CLONING OF NON-GLOBIN MESSENGER RNAs FROM MOUSE RETICULOCYTES

by

COLIN MAURICE CASIMIR

Presented for the Degree of Ph.D.

University of Glasgow,

May 1981.

ProQuest Number: 13818939

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 13818939

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

GLASGOW
UNIVERSITY
LIBRARY

11776 (copy 1)

CONTENTS

		Page No.
Summary		i
Acknowledgements		i.b.c.
List of Abbreviations		iv
List of Figures		vi
List of Tables		viii
INTRODUCTION:	CONTROL OF GENE EXPRESSION DURING DIFFERENTIATION AND DEVELOPMENT.	1
I.	MODULATION OF MESSENGER RNA ABUNDANCE DURING DIFFERENTIATION AND DEVELOPMENT.	2
I.A.	Analysis of messenger RNA populations.	2
I.A.1.	Comparison of differentiated cell types and tissues.	2
I.A.2.	m-RNA population changes during development.	7
I.A.3.	Hormonally controlled m-RNA population and *in vitro* differentiation systems.	13
I.A.4.	Analyses of nuclear RNA.	23
I.B.	Analysis of gene expression during differentiation.	34
I.B.1.	Globin gene expression during differentiation.	35
I.B.2.	Specific gene expression - hormonally controlled tissues.	45
I.B.3.	Specific gene expression during development.	51

Page No.

II.	DNA SEQUENCES MEDIATING IN GENE EXPRESSION.	61
II.A.	Role of repeated sequences in eukaryotic genomes.	61
II.B.	Regulatory genes.	68
II.B.1.	Evidence from classical genetics.	68
II.B.2.	Molecular analyses of thallassaemias.	78
II.C.	Detailed sequence analysis.	81
II.C.1.	Introns.	81
II.C.2.	Pseudogenes.	85
II.C.3.	Gene families.	86
II.C.4.	Regions displaying evolutionary conservation.	88
II.D.	Functional analysis of evolutionary conserved regions.	92
III.	CONTROL OF TRANSCRIPTION.	97
III.A.	Pol III genes.	97
III.B.	Role of chromosome structure.	101
III.B.1.	Classical lines of evidence.	101
III.B.2.	Analysis by nuclease digestion.	103
IV.	POST TRANSCRIPTIONAL CONTROL.	115
IV.A.	Translational control.	115
IV.B.	Messenger stability - role of poly A.	117
IV.C.	Role of splicing.	121

MATERIALS AND METHODS. 127

RESULTS. 143

I.	ANALYSIS OF RETICULOCYTE M-RNA AND PROTEIN SYNTHESIS.	144

Page No.

I.A.	Gradient fractionation of reticulocyte m-RNA.	144
I.B.	Hybridisation analysis of fractionated RNA.	147
I.C.	Analysis of reticulocyte protein labelled *in vitro*.	150
I.D.	Manipulation of the anaemic response.	153
II.	CONSTRUCTION OF CLONED PROBES.	156
II.A.	Preparation of cDNA for cloning.	156
II.B.	Construction of plasmid vectors.	158
II.B.1.	Preliminary characterisation of plasmid DNA.	158
II.B.2.	Experiments employing tailed plasmid.	159
II.B.3.	Experiments employing blunt end ligation.	162
II.C.	Optimisation of the technique of blunt-end ligation.	165
II.D.	Screening protocols and development of competition hybridisation assay.	168
III.	CHARACTERISATION OF NON-GLOBIN CLONES.	171
III.A.	Use of purified globin probes.	171
III.B.	Restriction analysis of non-globin recombinants.	173
III.B.2.	Tetranucleotide mapping of clone pNG 3.	177
III.C.	Southern blot analysis of non-globin recombinant clones.	182
III.D.	Quantitation of non-globin sequences.	187
III.D.1.	Quantitative estimation of relative abundance in reticulocyte m-RNA of sequences complementary to cloned cDNAs.	187
III.D.2.	Analysis of tissue specificity of non-globin m-RNAs from reticulocytes.	188

Page No.

DISCUSSION

I.A.1.	Gradient fractionation of RNA.	193
I.A.2.	Non globin RNA and protein synthesis.	200
I.A.3.	Cloning methodology and analysis of clones.	203
I.A.4.	Restriction data.	205
I.A.5.	Quantitative estimates of abundance and tissue specificity.	207

REFERENCES 218

Summary

The development of red blood cells in the mouse is a useful model system for studying gene expression during differentiation. Erythropoiesis entails the sequential expression of a coordinated set of gene products, culminating in the formation of a highly specialised cell, the erythrocyte. Many of these gene products can be used as recognisable markers of differentiation, for example, spectrin, glycophorin and the globins themselves.

To date, however, it has really been possible to analyse only globin gene expression in terms of its molecular mechanisms. This has been due to the fact that reticulocytes synthesise predominantly globins and therefore the messenger RNA population is similarly highly enriched in globin-specific sequences. In order to acquire a greater understanding of the genetic control mechanisms at work during erythroid differentiation, it would be highly advantageous therefore to be able to study concomitantly the regulation of other erythroid-specific gene sequences.

To this end genetic engineering techniques have been employed to construct a complementary DNA (cDNA) "library" of recombinant plasmids, using the total reticulocyte mRNA population as template. It was hoped therefore to select out clones from the library which contained "non-globin" sequences, with the aim of using these as molecular hybridisation probes to further investigate the process of differentiation.

As a preliminary study, the feasibility of using density-gradient centrifugation to provide a convenient method of enriching the mRNA preparations for "non-globin" sequences was investigated. By use of appropriate size markers, suitable centrifugation conditions were obtained; however, when messenger RNA processed in this fashion was analysed by molecular hybridisation techniques the "non-globin" sequence content remained only some 10% of the total messenger population. A closer analysis, however, indicated that mRNA prepared by standard methodologies was significantly lower than expected in "non-globin" sequences.

Such a conclusion was confirmed in a subsequent study of "non-globin" protein synthesis in reticulocytes during induced anaemia. Proteins were isolated from cells labelled *in vitro*, fractionated on Sephadex G100, and the "non-globin" fractions examined by electrophoresis on SDS/polyacrylamide gels. In addition, this analysis demonstrated that in immature reticulocytes

obtained earlier in the phase of anaemia, "non-globins" represented a significantly greater proportion of total protein synthesis than in their more mature counterparts (from which messenger RNA was usually prepared).

In order to utilise immature reticulocytes it was then necessary to boost yields of these cells from the low level (17%) obtained at this stage in the anaemia. By manipulation of the dosage of phenyl hydrazine administered to induce anaemia it was found possible to obtain blood containing a very high ($\geqslant$98%) proportion of circulating reticulocytes. Messenger RNA prepared from blood of this type was shown, by *in vitro* translation, to direct greater synthesis of "non-globin" proteins.

This mRNA was used therefore as a template for synthesis of cDNA and the single-stranded product converted to a double-stranded form by incubation with DNA polymerase. The double-stranded cDNA was then treated with S-1 nuclease to destroy the hairpin loop which remains joining the two strands together.

Attempts were made to insert this double-stranded cDNA into linearised pAT153 plasmid DNA by the complementary homopolymer extension technique. Unfortunately, this approach was not successful in producing recombinant bacteria and was thus rejected in favour of the technique of "blunt-end ligation". This methodology involves the covalent linking of flush-ended plasmid DNA and cDNA prior to bacterial transformation. Recombinants resulting from the use of this procedure were screened by filter-bound colony hybridisation (Grunstein/Hogness technique), to select out clones containing reticulocyte cDNA sequences.

From 100 clones so identified, ten putative "non-globin" recombinants were selected by a competition hybridisation assay. For this procedure duplicate nitrocellulose filters bearing DNA from colonies lysed *in situ* were challenged with either an unfractionated reticulocyte cDNA probe, or with a similar probe which had prev iously been hybridised, in solution, to a large excess of plasmid DNA containing and globin cDNA sequences. Colonies which showed evidence for competition, by reduction of hybridisation signal, were considered globin recombinants and were eliminated. Colonies unaffected by competition were considered putative "non-globin" recombinants and subjected to further study. These colonies were rescreened by the competition assay and additionally by a filter hybridisation using α and β globin cDNA sequences purified from a total reticulocyte cDNA.

The putative "non-globin" clones were then characterised by restriction

analysis to ascertain the length of the inserted cDNA sequence and confirm its insertion into the Bam H1 restriction site of the plasmid vector.

The restriction analysis was then extended by hybridisation to restricted recombinant DNAs, separated by agarose gel electrophoresis and transferred to a nitrocellulose filter. Such a filter was hybridised sequentially with a reticulocyte cDNA probe, an α and β globin cDNA probe and a probe prepared from α and β globin cDNA recombinant plasmids. Cloned cDNAs which showed evidence for positive hybridisation to globin probes were again eliminated.

By use of recombinant plasmid DNAs bound to individual filters and liquid scintillation counting, a similar methodology was employed to quantitate the levels, in reticulocyte mRNA, of messenger species complementary to the cloned cDNAs.

From the data acquired above, four clones were selected for a wider study. Utilising the quantitative hybridisation procedure described above, the representation of messenger RNAs homologous to the cloned sequences was appraised in a number of different tissues. Of the recombinants analysed in this way, two were found to specify sequences represented only in erythroid tissues; one in fact was observed solely in reticulocytes. The two remaining clones, by contrast, displayed a much wider distribution, showing significant levels of expression in adult liver and in fibroblasts as well as in Friend erythroleukaemia cells. In this last cell type a moderate rise in abundance of the two cloned sequences was observed when the cells were induced to differentiate in culture; they nevertheless displayed the greatest level of expression in reticulocytes.

Possible candidates for the identity of the messenger RNA sequences represented by the cloned cDNAs and the potential use of these clones to further investigate the genetic control of erythropoiesis are discussed.

Abbreviations

DNA	Deoxyribonucleic Acid.
RNA	Ribonucleic Acid.
dNTP	Deoxynucleoside triphosphate.
dCTP	Deoxycytidine triphosphate.
dGTP	Deoxyguanosine triphosphate.
A	Deoxyadenosine monophosphate in DNA.
C	Deoxycytidine monophosphate in DNA.
G	Deoxyguanosine monophosphate in DNA.
T	Deoxythymidine monophosphate in DNA.
DMSO	Dimethyl sulphoxide.
SDS	Sodium Dodecyl Sulphate.
EDTA	Ethylene diamine tetra acetic acid.
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethane sulphonic acid
Tris	Tris(hydroxymethyl)aminomethane.
Pol	DNA Polymerase.
BSS	Balanced Salts Solution.
P.B.S.	Phosphate Buffered Saline.
BSA	Bovine Serum Albumen.
^{32}P	Radioactive isotope of Phosphorous.
^{125}I	Radioactive isotope of Iodine.
^{35}S	Radioactive isotope of Sulphur.
^{3}H	Tritium.
nt	Nucleotides.
bp	Nucleotide base pairs.
TCA	Trichloracetic Acid.
TdT	Terminal Transferase.

cDNA	DNA copy of m-RNA(s) synthesised with reverse transcriptase.
sscDNA	Single stranded cDNA.
ds.cDNA	Double stranded cDNA.
m-RNA	Messenger RNA.
r-RNA	Ribosomal RNA.
t-RNA	Transfer RNA.

List of Figures

Figure 1. Polyacrylamide Gel Analysis of Mouse Reticulocyte m-RNA.

Figure 2. Density Gradient Fractionation of Reticulocyte m-RNA I.

Figure 3. Density Gradient Fractionation of Reticulocyte m-RNA II.

Figure 4. Density Gradient Fractionation of Reticulocyte m-RNA III.

Figure 5. Titration of "Non-Globin" m-RNA with Reticulocyte cDNA.

Figure 6. Size Analysis of "Non-Globin" cDNA.

Figure 7. Kinetic Analysis of "Non-Globin" m-RNA.

Figure 8. Globin m-RNA Sequence Content of "Non-Globin" cDNA.

Figure 9. Production of Circulating Reticulocytes During Induced Anaemia.

Figure 10. Fractionation of Acid-Acetone Precipitated Proteins.

Figure 11. SDS/Polyacrylamide Gel Electrophoresis of Fractionated Non-Globin Proteins.

Figure 12. Photomicrograph of Brilliant Cresyl Blue Stained Blood Smear.

Figure 13. Cell-Free Translations of Reticulocyte m-RNA.

Figure 14. Titration of S-1 Nuclease.

Figure 15. Analysis of the Products of d.s. c-DNA and S-1 Cleavage.

Figure 16. Preliminary Characterisation of Plasmid pAT 153.

Figure 17. Linearisation of Plasmid DNA.

Figure 18. Schematic Diagrams of Cloning Methodologies Adopted.

Figure 19. Time Course of Homopolymer Extension using Terminal Transferase.

Figure 20. Grunstein/Hogness Screening of cDNA Recombinants.

Figure 21. Competition Hybridisation Assay.

Figure 22. Competition Hybridisation of Putative "Non-Globin" Recombinants.

Figure 23. Hybridisation of "Non-Globin" Recombinants with Purified α and β globin cDNA Probes.

Figure 24. Restriction Digest of "Non-Globin" Recombinants using Hine II/Eco RI double digest.

Figure 25. Restriction Digest using Hin F/RI double digest.

Figure 26. Restriction Digest of Recombinants with Hae II.

Figure 27. Restriction Digest of Recombinants with Taq 1.

Figure 28. Dpn 1 Digest of Clone pNG 7.

Figure 29. Tetranucleotide Cutting Enzyme Analysis of Clone pNG 3.

Figure 30. Southern Blot Analysis of "Non-Globin" Recombinants.

Figure 31. Southern Blot Analysis (Contd.).

Figure 32. Tissue Specificity of "Non-Globin" m-RNAs from Mouse Reticulocytes. I.

Figure 33. Tissue Specificity of "Non-Globin" m-RNAs from Mouse Reticulocytes. II.

List of Tables

Table 1. Analysis of Non-Globin m-RNA Complexity.

Table 2. Analysis of Induced Anaemia.

Table 3. Comparisons of Ligase Efficiency.

Table 4. Comparison of Ligation Efficiency of Repaired and Phosphatased Plasmids.

Table 5. Ability of Exogenous Hae III Plasmid Fragments to Restore Ligation to Phosphatased Plasmid.

Table 6. Summary Table of Restriction Analysis of "Non-Globin" Recombinants.

Table 7. Distribution of Restriction Sites in Insert of Clone pNG 3.

Table 8. Quantitative Estimates of "Non-Globin" m-RNA Abundance in Reticulocytes.

Table 9. Quantitative Estimates of "Non-Globin" m-RNA Sequence Representation in Erythroid and Non-Erythroid Tissues.

INTRODUCTION

CONTROL OF GENE EXPRESSION DURING DIFFERENTIATION AND DEVELOPMENT

I. MODULATION OF MESSENGER RNA ABUNDANCE DURING DIFFERENTIATION AND DEVELOPMENT

A. Analysis of m-RNA populations.

1. Comparisons of differentiated cell types and tissues.

A simplistic approach to the analysis of differentiation would be to take the products of such events, two different tissues for example and test them for similarities and differences. It should be hoped therefore to observe different sets of genes or their products which are specific for a given tissue type. The possibility then exists that some common feature relating to expression of such gene products may be elucidated.

This type of rationale leads to intensive study of m-RNA populations by the technique of nucleic acid hybridisation. Such studies have been performed in a number of different systems, by comparing tissues as above (1),(2),(3) by studying development in simpler eukaryotes (4),(5) by manipulating cells in culture to mimic differentiation *in vitro*. (6),(7).

To begin comparisons, the structure of messenger populations must first be studied (1),(8). This was made possible by the advent of nucleic acid hybridisations performed in solution. Under conditions of large RNA excess the rate of its reassociation with a labelled DNA tracer has been shown to obey simple kinetic rules (9),(5). In particular, it has been shown that the rate at which a sequence hybridises is inversely proportional to its complexity,

(expressed in units of molecular weight) that is, the greater the total sequence diversity the slower the reaction progresses. More complex analyses of solution reactions have been possible because (to a first approximation at least) the cDNA tracer, copied from the m-RNA template, by reverse transcriptase maintains the same correct relative proportions of the different sequences within the m-RNA. The fraction of cDNA involved in a given transition can then be considered to be driven by a similar mass fraction of the RNA, hence the true kinetic complexity can be calculated.

Analysis of messenger populations by such techniques revealed an interesting and perhaps unexpected pattern. (1), (2),(3),(10),(11). A typical tissue or cell line of mammalian origin has been shown to contain somewhere between 11,000 and 15,000 different m-RNA sequences. However the bulk of the mRNA has been found to be composed of only a handful of sequences, the most abundant sequences being represented in many thousands of copies per cell. The majority of the sequence complexity is contained therefore in merely 10-30% of the total m-RNA mass. Moreover, such sequences will be present in as low as 5 or typically 20 copies per cell. (12),(13). Frequently a third abundance class of middle abundant sequences is identified (1),(3) present typically in one to a few hundred copies per cell. Hastie and Bishop (1) were particular to demonstrate that such a class of messenger does exist, but it should perhaps

be stated that the division into abundance classes is to some degree artificial (as may be expected). Some tissues, oviduct (3),(12) or pancreas (13) show very obvious evidence of abundance classes. Others, liver for example (3) show a pattern where the kinetic classes are much more difficult to identify. Computer fits of the data make it likely that such classes do exist in virtually all cases. It may be safe to say that within an abundance class the abundance of its individual members will vary less than abundances between classes.

Because of the complexity distribution, observed kinetic measurements can be underestimates - large complexities in small fractions of the RNA may be missed and the technique is most insensitive to changes of complexity in the lowest abundance class. Thus measurements where total sequence complexity is of prime interest have often been performed with single copy or unique DNA probes. Such probes can be obtained by preparative hybridisation, a technique which can also be used to obtain abundance class specific cDNA fractions. (1),(14),(10). A single copy DNA may be purified by reannealing total DNA to a low C_ot value (C_ot is conc. x time) at which only repetitive sequences will have hybridised (similarly for abundant sequences from m-RNA). The total mixture is then fractionated on the basis of differential binding affinities of single stranded and duplexed molecules, when bound to hydroxyapatite (HAP). Complexity estimates based on saturation values obtained

with uDNA probes tend to err in the opposite direction from kinetic estimates, as the extent of hybridisation is low (3-4% typically) and repeated sequence contamination will therefore amplify the observed complexity. For this reason many groups have adopted both approaches (11) (15),(16)(17).

Tissue comparisons of mammalian and avian species (1),(2),(3) have revealed a pattern which was not presupposed. As already stated the sequence complexity represented on the polysomes of a differentiated cell type may be a few percent of the total genomic complexity. It might therefore be expected that tissues express quantitatively quite different sets of sequences on their polysomes. The pattern observed however, has been one of extensive sequence overlap between different tissues (1),(2),(3) and at different developmental stages (4),(18),(19). In mammalian tissues Hastie and Bishop demonstrated, by preparing fractionated probes, that abundant sequences of liver and kidney were not absent from the heterologous tissue mRNA or from brain. However, quantitative variations could be quite dramatic. Abundant class cDNA from liver for example reacted 1000 times slower with kidney mRNA than with liver m-RNA. Abundant kidney m-RNA sequences were also found at only moderate abundance in other cell types. The converse situation however did not obtain. That is, rare sequences were found to be rare in all three tissue types and similarly middle abundant sequences were

at similar abundance (on average) in heterologous reactions. Some degree of qualitative variation has been demonstrated particularly in the abundant classes. Fractionated probes from this class most often do not reach 100% plateau values at saturation. Indicating that they contain sequences unable to hybridise to heterologous m-RNAs.

A similar pattern can be observed in comparison of avian oviduct and liver (3); m-RNA corresponding to the ten major egg white proteins constituting some 70% of oviduct m-RNA but also present in live m-RNA at a level representing rarest class messengers. Again the majority of less prevalent species were found in both tissues.

I.A.2. m-RNA population changes during development.

A more realistic approach to development can be obtained by attempting to analyse differentiation events as they occur *in vivo*. Such a study necessitates working with embryonic material and as such is very difficult in mammals. Some simple eukaryotes like the sea urchin however provide ideal material for this type of analysis as large numbers of embryos can be obtained at different developmental stages and the process of differentiation monitored as it progresses.

The use of this system has been brilliantly exploited by Davidson and his colleagues (4),(5),(20). The previously described complexity and abundance relationships in m-RNA have been shown to similarly apply also to organisms like the sea urchin (5) as well as to the mammalian systems already elucidated (1-3),(6),(12),(13). In an analysis of sequences expressed during embryonic stages and in adult tissues (4) these workers used an innovative approach to the problems of complexity estimation (kinetic versus saturation) to above . Their approach was to prepare a total unique DNA probe as described before, however this total DNA probe was then fractionated into sequences which were complementary to messenger RNA and sequences non complementary. A preparative hybridisation in large m-RNA excess followed by hydroxyapatite chromatography provided such probes. The m-RNA complementary DNA, referred to as m-DNA, can then be used in an analagous fashion to a cDNA probe; the crucial difference being the

equal sequence representation of all m-RNA species regardless of abundance. The preparation of such a probe for this work is vital as in the sea urchins some 99% of the total sequence complexity resides in only 10% of the m-RNA mass (4).

An m-RNA probe was then used to ascertain the messenger sequence overlap between gastrula stage embryos and adult tissues. A gastrula m-DNA was prepared as this corresponds to the first stage of the life cycle in which DNA transcription occurs. In more immature stages maternal messenger stored in the oocyte is utilised (4). This comparison with earlier stages than gastrula was also performed. The sequence complexity of gastrula m-RNA in 17×10^6 (4),(21) nucleotides which corresponds in this case to approximately 10,000 - 15,000 different sequences; the average messenger size being somewhat less than those already discussed. In total contrast, the m-RNA complexities of all the adult tissues, intestine, tube foot and coelomocyte were considerably lower representing only 1500 to 3000 different sequences. Moreover, all three adult tissues showed that 36% of the sequences expressed were also present in gastrula. By mixing experiments, it was also demonstrated that the sequences in common between gastrula and adult were identical in all three cases. A core of about 1500 sequences seem therefore to be ubiquitous in nature.

Sequences not held in common could be estimated using the fraction of the uDNA which did not hybridise to gastrula m-RNA. Total sequence complexity is then the sum of the m-DNA by hybridisation and the so-called null-DNA reaction.

It is worth noting that the rate of null DNA reaction in intestine was similar to that observed with m-DNA so both shared and specific sequences seem to be at comparable abundances.

Comparisons with other larval stages were equally illuminating. Pluteus stage larvae are more mature than the gastrulae and possess well defined skeletal and gut structures. The messenger RNA present on pluteus larvae polysomes however are identical to those found in gastrulae, although the total extent of overlap is only about 80%. The null DNA reaction failed to find any pluteus specific sequences. It would seem therefore that maturation from gastrula to pluteus is accompanied by loss of certain sequences.

This pattern of development whereby a gradual reduction of sequence complexity accompanies development was borne out by study of earlier stage embryos. Blastulae for example contain 70% of the gastrula sequences but in addition possess a further 10,000 or so sequences not found at later stages. The oocyte moreover, contained over twice the sequences complexity of the gastrula while at the same time displaying complete sequence overlap with all the gastrula sequences, including the 30% not seen in the blastula.

Despite a reduction in overall complexity the general pattern observed during sea urchin development is thus one where a fairly large body of sequences are found to be common to different developmental stages, each stage of development in addition contains sequences specific to itself.

One adult tissue showed itself to be quite different from the remainder and that was the ovary. This was found to have a total complexity somewhqt greater than the gastrula although again some 80% of the gastrula sequences were found to be shared. 20% of the total complexity was ovary specific.

Another simple eukaryote highly suitable for investigation of its differentiation programs is the slime mould **Dictyostelium discoideum** which has been extensively studied over many years (19),(22),(23),(24). The life cycle is somewhat extraordinary being normally uni-cellular,under certain conditions however, many single cells can be made to converge and aggregate to form a multicellular form referred to as a "slug". The slug migrates to a suitable area and then differentiates to form spore cells and stalk cells, the former producing the next generation. Despite this somewhat bizzare life habit, the genome composition is very similar to other eukaryotes which have been studied (19),(22), showing a genome with a typical single copy and repeated sequence pattern. In particular, the pattern of interspersion of repeated sequences between longer stretches of single copy DNA is observed (22).

Differentiation can be induced in Dictyostelium by starvation of a single amino acid. Studies of sequence changes have been performed by a number of groups (19),(23). The general trend is one of increasing sequence complexity, as development proceeds. During this process approximately 50% of the entire single copy genome is expressed (19),(23).

Differences in complexity are most marked when vegetative growing cells are compared with mature stalk and spore cells. Roughly 18% of the single copy DNA (assuming assymetric transcription) is found common to all stages of development, this represents some 5,000 sequences. (The total complexity of Dictyostelium is low). During differentiation marked qualitative changes do seem to occur as sequences representing 20% of the single copy genome appear whilst about half this number are lost as vegetative cells differentiate.

Kinetic estimates based on cDNA (23) corroborate such a pattern, although total sequence complexity estimates are a little lower. Comparison of vegetative cells and cultivation cells demonstrate a reduction of hybridisation rate of 3-4 fold if cDNA from vegetative cells is reacted with fully differentiated cell RNA. Approximately 27% of these sequences are specific to the later developmental stages.

That such a pattern is not atypical is borne out by similar findings in other simple eukaryotes (18),(25). During sporulation, Aspergillus begins to express a greater total sequence complexity, however sequences expressed in early vegetative cells also continue to be observed, as are sequences specific to spore cells. The increased complexity during conidia formation and sporulation can be observed as stage specific sequences exposed by fractionation of cDNA probes. Some 300 sequences were found to be specific to spores and another 1000 were expressed during differentiation but were

not found in the mature spore cell RNA or the vegetative cell RNA. Kinetic analysis of these stage specific sequences demonstrated that they still contained representatives of all three abundance classes.

As a fairly broad spectrum of new sequences seem to be expressed during Aspergillus development it is particularly interesting that a mutational analysis (26) revealed only two loci involved in conidia formation. Taken with the above data this would suggest that these loci represent regulatory genes controlling a fairly large number of structural genes. In support of such an interpretation one of these loci seemed to control the type of growth, switching between two distinct cell types observed in the conidiospore.

That a pattern of increasing sequence complexity during development has more general applicability and is not solely restricted to simple eukaryotes are the observations that in pancreatic development (13),(27) in the rat, which is amenable to such analysis, there appears to be a concomitant increase in sequence complexity as the organ develops and in uterine development although not strictly applicable there is a large increase in total sequence complexity (15),(28).

I.A.3. Hormonally controlled m-RNA populations and in vitro differentiation systems

Two possible approaches in mammals mimicking this type of developmental analysis of differentiation, have proven feasible. Firstly study of hormonally regulated or sensitive tissues have been investigated. Secondly some mammalian cell lines, when treated under various conditions are able to undergo changes both in morphology and gene expression which closely mimic changes occurring during in vivo differentiation. The most extensively studied of these are the Friend Cell (29),(30), a transformed erythroid precursor and the committed myoblast a precursor of muscle cells. Other examples are conversion of 3T3 fibroblasts to adipocytes (31) or transformation of teratocarcinoma cells into embryoid bodies (32). The former approach has the disadvantage of others thus far discussed, that they are tissues made up from a large number of cell types, which may generate artifacts, the latter systems suffer from the fact that most are transformed cells. In the case of the Friend Cell, the presence of the Friend virus (30) must be accounted for, moreover most of the inducers of differentiation in these systems are highly toxic in nature (DMSO, HMBA, BudR) and may in themselves cause changes unrelated to differentiation. Both techniques though do offer the possibility of observing a coordinated set of modulations of gene expression, which should go some way to providing an understanding of the differentiation process at this level.

The classic response to hormones is perhaps best exemplified by those tissues responding to steroids. The chick oviduct for example, shows dramatic dependence on oestrogen, both morphologically and with relation to the m-RNA population. The tissue does contain a number of cell types, which can complicate interpretation, but the overwhelming majority of cells are the tubular gland cells, which synthesise very large amounts of the egg-white proteins ovalbumin, conalbumin, ovomucoid and lysozyme. (3),(12),(33). Comparison of the m-RNA populations of mature oviduct with immature oviduct using cDNA probes (3),(12) has revealed striking differences in the population structure. In the adult, the bulk of the RNA is contained within a rapidly reassociating class of low complexity (3),(12). The immature oviduct lacks this class of m-RNA although exhibiting only two major transitions, not three as in the former. Such differences are almost certainly due to the hormone as these effects can be mimicked by administration of oestrogen to chicks, whereby the oviduct shows the typical response to hormone and the m-RNA population takes on the adult pattern. Subsequent withdrawal of hormone causes rapid reversion of the m-RNA population to the immature pattern (12).

