highly informative studies on the arrangement of the sequences for

18S and 28S rRNA in Xenopus laevis (Wallace and Birnstiel, 1966; Birnstiel et al., 1968 and Brown and Weber, 1968b), in numerous other eukaryotes (Sinclair and Brown, 1971) and the arrangement of tRNA genes of Xenopus laevis, all of which exploited the density difference between RNA:DNA hybrids and single-stranded DNA in CsCl. Investigations designed to ascertain the value of a modification of these techniques for determining the arrangement of tandemly linked families of eukaryotic genes were therefore begun. The ribosomal cistrons of chick were chosen as the experimental system because of the ready availability of large amounts of highly purified chick 18S and 28S rRNAs. If the procedure proved to be successful it was planned to use it for the analysis of the arrangement of keratin genes in the chick genome. During the course of this work the arrangement of the chick ribosomal cistrons, as determined by restriction analysis of total chick DNA and cloned chick ribosomal sequences using the Southern transfer technique, was published by McClements and Skalka (1977) allowing a critical evaluation of the hybrid-density technique to be made. The project was persued, nevertheless, in the belief that it might provide valuable information about average gene arrangements. These results could then be compared with those obtained from Southern transfer analysis, permitting more objective interpretations of the gene arrangement to be made.

(a) Requirements for the hybrid-density gradient procedure

For the technique to be useful for the

analysis of the average arrangement of a tandemly repeated family of genes, it was necessary to determine the relationship between the proportion of a single-stranded DNA molecule in hybrid form and the density of the hybrid structure relative to those of the two extremes of single-stranded DNA and 100% hybrid. Before this relationship could be investigated, however, it was necessary to determine the limiting densities and to be able to detect the hybrid, generated in the pre-hybridization reaction, in the density fractionating medium. Cs_2SO_4 was selected as the fractionating medium rather than CsCl since, in early studies, 100% hybrid was found to pellet at the bottom of the latter type of gradient (results not shown). Figure 5.9 shows that it was possible to obtain good separation of 100% hybrid from single-stranded chick DNA under the conditions of centrifugation chosen (density 1.5 g/cc, 28,000 r.p.m., Ti50 rotor, 20°C, 60-70 hours). The peak obtained for 100% hybrid was always quite broad which made the accurate estimation of this limiting density difficult. The breadth of this peak was presumably due to the diffusion of the small hybrid molecules around their actual banding density. So as to reduce the subjectivity involved in estimating densities from broad peaks, a line was drawn across the peak at half the peak height and the mid-point of this line was taken as being the actual banding position of the structure. This technique was used for all density estimates reported in this chapter.

Three methods for detecting the hybrid generated in the pre-hybridization reaction were attempted. The first involved the use of ^{125}I labelled rRNA to form the hybrid, fractionation of the hybrid on Cs_2SO_4 gradients followed by digestion of the unhybridized ^{125}I labelled RNA in each fraction with pancreatic RNA'ase A and TCA precipitation of the resistant material (Figure 5.10(a)). Much of the RNA pelleted as would be expected from its density (Szybalski, 1968), but much of it was also resistant to RNA'ase under the conditions of digestion used, causing high backgrounds of radioactivity especially in the fractions from the bottom half of the gradient. This resistant ^{125}I labelled material was presumably 28S rRNA since it has a high degree of secondary structure. Similar problems with the digestion of ^{125}I labelled 28S rRNA with pancreatic RNA'ase A have been encountered by another member of this Department (A. Robins, personal communication).

It was reported by Wallace and Birnstiel (1966) that hybrids made with radioactively labelled RNA bound to nitrocellulose filters while unhybridized RNA did not. When hybrids were generated with ^{125}I labelled rRNA and the fractions from the Cs_2SO_4 gradients filtered through nitrocellulose under these conditions, no unhybridized RNA bound to the nitrocellulose, but there was no discernable hybrid peak either (Figure 5.10(a)).

The third method was a slight modification

of that used by Birnstiel *et al*. (1968) in their study of the ribosomal cistrons of *Xenopus laevis*. Pre-hybridization was carried out using unlabelled rRNA, the mixture fractionated on Cs_2SO_4 gradients, the DNA from the resulting fractions was immobilized on nitrocellulose filters and the filters challenged with ^{125}I labelled rRNA. Figure 5.10(b) shows that it was possible to detect the hybrid using this technique. When single-stranded DNA was subjected to pre-hybridization conditions in the absence of RNA, centrifuged in Cs_2SO_4 gradients and the banding position of ribosomal sequences determined by filter hybridization, these sequences were found to be more dense than the bulk of chick DNA (Figure 5.11), but less dense than the hybrid shown in Figure 5.10(b).

These results suggested that it was possible to detect hybrid in Cs_2SO_4 gradients by filter hybridization but that it was not valid to assume that the density of single-stranded ribosomal sequences was the same as that of single-stranded total chick DNA. For example, the structural genes for 18S and 28S rRNA need not necessarily be of the same G + C content as the spacers, and since different molecular weight classes of single-stranded DNA were to be used in the evaluation of the hybrid-density gradient procedure, it would be necessary to determine the single-stranded density of ribosomal sequences for each size class. To achieve this, pre-hybridization was performed, the mixture centrifuged in Cs_2SO_4 gradients,

the gradients fractionated and the DNA from the fractions immobilized on nitrocellulose discs. The discs were then split in halves, one set of half filters being challenged with ^{125}I labelled rRNA to detect the hybrid and the other half being challenged with cDNA made to rRNA to detect the single-stranded ribosomal sequences.

This internal single-stranded density marker can only be accurate for a family of genes which are copied off the same strand of DNA. This would seem to be the case with the ribosomal cistrons of chick since the single-stranded density of ribosomal sequences did not appear to vary markedly regardless of whether the DNA used in the pre-hybridization reaction was 6 kb or $\geqslant$ 50 kb in length. If, however, the genes under study were transcribed off both strands, as is the situation with the Drosophila melanogaster histone genes (Lifton et al., 1977), the density of the sequences hybridizing to cDNA when short DNA (slightly greater than gene length) was used in the pre-hybridization reaction would be much lower than that observed when DNA of repeat unit length was used, providing all appropriate mRNAs were present in the pre-hybridization reaction. In fact, if long single-stranded DNA were used, the hybrid peak detected by hybridization using labelled mRNAs would be expected to be coincident with the peak detected by cDNA hybridization.