In accordance with this pattern the effects of testosterone on the rat ventral prostate gland are equally dramatic. In the normal male rat, the prostate synthesises a small group of secretory proteins the messengers for which account

for some 45% of the total m-RNA. If hormone is withdrawn by castration the level of these m-RNAs drops dramatically to about 0.03% after two weeks (34),(35).

Such a response however is by no means invariant, surprisingly in fact, the rat seminal vesicle although physiologically linked with the prostate responds quite differently to testosterone. The structure of the message population is very similar to that observed for the prostate, with the majority of the mRNA devoted to a few secretory products (35), yet withdrawal of hormone as above does not cause a dramatic fall in the abundance of these sequences which reduce in abundance by only some 3 fold. There was an order of magnitude shift in the level of middle abundant sequences numbering approximately 100-150, but this is minor in comparison with the effects of oestrogen on ovalbumin which may vary up to 400 fold in concentration. As by far the most striking effect of testosterone on the seminal vesicle is on total RNA content and synthesis it was supposed that the middle abundant sequences so affected by hormone might be ribosomal protein, as has been suggested for other middle abundant sequences (4),(36). However cross hybridisation experiments indicated that these sequences were specific to the seminal vesicle, contrasting with the bulk of the complexity in the rare class which was both unaffected by hormone and found also in the ventral prostate RNA at similar abundance (34),(35).

Another tissue where a large increase in RNA synthesis occurs is the rat uterus. In addition to a rapid assembly of polysomes in this tissue when hormone is administered an extreme change in the total complexity of the m-RNA occurs (15),(28). At 4 hr of stimulation with oestrogen there are approximately 8000 sequences expressed whereas at full development some 30-40,000 different sequences are being expressed. In this instance the complexity increase is almost certainly real, as saturation estimates were used to back up kinetic measurements of total sequence complexity. Monahan et al. (12) found a large difference in complexity for oviduct before and after oestrogen stimulation but subsequent study (33) indicates that this was artifactual, although probably does represent quantitative rather than qualitative shifts in the rare class messengers. Heterologous hybridisations in the uterus demonstrated that there was extensive sequence overlap but that 21% of adult sequences were not present in the immature organ. Interestingly, although the immature (4 hr. stimulated) uterus has an abundant class of messengers these proved not to be the same as the mature abundant sequences. Although data for totally hormone naive uterine tissue are difficult to obtain, it appears that this material contains mRNA which is distinct from that of the mature uterus.

Studies of _in vitro_ differentiation using cell lines again reveal that highly abundant m-RNA species are typical of the differentiated state. In the Friend Cell (7),(29)

for example comparisons of m-RNA populations before and after induction with DMSO show that a new class of highly abundant sequences appear, this being the build up of globin m-RNAs as the cells undergo erythroid maturation (14),(29),(37). Heterologous hybridisations have also revealed that no large scale qualitative changes seem to be occurring. Virtually all sequences of uninduced and induced cells are the same, although small abundance shifts can be identified and quantified using fractionated probes (14). In fact, by these methodologies, with the exception of the globins there seems to be little evidence for expression of new sequences during induction (7),(14),(29). Clones differ to some degree in their level of globin m-RNA content in the uninduced state (29),(37),(38) but usually it may be detected at middle to low abundance.

Although there is one instance of expression of new sequences during Friend Cell induction (39), the evidence points mainly to some subtle reduction in the level of certain middle abundant sequences (21), which shift to the rare class in induced cells.

One problem in analyses of this sort is the inevitable background of uninduced cells in an induced population, which could affect detection of qualitative changes.

This is true not only of the Friend Cell where differentiation is induced chemically but also in myogenic differentiation which occurs spontaneously in culture.

This behaviour can be manipulated, for if the cells are grown in medium containing BUdR then the myoblast cultures remain undifferentiated (6). If allowed to differentiate the myoblasts fuse to form syncytia which are multinucleate; such structures are called myofibrils or myotubes. Alternatively, some transformed myoblast cell lines exist which will still differentiate on attaining confluence in culture (40),(41).

The messenger population of the dividing myoblast is typical of those observed to date, with perhaps a slightly larger total complexity than usual for a cell line, numbering around 14,000 sequences (6). As described before for Friend cell differentiation, the muscle cell cultures show the development of a new highly abundant class of messengers, comprising about 20% of the total population in roughly six sequences. There also appears to be a marked qualitative change in complexity. Heterologous hybridisations however show that virtually all the myoblast sequences are still present in myotube cultures although at lower abundance still than that observed in myoblasts where they form the rare class of messengers. This interpretation was confirmed by single copy DNA hybridisation (16). The new class of abundant sequences in myotubes was found to be under quantitative regulation as these sequences could be detected at 40 fold lower abundance, in myoblasts. However, some contribution to this level comes from differentiated cells in the myoblast cultures (6).

An apposite comparison, which relates to studies of different cell types during development, has been to compare myoblast m-RNA with that of a pluripotent teratocarcinoma cell (10). Teratocarcinoma material in many ways should provide an ideal <u>in vitro</u> model for differentiation, as from an undifferentiated cell population a wide variety of differentiated cells may be derived. This has led to the belief that teratocarcinomas are derived themselves from transformed gamete progenitors. They certainly possess some unique features, of which, apart from expressing embryonic antigens, perhaps the most intriguing is their remarkable karyotype stability, even when grown as embryonal carcinoma cells in culture. Moreover, although transformed, embryonal carcinoma cells can give rise to normal structures and tissues in adult mice, when they have been injected into host blastulae (42).

Unfortunately, in culture the developmental plasticity of embryonal carcinomas cannot be easily manipulated, although induction of endoderm has recently been reported (32). Nevertheless some differentiated cell lines have been derived from pluripotent teratocarcinomas and some myoblasts lines are amongst these. It is an interesting comparison to observe the committed myogenic precursor with its pluripotent parent. In addition one committed precursor, the myoblast was compared with another committed cell, the Friend Cell which has already been described.

In terms of total sequence complexity the myoblast was shown to be expressing a wider range of sequences than either the Friend Cell or the teratocarcinoma, thus no simple relationship exists between number of sequences expressed and developmental state. These sequence differences lie at the rare end of the abundance spectrum. All the teratocarcinoma sequences were found in the myoblast line, however some 10% of erythroid sequences were not found in the latter cell type. Similarly a comparable fraction of myoblast c-DNA could not be driven into hybrid by Friend Cell m-RNA. In contrast though the sequences specific to the muscle cell were of low abundance whereas those specific to the erythroid precursor were found in the abundant m-RNA. These abundant Friend Cell sequences although not present in the committed myoblast were found to be present in the m-RNA of the undifferentiated teratocarcinoma, but at a lower concentration.

Taking the data obtained from the types of hybridisation analysis described above, it is difficult to arrive at firm conclusions. A number of different mechanisms seem to operate during differentiation. Terminal differentiation in most systems seems to be accompanied by an increased selectivity in the messenger populations and in most higher vertebrates at least, this is accompanied by the appearance of cell or tissue specific products which correlates with a high abundance of the messengers coding for such products. These

highly abundant sequences are commonly absent from other cell types, and whilst not entirely lacking are often found at markedly lower abundances. Thus predominantly quantitative changes seem to be regulating relative phenotypes.

Marked qualitative changes, particularly during development of lower eukaryotes, can also be observed and even in the sea urchin where very small complexities may be found in adult tissues, sets of gene transcripts of a specific nature, with regard to tissue or developmental stage can still be demonstrated.

The appearance of new sequences during development can occur within high abundance m-RNAs coding for tissue specific products as mentioned previously. Nevertheless, even minor changes in the rare class sequences which have also been found, represent profound changes in the number and types of sequences: a few percent of the m-RNA mass, representing many thousands of sequences. Thus large increases in m-RNA complexity seem to accompany developmental specialization as well as restriction of expressed sequences.

Notwithstanding these observations, it is still surprising the degree to which the different m-RNA population comparisons have demonstrated extremely extensive sequence overlaps, where often the majority of m-RNA sequences expressed in two totally unrelated cell types are held is common. Undoubtedly much of this transcription is aimed at sequences whose function is equally ubiquitous, that is to say obligatory

for the general maintenance of cellular functions; the so-called "house-keeping" functions. Numbers involved in quantitating exactly how many such genes would be required, seem to vary considerably (3),(4),(10) and it seems possible that many rare class messengers may lack physiologically meaningful functions. Whatever else, it indicates that phenotypes may be radically altered by expression of a small number of genes and that this may not necessitate "on/off" mechanisms but rather modulation of relative abundance.

As relative abundance is obviously of importance, it becomes politic to ask from where the observed abundance variation derives. More specifically, are the abundance classes observed a reflection of relative transcription rates or do selective post-transcriptional events regulate levels of different messengers? This can also be extended to other sequences which are specific to developmental stage or tissue types, that is if they are not expressed on the polysomes at a particular developmental stage, are the genes for these sequences transcribed or not?

The place to look for the answers to these questions undoubtedly therefore is within the nucleus, which owing to its evolutionary significance might be expected to play a crucial role in gene expression during differentiation.

I.A.4. Analyses of nuclear RNA

To analyse these problems, studies utilising similar approaches to those outlined for messenger population study have been undertaken. Analysis of nuclear RNA can be by use of single copy DNA probes (11),(16),(20),(41) or alternatively cDNA to nuclear poly A(+) RNA can be prepared (43),(44), although in this latter case it should be remembered that the probe is only representative of the 3' portion of the HnRNA template; assuming that this is the sole priming site for cDNA synthesis, an assumption which in itself may not be wholly valid.

The first feature of note with regard to nuclear RNA populations is that they exhibit a far greater complexity than that of polysomal m-RNA. (above refs.),(21),(45). This has been found to apply to a wide variety of organisms, including the sea-urchin (20),(21), various mammalian cell lines (11),(16),(41),(46). Such a pattern is not common to all eukaryotes, however and many of the simpler forms including fungi contain nuclear RNA which is by most criteria not distinguishable in sequence content from m-RNA (47).

That nuclear RNA is of greater complexity than m-RNA, by on average roughly five to ten times (11),(16),(43),(44), has led to much speculation regarding the nature of nuclear sequences which are not present on the polysomes but presumably turn over within the nucleus. Some of this confusion has now been resolved from study of specific transcripts and

use of cloned genomic DNA which will be discussed in more detail in a subsequent section. Particularly baffling was the very large total transcriptional activity of most nuclei, where on average 25-30% of the total unique sequence complements might be found in nuclear RNA. (11),(20). Either this was a reflection of a huge number of structural gene sequences which were transcribed only for the majority to be degraded again; alternatively much of this unique sequence representation must originate from non-coding regions of the genome. This latter was suggested by the more rapid rate of sequence divergence of nuclear RNA relative to polysomal RNA (46), and has since been found to be largely true, due to the presence of intervening sequences ("introns") in these transcripts. (48),(49),(50).

Other controversies still remain however and a number of factors are not resolved. Among these are the true poly A(+) content of nuclear RNA and the complexity of the poly A(+) and non polyadenylated compartments. One factor now beyond doubt is that nuclear messenger precursors are larger than the mature forms (48),(121),(122). Moreover the size increase, due to transcription of introns, of precursors over messengers is in good agreement with the complexity increases referred to above. Steady state nuclear RNA has variously been found to be much larger than m-RNA (44),(51) or of similar size (42),(43). Some correlation between size and complexity has been observed, in the two analyses of Jaquet and colleagues (16),(41) for example. This does not

account for the change in complexity of poly A(+) nuclear RNA observed by a number of workers (52),(53) when nuclear RNA is fully denatured (by heating in DMSO for example) prior to poly A(+) selection on oligo-dT-cellulose. After denaturation poly A(+) nuclear RNA has been shown to have a complexity very similar to polysomal mRNA in contrast to the greater observed complexities outlined previously. Whether the sequences which copurify on Oligo-dT cellulose arose by fortuitous aggregation, possibly due to repeated sequence elements in the RNA (46),(54) or reflect nicking of molecules during in vivo maturation or in vitro preparation is not clear. It is certain that many studies have been performed with nuclear RNA made by techniques not adequate to prevent degradation. The simplest interpretation of all this data is certainly that most high complexity poly A(-) RNA in the nucleus derives from spliced out intron sequences. This would of course favour the latter interpretation for the denaturation results. As such this is able to account for a number of the seemingly inconsistent observations regarding the RNA size, percentage of poly A(+) and relative complexities of poly A(+) and poly A(-) fractions, which can be explained primarily by relative rates of processing relative to export and degradation of intronic sequences.

Consistent with this view is the observation that poly A(+) contiguous sequences derived from nuclear RNA (16),(41),(44) are enriched in m-RNA and conversely that polysomal cDNA and m-DNA is completely hybridised by nuclear

poly A(+) RNA (7),(16),(20),(41),(43),(44). Additionally poly A(-) nuclear RNA shows little hybridisation to messenger cDNA (44). However when an m-DNA probe was used, which is representative of the entire m-RNA sequence, a significant fraction hybridised (16). This last point indicates that some of the RNA molecules may have lost their poly A tract due to degradation or suggests that transcription may be completed before polyadenylation occurs.

What has still not been established is whether sequences which may potentially be structural coding sequences are turned over in the nucleus as a means of selective gene expression. In support of such a hypothesis are a number of findings suggesting that some poly A(+) adjacent sequences from nuclear cDNA have no counterpart present in polysomal RNA (43),(44). Some caution is necessary before concluding that this represents potential m-RNA precursors not escaping from the nucleus. Firstly oligo-A tracts are known to be present in nuclear RNA (44),(46) and these may possibly prime cDNA synthesis. Additionally self priming activity of the RNA has previously been reported (44) (46) and undoubtedly regions of secondary structure exists within the RNA, furthermore much nuclear RNA is isolated as aggregates (41),(43),(51). As RNA/RNA duplex is the in vivo primer for reverse transcriptase it seems more than likely that self priming can account for a proportion of nucleus confined cDNA. This cannot explain the findings of Herman et al. (44) who sheared their HnRNA before selection on

Oligo-dT-cellulose. However it must be possible for a proportion of nuclear cDNA to run into intronic regions. These portions of the cDNA transcript would of course be nucleus confined.

A number of investigations however do provide evidence for qualitative regulation at a post-transcriptional level, although the mechanism involved is not clear. Two studies in particular - one in sea urchin (21) and one comparing nuclear RNA of different rat tissues (45) have shown that even greater overlaps in sequence representation occur at the nuclear level than at the cytoplasm. In the sea urchin for example an m-DNA probe made from blastula stage embryos hybridised to only 12% with intestine m-RNA. However when the reaction was driven with nuclear RNA from a variety of adult tissues including intestine complete reaction of the probe was observed. That this was physiologically significant was demonstrated by the fact that the blastula m-DNA sequences were present at the same concentration as sequences destined to appear on the polysomes.

In an analagous fashion single copy DNA saturation estimates in rat nuclear RNA have shown that brain nuclear RNA sequences show complete overlap with nuclear RNA of liver and kidney. However in this case distinctive differences in the total complexity of nuclear RNA were discovered with 4-6% of the probe hybridising to thymus nuclear RNA when the brain was able to hybridise over 15% of the probe. Thus a

common core of sequences is transcribed in all three of the above tissues and may also be common to all cell types. It should perhaps be noted that both these examples are slightly deviant from the norm in that both sea urchin gastrulae and brain possess extremely large complexities and might be expected therefore to show extensive overlap with other tissue types or developmental stages. Furthermore some evidence has already been shown that the general trend of sea urchin development is towards reducing complexity (4). As has also been described the polysomal population on sea urchin embryos also demonstrate a set of sequences of a specific nature (4). Evidence that this is reflected at a nuclear level has also been presented (55). Using single copy DNA not represented in gastrula nuclear RNA a significant proportion of new sequences were shown to be transcribed in intestine nuclear RNA despite the fact that this tissue has a much smaller polysomal RNA complexity (4). As the proportion of total sequence complexity found to be specific to intestine is not large, experiments like those described by Chikaraishi et al (35) would not be expected to reveal such differences. In study of the myogenesis system however, where there is a trend, unlike the sea urchin, towards a greater complexity in developing from a pluripotent precursor to a committed stem cell (10) a proportional increase in nuclear complexity has also been recorded (16). In other similar model systems for differentiation e.g. the Friend cell,

some enrichment in sequences abundant in messenger RNA has also been observed in nuclear RNA, suggesting that this abundance may be transcriptionally controlled (7),(48). Other work on avian erythropoiesis by Tobin and his colleagues (56),(57), has demonstrated both processes at work. In terminally differentiating erythroid cells the complexities of both nuclear and cytoplasmic RNA compartments was found to be surprisingly limited. The polysomes were found to express as few as one hundred sequences (only mammalian enucleate reticulocytes may have lower complexities than this). A coincidently small fraction of single copy DNA was found to be transcribed into nuclear RNA corresponding to only 4000 sequences of average size or less than one percent of the single copy genome. However the complexity difference between nucleus and cytoplasm was inordinately large; the nucleus expressing roughly forty times the number of sequences found in cytoplasmic poly A(+)RNA. It seems an inescapable conclusion here that nuclear polyadenylated sequences which could be potential messengers are turning over within the nucleus. More evidence in support of this finding is provided by the same group (57). In these experiments cDNA prepared from liver m-RNA was hybridised to total the RNA of avian erythrocytes. A small fraction of liver cDNA was indeed found to hybridise to avian nuclear RNA (poly A(+) and poly A(-)) and was found to represent something less than 100 sequences. Interestingly, these

sequences were found also on the erythroid cell polysomes as well as on the liver cell polysomes and seemed to derive from intermediate abundant liver m-RNA.

Another factor worthy of note is that the liver cDNA was present at similar copy numbers per cell in both nuclear and cytoplasmic RNAs of erythroid origin; the rarest class of normally observed erythroid messengers is in fact approximately one order of magnitude more prevalent in the cytoplasm than in the nucleus. Additionally the level at which these liver sequences were observed was extremely low at roughly 0.08 copies per cell, contrasting with 0.5 copies per cell for the rarest erythroid HnRNA sequences. The authors interpret this data as indicative of transcriptional control, and certainly the degree of sequence overlap is extremely small compared with those observed in the previously described data (21),(45). However it should be stressed that use of a cDNA probe will almost certainly bias results against finding high complexity components of small mass fraction and the results obtainable with a single copy m-DNA might be of interest. Nevertheless, as the sequences detected are of extremely low abundance this argues against any meaningful rate of liver sequence transcription in these erythroid cells. This could indicate very rapid turnover of these sequences after transcription, but as the authors point out it more likely represents a subpopulation of the erythroid cells, possibly of less mature developmental status,

than do express these sequences also found in liver. There is certainly no real evidence that any liver specific sequences are transcribed in the nuclei of erythroid cells but not subsequently expressed at a polysomal level.

The most detailed study with relevance to this issue is a recent one from Gros and his co-workers (58),(59). This study analyses the changes occurring during myogenesis as referred to earlier (6). Again a pluripotent terato-carcinoma cell line was used to compare with a myoblast line derived from a similar parent stock and these were compared to myotube cultures formed by differentation _in vitro_ (58). Relative complexities of nuclear RNA assayed with single copy DNA probes revealed a significant increase in nuclear complexity in the committed cells over that observed in the embryonal carcinoma. It was demonstrated here as before, that sequences specific to each developmental stage were detectable using c-DNA hybridisation to poly A(+) cytoplasmic RNA. To analyse whether the changes occurring at the nuclear level were due to de novo transcription of these same sequences, probes enriched in stage specific adult sequences were prepared by preparative hybridisations. No evidence for the presence of these sequences in the nuclear RNA of the earlier stages could be provided, strongly suggesting that these stage specific sequences arise from changes in transcriptional activity. Additional evidence suggesting this interpretation was evinced from changes in

chromatin nuclease sensitivity which has been correlated with transcriptional activity (60) (See Section III.B.2) and from in vitro translation (59) of RNA homologous to the specific cDNA preparations. The proteins seen to be coded for by these messengers did indeed turn out to be muscle specific proteins including myosin, actin and troponin

One feels that this investigation probably stretches the techniques utilised to the limits and whilst meaningful inferences are possible,if a more detailed understanding is to be derived it must be obtained from differing approaches. The study of RNA population has therefore been revealing of a number of points, not least that major phenotypic changes may be mediated by a relatively small number of changes at the level of gene expression. However, the fact that small changes are those most easily lost by this technique has meant that important data needs to be obtained from analysis of the expression of specific DNA sequences. Experiments adopting this general approach will be discussed in the following section.

Before leaving the analysis of nuclear RNA and its involvement in differential gene activity, attention must be drawn to one aspect beyond dispute. Whatever conflicting evidence for or against qualitative control at a post-transcriptional level, quantitative regulation most certainly does occur. When analysed with nuclear cDNA or single copy DNA the HnRNA most often hybridises with kinetics which show evidence for only a single abundance class (7),(11),

(20),(41), corresponding to an average representation of one copy per cell. Moreover hybridisation of abundant and rare cDNA from polysomal poly A(+) mRNA hybridises to nuclear RNA at similar abundances (7),(20),(41),(43),(44). Furthermore even when heterogeneity of abundance in nuclear RNA has been identified the spread of abundances is very limited compared with cytoplasmic variation (16),(30),(38),(56). Of course, whether this quantitative modulation is brought about by selective processing or transport or from differential cytoplasmic stabilities is a debate which will have to be reappraised in a subsequent section (IV. B and C).

I.B. Analysis of specific gene products.

Until recently, the only specific hybridisation probes available, have been derived from those gene transcripts which were present in extremely large amounts by virtue of their cell specific synthesis. Thus globin m-RNAs have been extracted from reticulocytes (61),(62), ovalbumin from oviduct (63), immunoglobulin from myeloma cell lines (64) fibroin from silk worms (65). Using preparative hybridisation it had also proved possible to isolate albumen m-RNA from liver (66) and the myosin chain m-RNAs from muscle cells (67). The pancreas specific m-RNAs are also studiable as in total they constitute about 90% of the total m-RNA (27) like globin and reticulocytes (9),(62),(68).

Whilst the study of such m-RNAs enables a detail of information to be achieved which is not possible when studying populations of RNA molecules, a necessary proviso has always been that genes of this nature, that is of extremely high abundance may be regulated in some aberrant fashion and are therefore not applicable to a general situation. Whether true or not, as tissue specific or tissue characteristic gene products, these probes have still provided meaningful information with regard to gene expression during differentiation

At the present time, however, such distributions are no longer necessary. The development and growth of genetic engineering has meant that virtually any gene or its m-RNA may be obtained pure and with the added advantage that quantities are no longer limiting. One particular

advantage that has accrued is the possibility of studying expression in detail, not just of single genes, but groups or families of genes having common structural (69),(70), developmental (71),(72) or functional (71),(73) relationships. When combined with the ability to study such genes at both the m-RNA level, to isolate them from genomic libraries (74),(75) and have access to their genomic organisation. The possibility of such dissection of differentiating systems should enable great advances to be made in the near future. This is especially true with the availability of sensitive blotting protocols (76),(77) to study DNA and RNA and the rapid sequencing techniques (78),(79). Already the sheer volume of new data available is astonishing, although it perhaps needs to be stressed that as yet not a great deal more is known about the mechanisms which regulate eukaryotic gene expression; it must surely be only now a matter of time before a degree of understanding is achieved.

I.B.1. Globin gene expression during differentiation.

One of the first questions therefore that required answering was could the prevalence of these abundant messenger RNAs derive from a similar elevation of sequence at the DNA level? For globin this question was answered fairly soon by saturation experiments using cDNAs transcribed from globin m-RNA hybridised to total genomic DNA (80),(81). Such experiments demonstrated that globin genes, as had been expected from human genetic observation, were coded for

by single copy genes, or if not strictly unique then the repetition frequency was extremely limited. It was certain at least that even the genes for proteins as abundant as globin were not derived from the repetitious portion of the genome, observable in reannealing kinetics of total genomic DNA (46),(49).

The sole true exception to this rule has been the histone genes, most particularly of the sea urchin, which have been shown to display fairly high repetition frequencies suggesting that approximately 500 copies of the genes may exist. In this case at least, the high abundance of the m-RNA and protein is facilitated by the presence of a high gene copy number. How universal this phenomenon turns out to be remains to be seen; the number of histone genes necessary however, seems to be decreasing as they are isolated from more advanced organisms. The frog Xenopus for example and the fruit fly Drosophila possessing only about 20-50 and 100 genes respectively. However the primitive eukaryote yeast has very moderate numbers of histone genes (82),(83), possibly only two copies per haploid gene.

A question which can be readdressed with these probes is whether these tissue specific gene transcripts can be found in other cell types of different lineage. A definitive answer might be expected from this approach, however the observed situation is still somewhat equivocal. With regard to globin gene expression for example Humphries _et al._ observed homologous sequences present in both nuclei and cytoplasms of a variety of cell types (84), including liver, brain and fibroblast lines. In complete contrast Groudine and Weintraub (85), Spector _et al._ (86), Jaquet _et al._ (41)

have all looked for the presence of globin transcripts and in the former two cases ovalbumin transcripts, in other cell types. Without exception no evidence for transcription of these specialised products in "non-expressing" tissues could be obtained. In the case of Spector and colleagues (86), no globin or ovalbumin mRNA sequences could be detected even though low levels of endogenous "sarc" related sequences were observed in the same cells. No globin was observed in either nuclei or cytoplasm of teratocarcinoma cells (41) under conditions which would be expected to detect one copy per 50 cells. However, Groudine and Weintraub (85) did observe hybridisable globin RNA in total RNA of RSV transformed fibroblasts, but not uninfected fibroblasts. Moreover the expression of globin sequences was correlated with the presence of the transforming gene and not with virus transcription or replication per se. Interestingly only embryonic and not adult globins were observed in these embryo fibroblasts. (It remains possible that this is connected with a comparable phenomena observed in the K562 human erythroid precursor cell line which will express foetal globins if induced to differentiate with haemin. (87),(88). This would be indicative of a relationship between foetal or embryonic stage cells and adult precursor stem cells. Such a connection would be supported by the observation that many hepatomas express the foetal gene product α-fetoprotein (133).)

If one accepts the possibility (as the authors do) that the globin observed in liver and brain represented blood cell

contamination of these RNAs then one possible way to resolve these seemingly contradictory sets of data is to propose that the globin observed in fibroblasts (84) was due to the fact that only transformed fibroblasts were studied. As demonstrated by Groudine and Weintraub (85) one form of transformation was able to cause expression of globin genes albeit not of adult types. However some Friend Cell clones have been shown to be somewhat leaky to adult globin gene transcription in the uninduced state (38). The lack of evidence for globin transcripts in embryonal carcinomas, may then lie in the unique nature of these cells, which despite being able to grow in culture show an extremely stable karyotype (42), and thus may not possess this type of transcriptional derepression suggested for other transformed cells.

One approach which only redress to specific probes can give is the study of primary sequence data. The globin m-RNAs have been some of the first to be sequenced (89),(90),(91),(92),(93). However, the only real gains from this work has been an awareness of non-translated portions of mammalian m-RNAs and perhaps more importantly that these non-coding portions of the m-RNA were better conserved throughout evolution than would be expected if they did not have some function. The potential of this type of analysis was not however fully exploitable until cloned DNA sequences could be combined with the new rapid sequencing techniques (78),(79) (the results of this type of study are discussed in section IIC).

The intrinsic purity of globin probes though, enabled unequivocal evidence to be obtained regarding the existence of high molecular weight nuclear precursors to globin m-RNA (94). Without doubt it has been established that β globin m-RNA exists as a nuclear species of approximately three times the size sedimenting at roughly 15S. (48),(49),(50),(94). In addition a similar sized precursor to human γ globin has been identified (95). This species of precursor has also been seen to be polyadenylated at its 3' terminus (48),(49),(94) however Bastos and Aviv (94) whose finding was one of the original observations also described an extremely large 27S molecule which contained globin sequences but lacked poly A. Despite the more sophisticated techniques now available and the possibility of using cloned probes, no other group has been able to find a similarly sized precursor (96),(97). The reconciliation of this point is important as it has implications for both the size of the globin primary transcript and the site of polymerase binding and furthermore relates to the role of poly A and the time and stage of its addition to message precursors. Some evidence against the notion of a very large β globin transcript have come from S-1 mapping data using end labelled fragments (98). These authors have shown that the 5' terminus of both the mature messenger and the 15S precursor have identical coordinates at the DNA level. This means that either transcription is initiated close to the messenger cap site or less likely that the large species if it exists possesses a large block of

non coding sequence at the 5' end of the precursor, which is removed to generate the 15S species.

The structure of the 15S precursor to globin m-RNA has also been well established by a number of techniques (48),(49),(50). These all demonstrated that the precursor was exactly homologous, to the globin gene sequences in the DNA and thus contained the introns (48),(49),(50). It became clear therefore that generation of mature messenger RNA, with the coding sequences lying contiguously, was a process carried out by RNA/RNA splicing and not by some alteration of DNA structure during transcription, for example.

Globin probes have also been used to study the accumulation of these sequences during erythroid maturation either _in vivo_, or _in vitro_ using Friend Cells as a model system. In the latter case accumulation of globin protein has been shown to mirror an equivalent increase in the concentration of the messengers (29),(30),(62),(99-105). The relative proportions of α and β however can vary significantly during the induction process (99),(101),(106), moreover the inducer used can markedly affect the relative rates of increase (101). Final ratios do not usually show such variation, values around unity prevailing. Though in the presence of haemin induction of β globin is about 3 times greater than that of α (101). Whether this uncoupling of α and β globin m-RNAs is physiologically meaningful, though is not clear. Studies of artificially stimulated erythropoeisis, using phenyl hydrazine to induce anaemia, have produced conflicting results (99),(107),(108). Cheng

and Kazazian (107) prepared cells from spleen during erythropoiesis and found an excess of β globin m-RNA, however balanced production of α and β m-RNA has been observed under similar conditions by Phillips (108) and most recently from bone marrow by Hunt and his colleagues (99).

The level at which the control of globin expression may be mediated during this induction of m-RNA is equally complex. A number of observations pointing to a change in the rate of transcription have been made, as undoubtedly the steady state level of globin in nuclear RNA increases during Friend Cell induction (7),(14),(29),(37),(38).

Another indication of transcriptional control in the globin system has come from study of the phenomenon of globin switching in sheep. Sheep and some other related species have the ability to express a new globin chain when put under erythropoietic stress or given exogenous doses of erythropoietin. This changeover is clearly attained by de novo expression of the β^c stress globin m-RNA and moreover no prior appearance at the nuclear level can be observed (109).

However one Friend Cell β clone, at least, has shown evidence of post transcriptional control, as little change at the nuclear level was observed (38). Furthermore when the induction of globin sequences in nuclear RNA is compared with that seen at a polysomal level a clear disparity exists; the concentration of globin in the cytoplasm being much greater relative to total poly A(+) RNA than in the nucleus (7), (14), (29), (37), (46).

Obviously some post-transcriptional mechanism is contributing to the observed high abundance of globin m-RNA in cytoplasms of induced Friend Cells. Whether this is due to relative rates of processing or transport from the nucleus or cytoplasmic stabilities has not yet been determined. However studies on the half-life of globin m-RNA during spleenic (62) or bone marrow erythropoiesis (110) or during Friend Cell induction has proved revealing (111),(112). Values for globin half life acquired from radio-labelling by pulsing spleen cells in culture or injecting mice directly and collecting bone marrow or circulating reticulocytes have indicated an approximate half life for globin m-RNA of about 17 hours. A similar value has been observed in induced Friend Cells (111),(113). The remaining components of the messenger population have either a very short (2-6 hrs) half-life or a smaller fraction 10-30% were shown to be much more stable, with a half-life of about 35 hrs. This fraction is crucial as the only way possible for globin, with a shorter half-life, to reach a greater than 90% proportion of reticulocyte m-RNA is for this component to become destabilised. This is particularly crucial as in the latter stages of erythropoiesis m-RNA synthesis ceases, stability must therefore be the major parameter determining abundance. If no change occurred then globin could only rise to 20% of total poly A(+) RNA (114). If this destabilisation model is correct then it would appear to occur at a late stage of development virtually at the reticulocyte stage or in the immediately previous orthochromatic

erythroblast, as Lowenhaupt and Lingrel (111) found no change in the stability of bulk non-globin m-RNA in inducing Friend Cells, which normally do not progress completely to become reticulocytes. Moreover these authors in fact observed a destabilisation of globin m-RNA itself, this however was due to the fact that in the early stages of induction by D.M.S.O. the m-RNA decays very slowly having a very stable turnover time of greater than 50 hrs. It may be this factor, coupled with the increased transcription rate which determines the final abundance of globin m-RNA. Such very stable globin m-RNA has not been observed in vivo as yet however. Perhaps the possibility of using specific non-globin probes with Friend Cell cultures, which under appropriate conditions, it appears can be made to develop into terminally differentiated reticulocytes (115), will be able to resolve this question.

This leads us to the question of whether globin m-RNA induction is obligatory for Friend Cell differentiation. A genetic analysis of this point by Harrison and his colleagues (reviewed in 30),(117) using mutants unable to differentiate in D.M.S.O. has tended to support this idea. Most of the non-inducible clones do not produce detectable amounts of globin m-RNA or other erythroid specific products. Interestingly, though many clones were inducible by different inducers, butyric acid or H.M.B.A. and this gives weight to the idea that different inducers do in actual fact act via different

mechanisms. Some mutant clones however, have been shown able to express a differentiation marker,the red cell specific membrane protein spectrin, whilst unable to induce globin. Thus some parts of the process of Friend Cell differentiation can be uncoupled from globin synthesis. Another mutant line has revealed a total inter-dependence between haem synthesis and globin gene expression during early induction. The same is true for terminal differentiation events; as no haemoglobin can form the cells do not go through terminal differentiation. That this is linked to haem expression was demonstrated by showing that supplementing the medium with haem as well as D.M.S.O. enables the cells to differentiate fully (118),(119).

I.B.2. Specific gene expression-hormonally controlled tissues.

The general pattern established for control of globin expression does in fact seem to have been corroborated by studies using other specific gene probes. For example nuclear precursors for a number of other messengers have been identified and these are all larger than the mature m-RNA and equally have been found to contain transcripts of the intervening sequences. Furthermore some of the processing intermediates also have been identified (120-124). Other investigations of the 5' and 3' termini of these precursors, in ovalbumin for example (121) using the S-1 mapping technique (98), have also confirmed the data from experiments on globin transcripts; namely that the largest species found in nuclear RNA show conservation of the termini from the DNA level through to the mature messenger. Thus the primary transcript does not appear to extend beyond the DNA equivalent of the Cap site at the 5' end and the poly A addition site at the 3' end. Revealingly, to date, no non-poly adenylated species of nuclear precursor have been identified. The addition of poly A at the 3' terminus must be one of the earliest post-transcriptional processing events; contrary therefore to expectations derived from study of nuclear RNA populations.

Studies on a variety of specific genes, particularly those under hormonal control have demonstrated a direct correlation between hormonal stimulation, increase of protein

product and level of cytoplasmic m-RNA. Much of this work has been carried out using steroid-hormone inducible sequences, such as the avian egg white protein genes. Study of two such sequences, ovalbumin and conalbumin, using isolated nuclei and tissue culture organ explants (125), to measure rates of synthesis and decay, have attempted to elucidate whether this messenger increase is controlled at the level of transcription. Whilst it was demonstrated that both genes were undoubtedly under transcriptional control (steps were taken to try to distinguish this from very rapid turnover of the primary transcript or blocking of polymerase elongation) other mechanisms were also controlling the level of m-RNA. Surprisingly perhaps, the behaviour of the ovalbumin and conalbumin genes to the same stimuli could be quite different.

Firstly, conalbumin was shown to have a finite level of constitutive transcription, ovalbumin on the other hand appears to be transcribed extremely rarely in the absence of hormone. This difference was accentuated in the time courses of induction by secondary stimulation (after prior oestrogen treatment and withdrawal). Under these conditions ovalbumin synthesis only begins after a lag of about 2 hrs whereas extensive conalbumin synthesis can already be seen by 30 mins and rises immediately on stimulation. Progesterone will also induce these sequences on secondary stimulation (although it cannot be used for primary stimulation). Under these conditions both sequences show increased synthetic rate, however conalbumin in particular shows a relatively

marked change in rate of turnover during induction and under suboptimal doses ovalbumin m-RNA is highly unstable. Thus it appears that progesterone at least, is additionally affecting the stability of both m-RNAs.

Other interesting observations were made regarding oestrogen receptor levels in the nucleus and cytoplasm and the level of induction; the nuclear concentration being proportional to synthetic rate. Additionally although showing a smaller degree of induction, the conalbumin gene seems to respond to lower levels of hormone than ovalbumin does. This may account for its higher constitutive levels of transcription. The time course of induction is also closely related to numbers of receptors, however the conalbumin gene began transcription at the same time as the receptor/hormone complex was found in the nucleus; the ovalbumin gene behaves differently as transcription is not activated immediately. Similar observations on control of ovalbumin have been obtained from other groups using more conventional techniques such as probing unlabelled de novo synthesised RNA.

Differential control of the ovalbumin gene by oestrogen or progesterone has been observed elsewhere (126). Use of a specific oestrogen inhibitor was able to reveal the effects of progesterone in isolation from oestrogen. Progesterone clearly affected translation in some manner, although the precise mechanism involved was not clear. A modulation of initiation rate of messenger on ribosomes seemed the most likely site of action. This is not a complete picture though

as in the prolonged absence of oestrogen, progesterone was unable to maintain ovalbumin m-RNA levels and a build up of monosomes and free ovalbumin m-RNA sequences occurred. Thus the continued efficiency of the translational machinery appears in some way to be dependent on oestrogen.

Another interesting comparison is one between the conalbumin gene and its identical liver analog transferrin (127). This latter whilst identical to conalbumin and similarly induced by steroid hormones is also regulated by serum iron levels (128). Monitoring of synthesis in isolated nuclei demonstrated transcriptional regulation in response to hormone or iron deficiency. In the presence of both stimuli a synergistic effect was also noted. By comparison with oviduct, the liver response is muted, however. Despite the interaction between iron deficiency and oestrogen in modulating transferrin expression, again different sites of action have been inferred from the fact that the presence of saturating amounts of ferritin did not block oestrogen stimulation.

Other steroid inducible genes shown to be regulated in this fashion are casein in mouse mammary gland (17),(129) or the androgen regulated kidney (132) and liver products - MUP (130) in the mouse and α_2 globulin of the rat (131). Glucocorticoids have also been shown to regulate tryptophan oxidase and Tyrosine amino transferase in the same fashion (131). Also in the liver, rates of synthesis of α-foeto-protein (133),(134) and serum albumin (132) have been shown dependent on m-RNA level. Control of α-feto-protein by

density dependence in a hepatoma line was additionally shown to affect rates of transcription.

Whether this relationship is true of peptide hormones remains unclear, however prolactin m-RNA levels are modulated by thyrotropin releasing hormone, in concert with prolactin synthesis (135).

Two observations nevertheless, supporting a slightly different pattern are worthy of note. The rat urinary protein α_2-globulin has been shown to be sensitive to a large number of different hormonal stimulation. Reinduction of synthesis in hypophysectomised animals required the presence of androgen, thyroid hormone, glucocorticoid and pituitary growth hormone. However, whilst the messenger RNA can be induced by the first three of this group, in the absence of growth hormone no protein synthesis occurs. Translation _in vitro_ and hybridisation studies (136) have demonstrated that functional message exists in the cytoplasms of animals so treated. Further treatment of these animals with growth hormone causes a rapid synthesis of α_2-globulin without affecting m-RNA levels. This is somewhat similar to the situation observed in chick oviduct stimulated with progesterone (126). Different mechanisms may be acting though, as in the latter case inefficient initiation by ribosomes seemed to be the governing factor. In the α_2-globulin example, activation of translation was accompanied by a shift from free polysome associated messenger to membrane bound polysomes. This could be due to an effect on elongation rate; the membrane association has been

postulated to occur from implantation of the nascent protein chain into the cell membrane.

The other example is that of procollagen synthesis in developing bone (137),(138). Whilst the level of messenger RNA here again correlates well with the rate of protein synthesis, the rate of m-RNA synthesis does not appear to be the dominant factor in establishing the m-RNA abundance. Rather, a decrease in the stability of non-collagen m-RNAs seem to be the cause. This is of course very similar to the proposed mechanism of terminal erythroid differentiation described earlier (114).

With regard to the future direction of this line of investigation, the ability to clone sequences of interest means that coordinately regulated groups of genes, like the major egg white proteins can be studied. This should provide greater insights into how tissues undergo specialisation. Such an approach is perhaps best exemplified by the isolation from a c-DNA libraryof avian liver sequences responding to oestrogen induction (71).

I.B.3. Specific gene expression during development.

Similar techniques have been used also in simpler eukaryotes to identify control mechanisms occurring during development. Some specific products, present in high abundance, sea urchin embryonic histones and Dictyostelium discoideum actin, for example have been analysed. Where specific probes have not existed one commonly used tool has been the application of cell free translation in vitro, using systems derived from wheat germ (400) or rabbit reticulocyte lysate (401). The advantages of this approach here are that a number of specific gene products can be analysed simultaneously, additionally as the m-RNA populations are predominantly of lower than average complexity a meaningful number of proteins can be analysed. Furthermore the nature of regulatory events can be usefully investigated by comparing protein patterns derived from in vitro translation and those derived from in vivo labelling of proteins. Quite sensitive analyses are possible, particularly if two dimensional gel electrophoresis (141) is employed.

A transition of particular interest occurs during sea urchin early embryonic development. During the blastula stage a change in the expression of histone m-RNA occurs. This results in a switch from a set of so-called "early" histone m-RNA to a new type referred to as "late" m-RNA. Each set of m-RNAs codes for all five histone variants but are thought to originate from different gene sets; the early m-RNA supposedly deriving from the highly repeated, clustered histone units. This implies that the "late" histone seen

throughout subsequent development may originate from an alternative source. To support this interpretation are the observations that the histone proteins synthesised at late blastulae stage change from a single form to a number of sub-types (139), possibly having future cell type specificities. Differences in m-RNA sizes correlating with this change have been identified in some species (139),(140). Hybrids formed under normal conditions also show reduced Tm values indicative of base mismatching, moreover stringent hybridisation to cloned "early" genomic DNA can produce only homologous hybrids; no hybridisation to late m-RNA is observed under these conditions (142).

The mechanisms determining this switch in histone gene expression are not so well defined conflicting reports exist regarding the time course of switching. Early and late histone transcripts have been observed in developmental periods outside their normal expression (149). Others however have been unable to find evidence of early histone m-RNA in late development, although a small amount of late histone m-RNA has been observed prior to mesenchyme blastula stage (142). What is perhaps rather more clear is that switching is mediated at the level of m-RNA synthesis. No translatable m-RNA has been extracted and shown active *in vitro* at times when no *in vivo* histone is present (140),(142).

Similar relationships have been implicated in the two major developmental changes of Dictyostelium which have been scrutinised. The first is the aggregation of vegetative cells and the subsequent production of spores, the latter is

the synchronous transition from spores, which germinate to form vegetative cells.

A fairly large number of changes have been identified during aggregation, however one of the most easily identified of these are the changes seen in the level of actin.

Typically very little actin is observable in gels of labelled protein until about 1 hr. into aggregation. A fairly large rise in actin level then occurs, however this rise is only transient and is followed by a swift decrease. Later levels rise again to an intermediate level but after 16 hrs of development actin decreases again to insignificant levels (143),(144). Investigation of these changes by in vitro translation (144) has demonstrated that all the modulation of actin gene expression are regulated by the concentration of actin m-RNA in the cytoplasm. No changes in half-life of actin m-RNA could be correlated with these changes so the authors conclude that actin m-RNA levels are probably regulated at a transcriptional level. In support of that conclusion was the sensitivity of this actin developmental programme to RNA synthesis inhibitors such as Actinomycin D. (144)

Analysis of a group of mutants for the aggregation processes revealed also that the final precipitous fall in actin synthesis may be controlled by cell-cell contacts, as mutants unable to aggregate maintained high levels of actin synthesis (140),(145). (Late aggregation variation in nine other proteins which fall in level at about the same stage as actin also relate to m-RNA synthesis and moreover

are also controlled by cell/cell contacts, as shaking to prevent aggregation deletes this developmental response

Two other developmentally regulated proteins of high abundance described in the same study show similar transcriptionally mediated regulation. One mutant strain developmentally incompetent was found unable to induce these proteins, which appear very early in the differentiation cycle. It is possible therefore that these proteins may have some developmentally important function (144),(145). More complete evidence for transcriptionally controlled developmentally regulated sequences has come from Williams and his colleagues (146). Hybridisation of pulse labelled nuclear RNA to such a sequence isolated from a vegetative cell c-DNA library revealed transcriptional rate changes correlated with developmental variations in m-RNA abundance.

During spore germination, a similar variety of developmentally mediated changes in protein synthesis occur (141),(147). Once again these changes were correlated with the presence or absence of translatable m-RNA (141). In the specific case of actin regulation, again variations in synthesis level were mainly accountable by m-RNA levels in the cytoplasm (147).

To redress the balance somewhat, a number of translationally controlled modifications of protein synthesis have also been observed in Dictyostelium. Lodish and his colleagues have identified six proteins which disappear very rapidly during differentiation (143),(144). This seems to be

mediated by a general drop in the rate of message initiation due to a reduction of availability in one of the initiation factors. As Lodish has described previously (148) such a change can have a specific effect on messengers with differing affinities for ribosomes. In addition Firtel has identified actin m-RNA presence in germinating spores before translation is detectable in vivo (147).

In other developmental systems under scrutiny, chorion production in Drosophila (149) and silk moth (70) and yolk protein production in Drosophila (150) presence or absence of m-RNA has again correlated with protein production. In yeast, the cytochrome C gene has been shown to be regulated transcriptionally (151).

Despite this definite trend, there may be one stage of development where translational controls or post-transcriptional controls actually are the dominant regulatory mechanisms. This is during the period of very early embryonic development.

During oogenesis in Xenopus laevis a major shift in the synthetic rate of histone occurs such that the egg is synthesising much more histone than the oocyte. By translating m-RNA extracted from oocytes, eggs and embryos in an in vitro system Ruderman et al (152) demonstrated that the total histone m-RNA population did not change during this period.

In the early mouse embryo translational control of three particular proteins is stimulated at the two cell stage of development. No changes in m-RNA population occur at this

time and translatable m-RNA was extractable from eggs although in vivo no synthesis of these proteins could be detected. Additionally, α-amanitin which is a specific inhibitor of polymerase II transcription, does not affect this change (152). Development beyond the two-cell stage however did seem to require embryo mediated transcription.

In concluding this section the overriding impression derived from study of the expression of these specific gene products must be that the dominant factor determining the level of cytoplasmic m-RNA coding for these proteins is modulation of the rate of transcription. Additionally it seems that many increases in transcription are further amplified by increased stabilities of the messengers involved. As the majority of these gene products are, by virtue of their tissue specificity, abundant proteins, this begs the question whether these genes are in fact controlled in a fundamentally different way from genes coding for low abundance products, which may also be of more ubiquitous representation. This is especially revealing as in the previous discussion on RNA populations, a general trend towards post-transcriptional regulation was described. Most of these investigations, being concerned with the major part of the sequence complexity amplify the effects of low abundance sequences.

A number of recent publications may shed more light on this situation in the future. Tilghmann and her colleagues, by constructing a cDNA library have selected out

sequences which are later developmentally regulated and rare in the m-RNA populations, their expression being confined to the early stages of muscle development. The regulation of such sequences may help to clarify this situation. The study of genes (and products) of this nature may also be revealing as they may function to control the expression of genes observed later in development and differentiation (153).

Two other groups have used recombinant DNA technology as direct extensions of their previous work on m-RNA population analysis. This enables specific probes, and the concomitant increase in detail that ensues to be exploited in examining the regulation of sequences of widely different abundance.

Analysing development in the sea urchin, using randomly selected clones from a library made at gastrula stage, Davidson and his coworkers described a number of interesting trends (154). One was a marked shift at the pluteus stage to expression of a small number of sequences at very high abundance. The gastrula by contrast had no extremely abundant sequences. Surprisingly perhaps, a large number of clones analysed contained sequences expressed uniformly throughout development. Most of these sequences were of the rare class and if more abundant at any stage were only modulated by a few fold in concentration. More abundant clones of gastrula and pluteus stage more frequently showed developmental stage abundance variation. However of 40

clones examined, only three patterns of behaviour were identified. Moreover the vast majority of these abundant pluteus sequences were fairly abundant in egg m-RNA. Whilst there was a late developmental shift up in abundance of these clones their general abundance relationships seemed to have been established in the egg where the majority of RNA is of maternal origin. Nevertheless during embryogenesis some sequences found in the egg are not detectable at later stages and conversely some sequences not detectable in the egg, rose to observable levels later during development. Preliminary evidence suggests, interestingly, that these highly prevalent pluteus sequences attain these levels by stabilisation of the m-RNA and a higher rate of transcription (154). Whether the same controls exist in the egg and during oogenesis however, remains to be seen.

The other example using random selection of clones, comes from a study by Darnell and colleagues (155). In this study nine cDNA clones corresponding to CHO cell m-RNAs were selected and analysed for rate of transcription and representation in nuclear and cytoplasmic RNA. The results of this study demonstrate that relative nuclear abundances for six of the clones were maintained in the cytoplasm. However, the range of nuclear concentration observed only varied by a factor of ten. The cytoplasmic concentration on the other hand differed by approximately 100 fold at steady state. Furthermore some m-RNAs were seen to have similar transcription rates but different cytoplasmic abundances and vice versa. This is probably the most direct

evidence at present of post-transcriptional modulations of gene expression. As the authors rightly point out their data may well be indicative of a mode of gene expression, such that the abundances of m-RNA exported from the nucleus relate closely to transcription rates but that differentiated cytoplasmic stabilities modulate the steady state levels. Such a model is of course testable and further evidence on this point will be of particular interest.

Perhaps slightly more contentious is their interpretation of sequence conservation data during processing of HnRNA to message. In this system only a 2-5 fold size increase in pulsed nuclear RNA was observed over the size estimates for the nine m-RNAs. Based on this data they have concluded that many more sequences must be present in nuclei than ever reach the cytoplasm i.e. that there is qualitative selection of sequences for transport to the cytoplasm. Accurate determination of primary transcript size is necessary for such a conclusion. Moreover results obtained from direct sequence analysis of genes in DNA from a number of sources has indicated that the average size difference between m-RNA and primary transcript may be significantly greater than the 2-5X observed by Darnell and his colleagues. If this can account for the observed discrepancy in transmitted and conserved sequences, then processing of some primary transcripts must be extremely rapid since this study utilised pulses of only 5 min. duration.

Undoubtedly, the ability to construct cDNA and genomic libraries will have marked effects on the study of

differentiation making it possible in particular to select out developmentally regulated or tissue specific sequences and more importantly perhaps, to extract messenger sequences under coordinate genetic regulation many of which may not be located on the same chromosome.

II. DNA SEQUENCES MEDIATING IN GENE EXPRESSION

A. Role of repeated sequences in eukaryotic genomes

It was well established a number of years ago that the genomic DNA of most eukaryotes contained a significant fraction of repetitious sequences. Some of this DNA was very simple in sequence and repeated in tandem arrays ; subsequently shown to be clustered at centromeric regions (46). More importantly some repeated sequences were observed to have moderate repetition frequencies in the region of 100-1000 times per haploid genome. A number of observations corresponded to suggest that these repeated sequences may play a role in differential gene expression.

Extensive sequence analysis of a number of eukaryotic genomes by Davidson and his colleagues (5) indicated that these repeated sequences were not clustered within the genome but extensively dispersed throughout it. Furthermore the repeated sequences, which averaged 300 bp in length, were interspersed between single copy DNA of approximately 1000 bp in length. That such an arrangement might be functionally significant was witnessed by the fact that this pattern was observed in organisms as disparate as the sea urchin and Xenopus laevis. One possible exception to this pattern though may be Drosophila which showed longer repeat elements and longer single copy regions.

Other observations were that these repeated sequences were not identical but showed evidence of sequence divergence from a common ancestral sequence (54). On the other hand the nature of the repeated sequences showed a surprising degree of sequence conservation during evolution, implying that these sequences were of functional significance. It was also established that the majority of structural genes were encoded by single copy DNA (5).

A number of models of gene expression using these observations as a basis have been published (5),(156),(157),(158).

Subsequent study of the RNA sequence content tended to corroborate those implications above. Unlike m-RNA for example, Hn RNA was shown to contain a significant repeated sequence fraction. Smith et al (54) by isolating the RNA reassociating at low C_ot, on HAP were able to show that these hybrids contained both repeated sequence and single copy DNA transcript covalently associated. This was evidenced by the RNAse sensitivity of a large proportion of the low C_ot duplex molecules. However if reassociation to higher C_ots was performed the sequences subsequently became nuclease resistant. Furthermore the size of the repeat sequence element was small, about 300 bp like the DNA repeats, whereas the HnRNA molecules were large, estimated to be on average 3Kb in size. Other investigators have found similar patterns (46).

However, the lack of resolution possible using total genomic hybridisation probes (54) plus the uncertainty about the true role for HnRNA (46) meant that interest in repeated sequences waned somewhat. More recently though, the possibilities of cloning and of rapid sequencing (loc cit) coupled with a number of other observations have reawakened interest in the role of repetitious DNA.

Klein et al (159) have used cloned repetitious DNA from the sea urchin to characterise the various repeated sequence elements. Reiteration frequencies were estimated for the different cloned sequences. A modal value of around 100-1000 copies per genome was obtained with a distribution from 3-12000. Checks were included to demonstrate that this random selection of clones was representative. The fact that each of these cloned repeats was different confirmed the original idea that the repeated sequences were a combination of a number of different repeat "families" (159). These families were investigated further by analysis of thermal stabilities of hybrids found between cloned repeat unit and total DNA. The vast majority of repeat

families showed that stable hybrids formed at high stringency and no extra sequences were hybridised at a criterion 10°C lower. However the melting temperature of these hybrids was about 11°C less than reassociated cloned strands, showing that intrafamilial sequence divergence does take place. In general the cloned repeats could be divided into three classes, those showing little or no divergence (greater than that existing due to population polymorphisms) those showing a greater divergence, but no difference between hybrids found at the two hybridisation criteria and those which did differ in this latter parameter: being the most widely divergent families. The majority of the families analysed fell into the second class.

The full significance of this division into repeated sequence families is better displayed by further analysis in the sea urchin HnRNA by Scheller et al. (160) and in maternal oocyte RNA by Costantini et al. (161). Parallel findings have also been obtained by Firtel and his colleagues (162),(163),(164) using genomic DNA libraries from Dictyostelium.

In the sea urchin, the representation of a number of cloned repeated sequences was analysed at different developmental stages, oocyte and gastrula and in adult intestine nuclear RNA. Differences in representation of over two orders of magnitude were observed, furthermore such variation in representation were observed to be highly specific to developmental stages, repeat families could be very highly represented at one stage and not at others. For example the gastrula contained approximately twenty times as many transcripts of one clone than did the intestine HnRNA. Thus the sets of abundant repeated sequence families differ in each of the three populations analysed. Perhaps unexpectedly both strands of each repeat were found to be present in all nuclear RNAs, most often both strands were represented at

roughly equal concentrations , but sometimes asymmetrical representation was observed. None of the cloned sequences was represented at anything like similar abundance in the cytoplasm and the small copy numbers could be accounted for by persistence of repeat containing maternal messenger.

Further evidence that the repeat elements were part of larger HnRNA molecules and not separate transcripts were obtained by hybridisation to different HnRNA size fractions. The most emphatic display of this pattern of transcripts however has been provided by Costantini et al. (161) using oocyte RNA which was self annealed to low R_ot values and analysed in the E.M. Both intra and intermolecular association are visible and many multimolecular hybrids are formed, indicating the presence of more than one repeated sequence element per HnRNA molecule. The region of hybridisation are bounded by long tails of unhybridised RNA, showing the covalent association with single copy sequences. Isolation of ^{125}I labelled repeat containing HnRNA using cloned repeat sequence elements was able to demonstrate that 85-90% of these molecules are single copy sequences. The repeated sequences like the single copy were also proportionally associated with poly A indicating that repeated sequence elements are associated with RNA destined to become messengers. What is unknown is whether the destined m-RNA has the repeat elements removed before translation, as do the messengers of the later cells. It seems unlikely that the intact molecules could be translated as they contain multiple "stop" codons in all reading frames (161).