Ideally the density gradients should

contain markers for both single-stranded sequence and 100% hybrid. Thus experiments were performed in which ^{32}P labelled 100% hybrid was included in the gradient to act as the second internal density marker. The banding position of this hybrid was then determined by Cerenkov counting of each fraction. It was found, however, that the ^{32}P labelled cDNA, present in the marker, bound to the nitrocellulose and masked the hybrid peak being detected by hybridization with ^{125}I labelled rRNA. Varying the settings on the scintillation counter for double labelling using ^{125}I and ^{32}P did not improve the situation since there were many more ^{32}P counts bound than ^{125}I counts hybridized. Owing to the unavailability of a γ counter, it was therefore not possible to include both density markers in the gradients. The density of 100% hybrid was therefore estimated in a number of independent experiments and the average of these taken as the true density.

(b) Hybridization conditions

For the hybrid-density gradient method to be of any use in determining the arrangement of tandemly repeated genes, it was necessary that the RNA:DNA hybridization was driven ahead of DNA reassociation during the pre-hybridization reaction and that the DNA was not broken down during this reaction. Since the DNA being used in pre-hybridization was long it was not reasonable to assume that the extent of reassociation obtained under these conditions would

be the same as that expected for molecules of length 0.5 kb at the same $C_o t$ value (0.5 kb was the length of the DNA fragments used in reassociation studies described in Chapter III). Figure 5.12(a) shows that when 20 µg each of 18S and 28S rRNA was incubated with 50 µg of single-stranded DNA of high molecular weight for 60 min, a considerable amount of DNA reassociation within the ribosomal cistrons occurred. This result conflicts with that shown in Figure 5.10(b) where the hybrid peak was sharp giving no evidence of reassociation within the ribosomal cistrons. The explanation for this lies in the fact that in many of the early pre-hybridization experiments, neutralization of the alkali denatured DNA was not performed with sufficient care leading to precipitation of a portion of the DNA. In these experiments vigorous agitation was used to redissolve at least some of the DNA. It was, therefore, possible that less than 50 µg of DNA was used in the experiment shown in Figure 5.10(b) and that this DNA, at the time of hybridization, was of somewhat smaller size than 50 kb. This alteration of conditions may have allowed hybridization to occur without DNA reannealing.

Figure 5.12(b) showed that when 60 µg each of 18S and 28S rRNA was hybridized to 50 µg of high molecular weight DNA, the amount of DNA reassociation occurring was markedly reduced. It was, therefore, apparent that by increasing the amount of RNA in prehybridization and decreasing the incubation time, or by decreasing the amount of DNA, alteration of the

density of the hybrid band due to DNA reassociation within the ribosomal cistrons could be avoided. The problem of reannealing effects was expected to be greater with DNA of length $\geqslant$ 50 kb than for smaller DNAs since the rate of reassociation is proportional to the square root of the DNA length. Accordingly, all subsequent pre-hybridization experiments employed 20 µg of a given DNA size class and 60 µg of whatever rRNA species was necessary.

After the initial problems with DNA neutralization had been overcome, Figure 5.13 clearly demonstrates that the treatments involved in the pre-hybridization reaction induced no appreciable DNA breakdown. Thus both the requirements of the pre-hybridization reaction were fulfilled.

(c) Hybrid-density studies using different single-stranded DNA size classes

For the hybrid-density method to be applicable to the study of eukaryote gene arrangement, it was necessary to determine the relationship between the proportion of a single-stranded DNA molecule in hybrid form and the density of that structure relative to single-stranded DNA and 100% hybrid. Once this relationship is determined for a known system, it could be used to analyse the average gene arrangement of any tandemly linked family of genes. The relationship was examined using the ribosomal cistrons of chick as a model system. 28S rRNA was used in pre-hybridization experiments with single-stranded DNA

size classes varying in length from just over structural gene size (5 kb) to repeat unit length (27 kb) (McClements and Skalka (1977)). In this size range any given length of single-stranded DNA should have carried only one structural gene for 28S rRNA. By dividing the length of the RNA by the length of the DNA used in pre-hybridization, it was possible to determine the average proportion of any DNA molecule containing ribosomal sequences which should be in hybrid form. 28S rRNA was used in this pre-hybridization step since it could be isolated in highly purified form by phenol extraction of large ribosomal subunits followed by molecular weight selection on sucrose gradients. It was important that no 18S rRNA was present since this would have led to anomalous hybrid densities. The proportion of the DNA molecule in hybrid form, as derived by experiment, was determined by dividing the density difference between generated hybrid and single-stranded sequences by that between 100% hybrid and single-stranded sequences.

Figure 5.14 shows the types of profiles obtained using size classes ranging from 5.9 kb to 14.6 kb. The sharp symmetrical peaks of cDNA hybridization, denoting the banding position of single-stranded ribosomal sequences, suggested that DNA reassociation within the ribosomal cistrons had not occurred under the conditions of hybridization used. The broad nature of the hybrid peak, however, rendered accurate hybrid-density estimates difficult. When

the theoretical and experimental values for the percentage of a DNA molecule saturated with RNA were calculated for the DNA size classes used in Figure 5.14, an overall trend of decreased percent saturation with increasing DNA size was evident, but there were few similarities in the actual estimates for any given single-stranded DNA size class (Table 5.2). It was concluded that the relationship between these two estimation procedures may not be a simple one and so the appropriate calculations were carried out with single-stranded DNA of many different sizes. The theoretical percent saturation was compared with that calculated from the experimental data for each size class (Figure 5.15). From the broad scatter of points it was apparent that no empirical relationship could be readily obtained from these data. This result suggested that the basic data used for determining the experimental values for the percent saturation were not sufficiently accurate to allow any meaningful comparison of these two estimation procedures.

(d) Experimental shortcomings

The hybrid-density gradient procedure required accurate estimates for the density of 100% hybrid, generated hybrid and single-stranded sequences. To make accurate density estimates, the peaks needed to be sharp and symmetrical. This was clearly not the case for 100% hybrid (Figure 5.9(a)) and generated hybrid (Figure 5.14) while the peaks for single-stranded sequences seemed quite adequate. There was,

however, some variation in the single-stranded density estimates for DNAs of different sizes (Figure 5.14), but it was difficult to determine whether this was due to the different G + C contents of different lengths of ribosomal sequence or some variation in density estimates from gradient to gradient. This latter type of variation could be induced by slightly different salt concentrations (other than Cs_2SO_4) in the different gradient buffers since the apparent density of single-stranded DNA when loaded in T.E. was markedly different from the observed when the DNA was loaded in pre-hybridization mix (results not shown). To reduce the effect of such systematic errors, it would have been best to include a 100% hybrid-density marker in each gradient but, as described before, this was not possible due to the ^{32}P labelled cDNA of the hybrid binding to nitrocellulose and masking the generated hybrid peak.

Broad single-stranded molecular weight size classes and inaccurate size estimates could also contribute to the observed discrepancies. The DNA size classes, while being much better defined than those obtained from alkaline sucrose gradient sedimentation, may still have been sufficiently broad to contribute to the errors observed. Such inaccurate DNA size estimates would contribute to errors both in experimental and theoretical estimations of the proportion of the DNA molecule in hybrid form.

Little is known about the nature of the

non-transcribed spacer sequences in the chick ribosomal cistrons. If they were highly reiterated then the methods of pre-hybridization described here would have been inadequate to drive RNA:DNA hybridization ahead of DNA reannealing resulting in complex structures containing hybrid, single- and double-stranded DNA. The only evidence that DNA reassociation did not occur comes from the observation that the peaks of cDNA hybridization after Cs_2SO_4 centrifugation (Figure 5.14) were sharp and symmetrical. There must, therefore, remain some uncertainty as to whether any DNA reassociation within the ribosomal cistrons occurred during pre-hybridization which could have contributed to the broad hybrid peak and perhaps inaccurate hybrid-density estimates. This particular problem, in principle, might have been dispelled by carrying out pre-hybridization in formamide under conditions which permitted RNA:DNA hybridization without DNA reannealing (Casey and Davidson, 1977; Vogelstein and Gillespie, 1977). Removal of formamide may have been a problem since dialysis may have led to losses of single-stranded DNA while ethanol precipitation may have led to some DNA breakdown. The presence of other large sources of error, however, indicate that this modification could only marginally improve the practical estimates of the percentage of the DNA molecule in hybrid form.

3. Concluding Remarks

From the results presented in this chapter, it

is apparent that it was possible to isolate relatively homogeneous single-stranded DNA size classes by preparative agarose gel electrophoresis, to detect hybrids, generated by the pre-hybridization procedure, in the density gradient and to apparently obtain complete rRNA:DNA hybridization without reannealing of sequences within the ribosomal cistrons. However, it was not possible to empirically determine the relationship between the density of the hybrid formed, relative to those of single-stranded DNA and 100% hybrid, and the proportion of the DNA molecule in hybrid form. There were a number of reasons for this failure, but the main contributing factor was the inaccuracy of the density estimates for 100% hybrid and generated hybrid. Both of these peaks were broad making density estimates difficult. The 100% hybrid peak was broad because the molecules involved were small and so tended to diffuse further around their actual banding density. The breadth of the generated hybrid peak is more difficult to explain since the molecules involved here were much longer. At least some of the variation would be expected to derive from the fact that the DNA fragments were generated by random shearing of the DNA. Thus fragments containing sequences complementary to 28S rRNA would be expected to carry a whole 28S rRNA coding sequence or any part of it. If the RNA was not broken down by the denaturation step prior to pre-hybridization, those DNA fragments containing only a fraction of the 28S sequence would be expected to carry large portions of single-stranded RNA rendering the structure more dense than expected. Alternatively, if the RNA was fragmented by boiling, then those DNA fragments carrying a fraction of the

28S coding sequence would only hybridize a fraction of the RNA molecule producing a hybrid of density lower than the average hybrid-density expected for DNA fragments of that size. In practice, both of these effects would be expected to be acting. This would be expected to lead to marked over-estimates of the hybrid-density for small DNA size classes. For longer DNA molecules, however, (about 10 kb) the effect would be expected to be negligible since the probability of a given DNA fragment having a complete sequence complementary to 28S rRNA would increase with increasing fragment size. It was, therefore, expected that the lower molecular weight DNA size classes would show broader peaks for generated hybrid than higher molecular weight size classes. While a sharpening of the hybrid peak with increasing molecular weight of the single-stranded DNA size class was not apparent from Figure 5.14, whenever high molecular weight DNA was used in the pre-hybridization reaction, the hybrid peaks obtained were quite sharp (Figures 5.10(b) and 5.12(b)).

The inability to determine a relationship between the proportion of a DNA molecule in hybrid form and the density of this hybrid relative to the densities of 100% hybrid and single-stranded sequence rendered the hybrid-density gradient procedure inadequate for determining the average gene arrangement of a tandemly repeated gene family. Although it could allow the determination of whether the genes of a tandemly linked family are transcribed off the same DNA strand, the large amounts of RNA required for pre-hybridization and the high degree of RNA purity needed would make the technique useful only for the study of those genes producing a highly abundant RNA species. The currently

available recombinant DNA technology, however, permits the same sort of information to be obtained with much more accuracy in less time. While certain improvements could be made to the hybrid-density gradient procedure (e.g., carrying out pre-hybridization under formamide conditions which permit RNA:DNA hybridization in the absence of DNA reannealing and the inclusion of a 100% hybrid-density marker in each gradient, generated hybrid being detected by γ counting), it seems unlikely that the technique could yield information about gene arrangement of sufficient accuracy to enable critical comparisons with gene arrangements determined by the recombinant DNA and Southern transfer approach.

CHAPTER VI

CONCLUDING DISCUSSION

CHAPTER VI

CONCLUDING DISCUSSION

The work described in this thesis can be divided into two distinct sections. The significance of each part and their limitations will be discussed separately below. Possible future research on the elucidation of the mechanism of control of feather keratin genes will also be discussed.

1. Structure of Feather Keratin mRNA

From the studies described in Chapter III and elsewhere (Kemp, 1975), it is apparent that keratin mRNA is comprised of an heterogeneous family of mRNA species. While this constitutes a very interesting system, the inability to isolate single pure mRNA species has made the analysis of keratin mRNA structure difficult. Ultimately, the best way to study mRNA structure is by nucleotide sequencing as has been done for human β globin, rabbit β globin and chick ovalbumin mRNAs (Chapter I.D.1), but with a complex system like the avian keratin mRNAs, this can only be achieved by cDNA cloning. Since recombinant DNA technology has only become available to Australian biochemists in the past twelve months, studies on the structure of keratin mRNA have had to be limited to an analysis of different regions of the total mRNA population using hybridization (Kemp, 1975; Kemp, Lockett and Rogers, in preparation) and DNA reassociation. Earlier studies (Kemp, 1975) had indicated that keratin mRNA contains some sequences which are unique and others which are reiterated in the

chick genome. This observation was confirmed, and in addition, by using short PolI-cDNA copied from keratin mRNA, it was shown that at least a portion of the unique sequences are located at the 3' end of the mRNA. By using PolI-cDNA to prime the synthesis of longer reverse transcripts and use of this elongated probe in reassociation kinetic analyses, it was possible to demonstrate that unique and reiterated sequences are covalently linked in keratin mRNA. Furthermore, density gradient analysis of reassociated duplexes containing keratin PolI- and AMV-cDNA revealed that the 150 nucleotide sequences adjacent to the 3' poly(A) tract of keratin mRNA, has a lower average G + C content than the whole mRNA. From the amino acid analysis of unfractionated feather keratin (Kemp and Rogers, 1972), and assuming equal representation of each base triplet coding for a given amino acid, it can be calculated that the keratin coding sequence of the mRNA should have a high G + C content (54.5%) (Lockett and Kemp, 1975). The results of the density studies on reassociated duplexes would, therefore, suggest that the sequences adjacent to the poly(A) tract are not translated into keratin and, therefore, constitute a 3' untranslated sequence analogous to those found in other eukaryotic mRNAs that have been extensively characterized (Chapter I.D.1).