In Dictyostelium a number of clones have been isolated displaying a similar organisation. Dictyostelium has a similar repeat sequence content to other eukaryotes. However the HnRNA size is significantly smaller than in higher eukaryotic species. This may have repercussions for some aspects of gene expression. Two clones have been identified which contain both single copy or repeated sequences. Mapping of these clones has shown that the repitititious elements are placed 5' to the coding sequences. For one clone a species

of RNA which made stable full length hybrids was identified, however a large amount of hybridisation to a heterogeneous number of RNAs was also observed. It should be emphasised however that the repeat elements in Dictyostelium are still present on messenger RNA in complete contrast to the sea urchin examples; there is some evidence that the repeat is larger in HnRNA however. Another divergent feature of this organisation relative to the sea urchin pattern is the asymmetrical transcription of the repeated sequences as well as of the single copy sequences.

Copy numbers for the Dictyostelium repeats are typical though for interspersed repeat families considered above, and the size of the repeated sequence element is also similar. Strong evidence for covalent association of the repeat with single copy sequence in RNA was provided by comparing reassociation of the cloned DNA with labelled poly (A)+ RNA. If the hybridisation was monitored by HAP repetitive kinetics are observed, if followed by RNAse digestion, predominantly single copy sequence kinetics result (163).

This organisation, whilst ubiquitous in the Dictyostelium genome (30% of poly A(+) RNA is thought to be arranged in this fashion) is not obligatory. One clone shown to encode a messenger containing an associated repeat element, also hybridised a totally unique sequence. In addition the multigene family of actin genes does not show the presence of a repetitiouselement. The immediate 5' sequence in this case was extremely A-T rich; the significance of which is not yet clear (165).

Further, it is not yet known whether changes in repeat sequence family representations in Dictyostelium show the developmental variations seen in the sea urchin. Equally, it is not known if the persistence of repeat sequences within Dictyostelium messenger RNA is related to the fact that gene sequences in this organism have been shown not to contain intervening sequences (with one recent exception).

Possibly, those m-RNAs not containing repeat elements are those which do contain introns and are therefore extensively processed before export from the nucleus, causing removal of any repeated sequences.

Another aspect of the repetitioussequence organisation of nuclear RNA and DNA has been reappraised using recent technologies. In particular the work of Jelinek and his colleagues has brought to light some interesting phenomena. Investigation of genomic repeated DNA has shown that a large percentage of this DNA contains inverted repeat sequences. These inverted repeat units and also DNA which reassociates at low C_ot have been shown to hybridise to HnRNA repetitive sequence (166),(167). In addition, if HnRNA is self annealed these repeat sequences can be isolated. Fingerprint analysis of the RNA repeats, DNA repeats and inverted repeat DNA all show the presence of the same oligonucleotides (167). The size of all these elements is around 300 bp like the major interspersed repeat DNA fraction. These observations have been made in both CHO cells (169), HeLa cells (166) and close homology has been demonstrated between the oligonucleotides in both the hamster and human nucleic acids (167).

Partial sequence analysis of these regions has shown that further homologies exist. Rubin et al. (170) and Houck et al (171) have characterised the major 300 hp repeat family from DNA. A common lineage has been identified for all these sequences due to possession of a characteristic restriction site for Alu I . This nuclease also cuts the majority of low C_ot repeated DNA and inverted repeat DNA in a diagnostic fashion. Furthermore the Alu family DNA repeats are closely related to the major repeat sequence from hamster DNA (167). Some human Alu family members have been sequenced after cloning (167) as have CHO analagous clones (169). Comparison of Alu family clones has shown the presence of a consensus sequence common to all members, however sequences removed from this region show varying degrees of divergence (167).

Additional homologies have been deduced from sequence comparisons, in particular pol III transcripts from cloned segments of genomic DNA (172) transcribed in vitro show strong sequence homology. A number of viral origins of replication also fit the consensus sequence. This is particularly interesting as SV 40 large T antigen is thought to act at this site and is probably responsible for mediating the late gene expression in SV 40 and Polyoma (173). The idea that these repeated sequence elements may play a role in gene expression has been reinforced by other findings on small RNA molecules associated with HnRNA. Jelinek and Leinwand (174) for example identified a small RNA in CHO and rat cells which copurified with HnRNA but which was not itself polyadenylated. Sequence analysis of this RNA has revealed a region of homology with the repeat sequences referred to above (169). Indeed there is some homology in this region to the human Alu family (167). The remaining sequence diverges strongly from these other repeated sequences. Therefore the small RNA is almost certainly transcribed from other regions of the genome.

The situation would be a lot clearer were it not for a veritable plethora of small RNAs currently under scrutiny. A 7S species of RNA of predominatly cytoplasmic origin has been described by Weiner (175), which is able to make strong hybrids with Alu family DNA. Small RNAs of predominantly nuclear localisation known as the U series have also shown some interesting properties. Apart from a high degree of conservation of sequence (176) and ubiquitous occurrence throughout eukaryotic organisms the U1 RNA has also been found associated with nuclear HnRNP particles (177),(178). Studies of sequence homologies have led Steitz and her colleagues to postulate that the U1 RNA may function in splicing of HnRNA, by virtue of an association between the 5' end of the U1 sequence with sequences at the intron/exon boundaries. The small RNA containing RNAse P of prokaryotes has been cited as a possible precedent here (172). Also

in favour of this interpretation is the existence of a solely nucleolar U series RNA which may therefore serve to process ribosomal RNA precursors (179).

As yet of course, the significance of these findings on repetitive sequences has not been fully elucidated. Perhaps the most intriguing observation is the polymerase III promotor activity of the Alu family sequence and their concomitant presence also within polymerase II transcripts (HnRNA). As this promotor activity was an in vitro observation its physiological significance has yet to be established as has the siting of Alu sequences with the β-globin sequence cluster.

However, the evolutionary significance of repeated sequence within the genome and the extremely high conservation of some of these sequences argues strongly for a central role in some biological process, even if this is merely as replication origins and their transcripts say putative primers for replication. Undoubtedly this field is one which will remain active in the near future and may still hold the key to some fascinating aspects of eukaryotic gene expression.

II.B. Regulatory genes. II.B.1. Classical genetic evidence.

Without doubt, one of the most significant factors in the elucidation of prokaryotic control mechanisms has been the use of genetics. It is genetic mutations more than any other factor which can provide the necessary link between structure and function. Regarding the problems facing study of these controls of gene expression in eukaryotes it is precisely this lack of mutant types which have failed to provide evidence for regulatory genes. The prime advantage of this genetic approach though is that no pre-conceptions about structure (at the DNA level) or mode of action need be made. The necessity to obtain mutants affecting expression of structural genes has been tackled basically in two ways. Firstly, using classical genetic techniques an emphasis has been laid on heavy mutagenisation of a well defined genetic locus to look for mutations

affecting expression which map outside the structural gene region. Ironically, this technique has proven less successful than might have been expected; the greatest success in discovering regulatory mutations having come from analysis of natural variation in different populations. The availability of highly inbred strains of mouse has proven invaluable in this respect.

The alternative approach is to abandon classical genetic analysis and essentially work backwards. The ability to clone and sequence DNA easily has meant that so-called surrogate genetics can be performed. In this instance DNA from a normal individual (or group) can be manipulated chemically, by restriction enzymes for example, and the resulting modified molecules assayed in vitro to see which modifications to the DNA structure affect their expression. Such analyses of course can only provide information about the nature of sequences lying very close (in genetical terms) to the structural genes under study. Within this limitation such an approach has proven quite fruitful in defining various control regions without shedding a great deal of new light on the basic regulatory mechanisms, which of necessity may act over somewhat larger distances.

Unfortunately, as yet there is still no unification within a single system where a genetically identified control locus can be subjected to scrutiny at a biochemical level. A number of systems are potentially interesting and one suspects that it may be in Drosophila, where the necessary precision of genetic analysis can be achieved, that the first breakthrough will occur. For the present the best characterised systems are simple eukaryotes, yeast having provided some of the most provocative information (180),(181),(182),(184). However until some notion of the general applicability of such mechanisms to higher eukaryotes can be ascertained, these findings will remain of dubious significance.

In the mouse, use of inbred strains has enabled a number of identifications of potential regulatory genes.

In nearly all these cases enzyme activity differences have been identified which cannot be accounted for by changes within the structural gene (or at least the coding portion). The best characterised system to date has been the analysis of the testosterone inducible acid hydrolases, β-glucuronidase α galactosidase and β galactosidase; the first of these has been described in most detail (185-190).

In studying inbred strains of mouse Paigen and his associates have isolated a number of types of regulatory mutants which they define in a number of ways. Systemic regulators change the levels of enzyme activity throughout the whole animal, whereas temporal regulation tends to affect only certain tissues and regulate enzyme activity changes with regard to particular developmental stages. A third class of regulation seems to affect the degree of response to hormone without affecting basal levels. Most of these types have been found associated with the hydrolases but their general applicability seems valid as other examples can be quoted, including some in Drosophila which act within these definitions.

The first identified locus in the β glucuronidase system was the Gus-r locus (which distinguishes it from the Gus-s structural locus). This gene regulates the degree of inducibility of the β glucuronidase by testosterone. A number of alleles have been observed at this site all of which showed preferential association with certain alleles at the Gus-s structural gene.

To date recombinants have not been obtained by laboratory crosses but the recombinant phenotype has been observed in vivo. Assay of enzyme synthesis has shown that Gus-r regulates the number of molecules of enzyme manufactured. As its close linkage to the structural gene, might infer, the regulation has been shown to be cis acting affecting the structural gene only on the chromosome which it is carried; heterozygotes show intermediate inducibility therefore. In heterozygote animals with both low and high inducibility alleles a predominance of one structural variant is observed.

Other regulatory sites in this complex have also been identified, Gus-t for example regulates a developmental shift in enzyme level seen during development only in mice of C3H strain. This gene affects only the activity in certain tissues, a third locus Gus-u appears to regulate enzyme levels, not developmentally but throughout the whole organism.

Similar control sites to these two have been found for β galactosidase and a temporal regulator of α-galactosidase defined. All these act in Cis and show very close genetic linkage to the structural gene. There is no fundamental reason in fact why they may not be located within it. However they are separable from the structural gene as different electrophoretic structural variants can show the presence of identical regulatory loci. All these identified sites regulate the number of molecules of enzyme synthesised. Other examples of regulatory genes behaving in this fashion have been isolated. Coleman (191),(192) has identified a linked LV site which regulates levels of δ-amino levulinate dehydrase (ALD) and again both high and low levels of enzyme were observed when identical gene products were synthesised.

Bernstine and his associates have identified a cis acting regulator of the mitochondrial malic enzyme (193-195). This has been shown to act on rate of enzyme synthesis in brain tissue, but does not affect heart. The two regulatory alleles can control identical structural gene products; the presence of structural electrophoretic variants has demonstrated the cis acting nature of the regulation.

Oddly, no other regulators of hormone inducibility have been discovered, with the possible exception of the Mup-a locus (196), which seems to control the relative induction of the two major urinary proteins MUP1 and 2. Good biochemical evidence is somewhat lacking in this system however.

Another pertinent feature is the predominance of cis acting regulatory sites closely linked to the structural gene. No cases of recombination have been identified in the laboratory; in the absence of a preferential association like that seen in the β-glucuronidase though, strains have shown opposite regulatory and structural linkages. The β galactosidase regulator for example Bgl-t has two alleles both of which have been found associated with either of the two structural gene variants.

The most significant finding regarding these cis acting regulators has been with regard to the Gus-r inducibility locus controlling β glucuronidase. Paigen et al (188) investigated m-RNA levels in testosterone induced mice of different Gus-r phenotype, by translation of isolated RNA in the Xenopus oocyte. This study revealed that Gus-r controls the level of m-RNA activity. As a translational regulation is less likely to be efficient in a heterologous system the most likely explanation is that Gus-r controls the rate of transcription of the Gus-s gene. To date, this is the only direct evidence for the role of a eukaryotic regulator.

In the mouse no definite cases of trans acting regulators seem to exist and the reason for this is not clear. Additionally two possibly distant regulatory sites which have been identified, in the development of H-2 antigens (197) and controlling expression of β-galactosidase (185) seem to act only on one chromosome or differently on the two products of the two chromosomes. No obvious mechanisms are able to explain this behaviour. Both systems require greater detail of genetic understanding however before any biochemical interpretations can be made.

Drosophila has been able to provide us with one seemingly trans acting regulatory site. This has been observed in the control of amylase in the midgut (198). Different Drosophila strains show three patterns of amylase activity

distribution within the dried gut. This distribution of amylase activity was found to be controlled by a single genetic locus mapping approximately two crossover units distal to the structural gene and thus fairly easily separable by recombination. In heterozygotes the allele of the control locus which promoted expression of amylase in the posterior midgut was found to be dominant over non-expression and thus all structural isozymes present were expressed.

The classic observation of a regulatory site in Drosophila is however again a cis acting control. This is a locus identified by Chovnick and his colleagues (199) after much intensive genetic analysis of flies bearing the rosy phenotype, which has been shown to be due to a mutation in the structural gene coding for Xanthine Dehydrogenase (XDH). This region has been extensively mapped using naturally occurring mutant strains, of which a number were identified due to changes in electrophoretic mobility, and also from mutagenesis by a variety of different techniques. The great advantage of using an organism like Drosophila is that rare crossover events can be identified by screening extremely large numbers of progeny, moreover the existence of well defined lethal mutations enables engineering of genotypes for such crossing experiments such that only recombinant phenotypes can survive. Thus very rare events can be observed without having to screen vast numbers of progeny. This is an analogous technique to plating bacteria on selective media.

Mutants showing variation in level of XDH activity were isolated however from naturally occurring strain polymorphisms, the mutagenisation producing mutants of "on/off" character was only able to generate structural gene mutants. These were useful however in defining the precise locations of the structural gene borders. Standard mapping techniques and examination of structural isozyme variants were able to identify the chromosomal location of the locus controlling high or low XDH expression. The regulatory site was shown to lie in the same region of chromosome three

as the XDH structural gene. Evidence that the level of XDH activity was not coded at the XDH locus came originally from the observation that high and low activities were both associated with a variety of electrophoretic variants; no systematic relationship could therefore be established between activity and structural gene identity. Experiments using antisera also demonstrated that relative activity could be correlated with numbers of enzyme molecules and not catalystic efficiency, again supporting the idea of separate identities. Initial experiments were unable to obtain recombinants though and when using a finer analysis on specially constructed strains nearly five million progeny were screened to identify 35 crossover types being recombinant, regulatory and structural gene loci. Further mapping of the site of the regulatory gene using lethal markers and structural gene mutants located at one extreme of the structural gene located the regulatory site to one side of the structural gene, separated from it by about 0.0034 map units. Conversion of this genetic data into actual lengths of DNA places the mutation site determining level of XDH expression approximately 3 Kb removed from one extremity of the XDH structural locus.

Finally, analysis of heterozygotes bearing electrophoretic variants of the structural gene and enzyme activity level demonstrated that the regulatory locus only controls expression of the structural locus to which it is genetically linked. This again is therefore cis acting regulation.

Undoubtedly now the techniques of molecular biology can be applied to study this relationship as accurate mapping of the site of mutation causing change in activity exists and moreover shows the distances involved to be amenable to such an analysis. To extract a recombinant phage from a genomic library bearing both structural and regulatory loci should now be comparatively straightforward, indeed DNA sequencing of the whole region should be eminently feasible. In combination with other recently evolved techniques some detailed understanding of how this gene exerts its effect on the XDH polypeptide production could then be obtained.

Two other systems which show promise also come from investigation of dipterans , however in these cases the molecular biology is perhaps better characterised than the genetics. In dipteran insects, a fair amount of information is now being accumulated on the chorion proteins and the genes which code for them (200-202). The chorion is secreted from three ovarian follicle cells and the chorion proteins form the shell around the mature dipteran egg. Detailed understanding of this system undoubtedly owes much to Kafatos and his coworkers (70),(203),(204) who have constructed recombinant cDNA libraries from the closely related silkmoth. The availability of these probes has made an analysis of the gene structure and expression possible.

Chorion morphogenesis has been shown to proceed in a highly ordered developmental fashion, with members of the two chorion multigene families strictly regulated by developmental stages. Various protocols have been developed to identify which proteins are synthesised at a given developmental stage and this has now been correlated with changes in m-RNA expression (200),(201).

Working on the development of chorions in Drosophila, Spradling, Waring and Mahowald (202) have identified that two major chorion proteins made at a particular stage of development correlate with the production of two specific m-RNA species. These RNAs were subsequently mapped against the polytene chromosomes and shown to be located in an area known to contain the mutational site for the trait occelliless. Female flies homozymes for this mutation manufacture abnormal chorions. Closer investigation of occelliless females demonstrated that the m-RNAs coding for these two proteins were produced in greatly reduced amounts as were the proteins. Isolation of electrophoretic variants of the two proteins enabled genetic localisation of the structural genes and this was found to be very similar to the location of the ocelliless mutation. Heterozygotes for the ocelliless mutation were also shown to produce much greater amounts of the two chorion proteins which were on the same chromosome as the wild type allele. Thus a cis action of ocelliless was identified.

The other system of promise is a mutant phenotype found in the silk moth Bombyx mori and also affects chorion production. This mutation known as Gr^B causes extreme reduction or absence of a number of the chorion proteins, additionally those affected are restricted to a given developmental stage in chorion synthesis. Kafatos and collaborators (204),(206) have also shown that heterozygotes for this trait synthesise the affected proteins at about half wild type levels. In vitro translation revealed that Gr^B mutants do not make m-RNA for the missing proteins. The coordinate loss of a number of proteins suggested that Gr^B was either a regulatory defect or a fairly extensive deletion.

By screening a cDNA library with probes derived from m-RNA of wild type and mutant phenotype, clones affected by the mutation were isolated. These clones were shown to code for the affected proteins by using binding DNA to filters and using these to extract the corresponding m-RNAs, which were then translated in vitro.

The isolated clones could then be used to analyse DNA of silk moths bearing wild or Gr^B phenotype. Southern blotting experiments revealed that the Gr^B phenotype is correlated with a large deletion of DNA presumably coding for the missing proteins. What is interesting about this finding is that a significant clustering of the chorion genes is indicated and perhaps more importantly that this clustering may relate to stage of developmental expression. In this context it is interesting to note that in Gr^B homozygotes (which do not produce late proteins) the early stage proteins whose synthesis usually stops, are not switched off. This switch mechanism could therefore be regulated by DNA sequences absent from the Gr^B genome. As such this is a very similar situation to that seen in the hereditary persistence of foetal haemoglobin (HPFH) syndrome seen in man which will be discussed below.

The last and most detailed example of the action of a regulatory gene (or genes) has been identified from yeast

genetics and intensive molecular biological investigation, it involves the control of the mating phenotype which is controlled by a locus known as MAT.

In a further example of yeast mutation, this time in the mitochondria the fusion of molecular biology and genetics is so complete it will be elucidated in Section IV.C. dealing with mechanisms of post-transcriptional control.

Two types of regulatory event are described by the MAT system, one is the ability of yeast strains to switch from expression of one of two different alleles at the MAT locus to expression of the other. The second is the means by which the MAT loci exert their affect on the genes responsible for the mating phenotype.

The ability to switch from one allele to the other at the MAT locus has been explained by the "cassette" model of regulation (180). A test of this theory (181) has confirmed most of the details, which envisaged the existence of two silent copies of the mating type gene in addition to the copy present at the MAT locus. The sequence at the MAT locus is the one expressed due to the presence of a proximally sited promotor. Mating type switching can be achieved therefore by substituting a copy of one of the silent genes, which may be of the alternative allelic type, at the active MAT locus.

II.B.2. Molecular Analysis of Thalassaemias.

Surprisingly perhaps the other instance of genetic and molecular biology coming together has been in humans. This is the long recognised clinical syndrome known as thallasaemia. The thallasaemias are a heterogeneous class of hereditary anaemias, particularly associated with malarial populations, and thought possibly to be common due to heterozygote advantage under such conditions (as in the case of sickle-cell disease).

The most commonly occurring thallasaemias are those affecting the β-like globin genes. This is most probably because the adult α type chains are expressed at all but the earliest stages of embryonic development and thus homozygous α thallassaemias tend to be fatal during foetal development. This condition being known clinically as Barts Hydrops fetalis. Many of the forms of β thallassaemia are fatal in the homozygous form too but as β expression begins only around the time of birth individuals so affected progressively develop symptoms during infancy.

A fair degree of additional heterogeneity exists within the definition of β-related thallassaemias. A form of the disease associated with loss of β and δ globin has been shown most often to originate from structural gene deletions. Evidence of this has been obtained from solution hybridisations and blot hybridisation using adult globin probes (207),(208). Other forms of β-thallassaemia are not so easy to explain (209-211). The β^+ thallassaemias are characterised by low levels of globin chain synthesis and often express fair quantities of messenger RNA. The β^0 condition however is characterised by absence of protein but quite variable amounts of message (209),(210). One particular β^0 has been consequently identified as due to a nonsense mutation, which can be complementedin cell free translation by a bacterial suppressor t-RNA (212).

Most cogent to this discussion however are the thallassaemias characterised by the condition described as hereditary persistence of foetal haemoglobin (H.P.F.H.). Individuals so affected do not express adult globins but suffer only mild anaemia as adults owing to a failure of

the foetal γ globin chains to cease synthesis at birth. In contrast, $\delta\beta$ thallassaemia is characterised by the normal reduced expression of foetal globins and this leads to markedly unbalanced α to β globin synthesis. Moreover H.P.F.H. is seen as a pancellular phenomenon, whilst synthesis is elevated in $\delta\beta$ thallassaemia in a heterocellular fashion. This latter is most likely due to selective survival of cells still expressing the γ chains (213).

Comparative study of the genomic DNA of H.P.F.H. and $\delta\beta$ thallassaemia sufferers might therefore be expected to help identify regions of DNA mediating the shut down of globin gene expression at birth. The results of such comparisons (208),(213) have shown that very large deletions of DNA characterise both syndromes. In H.P.F.H. the β and δ genes are absent and material to the 5' side of the δ gene stretching up to 9 Kb have been seen to be lost. Thus it seems that sequences mapping to the 3' side of the foetal genes are involved in their continued expression. This precedent is unique, no other examples of 3' regulatory sequences have been observed. Revealing in this instance is a $\delta\beta$ thallassaemic (208) which appeared to have the $^{A}\gamma$ gene deleted but left intact the $^{G}\gamma$ gene, which is situated further from the adult gene. In this instance although the supposed negative control region was absent normal reduction of γ synthesis had occured at birth.

Other relevant examples here are β thallassaemias where DNA deleted to the 3' side of the β-globin resulted in enhanced γ gene expression (211). Another unexplained case is a γ-β form of thallassaemia where the γ and δ genes are deleted, however a functional β gene plus bases of 5' flanking DNA are present yet no expression of this β gene occurs (214).

Data of this nature has led Flavell and his colleagues (213),(215) to propose a model that the regulation of the switch from foetal to adult gene expression is controlled by structural domains and thus mediated by chromatin organisation. In this context it may be worthwhile pointing out, as do these authors, that the interspersed pol III transcribed sequences identified by Duncan et al. (172) do delineate the regions of DNA specifying adult and foetal globins. Thus these may play a role in controlling this switch from foetal to adult globin expression. The role of chromatin structure in this context will be considered below (Section III.B.2).

It may be pertinent to make the observation that similar to the genetic examples referred to earlier these putative control regions act in cis only regulating the deleted chromosome and map many kilobases away from the genes whose expression they may be controlling.

II.C. DETAILED SEQUENCE ANALYSIS

The rationale behind most of the detailed study of DNA sequences is simple, by comparing the sequences of related genes within different species or different genes within related species one hopes to identify sequence elements displaying sequence conservation throughout evolution. Any sequence showing such behaviour can be considered to have some functional significance, the maintenance of which can be identified by the sequence conservation. Broadly speaking such an assumption has proven correct, in fact, sequence analysis has tended to reinforce the validity of this rationale. Naturally raw sequence data is of little value per se without some functional assay, but it does have value in determining which sequences may be worth analysing for function and can also provide new insights into the evolutionary relationships of different types of DNA sequence and the ways in which these evolve.

Sequence analysis has to date shown up three unexplained and unexpected phenomena. The first of these has been the discovery of intervening sequences (216-220) or so-called "introns"; stretches of non-coding DNA which interrupt the protein coding sequences or "exons" as they have been termed (221). The second is the existence of DNA sequences referred to as pseudogenes (222-226). These are typically gene copies showing relatedness to functional genes but which, for reasons unknown are no longer expressed. This fact is usually ascertainable from mutations affecting sites known now to be involved in gene expression or absence of open reading frames within the protein coding regions. Furthermore these pseudogenes show a marked divergence of sequences away from those known to be functional, such faster evolution is indicative of loss of genetic coding potential.

II.C.1. Introns.

The discovery of introns has been one of the most unexpected recent findings of molecular biology and has helped to reaffirm a different lineage for the prokaryotes and

eukaryotes. The fundamental difference highlighted is that of colinearity of gene sequence in DNA and in the m-RNA synthesised from it. The universality of this observation amongst the prokaryotes left the scientific community quite unprepared to find that this rule did not apply in general to eukaryotes.

Introns were first identified from the unexpected occurrence of restriction enzyme sites within the gene sequences, when from previous analysis of cDNA or m-RNA they were known not to exit at the level of messengers. This was found in, firstly a cloned gene, globin (230) and could possibly have been explained as artifactual, however subsequent findings on ovalbumin (218),(220) in total genomic DNA proved that this was the true natural biological organisation.

Discovery of various forms of sequence maturation in viral systems, adenovirus (227) and papovavirus (238),(239) demonstrated a degree of generality in the existence of introns. Further studies using the R-looping technique (230) which had already been used to show the existence of introns, were able to reveal that the nuclear 15S globin precursor RNA (48) and nuclear precursors to SV 40 m-RNA (231) contained the intronic sequences that were absent from the mature messenger. This data was subsequently corroborated by other methods. Thus as intron sequences were transcribed into RNA, the maturation process must involve a splicing reaction which excises the intronic region and joins the coding segments such that they become contiguous for translation.

At first it seemed possible that introns could conceivably be restricted to those gene products produced in high abundance. As more structural genes were cloned however (232-238) it soon became obvious that introns were the rule rather than the exception. It seemed possible that higher vertebrates could have had an obligatory requirement for intronic sequences; even the retroviruses for example (402) make spliced RNAs, although they do not actually possess equivalent structural gene interruptions. However,

not all coding DNA is organised in this fashion, the analysis of cloned interferon genes has demonstrated (239), (240) that both the leukocyte and fibroblastic forms lack intervening sequences.

Data acquired from non-vertebrate species has been far more difficult to interpret (241). Whilst yeast was shown quite early on to mature its t-RNA transcripts in an analogous though not identical fashion (242),(243) yeast nuclear genes in general were found not to contain introns (244),(245). An exception to this rule has been the single actin gene (246),(247) which to all intents and purposes fulfils all the requirements of a typical vertebrate structural gene. A similar situation obtained in Drosophila where the actin genes are multiple as in higher vertebrates. Rather perverse then perhaps is the finding that Dictyostelium which has also acquired multiple copies of the actin gene (165) has a form that lacks interruption (248). This would not be quite so difficult to reconcile were it not for the fact that Firtel (249) has identified a moderately expressed gene which does contain two intronic elements. It will be revealing to find out whether the sites of these actin gene introns split the coding sequence in the same place in the protein sequence in these different organisms and whether these sites demonstrate evolutionary conservation as in the case of the β-globin gene. Studies of β-related globin genes from man, mouse, rabbit, chicken (246) and frog (247) all show that the amino acid sequence is broken in precisely the same place in all these widely disparate organisms.

Such a high degree of evolutionary conservation has led a number of workers (250),(251) to suggest functional reasons for this "split-gene" type of organisation. The exact role of introns is still quite equivocal.