The complex reassociation kinetics observed using keratin AMV-cDNA as a probe have been explained by postulating that genes coding for keratin (and hence the mRNAs) contain a common sequence which is either short and faithfully conserved, or longer and mismatching. This interpretation is supported by the thermal denaturation studies (Figure 3.8). This is almost certainly the correct

interpretation for the majority of keratin genes, but the complexity of the keratin mRNA population (Kemp, 1975) makes it impossible to exclude the possibility that a subset of the keratin genes are uniquely represented in the chick genome.

Final analysis of keratin mRNA structure must rely on sequence analysis of cloned cDNAs. Recently double-stranded keratin cDNA has been cloned into the plasmid pBR322. Restriction endonuclease cleavage of genomic chick DNA coupled with Southern transfer studies (Southern, 1975b) using these pure probes, has revealed the existence of about 3 classes of mRNA. One class has been shown to reproduce the same complex autoradiographic banding pattern as total keratin AMV-cDNA when the filter washing procedures contain high levels of salt. When the stringency of the washing procedures is increased, the number of bands appearing decreases. These results suggest the existence of a large family of cross-hybridizing genes coding for keratin. The other classes appear to label fewer bands and the pattern does not vary greatly with filter washing conditions (R.B. Saint, personal communication). These observations, while being in good agreement with the results obtained by independent methods (Chapter III), also indicate which cloned cDNAs should be sequenced first to obtain the maximum information about keratin mRNA structure. With the availability of cloned cDNAs, it will also be possible to investigate the position and nature of the reiterated sequence in keratin mRNA. This could be achieved by carrying out reassociation kinetic analysis, or Southern transfer studies, on total genomic chick DNA using different restric-

tion fragments of the cDNA regions of the recombinant molecules as probes.

2. Studies on Gene Organization

The possibility of using a density gradient system, which exploited the density difference between RNA:DNA hybrids and single-stranded DNA, for the study of the arrangement of tandemly repeated genes was investigated. If the technique proved to be practical, it was to be used to study the arrangement of keratin genes in chick DNA. Since the keratin genes could only represent about 0.004% of the chick genome, it was necessary to partially purify these sequences to allow their easy detection in the hybrid-density gradient procedure. The presence of contaminating ribosomal sequences in keratin mRNA preparations also potentially posed problems in the interpretation of hybrid-density data obtained from studies involving keratin genes. It was, therefore, desirable to obtain a quantitative separation of DNA containing ribosomal and keratin sequences. Initial studies were therefore carried out to investigate the possibility of partially purifying sequences containing keratin genes by physico-chemical methods.

Thermal elution from HAP, while permitting an adequate separation of keratin and ribosomal sequences, resulted in substantial degradation and large losses of DNA. Therefore, this method could not be used in the preparative partial purification of keratin sequences (Chapter IV.C.1). CsCl gradient centrifugation using 200 µg of DNA per gradient, allowed a quantitative separation of keratin and ribosomal

sequences presumably with minimal DNA degradation (Tomizawa and Anraku, 1965). It has been shown previously that, by fractionation of chick DNA in CsCl gradients containing actinomycin D followed by a second CsCl step in the presence of netropsin sulphate, an eight-fold purification of keratin sequences was possible (Lockett and Kemp, 1975). Thus, by using CsCl gradients, to separate ribosomal and keratin sequences, followed by cycles of CsCl density gradient centrifugation in the presence of antibiotics, a quantitative separation of ribosomal and keratin sequences, as well as a substantial purification of DNA containing keratin sequences, could be achieved (Chapter IV.C.2).

The chick ribosomal cistrons were used as a model system for the evaluation of the hybrid-density gradient procedure as a method for the determination of gene organization. The reasons for this were two-fold:-

(a) Ribosomal sequences constitute a sufficiently large proportion of total chick DNA to permit their detection without any prior enrichment.

(b) Pure 18S and 28S rRNA species can be easily obtained in large quantities.

Excess 28S rRNA was hybridized to different size classes of single-stranded chick DNA under conditions which permitted only minimal DNA reassociation while allowing saturation of DNA coding for 28S rRNA with RNA (pre-hybridization). The resulting hybrids were centrifuged in Cs_2SO_4 gradients and the density of hybrid determined relative to the limiting

densities of single-stranded ribosomal sequences and 100% hybrid. From these values, and assuming a linear relationship between the proportion of the DNA in hybrid form and the banding position of this hybrid, relative to the two limiting densities, an experimentally derived value for the average percentage saturation of the DNA, containing ribosomal sequences, with RNA could be determined. Knowing the single-stranded size of the DNA used in the pre-hybridization reaction and the length of 28S rRNA, it was possible to estimate the average proportion of any given DNA molecule, containing ribosomal cistrons, expected to be in hybrid form. From this, a theoretical estimate of the percentage of the DNA molecule saturated with RNA could be obtained. From the comparison of these experimentally and theoretically derived estimates for a number of different single-stranded DNA size classes, it was apparent, from the scatter of points, that no meaningful relationship between these two estimation procedures could be drawn (Figure 5.15). Thus, the hybrid-density gradient technique appears to be insufficiently accurate for use in the analysis of the average arrangement of tandemly repeated genes.

Since the time of commencing this work, a number of new features of gene arrangement have been revealed. The sections below are designed to assess the degree of success that an accurate hybrid-density gradient procedure might have in detecting these arrangements.

(a) Intergenic spacer heterogeneity

The hybrid-density gradient technique could have, at best, only given information about

average gene arrangements. Thus, variations in the size of spacers, as found in the ribosomal cistrons of X. laevis and the 5S genes of this same toad (Chapter I.E.1.(a),(b)) could not have been detected.

(b) Intervening sequences in genes

The ability to detect intervening sequences in a tandemly linked set of genes by the hybrid-density gradient technique, would depend on the nature of the hybrid formed in the pre-hybridization reaction. If each part of the structural gene sequence hybridized to a complete mRNA molecule, then the presence of intervening sequences could be detected by comparing the densities of hybrids before and after RNA'ase treatment. Before RNA'ase treatment, hybrids would be expected to exhibit an anomalously high density while this density would be markedly reduced after RNA'ase treatment. If, however, one molecule of RNA hybridized to all structural gene sequences with out-loopings of DNA containing the intervening sequences (as is the situation in "R-loop" mapping), it would not be possible to detect the intervening sequences by the density procedure.