One hypothesis, first voiced by Gilbert (250) does have some experimental support (237),(252-254). This idea is that introns define functional units or domains within a given protein, such that new combinations of these units might be brought together during evolution by exon duplication events. Novel proteins could be created in this fashion and as each exonic element is functional in its own right, this should help to speed the evolution of proteins having new metabolic roles. Globin is one protein which provides support for this model, firstly a good correlation exists between the distribution of functions and of division within the gene sequence into exons (253). Secondly, in particular the central exon product has been shown to function as a haem binding peptide in isolation, when it can bind haem nearly as effectively as whole tetrameric haemoglobin (254). Additionally lysozyme has demonstrated a functional division between exons (252). The second exon for example, is mainly responsible for the catalytic activity, and includes the active site. The third exon adds extra substrate specificities. In support of such an hypothesis was their observation that a protein possessing a similar function in the dog shows homology, but has only two domains.

Separation of signal peptides from the rest of the protein by an intron has also been observed on a number of occasions, in ovomucoid (237)& conalbumin (255). Whereas in rat and human insulins although the structural C peptide is delineated by the large intron, the signal peptide is not (234), (235), (256).

More evidence on evolutionary relationships will be necessary before this theory can be established, one possible test might be to compare the haem binding regions of proteins other than globin. It would be revealing for example if these were also coded by a single exon, indeed one might expect a common evolutionary lineage for all such exons, if the theory is correct.

II.C.2. Pseudogenes

It has also been the study of globin genes from various species which has highlighted the existence of pseudogenes, examples of which exist in mouse (223), rabbit (222), goat (225) and human β globin gene clusters (226), and also mouse (257),(258) and human α gene clusters (224).

The majority of these pseudogenes are distinguished by marked sequence homology to functional globin genes. They are almost certainly themselves non functional by virtue of deletion, insertions, frame shifts or point mutations all of which render the gene untranscribable or impossible to translate, due to the presence of chain termination codons. Furthermore, where no such barriers exist, the characteristic splice junction sequences, so far shown to be necessary for processing, are seen to be missing (246). A related phenomenon seen in the mouse α pseudogene (257),(258) was the complete absence of introns, which had apparently been correctly removed according to the rules derived from study of other splice junctions (259),(260). In this instance the gene would be contiguous with its m-RNA, however no evidence for its transcription exists. Leder et al. have suggested a possible mechanism for generation of such a deletion of introns (261). Many other explanations are feasible; reverse transcription of m-RNA being one conceivable alternative.

Why pseudogenes exist is not really clear, some authors have pointed out (246),(262) that thus far pseudogenes always lie between the foetal or embryonic genes and those coding for the adult proteins. Maniatis and his colleagues have also presented evidence from sequence divergence patterns that a number of pseudogenes have had functional periods before acquiring mutations which subsequently inactivated them (224),(246). Proudfoot and Maniatis have thus suggested that minor globin genes, like the β minor and δ globins of the mouse and man respectively could therefore be evolving towards pseudogene status. Accordingly Martin et al (263) have identified a silent δ globin gene

in Old World Monkeys (all higher primate species possess δ genes).

Pseudogenes could therefore represent an obligatory stage in evolutionary change progressing through gene duplication, possibly due to unequal crossing over events, which have been shown to occur in the globin gene clusters of man (264),(224),

II.C.3. Gene families

A further observation, which incidentally has also partly derived from study of globins, is the existence of gene families. Recombinant DNA technologies have demonstrated that the majority of loci, contrary to what was indicated by hybridisation studies, coding for so-called unique genes, are in fact organised as families of related sequences. These families may take the form of multigene families with clustered organisations like the Drosophila chorion protein genes (201) or the mouse amylase genes (265) or may be dispersed throughout the genome like the Drosophila actin genes (248). More commonly, moderate repetition frequencies have been recorded, as in the α and β globin clusters, or the family of ovalbumin-like genes; two members of this family were only identified recently from analysis of overlapping clones isolated from a chicken genomic library (266) and are still of unknown function. In some examples, the family display differential expression of individual members as in the globins, or may show tissue specific variants, like the actins. The amylase genes of the mouse are odd in the fact that despite there being salivary and pancreatic isozymic variants, brought about by a gene duplication, the latter type has expanded to generate multiple copies. Individual strains of mice however express only a single member of this family.

Whether this organisation into gene families has any significance with relation to expression is unlikely, it more plausibly reflects gene duplication events referred to

earlier. However, a number of authors have pointed out that the mammalian globin gene clusters are all organised in such a fashion that the linear order of the genes along the chromosome mirrors their sequential expression during development. However, as the chicken has embryonic and adult genes arranged in a different fashion (266) with the adult genes flanked by the embryonic, the organisation of genes in the mammalian fashion cannot be indicative of general mechanisms for differential developmental gene expression. In fact it would seem that genomic proximity is in no way necessary to ensure coordinate expression. The α and β globin genes have been known for some time to reside on different mammalian chromosomes (267), yet are still coordinately expressed. This has occurred despite the fact that originally the genes must have been linked as was somewhat unexpectedly demonstrated by Jeffries et al. (268) from investigations of Xenopus globins. Nevertheless, some clustering of eukaryotic genes in relation to their developmental expression has been observed in Drosophila, and Bombyx (202),(205). Furthermore, a characteristic "back to back", divergent gene orientation in both the chorion genes (269) and in the so-called, major "heat-shock" loci have been noted (270).

II.C.4. Regions displaying evolutionary conservation.

Detailed sequence data now exists on approximately twenty eukaryotic genes from various sources as divergent as the sea urchin and man (246). Comparisons of such sequence data, often aided by computer, have revealed a number of highly conserved regions which may have functions in regulating transcription. Major technical breakthroughs have enabled tentative testing of the functionality of these regions both in vivo and in vitro. It is this link which will probably be vital in establishing how such mechanisms operate.

By analogy with prokaryotes, eukaryotic promotor sequences able to bind RNA polymerase would be expected to be found mapping to the 5' side of the initiation site of transcription. From S-1 mapping data (272) and other such techniques, or cloning of c-DNAs the 5' termini of a number of messenger RNAs has been established. The terminal nucleotide is recognisable by virtue of its "cap" (reviewed in 273), a modified guanine base. Study of DNA sequences around the region of initiation has identified a conserved sequence which Birnstiel and his colleagues (274),(275) termed a cap sequence. Whilst this sequence is not ubiquitously conserved related genes show related consensus sequences, as in the case of the histones (274) or the globin genes (246). Moreover, the capped nucleotide has almost always been an A residue (246),(262).

Further upstream (5' to the initiation site), analysis of cloned Drosophila genes by Hogness and Goldberg revealed the presence of a sequence TATAAATAG residing approximately 30 nucleotides from the cap site. What is immediately striking is the similarity of this sequence to the prokaryotic promotor sequence known as the Pribnow box (276). The eukaryotic version has since been multifariously referred to by its discovers names or by a number of onomatapoea resulting from the conserved sequence. Comparisons by groups working on egg white protein genes (276), globin genes (261),(246) and histone genes (274),(275) have all

been able to identify a similar such sequence preceding the initiation of transcription and to the list given by Maniatis and his group (246) can be added the major heat shock protein gene Hsp 70 (277) and even the yeast actin gene (243). Apart from good sequence conservation, (though between them the globins can only muster universal appearance of ATA at the equivalent site) particularly noteworthy is the extreme conservation of the distance of this Hogness box region from the initiation site. The variation in position stretches only five nucleotides in all the examples so far considered (246).

Two groups have identified a further region of sequence conservation, at a position similar to the prokaryotic recognition site, approximately ten base pairs upstream from the promotor (274),(276). In addition however, a eukaryotic region further removed from the Hogness box, which displays sequence homology to the recognition site has also been located, some fifty base pairs upstream. This has been termed by Proudfoot _et al_ (262) a CCAAT ("cat") box. This sequence is extremely well preserved within the globin family and the CAA element is common to virtually all examples known (246). Some other gene-specific sequences have been inferred as possible regulatory elements for example, the mammalian adult β-globin sequences share a lot of homology in their 5' flanking sequences, not common to the embryonic or foetal β globin genes (246). Amongst the histone clusters of two widely diverged genes, genes coding for the same histone types show some strong homologies (274). Particularly the H2A histones possess a run of 30 nucleotides of continuous homology. Another sequence which may be histone specific has also been identified some 15 nucleotides upstream from the TATA box. This GATCC region does not seem to have any equivalents amongst other eukaryotic genes. Interestingly, Efstratiadis _et al_. (246) have identified a potential CAAT box, in the sea urchin histones, but this is rather further from the initiation codon than is usual. Amongst other non-vertebrates another CAAT is found in a similar upstream

position, this time in the silk fibroin gene (278). However, no such sequence can be found in the Drosophila heat shock flanking region whereas an impressive homology does exist between this gene and the fibroin gene at a very similar site to the globin CAAT box (277). Unlike other eukaryotes the Hsp 70 gene also shows a region of potential secondary structure twelve nucleotides to the 5' side of the TATA box. This region also demonstrates dyad symmetry (277); such features are common in prokaryotic DNA/protein interactions (274).

The histone genes have also revealed dyad symmetry but this time located in the 3' flanking region (279),(275). In general the 3' flanking region showed much greater divergence of sequence in closely related genes than does the 5' sequence. One ubiquitous feature of the 3' flanking sequence is the element AATAAA in the DNA just upstream from the poly A addition site, revealing (246),(276),(280) this sequence is lacking from the 3' flanking region of the histones, which of course are not polyadenylated (279). Furthermore the curious finding that the Dihydrofolate reductase (DHFR) messenger can terminate at alternative sites (281) enhances the idea that this sequence plays a role in polyadenylation as the alternative addition site is preceded by AATAAA (281).

The secondary structure region seen in the histones falls into two classes of conserved sequence, one common to the histones H2A, H2B and the other to the remainder. The dyad symmetry enables formation of a loop and stem structure, which lies a few nucleotides upstream from an additional stretch of conserved sequence. Birnstiel and his colleagues (279) have suggested that these regions might act as processing sites to cleave a polycistronic histone transcript into individual messengers; the histones lack introns within the coding sequence (282).

A notable feature of the flanking regions downstream from the coding sequence is the complete lack of relationship

to any termination signals observed in the prokaryotes. This is in marked contrast to the putative promotor sites at the 5' side and thus may relate to some RNA processing functions, such as polyadenylation, unique to eukaryotes.

II.D. Functional analysis of evolutionary conserved regions.

A variety of means now exist with which to test the functionality of these putative regulatory signals. The ability to test both *in vivo* and *in vitro* has been illuminating, although as yet some of the precise mechanisms remain to be understood in detail.

Until recently all investigations of *in vitro* transcription had been attempted with bacterial RNA polymerase. An approach of doubtful applicability, as there has been no evidence that this enzyme was able to recognise eukaryotic promotor signals. The availability of specific cloned DNA fragments redirected efforts to enable meaningful *in vitro* transcription studies to be performed. Two cell free transcription systems have been evolved (283),(284), one depending on exogenous, purified polymerase II from eukaryotic cells (283), the other using endogenous polymerase activity with a cell free concentrate (284). Both systems are somewhat crude in that they use whole cell extracts to provide the initiation specificity. However they are both capable of transcribing adenovirus (283) and globin genes (285) from an initiation site identical to that utilised *in vivo*. No doubt future work will be directed towards identifying materials within the extract which donate specificity of transcription; some way of removing the endogenous pol III activity would be of obvious benefit.

Using such systems, the transcription of globin genes (262), adenovirus genes and conalbumin genes (286) have been studied. Transcription of mutants created *in vitro* by restriction and deletion has indicated that the TATA box sequence does play an important role in promotion. Chambon and his coworkers for example found that deletion of the 5' region beyond this region has no effect on transcription, however extension of the deletion into the region of the TATA box can reduce transcription to less than one percent of its original efficiency (286),(287). Such a change can

be brought about by a change of 3 bases in the end point of the deletion. Moreover, the same is true for both adenovirus genes and the conalbumin gene neither of which shows any homology in the 5' upstream sequences except for the TATA box region. Additionally, juxtaposition of the TATA box with plasmid sequences instead of the true m-RNA sequence does not abolish transcription although it does both reduce it and cause variation of the start point. Sequences around the capping site are therefore contributory to transcriptional efficiency, but not vital. Indeed Proudfoot _et al_ (262) have found an increase in efficiency from some deletions in this area. These authors have also found the CAAT box which they identified to be inessential for _in vitro_ transcription. Interestingly, where deletion of substitution of the capping region occurred, initiation was favoured at the A residue closest to the correct distance from the TATA box. Perhaps the most impressive evidence to date for the role of the TATA box has come from Chambon's group (286),(288) who rather than deleting the region have altered its sequence, by _in vitro_ site-directed mutation. A change in the Hogness box from TATA to TAGA has been observed to cause a dramatic loss of transcriptional efficiency _in vitro_.

These results must be tempered however by some of the more obvious objections to the fidelity of the _in vitro_ condition. Whilst it is particularly interesting that the δ globin gene, which shows some sequence variation at the 5' flanking region compared to the other globin genes was less well transcribed than the β, as is the case _in vivo_, none of the globinstranscribe as well as the adenovirus template (262). Furthermore the conalbumin gene has been shown to transcribe as efficiently as the adenovirus template, ovalbumin though was significantly poorer and similar to the transcription of the related "x" gene. In the intact cell however, ovalbumin is transcribed much more efficiently than conalbumin and forty times more rapidly than the x gene (286).

That the results of the *in vitro* investigations might not be the whole story has been corroborated by somewhat different findings using heterologous *in vivo* systems. For example Benoist and Chambon (289) have demonstrated that deletion of the TATA box in the SV 40 early gene promotor does not abolish T antigen expression in permissive host cells. It does remain possible here that some alternative minor promotor is responsible, however Grosschedl and Birnstiel using cloned sea urchin histone genes injected into Xenopus oocytes (290) have shown that transcriptional efficiency is not grossly affected by similar such deletions.

Alterations in the cap-site sequence were also ineffective in destroying transcription, although again initiation proceeded from a site corresponding to the normal distance 3' to the Hogness box.

Preliminary evidence from *in vivo* observations also have suggested that regions further removed on the 5' side from the initiation site may be involved. Two examples of deletion of upstream CAAT sequences affecting transcription are known (290),(291). In the former case, however, in the histone H2B gene this deletion increased transcriptional activity whereas in the latter case the transcription of globin genes was markedly reduced. Other upstream regulatory regions have been implicated in results from SV 40, histone gene transcription and Drosophila (291). Some of these sites are up to 200 bases removed from the initiation site and so novel mechanisms of action may have to be invoked.

A great deal of work is obviously required to refine these techniques especially the *in vitro* systems to acquire knowledge of factors mediating transcriptional specificity. Heterologous systems also have obvious limitations. Complementary application of both approaches at this stage is vital and have provided some degree of insight. The role of the TATA box for example has been shown to be related to that seen in prokaryotes but not directly comparable. Two examples of genes that lack this region

(292),(293) both viral systems and evidence from *in vivo* deletion mutants suggest that its presence is not obligatory for transcription. However absence of such a transcriptional signal has almost invariably been associated with heterogeneity at the start point of transcription (289),(292),(293). Thus it seems likely that TATA sequences act to align the polymerase with the transcriptional start point. In support of such a view are the observations that the presence of Hogness boxes upstream from unrelated sequences can mediate transcription *in vivo* and *in vitro* (286) from a start site very similar to that predictable from the start point in the natural gene.

These approaches will undoubtedly provide valuable information on the role of primary sequence elements in controlling transcription, the mechanisms elucidated must have the particular limitation that in crossing so many species barriers or constructing as yet crude *in vitro* systems only generalised transcription mechanisms will be revealed. Whilst these are important aspects it may well be impossible to approach the questions of tissue specific or developmentally regulated control of gene expression. This whole aspect may be better served by other approaches. The possibility of transferring eukaryotic genes into different cultured cells may be one such viable proposition (294). Here it has been observed for example, that rabbit globin genes can be faithfully transcribed in both mouse erythroleukaemia cells as one might expect and also in mouse fibroblasts (295),(296),(297). Cloned ovalbumin genes however are transcribed from an aberrant start point in the latter such cells (286),(298).

Some of the problems associated with modes of integration in these transfer systems can be obviated by transfer of cloned circular plasmid which seems capable of epigenetic replication in eukaryotic cells (299). The use of SV 40 vectors promises much (297),(300),(301) especially as the endogenous eukaryotic globin promotor sequences seem capable

of directing globin gene synthesis in hybrid molecules of this sort (302). New in vitro systems using SV 40 mini-chromosomes could be particularly fruitful, for few if any regulatory interactions occurring in eukaryotic cells are mediated via naked DNA but by that intractable material chromatin. The structure of this nucleic acid protein mixture, being unique to eukaryotes, could be at the very heart of differential gene expression. The role of chromatin structure will be discussed further below.

Before going on to discuss this aspect I should like to consider the best understood example of factors affecting transcriptional regulation. This is the system of pol III transcription which has provided an object lesson in how to expect the unexpected.

III. CONTROL OF TRANSCRIPTION

III.A. Pol III genes

The use of the *in vivo* and *in vitro* systems approach highlighted above on elucidating transcriptional mechanisms has nowhere been implemented to greater effect than in the control of transcription of pol III dependent genes.

Both eukaryotic and prokaryotic cells are now known to contain three functionally and structurally distinct classes of RNA polymerase. The first class Pol I normally functions to transcribe the ribosomal RNA genes. Pol II has been shown to be the polymerase responsible for messenger RNA transcription and is therefore the subject of much intensive recent study. This polymerase is characterised by its particular sensitivity to α amanitin (303). The final class Pol III, transcribes the small RNAs, t-RNA and 5S RNA which are repetitious genome sequences and it thus may function also to transcribe other small RNAs such as the U series of small RNAs or the Alu family sequences of the type found in the Pol III transcripts distributed within the β-like human globin cluster (172) (and see Section II.A). The significance of Pol III transcription may therefore be of greater importance to control of gene expression than might at first sight be expected.

Cell free systems, like those referred to earlier, had been elucidated for Pol III transcription somewhat prior to those capable of accurate Pol II initiation (304), (305), (306). Purified DNA templates such as cloned 5S genes (304), (305) could be transcribed *in vitro* in the presence of cell-free extracts from cell cytoplasms (305) or oocyte nuclei (304).

In order to identify regions responsible for controlling 5S gene transcription deletion mutants constructed by similar methodologies to those described previously were analysed by assay in the cell free system. Transcripts of the gene initiated correctly can be easily identified by size

as this system shows correct termination. By analogy with findings from prokaryotic systems the regions flanking the 5' side of the 5S gene were analysed for deletions affecting transcription. Deletions within this region were surprisingly, found to be ineffective in altering the in vitro transcription; all such mutants were able to synthesise 5S RNA (307). Replacement of sequences at the 5' border of the gene by plasmid sequences did result in some newly occurring heterogeneity of transcriptional initiation site similar to that observed with deletion of capping signals in the Pol II transcription systems. In this instance transcription was initiated at G residues, as in the case in the intact gene. Additionally transcription could be observed when extra nucleotides were inserted into the 5' side of the gene sequences.

When deletion from the 5' side of the 5S gene were extended beyond nucleotide +50 into the gene sequence however a dramatic change occurred such that deletions extending to +55 nucleotides abolished transcription totally in vitro. Thus it appeared that a region at the centre of the gene was controlling transcription which would then initiate at a roughly constant distance from this control sequence, the precise initiation point depending on the precise sequences at this distance. Using deletions mapping in from the 3' margin of the 5S gene, Brown and his colleagues (308) were further able to characterise the exact site of the control region. The border to the 3' side of the control region was found to map at +83 nucleotides from the normal initiation site. It is not yet clear whether all the nucleotides from +50 to +83 are responsible for the control. Analysis of sequence homologies amongst other Pol III transcribed genes pointed to an area at the 5' side of the control region. Subsequent findings however cast doubt on the validity of this observation.

Powerful evidence for the role of the control region in controlling transcription was provided by excising this

region and recloning it in plasmid pBR 322. This region on its own was shown able to direct transcription of RNA *in vitro* whereas a gene constructed from deletion of a similar region was quite unable to do so.

The nature of this transcriptional control has turned out to be of a novel nature from work continued by Pelham and Brown (309) and Roeder and his colleagues (310),(311). A major finding came from Roeder's continued analysis of factors within the nuclear extract able to confer specificity of initiation on Pol III transcription *in vitro*. Such a study was greatly facilitated by the existence of an ideal system for assay of complimenting activity. This was an extract made from the mature eggs of Xenopus in the same way as the oocyte extract, the crucial difference being that the former cells do not synthesise 5S RNA. Fractions of partly purified material from oocytes were therefore used to test their ability to stimulate 5S transcription in an egg cell free extract, containing cloned Xenopus 5S genes. Ion exchange fractionation and electrophoretic analysis of complimenting activity revealed that it resided in a protein of about 37,000 molecular weight. DNA binding studies using this purified protein were undertaken, using DNAse 1 cleavage to detect areas of the DNA protected by virtue of bound transcription factor. This located binding of the factor to a region within the 5S gene mapping between nucleotides +47 and +96. This region of course correlates very well with the control region defined by Brown and his group as instrumental in directing 5S gene transcription.

Particularly cogent was the finding that this 37K protein did not affect transcription of t-RNA genes also transcribed by Pol III but was specific to 5S genes. However whilst the purified factor is able to affect transcription of 5S genes of both oocyte type and somatic cell type *in vitro*, no evidence for its presence in non-oocyte cell extracts has been found (311), despite the fact that such cells transcribe the somatic type 5S genes.

The factor is therefore not responsible for differential control of expression of the two types of 5S genes, although it is possible that it bears different affinities for DNA of either type. More likely however is that this regulation is mediated by some different process(es). The factor was also found to be absent from egg extracts when assayed by specific antiserum (311).

A further extremely novel finding however was that the same protein has been identified associated with 7S cytoplasmic ribonucleoprotein particles which contain 5S RNA (309-311). Pelham and Brown (309) and Howden and Roeder (311) have both shown that the 37K transcription factor protein is able to bind the 5S RNA transcript as well as the DNA sequence. Furthermore, addition of 5S RNA to a cell free transcription system was able to abolish transcription of the genes. It appears that the RNA/protein interaction is less tightly bound, so most likely occurs only at high RNA concentrations. Thus it seems possible that 5S RNA synthesis in the oocyte may control its own transcription by sequestering the transcription factor as the concentration of the 5S RNA increases, finally switching it off at egg maturity.

Such a feedback control, if it does exist would be extremely elegant in operation, and it remains to be seen whether this mechanism does actually operate *in vivo* during development. One observation which is hard to explain if this feedback regulation does occur was that the 7S RNP particle was not able to protect DNA in a binding assay, yet was able to direct transcription of 5S genes *in vitro*. Binding would be established as would be expected if the 7S RNP was previously treated with RNase.

It may be possible that the transcription assay for some reason potentiates disassociation of the 7S RNP complex facilitating binding, perhaps alternatively the extract contains enough endogenous RNAse activity to release sufficient factors to stimulate transcription.

III.B. Role of chromosome structure

1. Classical lines of evidence

Control of transcription at a chromosomal level has been implicated by two classical observations. The first of these is the characteristic "puffs" of the Drosophila (and other Dipteran) polytene chromosomes, which have been demonstrated by Beerman and Clever to show both tissue specific and developmental patterns of appearance (312). Furthermore new puffed regions have been identified on the administration of ecdysone the Dipteran maturation hormone. Puffs have additionally been shown to be the sites of RNA synthesis. But perhaps the best recent evidence for puffs being the regions of active transcription have come from study in the heat-shock system. The induction of the heat-shock response is easily brought about by culturing Drosophila larvae at elevated temperatures for a short period (40 minutes at 37° centigrade for example (313) as opposed to the normal environment of 25°C). After such treatment the observed cytological changes are (313): regression of pre-existing puffs and the appearance of eight to nine new puffed regions; the extent of heat shock correlates with the size of the puffs. In addition heat shock leads to the appearance of a similar number of new proteins and polysome associated RNAs. From these RNAs have been identified species which hybridised specifically to heat shock puffs when analysed by <u>in situ</u> hybridisation. These m-RNA species also direct the synthesis of heat shock polypeptides in <u>in vitro</u> translation systems. Puffing can also be inhibited by transcriptional inhibitions such as actinomycin-D and α-amanitin, moreover the immunofluorescent detection of RNA Pol.II shows a preferential association of the polymerase with heat shock loci and an inverse of level in the nucleoplasm, correlating with a decline in non-heat shock regions (313).

The heat shock system provides a link with the second observation as ,if heat shocked cell cytoplasmic extract is

incubated with control nuclei heat puffs can be induced (314). A similar phenomenon occurs when steroid hormones interact with their target cells. Both oestrogen (315) and androgen (316) have been shown to interact with a cytoplasmic receptor protein. The binding of hormone causes two particular changes in the receptor first a change in conformation seems to occur, as the sedimentation coefficient of the receptor changes in some cases, secondly the binding of hormone causes a major translocation of receptor/hormone complexes to the nucleus, where they bind to chromatin. The receptors have been studied in some detail and seem intimately involved with steroid action, cell lines unable to respond to hormone for example have been found to lose receptor or possess receptors unable to bind hormone. A naturally occurring mutation in mice causing feminisation of males (T fm) (317) seems to be caused by loss of receptor activity. Additionally, so-called anti-inducers have been identified which although able to bind the receptor protein are unable to cause translocation and thus do not elicit a hormonal response (316).

In theory, if the binding of receptors to DNA or chromatin is able to activate transcription then it should be relatively simple to isolate the regions of genetic information which specifically interact with the receptor/ steroid complex to illicit de novo transcription. Unfortunately, the binding capacity of the nucleus for receptor complex is virtually infinite (316) and thus a large degree of non specific binding undoubtedly occurs. Factors like these complicate analysis but do not invalidate the theory that receptor complex interaction with specific sites mediates to produce transcriptional activity. Non specific binding of specific regulatory proteins was previously observed in E.Coli (318) in the case of Lac repressor binding to whole DNA versus binding to specific operator sequences.

That hormones do cause changes in chromosome superstructure or architecture can be tentatively concluded from experiments using E.Coli RNA polymerase. Despite the pitfalls of such analysis, after removing all the known artifacts of such investigations (and there have been many) O'Malley and his co-workers (319) have been able to conclude that the accessibility of the ovalbumin gene to polymerase is greatly enhanced by prior administration of oestradiol. By contrast, unstimulated chromatin supports transcription in this system at a greatly reduced level, much lower in fact than random transcription of all DNA sequences would predict. However it may be worth noting that some transcription does occur under these conditions and indeed that some untreated heat-shock puffs have been observed to accumulate label in the presence of radioactive RNA precursors (314). Thus even transcriptional controls via chromatin structure modulation may in fact not truly be the classically defined purely qualitative "on/off" type interactions.

III.B. 2. Analysis by nuclease digestion

In order to probe the relationship between chromatin structure and gene activity more precisely a number of workers have utilised the approach of nuclease digestion of chromatin (60),(320),(322). This might attempt to identify different susceptibilities to attack, by nucleases, of differentially active regions of chromatin. Alternatively, proteins involved in gene regulation, such as hormone/ receptor complexes or the proteins via which they interact with DNA, might be released from active regions and thus identified.

No-where has this approach been exploited more adroitly than by Weintraub and his colleagues (322-326). Using low levels of DNase I (isolated from bovine pancreas) digestion Weintraub and Groudine (60) have shown that the globin genes in expressing tissues were more sensitive to digestion than

in non-expressing tissues. Furthermore the ovalbumin gene was found to be preferentially digested in oviduct tissue but not in red blood cells or fibroblasts. Subsequently a number of other laboratories confirmed these findings (327),(328),(329). However, two major reservations soon became obvious when trying to correlate gene activity and DNase I sensitivity. Firstly genes transcribed at vastly differing rates were equally sensitive to DNase I (330) thus preferential digestion was not a direct function of transcriptional activity. Secondly both the β globin and ovalbumin genes, in mature red blood cells of the chick and in hormone withdrawn animals (60),(331) respectively were found still to be in a DNase I sensitive chromatin conformation. As mature chick red blood cells condense their chromatin and shut down all transcriptional activity (rather than extrude the nucleus, as in mammals) the preferential digestion can only be indicative of a transcriptional potential of certain genes.

This interpretation was further vindicated by the finding that in mouse erythroleukaemia cells the globin genes were DNase I sensitive both prior to and after induction. In fact the examples of cell lines with low and high uninduced globin levels were of identical sensitivity, as was an additional non-inducible variant (332). Thus despite good evidence for transcriptional regulation of globin in these cells no modulation of chromatin structure appears to occur. However a distinct lineage specific relationship was revealed by this study as both adult liver and hepatoma did not possess preferentially digested globin genes whereas in the erythropoietically active foetal liver the globin genes were sensitive to DNase I. A further intriguing observation from the same study was that a lymphocytic cell, a myeloma expressing immunoglobulin did not show DNase sensitivity of the globin genes whereas a lymphoblast line did show digestion of the globin sequences. The implication here is that differentiation may involve a general closing down of potential transcriptional activity of certain regions of the chromatin. A haemopoietic stem cell would therefore maintain globin genes in a potentially active state and these would be activated or not depending on the differentiation

pathway followed. If the cell developed along the myeloid lineage the genes would eventually be removed from the potentially active compartment and thus lose their sensitivity to DNAse, presumably as they matured from lymphoblast to lymphocyte. Such an interpretation whilst undoubtedly quite attractive is somewhat confounded by the fact that the globin genes in mature red cells are still DNAse I sensitive. However, as other transcriptional factors are most likely lacking in such cells it may not be necessary here to inactivate the genes. Indeed, from work using monoclonal antibodies Greaves et al (333) have identified markers similar to erythroid specific antigens as well as myeloid markers on early haematopoietic precursors.

Another possibly relevant finding is that of Groudine et al (322) that one of the endogenous viral genomes of the chicken is sensitive to DNAse I. Expression of these genes is under cellular gene control identified as a Mendelian locus Gs which is dominant to an alternative allele gs. As gs/gs homozygotes do not express the viral genome it is perhaps strange that the viral genome is in fact DNAse sensitive, however expression of the virus genes was observed in erythroblasts although not in chick embryo fibroblasts. The latter cell expresses in the Gs/Gs genotype. Thus the two alleles seem to affect the tissue specificity of the endogenous viral genome expression, without altering the chromatin confirmation. Interestingly, Gs/gs heterozygotes express intermediate levels of viral specific sequences, thus this regulatory locus like the others previously identified acts in a cis fashion, only determining expression of genes on the same chromosome.

A further refinement of the DNAse I technique has revealed a slightly different picture however. An original observation of Weintraub and Groudine (60) was that the embryonic globin genes of the chick were insensitive to DNAse I digestion at times when the adult genes were being expressed. This observation like the others quoted so far

relied upon a solution hybridisation assay for its interpretation. Thus sensitivity was monitored by gradual loss of hybridisability on exposure to DNAse I, owing to cleavage into pieces too small to hybridise efficiently under these conditions. The refined technique exemplified in Stalder et al (323) utilised analysis of DNAse sensitivity by Southern blotting type experiments where mild DNAse digestion was followed by complete restriction enzyme cleavage. Preferential cleavage could then be observed as loss of a characteristic hybridisation band. This assay is far more sensitive as a single double strand scission by DNAse will cause loss of the characteristic sized band.

Under these conditions a closer examination of the switch from embryonic β globin gene expression to adult gene expression was undertaken. The blot assay revealed a number of new distinctions. Firstly, in erythroblasts expressing the adult gene at 12 days of foetal development, the embryonic genes were found still to be relatively sensitive to DNAse I digestion. This degree of sensitivity however was not as great on the fully active adult genes. With increasing developmental stage this difference becomes more pronounced. By contrast, the ovalbumin gene is at all times more resistant than either of the β globin genes at all stages of development in red blood cells.

A further surprise though was the finding that the adult β genes were equally digestible with DNAse I as the embryonic β genes in embryonic cells expressing only the latter. Thus it appeared that the adult genes may have been selected for expression prior to the event and existed in a "pre-activated" state. More recently however, it appears that a more prosaic consideration may be responsible that is in the chicken genome unlike the mammalian β-like globin gene cluster the orientation of the genes does not follow their order of expression during development but the adult genes appear to be flanked on either side by the embryonic (267). Therefore a purely structural consideration might constrain the adult genes in a DNAse sensitive configuration when both the

embryonic genes are being expressed. This interpretation has been reinforced by findings in a subsequent study by Stalder et al. (324). Here using probes from both coding regions, and flanking regions differing degrees of sensitivity to DNAse could be observed. The coding region was exceptionally sensitive however quite distant regions from the actual structural gene regions still displayed a moderate level of sensitivity. In one case the transition from moderate sensitivity to resistance (defined by the level of digestion from non-expressed genes like ovalbumin) was mapped to a region 7 Kb upstream from one of the embryonic genes. It appears likely that all the β-like globin genes are contained within what the authors describe as a "chromosomal domain", which may correspond to a looped out configuration like that observed in Drosophila "puffs" or Xenopus lampbush chromosomes. Additional evidence for higher order domains within chromatin structure have also come from related investigations by Elgin and her group (334),(335). Again using a combination of DNAse (micrococcal) and restriction enzymes to locate a sequence specific organisation.

Another significant finding was that of hypersensitive regions of the chromatin which caused the appearance of tissue specific cleavages. These could be recognised as sub-bands on the blot where one cut was a restriction enzyme cut and the other a double strand DNAse I scission. Specific cleavage sites were observed for the α genes which were not found in nuclei from brain or in a non erythroid cell line. These hypersensitive sites were located in non coding regions and were identified on both the 3' and 5' sides of the α gene. Additionally in the β globin domain cutting sites specific to the embryonic gene and the adult gene were revealed. These were shown to be directly linked to expression of the genes as the embryonic hypersensitive site was lacking from adult gene expressing red cells and vice versa. Similar types of cleavage have been identified elsewhere by Chambon (336) about 1 Kb to the 3' side of the conalbumin gene and by Wu (337) to the 5' side of two

heat-shock loci. In this latter case these sites were sensitive to DNAse I prior to heat shock, in the former the hypersensitivity was again tissue specific and was found only in oviduct and liver where conalbumin is transcribed. Wu has suggested that this difference might relate to genes which are developmentally regulated and those like heat-shock loci which are not. Another possibility of course is that the hypersensitivity occurs for a variety of conceivably unrelated reasons, certainly no particular correlation with site relative to the affected structural genes can be observed. One interesting precedent does exist though and this is in the hypersensitivity of a region of the SV 40 genome which has been shown to be both the replication origin and the transcriptional origin (338),(339). This again raises the possibility that origins of replication may play some role in regulation of gene activity as indicated earlier (Section II.A.) from the work of Jelinek and others. It would be interesting to locate hypersensitive regions in a well mapped genome region like the human β-like globins and see if they show any relation to the Pol III transcript regions of Duncan _et al_ (172) or conversely to obtain some more detailed sequence information on some of the already identified hypersensitive chromatin sites. Possibly these regions may correlate with the "remote control" loci identified by surrogate genetic techniques which only appear to function _in vivo_ and not _in vitro_ (291).

These results beg the question as to what confers preferential DNAse I sensitivity on regions of expressed chromatin. One possible explanation is provided by the observation that DNA and histones were organised into nucleosome or nu body structures (340),(341). Such an interpretation was reinforced by the observation that no similar bodies can be identified at the highly transcribed ribosomal RNA genes of Xenopus (342). Less rapidly transcribed genes however have displayed the presence of nucleosomes in the electron microscope and further nuclease studies have indicated that micrococcal or staphlococcal

nuclease generate characteristic nucleosome associated "ladders" from active genes. These ladders form due to the lower susceptibility to cleavage of the nucleosome associated "core" DNA relative to the "linker" region (322),(343).

Recently however, Weisbrod and Weintraub (325) have identified two nuclear proteins which could be eluted from chick chromatin under fairly low salt conditions. DNAse I digestion of trout testis chromatin has also been found to release a very similar protein (344), as did digestion of heat shock puffs (345). These proteins on closer examination were found to correspond to proteins 14 and 17 of the "high mobility group" of Johns et al (346) or HMG proteins as they are now referred to. HMG14 and 17 have to date been shown to have the unique property of conferring DNAse I sensitivity on the nucleosomes to which they bind. The most surprising feature of this interaction has been the ability of the HMGs to reconstitute with depeleted chromatin to restore DNAse sensitivity, which is lost when they are eluted. This interaction is quite specific, only sequences sensitive before elution, regain it on reconstitution (325). The implication from this degree of specificity of binding undoubtedly is that active nucleosomes possess some other character which enables the HMG proteins to recognise them. In a closer study of nucleosome proteins, isolated from active and inactive chromatin by using an HMG affinity column (326). No differences in constitution could be observed although many properties were tested. This study however did demonstrate that in the nucleosome core, histone stoichiometry was maintained in both active and inactive chromatin, thus HMG14 and 17 do not replace any of these histones within the nucleosome. In particular, the somewhat vexed question of histone acetylation (see Ref. 347 for review) and gene activity was approached. The results were inconclusive for whilst growth in butyrate caused an increase of total

binding to the HMG column the bound fraction actually contained less acetylation than the unbound fraction. In control cells electrophoretic analysis of histones from the bound and unbound fractions displayed no significant differences in acetylation. Therefore it appears that whilst specific acetylation cannot be ruled out, the bulk acetylation of nucleosomes does not confer specificity of HMG binding and therefore DNAse sensitivity.

One observation of particular note from this study was the relationship of HMG binding to the moderately sensitive flanking region of chromatin around the active globin genes. An interesting correlation was found with true DNAse I sensitivity and the presence of HMG14 and 17, the moderately sensitive region on the other hand were devoid of nucleosomes of this sort. Thus it appears that this degree of sensitivity is mediated by other factors, possibly higher order chromosome structures. Like the DNAse I sensitivity itself however the HMG distribution seems not to confine itself purely to transcribed regions of chromatin for both the spacer region between the two α-globin genes as well as the genes themselves were associated with HMGs.

If the histones of the nucleosome core do not contain the specificity for reconstituting HMG14 and 17, then possibly the DNA with which the nucleosomes are associated might have this property. One interesting observation supporting such an interpretation was that free DNA, whilst having a lower affinity for HMG14 and 17 than active but HMG depleted nucleosomes, actually has a greater affinity for the two HMG proteins than do nucleosomes from inactive chromatin (325).

A number of laboratories have now identified a rather more persuasive correlation however. Restriction endonucleases which cut at sites containing the CpG moiety show varying abilities to cut when the C residue is methylated. In this way the low representation of CpG in eukaryotic genomes has recently been reaffirmed (348). Closer examination using specific

sequence probes (349),(350),(351) have shown interesting variability in some CpG sites in DNA of different avian or mammalian tissues. For example, methylation of virtually all these variable sites occurs in sperm. However in oviduct certain sites around and within the ovalbumin gene lack methylation whilst in erythrocytes these sites are methylated. Despite the fact that a number of inconsistencies exist a general trend towards undermethylation at variable sites seems to correlate with tissue specific modes of gene expression. For example in the oviduct ovalbumin ovomucoid and conalbumin all show undermethylation at such restriction sites, whereas only conalbumin displays undermethylation in liver, where it, unlike the other two egg white proteins, is also expressed (349). Some interesting exceptions exist, one site near the ovalbumin gene is found to be undermethylated in erythrocytes but not in oviduct and similar but reversed relationships exist for the adult β globin gene (349).

The most impressive relationship though has been between undermethylation, usually flanked by fairly well methylated DNA regions, and the sensitivity of these regions to DNAse I. Chambon has observed for example that some residual methylation occurs in oviduct, this tissue is only in fact 80% tubular gland cells and thus only 80% of cells express ovalbumin. This gene shows a corresponding 20% of resistance to DNAse I digestion (336). Conalbumin similarly demonstrates a close relationship between undermethylation at variable sites and DNAse sensitivity. Particularly important is that most of the exceptional cases of methylation or undermethylation lie in regions which do not display DNAse sensitivity. A very similar picture for the regions surrounding the chick α-globin genes has come from Weintraub (352) who has shown that a very close identity in map regions affected occurred when the α-gene region was subjected to DNAse I digestion, HMG14 and 17 binding and methylation sensitive restriction studies. Thus it would seem that all these features are

combined in some way to define regions or domains of the chromosome which are potentially active in transcription of the enclosed structural gene sequences. This latter point cannot be stressed enough as these parameters do not show any direct relationship to transcriptional activity. Thus at least one other parameter must determine whether such potentially active genes are indeed transcribed. Methylation might have seemed an eminently suitable candidate as in prokaryotic systems methylation at a single base has been found to drastically alter the affinity of such regions for specific binding proteins. Such a relationship has not as yet been identified and in the case of ovalbumin at least Chambon and his colleagues have found no change in any of the variable sites around the ovalbumin gene during hormonal stimulation (336). Furthermore it has been indicated from study in the chick globin system that undermethylation may precede the establishment of a DNAse sensitive structure (353). One enticing possibility is hinted at from this connection and that comes from the finding that to date although methylases of DNA have been identified no demethylases have been found. For regions of undermethylation to be established therefore a new round of cell division would have to occur. A number of previous investigations have related commitment to differentiation along a given pathway to the necessity to pass through a replication cycle. There is also recent evidence that the switch to late gene expression in adenovirus requires DNA replication per se as an obligatory pre-requisite (354). If such a relationship does exist, then a likely candidate for regions of DNA able to direct the establishment of new undermethylated chromosomal domains, would naturally be replication origins. These regions have been referred to a number of times in this discussion and it would seem an unlikely coincidence if they did not play some role in gene expression.

If the initiation of transcription cannot be fully explained by the aforementioned parameters then other factors must be involved. By virtue of the knowledge acquired from

prokaryotes and possibly now also from the 5S gene examples, the hunt for DNA binding proteins capable of recognising sequence specific signals has been a popular area of investigation for many years. Two recent examples indicate that this supposition may not be unfounded. Gehring and his group have identified a DNA binding protein extracted from Drosophila eggs which has been found to bind a single specific sequence from a genomic library (355). This protein represents an extremely small fraction of the total DNA binding protein and could conceivably function in the regulation of gene expression. The other report concerns a developmental mutant of cartilage function in the chick, which seems to be caused by a mutant form of chromatin associated protein which is normally only observed during the precartilage stage of development. The normal protein is present in the chromatin of heterozygotes, however the mutant form is not and thus may be unable to bind the chromatin successfully. Correlating with the loss of this binding in the mutants homozygotes was the premature appearance of a low molecular weight protein only normally associated with differentiated cartilage cells (356).

Such evidence of course is only circumstantial and merely sets the precedent for future investigation. The importance of correlating biochemical data with genetic information is once again emphasised and thus the possibility of investigating some of the many Drosophila developmental and homeotic mutants using the new sophisticated techniques could eventually prove most illuminating. Work of this sort is being instigated now by Hogness. Nevertheless, the ability to look closely at defined regions of DNA perhaps assembled into mini-chromosomes utilising more highly specified <u>in vitro</u> systems should enable critical appraisal of some of these DNA protein interactions. The hormone receptor complex is a particularly cogent example of a protein which

could be investigated in this way both *in vitro* and by gene transfer *in vivo*.

Another popular current line of investigation relates to nucleosomes. Not any longer their extreme presence or absence but more subtle variations like phasing which may be able to alternately cover or expose regions of significant sequence organisation. In this context, it is notable that the hypersensitive SV 40 region around the replication/ transcription origin appears to be nucleosome free (357) and the important regulatory sites around the 5S gene seem to be similarly arranged (358). As yet nuclease hyper-sensitivity is the only direct link we have with active transcription so this may have important repercussions. Changes in nucleosome phasing have been observed during development (343),(359) and the intriguing possibility that regulatory proteins may interact with multiple sequence elements brought into register by nucleosome phasing arrangements has been proposed (360).

IV. <u>POST-TRANSCRIPTIONAL CONTROL</u>

When previously considering evidence for post-transcriptional controls during differentiation and development I have made no distinction between the various levels at which such controls might be instigated. A somewhat crucial question with regard to post-transcriptional control is whether controls of this nature are exercised in the cytoplasm by messenger stabilities or translational control, or whether events in the nucleus are able to regulate the flow of message into the cytoplasm. A possible candidate here would be differential rates of splicing. The role of splicing will be considered below. First I shall look at translational control and the role of messenger stabilities; in regard to the latter some evidence relates to how turnover may be controlled.

IV.A. <u>Translational control</u>

The control of mRNA translation is an extremely versatile stage at which to regulate protein synthesis in response to fluctuating environmental conditions. The response can be widespread, coordinated, and furthermore extremely rapid. Moreover recovery from inhibition of protein synthesis could be accomplished very rapidly assuming m-RNA molecules are still available for translation. Possibly the best example of translational control is that of the mature unfertilised egg cell. Costantini <u>et al</u>. (161) have demonstrated the unique nature of stored maternal messenger from both sea urchin and Xenopus. The breakdown of nuclear-cytoplasmic membranes in these cells means that this control may be exercised not purely in a translational control but as a maturation arrest of m-RNA molecules. The structure of egg RNA is very similar to that indicated from more indirect means for nuclear HnRNA. Thus the egg may exercise its translational control at the processing stage. Certainly there is evidence that the egg can translate competent m-RNA. Globin m-RNA for example has been translated in a mouse ovum (361).

One of the clearest examples of translational control in a coordinated response to external stimuli is in the control of protein synthesis in fibroblasts deprived of a surface on which to grow (362),(363). This highlights in fact a number of coordinated multilevel changes. Messenger production reduces five fold, although HnRNA synthesis remained normal. Protein synthesis declined (362) but m-RNA was shown to remain in the cytoplasm (363). No major sequence changes occurred but m-RNA extracted from such cells cannot translate efficiently in cell free systems. This RNA although normally labile turned over much more slowly. One possible implication from this work is a link between translation and stability such that active translation reduces the stability of the message. Other systems though have indicated an opposite interpretation that non-polysomal m-RNAs are in fact being turned over. The arrest in translation of messenger in suspended fibroblasts appears not to stem from an initiation defect as during recovery there is a dramatic switch to large polysomes as messengers are translated once more (363).

The previous examples illustrate a general translational control mechanism however some specific cases have already been referred to. (Section I.B.3). Precise mechanisms for how such controls are exercised are unknown, some potential control RNAs have been identified (364),(365) but as yet no clear picture has emerged.

IV.B. Messenger stability - role of Poly A

The widely different abundances of messenger RNAs in the cytoplasm of a number of cell types has already been highlighted. The fact that nuclear RNA abundances vary less than cytoplasmic suggests that relative cytoplasmic stabilities may play an important role in determining abundance. One situation, the reticulocyte has shown how the half life of different messengers can play a dominant role. How does this apply to nucleated cells? Studies by Penman and colleagues (366) and by Perry (367)(368) show a somewhat clouded picture. Penman's group compared the reassociation kinetics of cDNA prepared from steady state and actinomycin blocked cells. The results demonstrated in Drosophila cells a striking correlation between abundance and stability, such that blocked cells showed an enrichment for low complexity prevalent messenger species. Equally in mouse L cells Meyuhas and Perry (368) were able to show a similar correlation. This was improved by making fractionated cDNA probes. Interestingly this study revealed that abundance, low complexity, and small messenger size generally correlate well. Thus, small messengers were both stable and abundant but large messengers tended to be rare and part of the complex class. This relationship was of course only a broad generalisation but suggests that turnover could be related to both random events, mediated by target size and specific mechanisms which would allow larger m-RNAs like ovalbumin to accumulate.

However, HeLa cells do not seem to show a similar relationship (366). Perry has suggested this might be due to the comparison of steady state and pulse labelled material rather than long term chased and pulsed RNA. The presence of a fairly large turnover in the steady state population could mask the stability/abundance relationship. It would

seem that this point requires some further investigation most likely the use of cloned probes such as in Darnell's study of CHO cells (155) could help to clarify the interrelationships of transcription, processing and decay rates. The evidence from abundant, specific probes has certainly given support to the idea that stability plays an important role in determining cytoplasmic abundance.

In trying to determine what factors might control stability one aspect repeatedly emerges but an unequivocal answer has remained elusive. This has been the role of polyadenylation of m-RNAs and stability. Two features lend support to a connection firstly, the rapidly degraded m-RNAs of prokaryotes lack poly A, and secondly the histone m-RNA of eukaryotes similarly lacks poly A and also shows a short half-life (369).

During early development however in sea urchins and in Xenopus most of the histone is made from stored, maternal messenger which must therefore be stable. It is logical therefore to ask the question whether this histone messenger is polyadenylated. Investigations in unfertilised eggs and oocytes have indicated a higher proportion of poly A(+) histone messenger (152), however it seems that the three hours half-life of the m-RNA observed during cleavage would just be sufficient during oogenesis to supply the needs of the embryo (369). In analysing the switch from a low to high level system of histone m-RNA when Xenopus oocytes become mature eggs, Ruderman et al. (152) concluded that a relationship between stability and translation existed. This synthetic rate change of 50 fold is not determined by changes in total m-RNA content. However a change in the distribution of poly A(+) and poly A(-) histone m-RNA does occur; at all times active m-RNA for histone was located in the poly A(-) fraction as assayed by in vitro translation. Thus it appeared that stability was linked to translation, only translated messengers being degraded, the rapid switch in synthetic rate being determined by deadenylation of m-RNA

thereby establishing a link between the reduced stability and poly A(-) status. No good evidence exists elsewhere that poly A(-) m-RNA translates or initiates better than poly A(+) although it is notable that the poly A(-) fraction of cells is only found associated with polysomes whereas poly A(+) messengers are found as free mRNPs as well (370). Thus deadenylation could occur during sequestration into the polysome fraction.

Other evidence linking polyadenylation to stability has come from correlations between decay rates of m-RNA and poly A tail shortening (371),(372). However in other systems no relationship could be found (373). Nevertheless, newly synthesised molecules do seem to have larger poly A regions than those that have "aged" in the cytoplasm.

The most direct evidence on the role of polyadenylation in determining stability has been acquired by Huez and his coworkers (374),(375),(376). Globin m-RNA from mammals when injected into Xenopus oocytes shows a marked stability and can be translated for some time. By contrast, if the message is deadenylated before injection rapid degradation ensues after a short period of functional translation. Subsequent readenylation of globin m-RNA is able to restore its stability in this system. Similarly histone m-RNA extracted from HeLa cells and injected into oocytes direct systems for a period which is then followed by a rapid decline. On addition of a stretch of 40-50 Adenylate residues the injected RNA was stabilised such that histone synthesis was still detectable 43 hours later. Poly A not covalently attached to the histone was unable to bring about this stabilisation.

Nevertheless when a similar study was attempted with fibroblast interferon removal of the poly A tract usually associated with this m-RNA did not affect its stability in the Xenopus oocyte (377). Unless the oocyte can recognise

interferon m-RNA and readenylate it in a way not possible for the other messengers, the connection between poly A and stability cannot be absolute. Here of course the problems associated with employing heterologous systems could be manifold. Under these conditions for example globin m-RNA seems more stable than in reticulocytes, although in these cells in fact it may have been destabilised (111). The relationship between polyadenylation and stability may remain obscure until a naturally occurring change in stability similar to that observed for globin m-RNA in Friend cells (111), (112), can be shown to derive from some specific change such as a deadenylation event.

Polyadenylation may in fact have a quite different role. Darnell for example (378),(379) has propounded the idea that poly A addition selects out sequences for export to the cytoplasm from those to be turned over within the nucleus. This work on transcription in adenovirus has shown that poly A addition is very rapid and more interestingly although termination of the polymerase always occurs at the same place, the concentration of different messenger transcripts could relate to the choice of polyadenylation site.

IV.C. Role of splicing

Of rather more interest recently though has been the role of that originally novel and perplexing observation that eukaryotic genes contain interruptions in the coding sequence. Study in a number of viral systems soon revealed that m-RNAs were maturation products of higher molecular weight precursors, which had the intronic elements removed by RNA/RNA splicing (227),(228),(229). Since these observations were soon confirmed for cellular genes, the possibility that splicing may play a central role in determining gene expression has been pursued with much vigour. Splicing could be thought to control gene expression post-transcriptionally in a variety of ways. Firstly, qualitative regulations could be made on the basis of splice/no splice decisions or quantitative modulations could be utilised in splice choice determinations like the alternative splicing modes for the two T antigens in SV 40 (380),(381). Alternatively quantitative variations could be instigated by different processing rates for different messengers, perhaps dictated by variation in splice sequences, showing different affinities for the splicing machinery.

To evaluate such possibilities the mechanisms of splicing have come in for some close investigation. At first the possibility of studying splicing in yeast t-RNAs was extremely attractive due to the ability to obtain mutants (382),(383) and genetic in vitro splicing systems. Unfortunately it soon became clear that factors governing the splicing of t-RNAs were different from those pertaining to pol II transcripts. A number of investigators have alternately tried to invoke secondary structure models governing the processing of structural gene transcripts (384),(385). No satisfactory relationship has yet been found which can unify splicing of m-RNAs from widely diverse species into a single mechanism based on such considerations. Unexpectedly perhaps, the only durable interpretation of obligatory splicing signals has come from Chambon's group

and their studies on intron/exon sequence junctions (286), (289). The only inviolable rule seems to be the GT/AG dinucleotide conservation observed at these regions. Such considerations make the mechanism of splicing obscure in the extreme but the model proposed by Steitz and her colleagues (177),(178) involving the role of the U1 snRNA is at least attractive if not wholly convincing; problems of stabilising the RNA/RNA duplex with such short homologies seem to exist (386). However, study of U1 related sequences within introns of the gene has revealed not only that multiple splice donor sequences exist but that they seem also to function _in vivo_ (387). A similar pattern seems to occur in the 15S globin m-RNA precursor (388).

Some of the stability problems involved in U1 association with splicing sites could be resolved perhaps by reference to others of the family of snRNAs which exist. Utilisation of different sequence variations of the U series RNAs could be one method of varying the processing rates of different messengers in different cell types by virtue of varying degrees of homologies between the snRNA and the splice junction sequence.

The obligatory nature of splicing for messenger RNA precursors which normally contain introns has come from a number of investigators. Hamer and Leder (389) constructed a number of SV 40/globin hybrid molecules and compared the efficiency of the different genomes to generate stable globin containing m-RNA. Only molecules containing a functional splice were able to generate stable RNA. When, for example, the globin sequence was inverted such that it was read the antisense strand no functional messenger could be formed. The fact that intronic sequences differ so much in content and length has indicated that the junction sequences are the crucial signals in achieving successful splicing. The previous authors for example have shown, that only 18 nucleotides of the large globin intron including the junction sequences were required to generate a functional splice.

Deletions in the intron/exon junction can be crucial to expression however (390),(391). In the latter case removal of an SV 40 splice junction in vitro and subsequent infection caused the destabilisation of a normally stable messenger. Other data on SV 40 from Weisman and his colleagues (246) indicated that preservation of nine nucleotides on the 3' side of the SV 40 leader splice junction are able to direct the splicing of the transcript, if however this is reduced to seven nucleotides large amounts of unspliced m-RNA are formed.

The extremely close relationship between splicing and m-RNA formation would imply this may be a process where major regulations of gene expression could occur. Splicing choice differences within the SV 40 T-antigen messenger precursor obviously is instrumental in determining the relative levels of large T-antigen and small T-antigens (380), but no evidence exists as yet that this ratio can be modulated during development. Whether higher organisms use this type of genetic organisation typified by the viral systems is also highly debatable. Some evidence from the sea-urchin (392) suggests that four differently sized transcripts can be formed from a single gene. Whether the case for the same gene product is not known. Such molecules could of course be generated by terminator read through to give heterogeneity in size such as that observed for D.H.F.R. (281).

A great deal of circumstantial evidence now points to the use of different splicing choices within the immunoglobulin system to generate a number of alternative messenger RNAs. Hood and his coworkers have highlighted two processes where splicing may well dictate messenger identity (393),(394). In the heavy chain Ig M immunoglobulins, molecules secreted into the medium possess a different 3' terminus from those which are implanted into the cell membrane. Both these forms of the immunoglobulin display the same idiotype and thus must have the same variable region. This mechanism cannot

be the same as the heavy chain switch of content regions (achieved by DNA deletions) as both Ig M forms are co-expressed. Mapping of this genomic region has indicated the presence of an additional coding segment located distal to the normal 3' end of the Ig M constant region which is able to code for the C-terminal region of the membrane bound form. This was indicated by the high content of hydrophobic amino acid coded by this region; a feature typical of membrane associated proteins. The model proposed therefore is that splicing choice determines whether the complete transcript, which includes the membrane exon, is processed using the proximal or distal C-terminal domain exons; to generate secretory or bound forms respectively. A somewhat novel arrangement is involved in producing the membrane associated form as the donor sequence for the last splice is actually part of the secretory form exon and thus lies within a coding segment.