(c) Gene rearrangements

If gene rearrangement is a prerequisite for gene expression, the mode of detection of these changes would be dependent on the nature of the re-arrangements. Suppose that the inactive arrangement of the gene family under study is a random distribution

around the genome while in the active arrangement, these genes are tandemly linked. In this situation the rearrangements could be detected by using high molecular weight single-stranded DNA form producing and non-producing tissues in the pre-hybridization reaction and determining the density of the hybrid in Cs_2SO_4 gradients. If this mode of gene rearrangement occurs, hybrids formed using DNA from non-producing tissues would be expected to exhibit a density close to that of single-stranded sequences while hybrids formed using DNA from producing tissues would show a distinctly higher density. If, on the other hand, fragments of the genes existed as a tandem array in tissues where the genes were inactive, and the remaining fragments (whether originating from another tandem bank of fragments or from random sites in the genome) were inserted next to the first set of fragments to produce functional genes, this could be detected by hybrid-density gradient analysis before and after RNA'ase treatment. Where gene fragments were tandemly linked, the hybrids before RNA'ase treatment would be expected to exhibit an abnormally high density which would be observed to decrease after RNA'ase treatment. Where the whole structural gene is present in a tandem array of such genes, however, no such density difference before and after RNA'ase treatment would be expected. All of these density studies would, of course, have to be performed using single-stranded DNA of minimum length about 10 kb (Brown and Weber, 1968b).

While it is clear that an accurate hybrid-density gradient procedure might have been able to detect some of the recently discovered anomalies of tandemly repeated gene arrangements, no details about such arrangements could have been gleaned. When the amount of DNA, time, effort and detail of the results obtained for the density-gradient procedure is weighed against the same criteria for Southern transfer analysis and recombinant DNA techniques, it is obvious that the hybrid-density gradient procedure, even if all was going well, should not be perservered with.

It is apparent, therefore, that future efforts to understand the arrangement of keratin genes in the chick genome will rely heavily on recombinant DNA technology. With the isolation of genomic fragments containing keratin genes, it should be possible to definitively determine whether the keratin genes are linked, their relationship to each other, their polarity in the DNA and to identify, by DNA sequencing techniques, putative elements leading to the co-ordinate control of these genes.

Analysis of keratin gene organization, however, can only provide limited insight into the events surrounding the onset of terminal differentiation in the embryonic feather. Ultimately it will be necessary to understand the molecular events occurring in the chromatin at the time of differentiation and the mechanism by which the dermis can exert its control over epidermal differentiation.

Adams, J.M. and Cory, S. (1975). Nature 255, 28-33.

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

Baralle, F.E. (1977a). Cell 10, 549-558.

Baralle, F.E. (1977b). Nature 267, 279-281.

Baralle, F.E. (1977c). Cell 12, 1085-1095.

Baralle, F.E. and Brownlee, G.G. (1978). Nature 274, 84-87.

Bartels, E.A. (1943). Arch. f. Entwicklmech 142, 763-816.

Bastos, R.N. and Aviv, H. (1977). Cell 11, 641-650.

Beckingham Smith, K. (1973a). Dev. Biol. 30, 249-262.

Beckingham Smith, K. (1973b). Dev. Biol. 30, 263-278.

Beckmann, J.S. and Daniel, V. (1974). J. Mol. Biol. 89, 355-362.

Berget, S.M., Moore, C. and Sharp, P.A. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 3171-3175.

Berk, A.J. and Sharp, P.A. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 1274-1278.

Berns, A.J.M., Strous, G.J.A.M. and Bloemendal, H. (1972). Nature New Biol. 236, 7-9.

Birnstiel, M.L., Chipchase, M. and Speirs, J. (1971). Progress in Nucl. Acids Res. and Mol. Biol. 11, 351-389.

Birnstiel, M., Speirs, J., Purdom, I., Jones, K. and Leoning, U.E. (1968). Nature 219, 454-463.

Birnstiel, M., Telford, J., Weinberg, E. and Stafford, D. (1974). Proc. Nat. Acad. Sci. U.S.A. 71, 2900-2904.

Bonner, T.I., Brenner, D.J., Neufield, B.R. and Britten, R.J. (1973). J. Mol. Biol. 81, 123-135.

Brack, C. and Tonegawa, S. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 5652-5656.

Bratosin, S., Horowitz, M., Laub, O. and Aloni, Y. (1978). Cell 13, 783-790.

Brawermann, G. (1974). Ann. Rev. Biochem. 42, 621-642.

Breathnach, R., Mandel, J.L. and Chambon, P. (1977). Nature 270, 314-319.

Britten, R.J. and Kohne, D.E. (1968). Science 161, 529-540.

Brown, D.D. and Weber, C.S. (1968a). J. Mol. Biol. 34, 661-680.

Brown, D.D. and Weber, C.S. (1968b). J. Mol. Biol. 34, 681-697.

Brown, D.D., Wensink, P.C. and Jordan, E. (1971). Proc. Nat. Acad. Sci. U.S.A. 68, 3175-3179.

Brown, D.D., Wensink, P.C. and Jordan, E. (1972). J. Mol. Biol. 63, 57-73.

Brownlee, G.G. and Cartwright, E.M. (1976). J. Mol. Biol. 107, 527-547.

Brownlee, G.G., Cartwright, E.M. and Brown, D.D. (1974). J. Mol. Biol. 89, 703-718.

Brownlee, G.G. Cartwright, E.M., Cowan, N.J., Jarvis, J.M. and Milstein, C. (1973). Nature New Biol. 244, 236-240.

Brownlee, G.G. and Cartwright, E.M. (1977). J. Mol. Biol. 114, 93-118.

Bukhari, A.I., Shapiro, J.A. and Adhya, S.L. eds. (1977). "DNA Insertion Elements, Plasmids and Episomes" (New York : Cold Spring Harbour Laboratory press).

Callan, H.G. and Lloyd, L. (1960). Phil. Trans. Roy, Soc. B. 243, 135-219.

Carroll, D. and Brown, D.D. (1976a). Cell 7, 467-475.

Carroll, D. and Brown, D.D. (1976b). Cell 7, 477-486.

Casey, J. and Davidson, N. (1977). Nucl. Acids Res. 4, 1539-1552.

Catteral, J.F., O'Malley, B.W., Robertson, M.A., Staden, R., Tanaka, Y. and Brownlee, G.G. (1978). Nature 275, 510-513.

Cheng, C.C., Brownlee, G.G., Corey, N.H., Doel, M.T., Gillam, S. and Smith, M. (1976). J. Mol. Biol. 107, 527-545.

Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977). Cell 12, 1-8.

Clarkson, S.G. and Birnstiel, M.L. (1973). Cold Spring Harbour Symp. Quant. Biol. 38, 451-459.

Cohen, S. (1962). J. Biol. Chem. 237, 1555-1562.

Cohen, S. (1965). Dev. Biol. 12, 394-407.

Cohen, S. and Carpenter, G. (1975). Proc. Nat. Acad. Sci. U.S.A. 72, 1317-1321.

Cohen, S. and Stastny, M. (1968). Biochim. Biophys. Acta 166, 427-437.

Cohn, R.H., Lowry, J.C. and Kedes, L.H. (1976). Cell 9, 147-161.

Crawford, R.J. and Wells, J.R.E. (1976). Mol. Biol. Reports 3, 167-173.

Crawford, R.J., Scott, A.C. and Wells, J.R.E. (1977). Eur. J. Biochem. 72, 291-299.

Curtis, P.J. and Weissmann, C. (1976). J. Mol. Biol. 106, 1061-1075.

De Larco, J. and Guroff, G. (1973). Biochem. Biophys. Res. Commun. 50, 486-492.

Denhardt, D.T. (1966). Biochem. Biophys. Res. Commun. 23, 641-646.

Dhouailly, D. (1967). J. Embryol. Exptl. Morphol. 18, 389-400.

Dhouailly, D., Rogers, G.E. and Sengel, P. (1978). Dev. Biol. 65, 58-68.

Dodson, J.W. (1963). Exptl. Cell Res. 31, 233-235.

Doel, M.T., Houghton, M., Cook, E.A. and Carey, N.H. (1977). Nucl. Acids Res. 4, 3701-3713.

Dreyer, W.J. and Bennett, J.C. (1965). Proc. Nat. Acad. Sci. U.S.A. 54, 864-869.

Dugaiczyk, A., Woo, S.L.C., Lai, E.C., Mace Jr., M.L., McReynolds, L. and O'Malley, B.W. (1978). Nature 274, 328-333.

Efstratiadis, A., Kafatos, F.C. and Maniatis, T. (1977). Cell 10, 571-585.

Erikson, R.L. and Szybalski, W. (1964). Virology 22, 111-124.

Fell, H.B. (1957). Proc. Roy. Soc. (Biol.) 146, 242-256.

Fell, H.B. (1962). J. Embryol. Exptl. Morphol. 10, 389-409.

Fell, H.B. and Mellanby, E. (1953). J. Physiol. (London) 119, 470-488.

Fitton-Jackson, F.L. and Fell, H.B. (1963). Dev. Biol. 7, 394-419.

Forget, B.G., Marotta, C.A., Weissman, S.M., Verma, I.M., McCaffrey, R.P. and Baltimore, D. (1974). Ann. N.Y. Acad. Sci. 241, 290-309.

Fraser, R.D.B., MacRae, T.P. and Rogers, G.E. (1972). "Keratins, Their Composition, Structure and Biosynthesis" C.C. Thomas, Springfield, Illinois.

Furuichi, Y., Morgan, M., Muthukrishnan, S. and Shatkin, A.J. (1975a). Proc. Nat. Acad. Sci. U.S.A. 72, 362-366.

Furuichi, Y., Morgan, M., Shatkin, A.J., Jelinek, W., Salditt-Georgieff, M. and Darnell, J.E. (1975b). Proc. Nat. Acad. Sci. U.S.A. 72, 1904-1908.

Gambino, R., Kacian, D., O'Donnell, J., Ramirez, F., Marks, P.A. and Bank, A. (1974). Proc. Nat. Acad. Sci. U.S.A. 71, 3966-3970.

Gilbert, W. (1978). Nature 271, 501.

Gilmore-Herbert, M. and Wall, R. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 342-345.

Glover, D.M. and Hogness, D.S. (1977). Cell 10, 167-176.

Goddard, D.R. and Michaelis, L. (1934). J. Biol. Chem. 106, 605-614.

Gould, M.J. and Hamlyn, P.H. (1973). FEBS Letters 30, 301-304.

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.

Gross-Bellard, M., Oudet, P. and Chambon, P. (1973). Eur. J. Biochem. 36, 32-38.

Grunstein, M., Schedl, P. and Kedes, L. (1973). In "Molecular Cytogenetics". B. Hamkalo and J. Papaconstantinou eds. (New York Plenum Press) pp.115-122.

Gulati, S.C., Kacian, D.L. and Spiegelman, S. (1974). Proc. Nat. Acad. Sci. U.S.A. 71, 1035-1039.

Hamlyn, P.H., Gillham, S., Smith, M. and Milstein, C. (1977). Nucl. Acdis Res. 4, 1123-1134.

Harrap, B.S. and Woods, E.F. (1964a). Biochem. J. 92, 8-18.

Harrap, B.S. and Woods, E.F. (1964b). Biochem. J. 92, 19-26.

Harrap, B.S. and Woods, E.F. (1967). Comp. Biochem. Physiol. 20, 449-460.

Holtzer, H., Mayne, R., Weintraub, H. and Campbell, G. (1973). In "The Biochemistry of Gene Expression in Higher Organisms" (Pollak, J.K. and Lee, J.W. eds.) pp.287-304. Australia and New Zealand Book Co., Sydney.

Holtzer, H., Sanger, J.W., Ishikawa, H. and Strahs, K. (1972). Cold Spring Harbour Symp. Quant. Biol. 37, 549-566.

Hoober, J.K. and Cohen, S. (1967). Biochim. Biophys. Acta 138, 347-356.

Hood, L., Campbell, J.H. and Elgin, S.C.R. (1975). Ann. Rev. Genet. 9, 305-353.

Hood, L., Loh, E., Hubert, J., Barstad, P., Eaton, B., Early, P., Fuhrman, J., Johnson, N., Kronenberg, M. and Schilling, J. (1976). Cold Spring Harbour Symp. Quant. Biol. 41, 817-836.

Hourcade, D., Dressler, D. and Wolfson, J. (1973a). Proc. Nat. Acad. Sci. U.S.A. 70, 2926-2930.

Hourcade, D., Dressler, D. and Wolfson, J. (1973b). Cold Spring Harbour Symp. Quant. Biol. 38, 537-549.

Hozumi, N. and Tonegawa, S. (1976). Proc. Nat. Acad. Sci. U.S.A. 73, 3628-3632.

Jacq, C., Miller, J.R. and Brownlee, G.G. (1977). Cell 12, 109-120.

Jeffereys, A.J. and Flavell, R.A. (1977). Cell 12, 1097-1108.

Karp, R. and Hogness, D.S. (1976). Fed. Proc. 35, 1623.