A similar system seems to have evolved for Ig D expression which enables two types of constant regions to be co-expressed with a single idiotype. Blattner and his colleagues (395),(396) have shown that the Ig D constant region is coded for only 2 Kb downstream from the μ constant region contrasting markedly to the widely separated remaining heavy chain C regions involved in the heavy chain switch. These authors have similarly suggested that Ig D can be expressed by different splicing arrangements of a transcript which includes both μ and k constant regions.

Both these groups though have pointed out that this splicing choice could be a passive consequence of changes in poly A addition site or polymerase termination. Use of alternative poly A addition sites is a phenomenon highlighted earlier and thus may make this the more likely mechanism. In this case read through to the distal terminus could generate the alternative messenger form and thus the primary transcripts would be of variable sizes. This would be one possible way of distinguishing between these two processing pathways.

However widespread the use of splicing as a vehicle for controlling gene expression turns out to be, itsperturbation has already been shown to cause a genetic defect. Faulty splicing patterns seem to be responsible for two known cases of β thallassaemia (397). It is notable that these consist both of a qualitative and a quantitative change in processing in each case respectively. This at least may enable some of the mechanics of the splicing system to be elucidated.

The most remarkable instance of splicing linked gene expression has come from genetic and molecular biological investigation of mutants of yeast cytochrome b. This gene has been shown to consist of five exonic regions and four introns (398). One particular class of mutants comprising a single complementation group have been identified as mapping within the second intron; these are referred to as box 3 mutants (399). Analysis of certain box 3 mutants has revealed that they prevent splicing of the cytochrome b m-RNA precursor which thus accumulates. Moreover these mutants can be complemented in trans, which suggests the presence of a diffusible gene product instrumental in splicing. Sequence analysis of wild type and mutants, particularly of chain terminator type has demonstrated that the second intron (box 3 region) possesses an open reading frame.

This region seems to be translated, as nonsense mutants accumulate novel polypeptides which demonstrate colinearity with the gene sequence. Thus it appears that the unspliced box region transcript is first translated into a protein which is then able to function as a splicing enzyme called a"maturase" by the authors, which removes the second intron in which it is encoded. Removal of this intron aligns exonic regions of the cytochrome b gene transcript which can then be translated into functional cytochrome. The existence of mutants which make inactive cytochrome b

but are able to compliment box 3 mutants argues strongly for this interpretation. One advantageous spin off from this organisation is that cytochrome b systems should be self regulatory. The concentration of the maturase will determine the efficiency of splicing and thus the concentration of cytochrome. It is clear however that a small quantity of maturase is sufficient to splice many m-RNA precursors.

A particularly relevant factor is that the box 3 coded maturase looks to be pleiotropic as such mutants make no cytochrome oxidase subunit 1 either, so this could be identifying a fairly general mechanism. On the other hand it seems that the box 7 mutants of cytochrome b could also function as maturases for one of the additional introns of this gene as they are complemented by and complement box 3 mutants; they should therefore code for a different maturase. There is also evidence that other split genes in yeast mitochondria may encode maturases within introns, the oxi 3 gene being a likely candidate.

The generality of this type of regulation in a wider context is dubious but not totally implausible. Although it is simple for mitochondria to translate unspliced messenger precursors there is certainly no precedent for this in nuclear genes. However as splicing is obviously efficient in terms of splices per maturase molecule, concentrations at steady state could be vanishingly small. Nevertheless, it is revealing that one splice is performed in generating the maturase, so some other, perhaps less specific system looks to be involved as well. Finally, the intron/exon boundary sequences whilst showing some features common to those of nuclear transcripts do not appear to follow identical splicing rules and thus may be mediated by a separate processing machinery. As mitochondria would also appear exceptionally to utilise a variant genetic code however, these factors may not be mutually exclusive.

MATERIALS AND METHODS

Production of Reticulocytes

Reticulocytes were obtained originally from the blood of mice made anaemic by intraperitoneal injection of phenylhydrazine hydrochloride (0.1 ml; 2.5%). Cells were harvested after filtration through glass wool from mice bled on the fifth day after injection by centrifugation (MSE Mistral 4 x 250 ml 15 mins. 1500 rpm). The white cell fraction was removed by aspiration from the top of the pellated red cell fraction. The remaining red cell fraction was resuspended in Hanks B.S.S. and the process described above repeated. The cells were washed in this manner once more and pelleted again before lysis.

For protocols where immature cells were required the injection regime was modified as follows: acetyl phenylhydrazine was injected (0.2 ml 2.5% in P.B.S.) as before but the procedure was repeated on the second day after injection. Cells were then harvested on the subsequent day.

Reticulocytes were then purified from this blood by centrifugation through Ficoll/Triosil (6%/16.7%). The density of which was calculated to allow mature red blood cells and reticulocytes to pellet but leave the lymphocytes at the blood/Ficoll interphase. Granulocytes were distributed within the Ficoll/Triosil layer but did not pellet.

Reticulocytosis was estimated by supra-vital staining with brilliant Cresyl blue. Ten drops of packed cells were diluted with ten drops of B.S.S. and an equal volume of stain added. The mixture was incubated for 10 mins at 37°C and then used to prepare microscope slides. The stained preparation was diluted to give approximately 100 cells per field when viewed under oil immersion. The reticulocyte count was estimated from analysis of ten random fields and is expressed as the percentage of non nucleated cells found to be characteristically stained. (See Fig. 12).

White cell contamination was estimated from May-Grunewald Geimsa staining (kindly performed by S. Hughes) and was never greater than 0.1%.

Isolation of messenger RNA

RNA was prepared from reticulocyte rich blood by repeated phenol/chloroform extraction in ANE/SDS buffer as described (62). Where polysomes were prepared cells were lysed in 1 mM $MgCl_2$ and mitochondria and cellular debris pelletted at 8000 g. Polysomes were then pelletted by centrifugation for 1-2 hr at 120,000 g. The polysome pellets were resuspended in ANE/SDS and the RNA obtained by repeated phenol/chloroform extraction.

RNA from younger reticulocytes was prepared by modification of the methods described in (403) and (404). Cells were resuspended in T.N.M. buffer (100 mM Tris HCl, pH 7.4 150 mM NaCl; 3 mM $MgCl_2$) and lysed by addition of Nonidet NP 40 to a final concentration of 0.5%. The lysed cells were then centrifuged at 800 g as above to remove macroscopic debris. The supernatant was then made 10 mM in EDTA and 1% in S.D.S. Caesium Chloride (B.R.L. ultrapure reagents) was added to a final concentration of 1.4 g/ml and the solution centrifuged at 10,000 g for 30 min at 20°C (Sorvall HB-4 rotor). The supranatant was removed by passing a syringe needle through the protein SDS pellicle as indicated (403)diluted with 2-3 vols of distilled deionised water and precipitated with 7-8 vols of ethanol. The precipitate was recovered by centrifugation at 10,000 g (10'0°C) the ethanol decanted, and the pellet dried and then resuspended in T.N.M./EDTA/SDS as described above. The solution was then re-extracted with caesium chloride and ethanol precipitated.

The total RNA prepared by either method was then fractionated by affinity chromatography on Poly-U-Sepharose (405)or where small yields were expected by oligo-dT-cellulose (406).

The poly A(+) RNA was then desalted on Sephadex G-50 columns,the required fractions pooled and freeze dried.

For gradient fractionation the RNA was bound once to an oligo-dT-cellulose affinity column, precipitated and

loaded onto the top of 5-20% sucrose gradients carrying a 60% sucrose pad at the base. Centrifugation conditions were as detailed in appropriate figure legends.

Analysis of affinity chromatography was performed by electrophoresis on 2.6% acrylamide disc gels. The gels were run in buffer containing 36 mM Tris base, 30 mM NaH_2PO_4, 1 mM EDTA, 0.2% SDS, using bromophenol blue as tracer dye run in a parallel gel. After electrophoresis the gels were scanned using a Joyce-Loebel gel scanner to detect the presence of RNA by ultraviolet absorbance.

Density Gradient Centrifugation

Fractionation of messenger RNA was performed on 5-20% sucrose gradients in N.E.T.S. buffer (100 mM NaCl, 10 mM Tris - HCl pH 7.4, 1 mM EDTA 0.1% SDS). The gradients were formed by underlaying progressively denser sucrose solution in a series of defined steps and allowing the gradients to form by overnight diffusion. Gradients were used immediately after this. The quality of the gradients was ascertained (sometimes from a parallel gradient) by measurement of refractive index using a refractometer. Unloading was by upward displacement using an M.S.E. unloading cap. Fluorochemical (FC43, 3M Co) was employed, owing to its chemical inertness and high density, for displacement of the gradient. The presence of unlabelled RNA was monitored with an L.K.B. Uvicord spectrophotometer and labelled RNA by liquid scintillation counting.

For estimating the size of c-DNA probes, alkaline sucrose gradients (0.9 M NaCl, 0.1N NaOH) 4-11% sucrose were employed; prepared as described above. Analysis was performed on 12 ml gradients run in an IEC B-60 with a 6 x 14 rotor. Precise conditions of centrifugation are described in the relevant figures. S-values were calculated by computer from distances migrated.

Solution Hybridisations

These were performed in sealed glass capillaries at $42^{o}C$ in buffer containing 0.5 M NaCl, 100 mM Hepes pH 7.0, 5 mM EDTA and 50% v/v formamide (Fluka). Appropriate volumes of mRNA, cDNA and E.coli carrier t-RNA were mixed from solutions in distilled water and lyophilised to dryness. The pellets were resuspended in a suitable volume of hybridisation buffer (0.5 - 5.0 µl) and sealed inside glass capillaries. The reaction mixes were boiled for 5 mins and quenched in ice before beginning the hybridisation. The extent of hybridisation was measured using S-1 nuclease. The contents of each capillary were expelled using 250 µl S-1 buffer (0.07 M Na acetate, 0.14 M NaCl, 5 mM $ZnSO_4$, pH 4.5) supplemented with 14 µg/ml denatured calf thymus DNA. To the mixture was added 100 µl S-1 enzyme (Sigma) (20 units) and the reaction incubated at $37^{o}C$ for 1 hr. To estimate the total number of counts 100 µl was removed to a counting vial. To estimate the acid soluble material 200 µl were transferred to a tube containing 50 µl 1 mg/ml BSA chilled on ice. To this was added 50 µl of PCA (6N) and the tubes were incubated on ice for 20 mins. The precipitable material was removed by centrifuging at 2500 rpm for 15 mins. 200 µl of the supernatant was used for scintillation counting. The percentage of hybridised material (H) was calculated from $H\% = 100 \times (1 - \frac{0.85S}{T})$ where S and T are the acid soluble and total counts respectively. The factor of 0.85 is calculated from the relative volumes counted, corrected for differential efficiency of counting in acid media.

Synthesis of cDNA probes

For solution hybridisation cDNA was synthesised by a method adapted from Baltimore (407). The reaction mix was as follows: 25 mM Tris pH 8.2, 50 mM Mg acetate, 10 mM D.T.T., 200 mM NaCl, Oligo dT_{12-18} 200 µg/ml (P.L.Biochemicals), Actinomycin D 75 µg/ml, unlabelled dNTP's at 2 mM and labelled dNTP at 0.2 mM. Template concentration was 10 µg/ml and the reaction contained 100 µ/ml reverse transcriptase. The reaction mix was incubated for 2 hr at 37°C incorporation stopped by addition of EDTA to 0.01 M and SDS to 1%. The cDNA was separated from unincorporated nucleotides by Sephadex G-50 chromatography. The excluded material was made 50 mM in NaOH, 25 mM in E.D.T.A. and the template RNA hydrolysed at 37°C for 4 hrs. The solution was then neutralised in the presence of 50 mM Tris and phenol red. The c-DNA was finally desalted on Sephadex G-50 and lyophilised. The c-DNA was stored at -20°C in sterile water.

Probes synthesised as above had specific activities in the range 2-8 x 10^6 cpm/µg using ^{3}H dCTP (Radiochemical Centre). Their sizes were estimated by centrifugation in alkaline sucrose gradients. (Fig. 6a/b).

For Southern blotting experiments probes were synthesised to approximately 6 x 10^8 cpm/µg. ^{32}P dCTP (400 ci/mmol) being added to a typical reaction mix containing 10 µg/ml m-RNA, 0.5 mM dNTP's (Boehringer Corp. Ltd.), 40 µg/ml oligo dT_{12-18}, 100 µCidCTP^{32}P, 140 mM KCl, 50 mM Tris pH 8.0, 10 mM $MgCl_2$ 30 mM β-mercaptoethanol and 5-10µ of A.M.V. reverse transcriptase. This mix was incubated for 1 hr at 42°C and the template then hydrolysed as above but at 50°C for 30 min. The probe was then isolated on G-50 Sephadex and concentrated if necessary by lyophilisation.

Synthesis of cDNA for cloning

Originally this was performed as described by Humphries et al (404). The first strand synthesis was performed as for solution hybridisation probes. The second strand was synthesised in a reaction mix containing 30 mM Tris HCl pH 7.5, 4 mM $MgCl_2$, 0.5 mM β-mercaptoethanol, 1 mM dNTPs and 9 μ of Klenow fragment of DNA polymerase I (B.C.L.) per microgram cDNA template. The mix was incubated for 5 hrs at 20°C then phenol extracted, dialysed extensively against 1 mM Tris HCl pH 8 and incubated with S-1 nuclease at 5μ/ml in buffer containing 300 mM $NaCl_2$, 30 mM Na acetate, pH 4.5, 3 mM $ZnSO_4$ to remove the hairpin loop. The mixture was again phenol extracted and dialysed as above, before being concentrated by lyophilisation.

For its convenience, a modified protocol of Schimke (408) was adopted for later synthesis. First strand synthesis was as described for Southern blotting probes except that template concentration was 25 μg/ml and 7μ/μg of reverse transcriptase were used. Final specific activity was adjusted in this case to 10^6 cpm/μg. After incubation with reverse transcriptase the reaction was stopped by boiling for 3' and rapid quenching on ice. The mix was then diluted with an equal volume of 200 mM Hepes pH 7.0 containing 0.5 mM dNTP's. To this was added 35 μ of DNA polymerase I Klenow fragment and the mix was incubated for 2 hr at 15°C. The reaction was stopped by addition of SDS to 0.1% and the unincorporated nucleotides removed by chromatography on Sephadex G-50 in 20 mM NaCl. The excluded material was treated as described previously to remove the hairpin. The products of these reactions were analysed by alkaline sucrose gradient centrifugation in 4-11% gradients. (Fig. 15). Finally, to ensure perfectly flush ends, the c-DNA was incubated with DNA polymerase I as described for plasmid repair.

Preparation of Plasmid Vectors for Cloning

Covalent closed circular pAT 153 plasmid DNA (25 µg) was linearised using Eco RI, Bam HI or Sal I and the digests checked by analysis on 1% Agarose gels. (Fig. 18).

The 5' overhanging ends were then made flush by "filling-in" using DNA polymerase I, Klenow fragment under conditions described by Chambon ([409]). (20' at 14°C) in a reaction mix containing 30 mM Tris-HCl, pH 8.0, 4 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mM each of dNTP's and 2.5 µ/µg of Klenow fragment of polymerase I. The terminal 5' phosphates were then removed using alkaline phosphatase (Boehringer Corp. calf intestinal grade 1) (0.05 µ/µg) in 50 mM Tris (pH 8.0) at 37°C for 30 min. The reaction was stopped by addition of phenol/chloroform and extracted twice followed by ether extraction and then ethanol precipitation.

Plasmid DNA treated as above was ligated to equimolar quantities d.s. c-DNA by blunt-end ligation using T_4 DNA ligase (N.E. Biolabs.), under conditions containing 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 1 mM ATP, 50 ng of plasmid DNA, 2-8 ng of ds c-DNA and 1 µl of ligase in a total volume of 10 µl. The reaction mix was incubated at 12°C for 20 hrs when a second addition of ligase was made and the reaction allowed to proceed for a further 24 hrs. The ligation mix was then stored for future use at -20°C or kept on ice for immediate use in transformation.

Preparation of Plasmid DNA

Plasmid was initially prepared utilising the method described by Birnie ([410]). Small overnight cultures were incubated from plates or stabs and used to seed bulk cultures grown in L-Broth (1L contains 20 g Bactotryptone, 10 g yeast extract, 20 g NaCl) which were allowed to grow to late log-phase (A_{600} = 0.6) whereupon chloramphenicol was added to 0.2 mg/ml and the cultures grown for a further 17 hrs.

Cells were then harvested by centrifugation 8.5K 10' (Sorval GS3) at $4^{o}C$. The pellet was resuspended in 0.05 M Tris pH 8.0 and the cells repelleted. The cells were then lysed using 0.05 M Tris/25% sucrose and lysozyme 1 mg/ml (Sigma), by incubation at $4^{o}C$ for 5' followed by incubation at $30^{o}C$ for 15 mins after addition of EDTA to approximately 0.08 M. 0.7 vol 36% Triton-X100 was added to complete lysis. Chromosomal DNA was removed by centrifugation at 30 K for 1 hr at $20^{o}C$. The supernatant was carefully decanted to a vol of 30% NaCl added and the plasmid DNA was concentrated by precipitation at $4^{o}C$ in the presence of 10% (w/v) P.E.G. The plasmid DNA was then extracted by centrifugation (8.5 K 15' Sorval G5A) and loaded onto Casium chloride isopycnic gradients in 0.1 x SSC containing Ethidium Bromide (0.3 mg/ml). The plasmid was banded by centrifugation at 37.5 K rpm for 64 hrs (MSE 10 x 10) at $20^{o}C$. Unloading was by peristaltic pump using upward suction.

As yields were rather poor (100 µg from 1.6 L of culture), although of high purity (see Fig. 18), where large quantities were required, the alkaline extraction technique of Birnboim and Doly (411) was employed. The basis of this technique is denaturation and rapid renaturation which precipitates the chromosomal DNA but enables the plasmid to "snap-back" and thus it remains in solution. For a one litre culture such as above, the cells were resuspended in 40 mls/gm solution (25 mM Tris, pH 8.0, 50 mM Glucose, 10 mM E.D.T.A., 5 mg/ml lysozyme) and placed on ice for 30 min, before addition of 80 mls of alkaline S.D.S. (0.2 N NaOH, 1% S.D.S.). 60 mls of 3MNa acetate (pH 4.8) are then added and the solution gently mixed. Chromosomal DNA is pelleted by centrifugation at 8000 rpm $0^{o}C$ 10 min. The supernatant is then used to precipitate the plasmid DNA by addition of 320 mls of ethanol. The plasmid is pelleted and resuspended in 40 mls of 0.2 M Na acetate, 0.1 M Tris HCl, 1 mM E.D.T.A., 0.2% SDS (pH 8.0), then phenol/chloroform extracted twice,

followed by ethanol precipitation. The pellet is then RNAse treated (50 mM Tris HCl, pH 8.0, 1 mM E.D.T.A.) in 5 mls 200 µg/ml RNAse A at 37°C for 1 hr, followed by room temperature precipitation with 1.2 vols Ethanol 0.2 M Na acetate. The pellet is then resuspended in 2 mls Na acetate/Tris/ E.D.T.A./SDS and treated with Proteinase K (50 µg/ml) at 37°C for 30 min. After proteinase treatment the plasmid DNA is re-extracted with phenol/chloroform and precipitated with 3 vols of cold ethanol. The pellet is then resuspended in 0.2 M Na Acetate and finally re-precipitated at room temperature using 1.2 vols of ethanol.

Small cultures of bacteria were used (2 mls) in a scaled down version of this method for making rapid preparation of DNA for simple analysis. In this case the RNAse step is omitted but phenol/chloroform extraction was found to give improved results in restriction analysis.

As the previous methodology was somewhat tedious and involved, for preparations requiring a number of different samples the modifications of D.Isch-Horowitz (pers. comm.) were utilised. This is essentially an amalgam of the two previous methodologies. The alkaline extraction is followed as far as the neutralisation step, however the neutralising salt is Potassium acetate as opposed to Sodium acetate, which helps to pack down the chromosomal DNA as the potassium salt of S.D.S. is highly insoluble. The supernatant is then decanted through gauze, to remove any floating precipitate and the plasmid precipitated by addition of 0.6 vol of isopropanol. After centrifugation (8 K rpm 10 min 4°C Sorvall G.SA) to pellet the plasmid DNA, it was re-suspended in 13.4 mls Tris/EDTA (10 mM: 10 mM) to which was added 14.4 g CsCl and 1.4 ml 5 mg/ml Ethidium bromide. This solution was split into two equal portions and centrifuged as previously described (I.E.C. B60 10 x 10 rotor, 40 K, 42 hrs, 20°C).

Transformation

Competent cells were prepared by the Calcium/Rubidium technique. E.coli HB101 were taken from an overnight culture and seeded into a 200 ml culture in L-Broth. Growth was continued until the culture reached an optical density of 0.6 O.D. (A_{600}), when the cells were harvested by centrifugation at 4,500 rpm for 10' at 4^oC. (Sorval G5-A rotor). The bacteria were then resuspended in 100 mls of buffer containing 50 mM $CaCl_2$, 50 mM RbCl, 10 mM Tris pH 7.9 and incubated on ice for 20'. The cells were pelleted once again then gently resuspended in 20 mls of buffer as described above supplemented with 15% v/v glycerol (sterile) and 1 ml aliquots distributed into Eppendorf tubes on ice. The competent cells were stored at -70^oC for up to 4 months.

For transformation aliquots were thawed by warming at 37^oC just sufficient to melt the bacterial suspension which was mixed gently by inversion and kept on ice. 200 µl of cells were added to 50 ng of plasmid DNA (in 10 mM Tris pH 7.9, 20 mM $MgCl_2$, 10 mM $CaCl_2$) and incubated for 20 min at 0^oC, the mixture was then heat-shocked for 5 min at 37^oC when 1 ml of L-Broth was added. The suspension was further incubated for 1 hr at 37^oC with gentle shaking then plated (100 λ per plate) onto L-Agar plates containing 100 µg/ml Ampicillin.

Storage of Transformants

Recombinant colonies were transferred individually to 96 well microtitre plates using sterile toothpicks. Each well contained 120 µl of medium consisting of L-Broth supplemented with K_2HPO_4 (6.3), KH_2PO_4 (1.8), Na Citrate (0.45), $MgSO_4.7H_2O$ (0.09), $(NH_4)_2SO_4$(0.9), glycerol (44) (where the numbers in brackets are grams per litre) and 100 µg/ml Ampicillin. The microtitre plates were stored frozen at -20^oC in sealed plastic bags. Replicas of the plates could be made using Cooke transfer plates to inoculate the fresh microtitre plate. Plates were incubated overnight at 37^oC before permanent storage.

Screening of Recombinant Colonies

This was performed by modification of the standard Grunstein Hogness technique (412). Microtitre plate stocks were replica plated onto sterile nitrocellulose filters (millipore) overlaying antibiotic supplemented L-Agar petri plates, using Cooke transfer plates. This enabled transfer of all 96 colonies in one step. The transfer plate was applied to the surface of the nitrocellulose with slight pressure and then removed. The transferred spots of bacterial suspension were allowed to dry by absorption and the plates were then incubated overnight to allow colony formation. A sterile needle was used to key the filters.

Lysis was as follows: all procedures being performed by applying the nitrocellulose filters to the pre-soaked surface of Whatman 3MM paper. The filters were treated with 10 mg/ml lysozyme (Sigma Grade 1) in 10 mM Tris pH 8.0, 4% Sucrose for 5 min followed by two 5 min treatments with 0.1% Triton X100, 0.5N NaOH. After lysis the filters were neutralised with three 10 min washes on 1M Tris-HCl pH 7.6 and a final treatment with 1.5 M NaCl 0.5 M Tris HCl pH 7.6. Finally the filters were baked at 80°C for 2 hours to fix the DNA.

Baked filters were then prehybridised in a solution containing 3 x SSC (SSC is 0.15 M NaCl, 0.015 M Na Citrate), 1 x Denhardt's solution (0.2% of each polyvinyl pyrolidine, Bovine serum albumin and ficoll) for at least 3 hrs at 68°C. Hybridisations were performed in the same solution supplemented with 100 µg/ml Salmon Sperm DNA and $5 \times 10^5 - 10^7$ cpm of ^{32}P labelled cDNA, under the conditions described above. Hybridisation was for at least 16 hrs in sealed plastic bags containing approximately 4 mls of buffer. After hybridisation the filters were washed (approximately 125 ml per filter) with three changes of 0.1% SSC, 0.1% SDS at 42°C for 30 mins each wash. The filters were then allowed to air dry and subjected to autoradiography at -70°C with Dupont "Cronex" intensifying screens.

Southern Blotting

Restriction fragments were transferred to nitrocellulose filters from 2% agarose gels run in 40 mM Tris-HCl pH 7.8, 20 mM Na Acetate, 2 mM EDTA by the technique of Southern (76). The gel was treated with 0.5 M NaOH, 1.5 M NaCl for 20 mins and then neutralised with two washes of 20 min each in 3 M NaCl, 0.5 M Tris pH 7.0. The gel was then placed over two "wicking sheets" of 20 x SSC flooded 3 MM paper, standing with the 3 MM in contact with a reservoir of 20 x SSC. A nitrocellulose filter cut to size was placed over the gel, after being thoroughly soaked in buffer and any air bubbles carefully removed. The nitrocellulose was then protected by two similarly cut pieces of (Whatman No.1) filter paper over which was then placed a stack of paper towels (about 5 ins. thick) suitably weighted to gently compress the towels. After allowing the transfer of nucleic acid to proceed overnight, the filter was baked at 80°C for 2 hrs. Prehybridisation was as described for the filter hybridisation protocol but hybridisation was performed in buffer containing 20 mM Na_2PO_4 and 10% w/v Dextran Sulphate (Pharmacia) in a volume of approximately 4 mls for 16 hrs. The filter was then washed with two changes of 125 mls 2 x SSC for 15 mins each at room temperature followed by four changes of 0.1 x SSC 0.1% SDS at 65°C. After washing the filter was air-dried and autoradiographed as described above.

Nick Translation

Nick-translated probes were prepared by the method of Rigby et al (413) with some modifications. DNase I was activated as described by thawing 50 µl of a stock solution of 1 mg/ml into 0.45 ml of 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mg/ml BSA at 0°C for 2 hrs. The reaction mix contained 20 µM dNTP's (unlabelled), 50 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 133 ng/ml activated DNAse I, 100 µCi ^{32}P dCTP, 0.5 µg DNA and 0.5 units of DNA

polymerase I (Boehringer Corp. Ltd.). The mix was incubated for 1 min at room temperature before addition of polymerase when it was transferred to a 14°C water bath and incubated for 4 hours. To stop the reaction EDTA was added to 130 mM final concentration and the mix incubated at 68°C for 15 mins. The nick translated DNA was separated from unincorporated nucleotide on a 20 x 1 cm Sephadex G-50 column run in 1 x SSC buffer. The excluded material was pooled and concentrated, if necessary, by lyophilisation. Probes prepared in this fashion were of specific activity 1-2 x 10^8 cpm/µg.

Restriction Analyses

Restriction endonucleases were used in accordance with the manufacturers recommendation (B.R.L. or N.E. Biolabs). Incubations were for 1 hr at 37°C, except for Southern blotting experiments where genomic DNA was digested overnight (500 µg/ml). Restriction products were analysed on 1% or 2% agarose gels or 6% polyacrylamide run in Tris, Na Acetate, EDTA buffer (pH 7.8). DNA bands were visualised by staining the finished gel in Ethidium Bromide (1 µg/ml) and observing under ultraviolet light. Photography was by Polaroid 540 pos/neg or 54 fast pos only film; exposures varying from 0.5 - 4 min (540) and 1 - 10 secs (54).

Filter Hybridisation of Cloned Sequences

For titration of cloned sequences in cDNA of reticulocytes and other tissues plasmid DNA was bound to nitrocellulose filters (1.3 cm dia. Millipore Corp. Ltd.) as follows: the DNA was suspended in 1 x SSC such that 0.4 ml contained the requisite quantity of DNA to bind to the filter. 1 M NaOH (0.4 ml) was then added to denature the plasmid and this was incubated at room temperature for at least 30 mins.

Presoaked filters were then assembled in "swinnex" holders, fitted to a suction apparatus and syringes were

attached above to form reservoirs. The filters were then washed with 5 ml 6 x SSC by applying suction from a vacuum pump; the flow rate was adjusted such that 5 mls flowed through in approximately 4-5 min.