Kedes, L.H. and Birnstiel, M. (1971). Nature New Biol. 230, 165-169.

Kedes, L.H., Cohn, R., Lowry, J., Chang, A.C.Y. and Cohen, S.N. (1975). Cell 6, 359-370.

Kemp, D.J. (1975). Nature 254, 573-577.

Kemp, D.J. and Rogers, G.E. (1972). Biochemistry 11, 969-975.

Kemp, D.J., Dyer, P.Y. and Rogers, G.E. (1974a). J. Cell Biol. 64, 1114-1131.

Kemp, D.J., Partington, G.A. and Rogers, G.E. (1974b). Biochem. Biophys. Res. Commun. 60, 1006-1014.

Kemp, D.J., Schwinghamer, M.W. and Rogers, G.E. (1974c). Mol. Biol. Reports 1, 441-446.

Kemp, D.J., Walker, I.D., Partington, G.A. and Rogers, G.E. (1975). In "The Eukaryote Chromosome" Ed. W.J. Peacock and R.D. Brock, A.N.U. Press, Canberra. pp.205-216.

Kersten, W., Kersten, H. and Szybalski, W. (1966). Biochem. 5, 236-244.

Kinniburgh, A.J., Mertz, J.E. and Ross, J. (1978). Cell 14, 681-693.

Kischer, C.W. and Furlong, N.B. (1967). Proc. Soc. Exptl. Biol. Med. 124, 1188-1190.

Kitano, Y. and Kuroda, Y. (1967). Exptl. Cell Res. 48, 350-360.

Kitos, P.A., Saxon, G. and Amos, H. (1972). Biochem. Biophys. Res. Commun. 47, 1426-1437.

Klessig, D.F. (1977). Cell 12, 9-21.

Kohne, D.E. (1970). Quart. Revs. Biophys. 3, 327-375.

Lai, E.C., Woo, S.L.C., Dugaiczyk, A., Catterall, J.F. and O'Malley, B.W. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 2205-2209.

Laird, C.D. (1971). Chromosoma 32, 378-406.

Lavi, S. and Groner, Y. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 5323-5327.

Levy, S., Childs, G. and Kedes, L. (1978). Cell 15, 151-162.

Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977). Cold Spring Harbour Symp. Quant. Biol. 42, 1047-1052.

Lillie, F.R. (1965). "Lillie's Development of the Chick" (Hamilton, H.L., ed.) pp.546-606. Holt, Rinehart and Winston, New York.

Lizardi, P.M. and Brown, D.D. (1975). Cell 4, 207-215.

Lockett, T.J. and Kemp, D.J. (1975). Proc. Aust. Biochem. Soc. 8, 107.

Mandel, J.L., Breathnach, R., Gerlinger, P., Le Meur, M., Gannon, F. and Chambon, P. (1978). Cell 14, 641-653.

Maniatis, T., Kee, S.G., Efstratiadis, A., Kafatos, F.C. (1976). Cell 8, 163-182.

Marotta, C.A., Wilson, J.T., Forget, B.G. and Weissman, S.M. (1977). J. Biol. Chem., 252, 5040-5053.

Marushige, K., Brutlag, D. and Bonner, J. (1968). Biochemistry 7, 3149-3155.

Matthyssens, G. and Tonegawa, S. (1978). Nature 273, 763-765.

Matulionis, D.H. (1970). Z. Anat. Entwickl. Gesch. 132, 107-157.

Maxam, A.M. and Gilbert, W. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 560-564.

McClements, W. and Skalka, A.M. (1977). Science 196, 195-197.

McConaughy, B.L., Laird, C.D. and McCarthy, B.J. (1969). Biochemistry 8, 3289-3295.

McDonell, M.W., Simon, M.M. and Studier, F.W. (1977). J. Mol. Biol. 110, 119-146.

McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978). Nature 273, 723-728.

Milstein, C., Brownlee, G.G., Cartwright, E.M., Jarvis, J.M. and Proudfoot, N.J. (1974). Nature 252, 354-359.

Modak, M.J., Marcus, S.L. and Cavalieri, L.F. (1973). Biochem. Biophys. Res. Commun. 55, 1-7.

Morris, C.P. and Rogers, G.E. (1979). Mol. Biol. Reports (in press).

O'Donnell, I.J. (1973a). Aust. J. Biol. Sci. 26, 401-413.

O'Donnell, I.J. (1973b). Aust. J. Biol. Sci. 26, 415-437.

O'Donnell, I.J. and Inglis, A.S. (1974). Aust. J. Biol. Sci. 27, 369-382.

Partington, G.A., Kemp, D.J. and Rogers, G.E. (1973). Nature New Biol. 246, 33-36.

Pellegrini, M., Manning, J. and Davidson, N. (1977). Cell 10, 213-224.

Perry, R.P. (1976). Ann. Rev. Biochem. 45, 605-629.

Perry, R.P., Bard, E., Hames, B.D., Kelley, D.E. and Schibler, U. (1976). Progress in Nucl. Acids Res. and Mol. Biol. 19, 275-292.

Perry, R.P., Kelley, D.E., Friderici, K. and Rottman, F. (1975). Cell 4, 387-394.

Pinder, J.C., Staynov, D.Z. and Gratzer, W.B. (1974). Biochemistry 13, 5373-5378.

Portmann, R., Schaffner, W. and Birnstiel, M. (1976). Nature 264, 31-34.

Prensky, W. (1976). Methods in Cell Biology 13, 121-152.

Proudfoot, N.J. (1976a). J. Mol. Biol. 107, 491-525.

Proudfoot, N.J. (1976b). Nucl. Acids, Res. 3, 1811-1821.

Proudfoot, N.J. (1977). Cell 10, 559-570.

Proudfoot, N.J. and Brownlee, G.G. (1974). Nature 252, 359-362.

Proudfoot, N.J. and Brownlee, G.G. (1976). Nature 263, 211-214.

Proudfoot, N.J. and Longley, J.I. (1976). Cell 9, 733-746.

Proudfoot, N.J., Gillam, S., Smith, M. and Longley, J.I. (1977). Cell 11, 807-818.

Rabbitts, T.H. (1974). FEBS Letters 42, 323-326.

Rabbitts, T.H. and Forster, A. (1978). Cell 13, 319-327.

Rawles, M.E. (1963). J. Embryol. Exptl. Morphol. 11, 765-789.

Rawles, M.E. (1965). In "Biology of Skin and Hair Growth". (Lyne, A.G. and Short, B.F. eds.) pp.105-128. Angus and Robertson, Sydney.

Regier, J.C., Kafatos, F.C., Kramer, K.J., Heinrikson, R.L. and Keim, P.S. (1978a). J. Biol. Chem. 253, 1305-1314.

Regier, J.C., Kafatos, F.C., Goodfliesh, R. and Hood, L. (1978b). Proc. Nat. Acad. Sci. U.S.A. 75, 390-394.

Rochaix, J.D., Bird, A. and Bakken, A. (1974). J. Mol. Biol. 87, 473-487.

Romanoff, A.L. (1960). "The Avian Embryo". MacMillan, New York.

Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M. and O'Malley, B.W. (1978). Cell 15, 671-685.

Rothberg, S. and Ekel, T.M. (1971). Nature 229, 341-342.

Sanger, F. and Coulson, A.R. (1975). J. Mol. Biol. 94, 441-448.

Schachat, F.H. and Hogness, D.S. (1973). Cold Spring Harbour Symp. Quant. Biol. 38, 371-381.

Schaffner, W., Gross, K., Telford, J. and Birnstiel, M. (1976). Cell 8, 471-478.

Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978). Cell 14, 655-671.

Schultz, G., Beato, M. and Feigelson, P. (1972). Biochem. Biophys. Res. Commun. 49, 680-689.

Seidman, J.G., Leder, A., Edgell, M.H., Polsky, F.I., Tilghman, S., Tiemeier, D.C. and Leder, P. (1978a). Proc. Nat. Acad. Sci. U.S.A. 75, 3881-3885.

Seidman, J.G., Leder, A., Nau, M., Norman, B. and Leder, P. (1978b). Science 202, 11-17.

Sengel, P. (1957). Experimentia 13, 177-182.

Sengel, P. (1971). Advances in Morphogenesis 9, 181-230.

Shain, W.G., Hilfer, S.R. and Fonte, V.G. (1972). Dev. Biol. 28, 202-218.

Shine, J. and Dalgarno, L. (1975). Nature 254, 34-38.

Sinclair, J.H. and Brown, D.D. (1971). Biochemistry 10, 2761-2769.

Smith, G. (1973). Cold Spring Harbour Symp. Quant. Biol. 38, 507-513.

Southern, E.M. (1971). Nature New Biol. 232, 82-83.

Southern, E.M. (1975a). J. Mol. Biol. 94, 51-69.

Southern, E.M. (1975b). J. Mol. Biol. 98, 503-517.

Sugimoto, M. and Endo, H. (1969). Nature 222, 1270-1272.

Sugimoto, M., Tajima, K., Kojima, A. and Endo, H. (1974). Dev. Biol. 39, 295-307.

Sures, I., Maxam, A., Cohn, R.H. and Kedes, L.H. (1976). Cell 9, 495-502.

Symons, R.H. (1975). Mol. Biol. Reports 2, 277-285.

Szybalski, W. (1968). In "Methods in Enzymology". (Grossman, L. and Moldave, K., eds.) Vol. 12B pp.330-360. Academic Press, New York.

Tartof, K.D. and Dawid, I.B. (1976). Nature 263, 27-30.

Taylor, J.M., Illmensee, R. and Summers, J. (1976). Biochim. Biophys. Acta 442, 324-330.

Tiemeier, D.C., Tilghman, S.M., Polsky, F.I., Seidman, J.G., Leder, A., Edgell, M.H. and Leder, P. (1978). Cell 14, 237-245.

Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J.V. and Leder, P. (1978a). Proc. Nat. Acad. Sci. U.S.A. 75, 725-729.

Tilghman, S.M., Curtis, P.J., Tiemeier, D.C., Leder, P. and Weissmann, C. (1978b). Proc. Nat. Acad. Sci. U.S.A. 75, 1309-1313.

Tomizawa, J. and Anraku, N. (1965). J. Mol. Biol. 11, 509-527.

Tonegawa, S., Hozumi, N., Matthyssens, G. and Schuller, R. (1976). Cold Spring Harbour, Symp. Quant. Biol. 41, 877-889.

Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 1485-1489.

Vogelstein, B. and Gillespie, D. (1977). Biochem. Biophys. Res. Commun. 75, 1127-1132.

Vogt, V.M. (1973). Eur. J. Biochem. 33, 192-200.

Voitkevich, A.A. (1966). "The Feathers and Plumage of Birds" Sidgwick and Jackson, London.

Walker, I.D. and Bridgen, J. (1976). Eur. J. Biochem. 67, 283-293.

Walker, I.D. and Rogers, G.E. (1976a). Eur. J. Biochem. 69, 329-339.

Walker, I.D. and Rogers, G.E. (1976b). Eur. J. Biochem. 69, 341-350.

Wallace, H. and Birnstiel, M.L. (1966). Biochim. Biophys. Acta 114, 296-310.

Watterson, R.L. (1942). Physiol. Zool. 15, 234-259.

Wei, C.M. and Moss, B. (1975). Proc. Nat. Acad. Sci. U.S.A. 72, 318-322.

Weinberg, E.S., Birnstiel, M.L., Purdom, I.F. and Williamson, R. (1972). Nature 240, 225-228.

Weinberg, E.S., Overton, G.C., Shutt, R.H. and Reeder, R.H. (1975). Proc. Nat. Acad. Sci. U.S.A. 72, 4815-4819.

Weinstock, R., Sweet, R., Weiss, M., Cedar, H. and Axel, R. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 1299-1303.

Wellauer, P.K. and Dawid, I.B. (1977). Cell 10, 193-212.

Wellauer, P.K., Dawid, I.B., Brown, D.D. and Reeder, R.H. (1976a). J. Mol. Biol. 105, 461-486.

Wellauer, P.K., Reeder, R.M., Dawid, I.B. and Brown, D.D. (1976b). J. Mol. Biol. 105, 487-505.

Wensink, P.C., Finnegan, D.J., Donelson, J.E. and Hogness, D.S. (1974). Cell 3, 315-325.

Wessells, N.K. (1961). Exptl. Cell Res. 24, 131-142.

Wessells, N.K. (1962). Dev. Biol. 4, 87-107.

Wessells, N.K. (1965). Dev. Biol. 12, 131-153.

White, R.L. and Hogness, D.S. (1977). Cell 10, 177-192.

Wolf, H. (1975). Analytical Biochem. 68, 505-511.

Woodin, A.M. (1954). Biochem. J. 57, 99-109.

Woods, E.F. (1971). Comp. Biochem. Physiol. 39A, 325-331.

Wu, M., Holmes, D.S., Davidson, N., Cohn, R.H. and Kedes, L.H. (1976). Cell 9, 163-169.

APPENDIX - PUBLICATIONS

PAPERS PRESENTED AT MEETINGS

The organization of keratin genes in the chick genome. (1975) Proc. Aust. Biochem. Soc., 8, 107.

Unique and reiterated sequences in feather keratin mRNA. Proc. Aust. Biochem. Soc., 11, 78.