The plasmid DNA was then neutralised by addition of approximately 2 mls of neutralising solution (0.2 M HCl, 9 x SSC and phenol red indicator). The neutralised DNA solution was then applied to the filters as before which were then washed with a further 5 ml of 6 x SSC and baked for at least 2 hrs at 80°C to fix the DNA.

Nitrocellulose filters (1.3 cm Millipore) with bound plasmid DNA (2 µg) were prehybridised overnight in a solution containing 5 x SSC, 5 x Denhardt's solution, 50 mM Na_2SO_4, 10 g/l glycine, 250 µg/ml carrier DNA and 50% v/v formamide (BDH analar). Hybridisations were performed in 200 µl of buffer as above but without glycine or Na_2PO_4 and with 1 x Denhardt's solution and 100 µg/ml Salmon sperm DNA. Labelled cDNA probe (^{32}P specific activity approximately 6-7 x 10^6cpm/µg was added to give between 2.5 x 10^5 and 1.8 x 10^6 cpm per well. The hybridisations were carried out in 24 well microtitre plates sealed with teflon tape (3 M Co. Ltd.) placed inside similarly sealed sandwich boxes, incubated for 48 hrs at 42°C with swirling. After hybridisation, the filters were washed extensively (4 changes of 125 mls per 18 filters; 20 mins each wash) in 0.1 x SSC, 0.1% SDS at 60°C, air-dried and the hybridised material estimated by liquid scintillation counting (5 mls Packard MI 96 scintillant).

Protein Fractionation

Packed cells from approximately 30 mice were obtained as described above and 1 ml mixed with 3 mls of Ham's DF12 containing 100 μCi ^{35}S methionine. The cells were gassed and incubated for 1 hr at 37°C during which incorporation of radioactivity was approximately linear. After incubation the cells were pelleted once again and washed twice with Hank's BSS. The cells were then lysed by osmotic shock by addition of 10 mls of distilled water. The haemolysate was centrifuged to remove macroscopic debris (6000 rpm, 5', 4°C). The supernatant was added drop by drop to a vortexing solution of 200 mls acid-acetone (1.5% v/v HCl). The precipitate was allowed to settle out, the supernatant poured off and the protein concentrated by spinning at 6000 rpm for 5 min. The pellet was washed twice with copious quantities of acetone to remove any residual haem and dried under a nitrogen stream. Protein was stored at -20°C until required.

The total reticulocyte proteins were fractionated on a 2.5 x 100 cm G-100 Sephadex column run at a flow rate of 15 ml/hr as described by Burka (414). The column was equilibrated with 6 M Urea, 1% β-mercaptoethanol. 2.5 mls fractions were collected and the eluate monitored at 280 nm using an LKB uvicord. The void volume was indicated by the presence of blue dextran added to the sample. Aliquots of the fractions (0.5 ml) were used for liquid scintillation counting to detect the presence of labelled proteins.

RESULTS

I. ANALYSIS OF RETICULOCYTE M-RNA AND PROTEIN SYNTHESIS

A. Gradient Fractionation of Reticulocyte m-RNA

As documented elsewhere (39)(61)(68) the m-RNA of reticulocytes comprises some 80-90% globin m-RNAs. It was hoped therefore to make some simple one step purification which might enrich for non globin m-RNA sequences. The globin m-RNAs have already been shown to sediment with a value of 9S (37)(61), bulk messenger by contrast shows a median value of around 18S (41). The rationale of this approach was therefore to attempt a separation by molecular weight using sucrose gradients.

An initial gradient calibration experiment is shown in Fig. 2. Retic ribosomal RNA was used as markers and the position of globin m-RNA in the gradient was indicated by trace amounts of in vitro (^{125}I) labelled mouse globin m-RNA. To achieve a satisfactory degree of purification it was necessary to achieve a greater separation of the 9S and 18S peaks while at the same time preventing the 28S material from pelleting. In addition it was desired to concentrate the non globin fractions into a smaller region of the gradient to maximise recovery of what could be only small amounts of RNA. To try and facilitate this separation the basic 5% - 25% sucrose gradient was modified by inclusion of a heavy sucrose pad at the base. A variety of concentrations were tried, optimum results being obtained with a pad of 60% (w/w) sucrose. The calibration of such a gradient is shown in Fig. 3.

FIGURE 1. Polyacrylamide Gel Analysis of mouse Reticulocyte m-RNA.

Samples were loaded onto the top of 4.5% polyacrylamide gels and electrophoresed until the Bromophenol Blue marker dye had reached the bottom (8mA; 5 hrs). The gels were scanned using a Joyce-Loebel gel scanner.

a) 40 μg RNA unbound after one passage over Poly-U-Sepharose.

b) 4 μg RNA bound after two passages.

c) 4 μg RNA not bound on second passage.

Gels b) and c) were scanned at 3 x sensitivity relative to gel a).

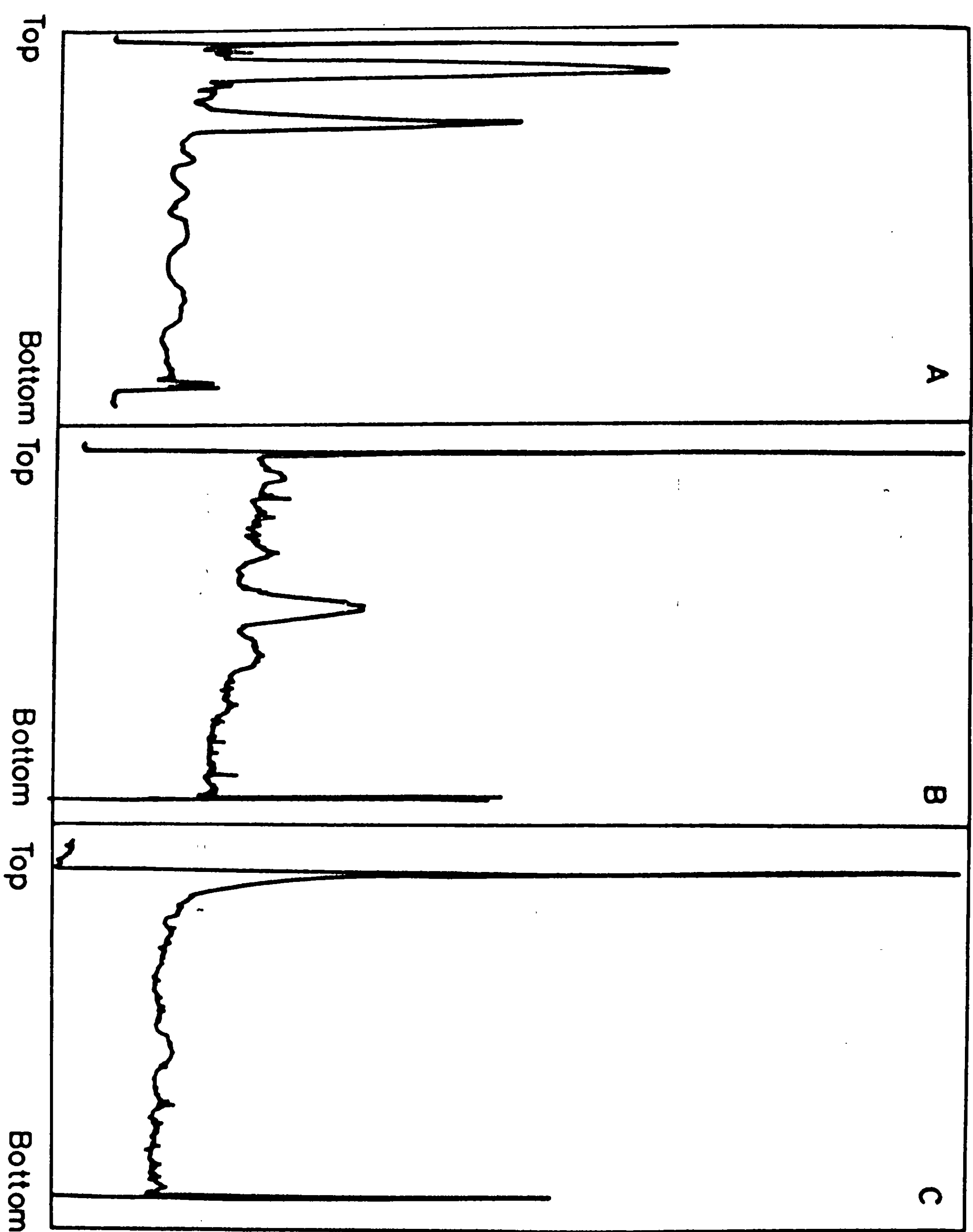
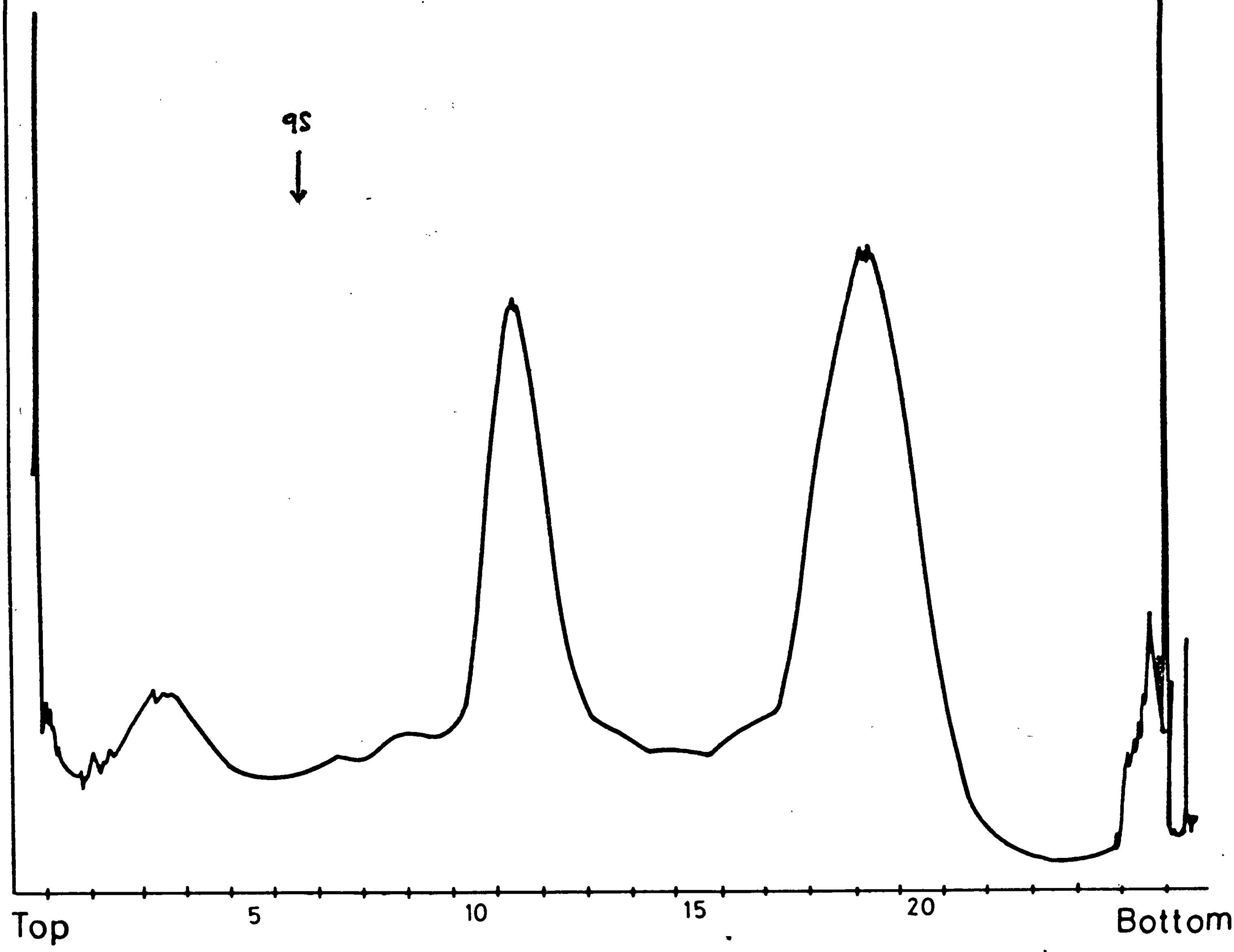

Absorbance 254nm
Top
Bottom
A
Top
Bottom
B
Top
Bottom
C

FIGURE 2. Density Gradient Fractionation of Reticulocyte m-RNA.

RNA (40µg reticulocyte poly A(-); 5000 cpm 9S globin m-RNA ^{125}I) was loaded onto 12 mls, 5%-20% sucrose gradients in NETS buffer (100 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1% SDS) and centrifuged for 5 hrs at 40,000 rpm, 20°C in an IEC 6X14 swing-out rotor. The gradients were unloaded by upward displacement, the presence of the unlabelled RNA being monitored using an LKB Uvicord. Fractions were collected and analysed for m-RNA content by liquid scintillation counting.

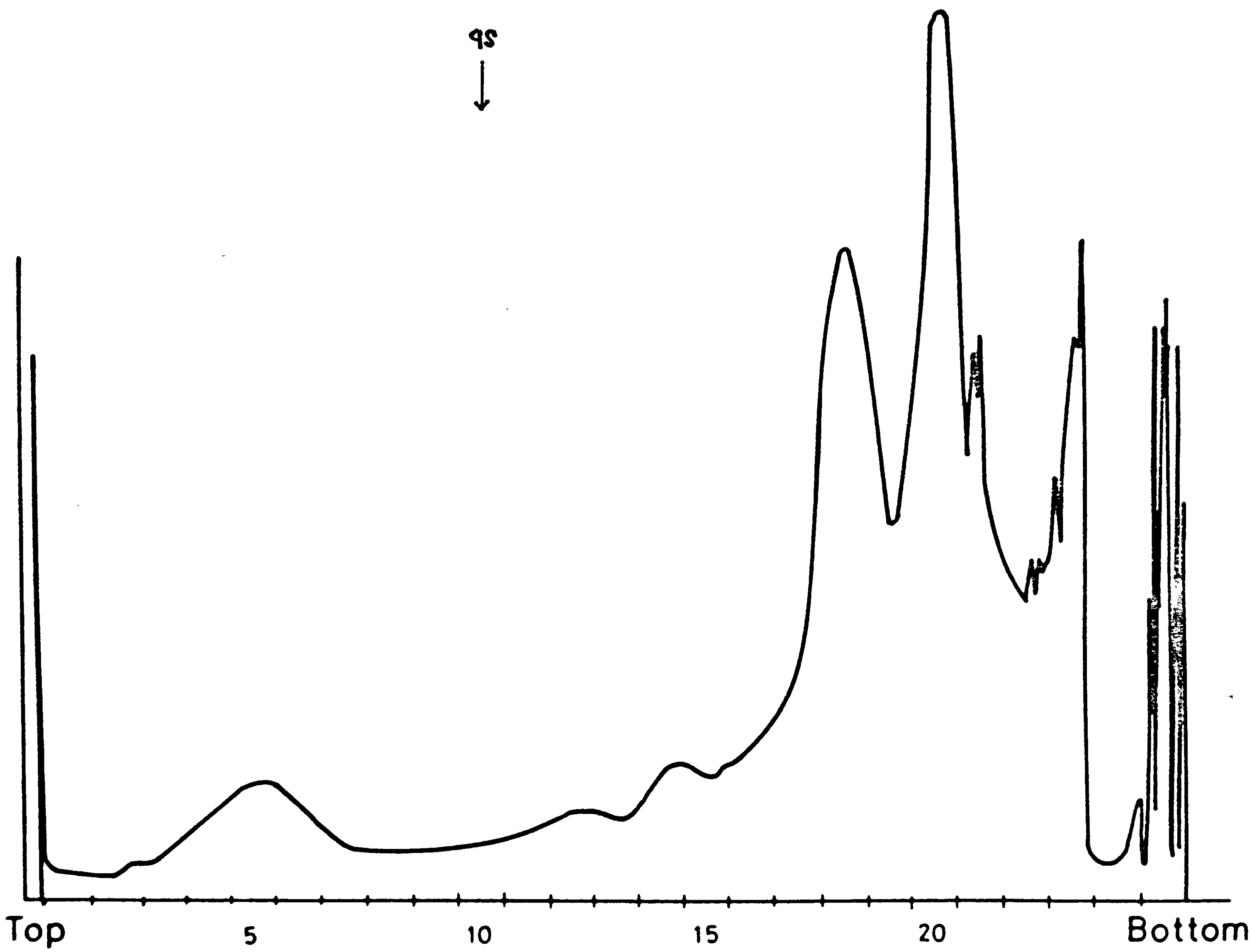

9S
Absorbance 254nm
Top
5
10
15
20
Bottom

The first attempts to recover RNA by this method proved problematical as recovery from the heavy sucrose by ethanol precipitation was found to be poor (less than A way of circumventing this problem was to isolate the poly A+ RNA from the gradient by affinity chromatography. This had the combined advantages of being insensitive to m-RNA concentration, therefore allowing dilution of the sucrose solution, and also effectively removes any gradie impurities. Chromatography on Oligo-dT-cellulose (406) wa considered most appropriate as only small amounts of mRN were expected to result from good purification. The fina material can then be concentrated by lyophilisation rathe than precipitation which might reduce yields (Poly-U-sepharose owing to its higher binding affinity requires elution in Formamide and therefore precipitation is obligatory.)

An example of the finalised technique is shown in Fig. 4. The RNA for the gradient was cycled once on Poly-U-sepharose (405) and loaded onto the top of a 12 mls gradient with a 60% pad at the base. To maximise the sharpness of the density interface the pad was added to step formed gradients immediately before sample loading. The presence of residual ribosomal RNA was hoped to act as carrier during the centrifugation and subsequent handling. Identical marker gradients were run in paralle. Approximately 10.OD. units (400μg) of RNA could be processed in a single tube. After unloading the pooled non-globin

FIGURE 3. Density Gradient Fractionation of Reticulocyte m-RNA.

Gradients were prepared as for Figure 2 but they were underlayed immediately before centrifugation using a 2 ml pad of 60% sucrose in NET. Centrifugation conditions were as described previously except that the separation was for 10 hrs.



Absorbance 254nm
Top
5
10
15
20
Bottom

fractions were diluted with an e ual volume of N.E.T.S. buffer and brought to a final salt concentration of 0.5 M for binding to Oligo-dT-cellulose.

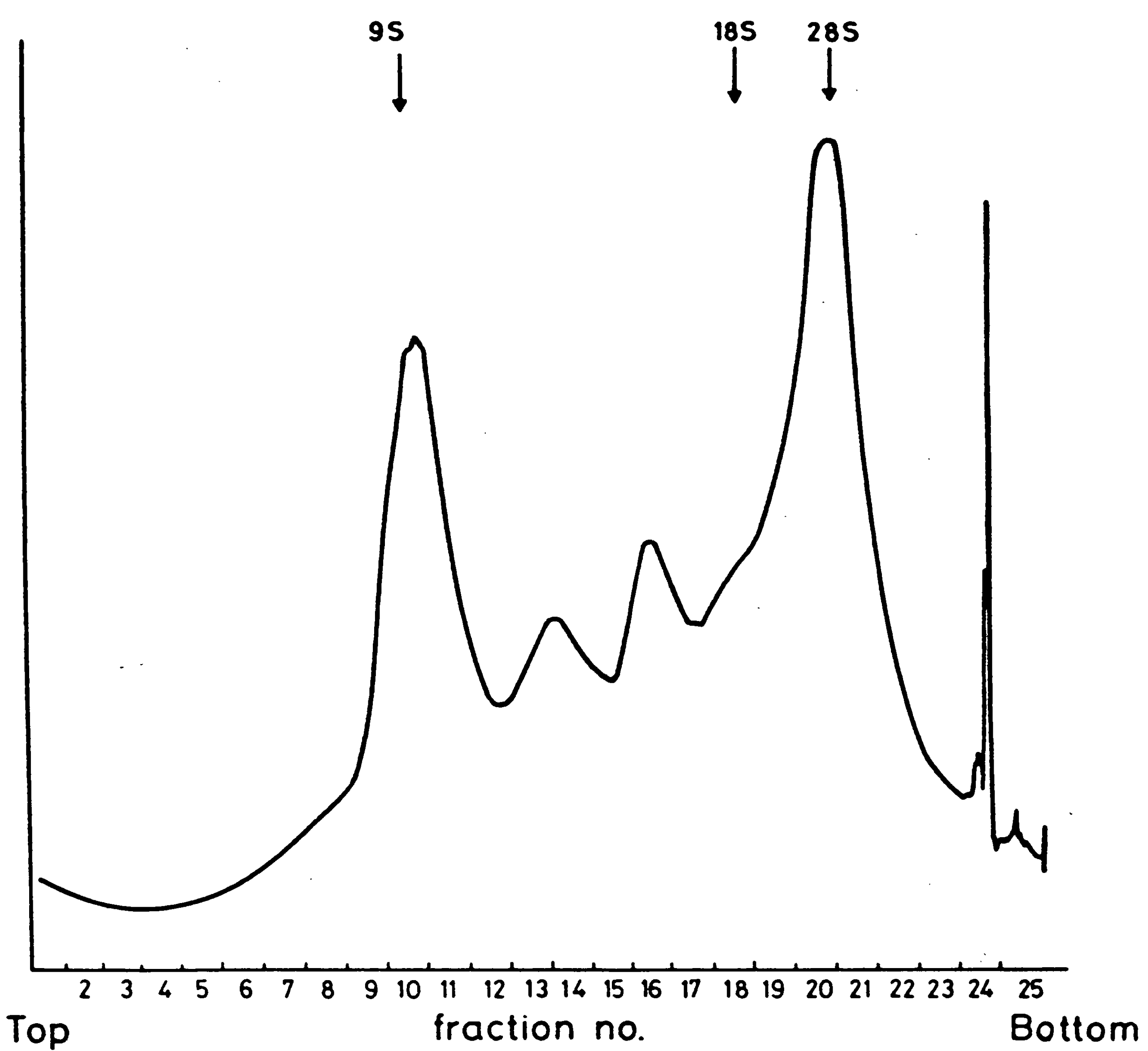
Recovery of RNA from the gradient was approximately 10-12% of the total applied which accorded well with expected recoveries.

FIGURE 4. Density Gradient Fractionation of Reticulocyte m-RNA.

Poly-A containing RNA (400μg) selected from a single cycle of affinity chromatography was loaded onto 5%-20% sucrose gradients with a 60% sucrose pad at the base and centrifuged as described in the legend to Figure 3.

Material from fractions 15-22 was pooled and poly-A containing RNA isolated by addition of NaCl to 0.5M and dilution with an equal volume of Binding Buffer, (0.5M NaCl, 10mM Tris-HCl, pH 7.4, 1mM EDTA, 0.1% SDS) followed by two cycles of binding to Oligo-dT-cellulose (2 ml bed volume). The bound RNA was eluted with sterile water, concentrated by lyophilisation and desalted on a 20X1 cm Sephadex G-50 column.

RNA sedimenting at 9S was recovered from fractions 9-11 by precipitation with ethanol.


9S
18S
28S
Absorbance 254nm
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Top
fraction no.
Bottom

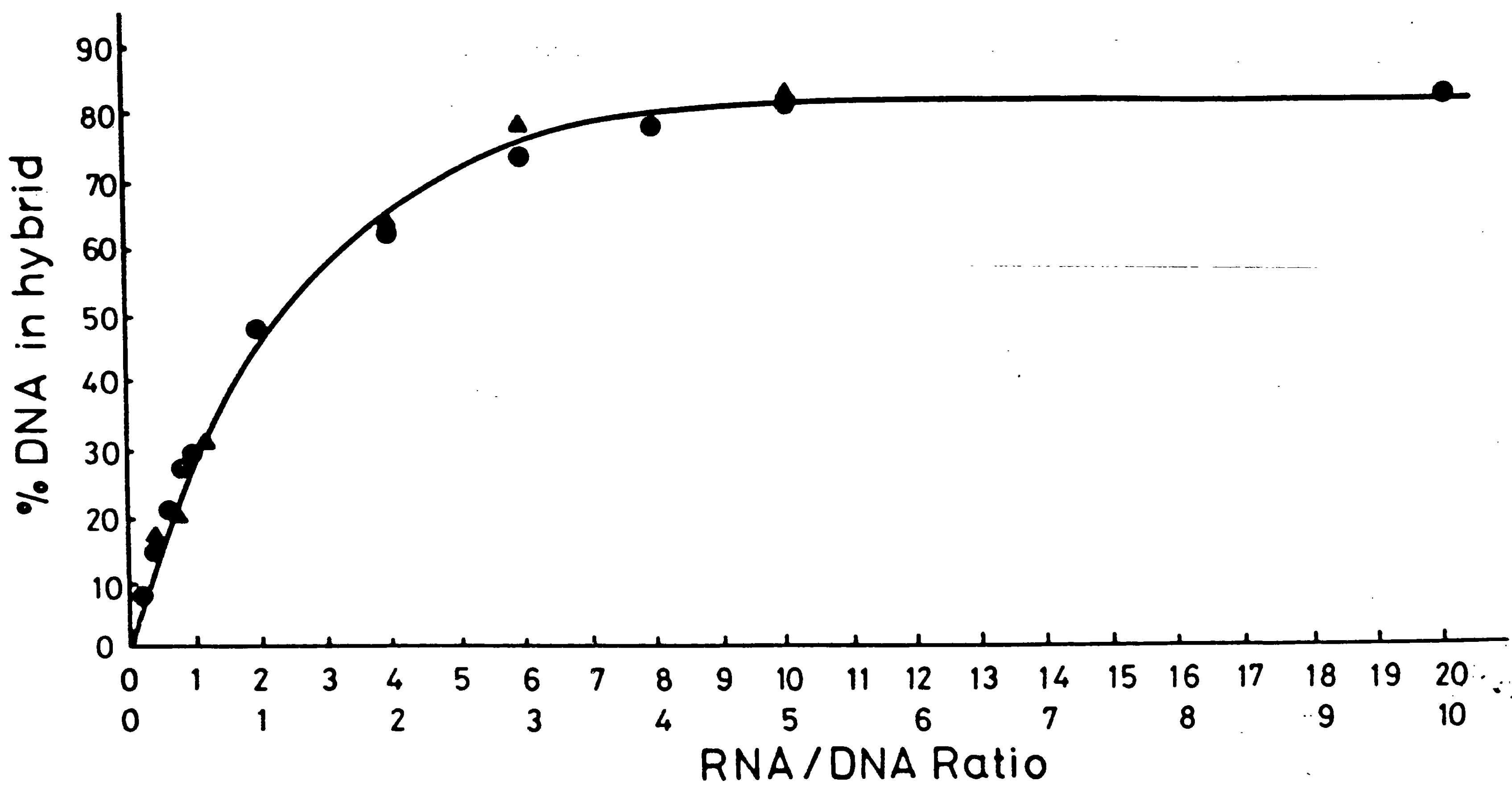
B. Hybridisation Analysis of Fractionated RNA

The size fractionated putative non-globin mRNA was first analysed by molecular hybridisation, utilising a titration procedure (37)(38). In this approach a standard titration curve is prepared by hybridising a fixed mass of cDNA to its template RNA at increasing mRNA/cDNA ratios. The same procedure is performed with a test RNA and the original cDNA. By comparison of the two curves, the concentration of cDNA sequences in the test RNA can be determined. The results of this type of analysis are shown in Fig. 5: the standard curve was constructed using unfractionated reticulocyte mRNA and hybridising it to a cDNA made from it. The "non-globin" RNA was similarly hybridised to this cDNA. The exact concentration of globin sequences in the "non-globin" mRNA can be calculated from the required shift in the mRNA/cDNA scale which will enable the two titration curves to be superimposed. The result indicated by Fig. 5 was that a two-fold change in scale was required to align the "non-globin" titration curve with the standard curve. The estimated globin content of the "non-globin" m-RNA was thus still approximately 50%. Whilst this was still a large fraction of the "non-globin" RNA it did imply a five-fold purification of the "non-globin" sequences. From such an analysis however, it is not possible to conclude categorically that the 50% of non-globin RNA sequences are all derived from messenger.

FIGURE 5. Titration of "non-globin" m-RNA with Reticulocyte c-DNA.

Titration was performed as described in Materials and Methods. Hybridisation values are corrected for S-1 resistant background and differences in final plateau values for the two c-DNA probes.

Upper scale and -●- :"Non-globin RNA
Lower scale and -▲- : Standard curve.

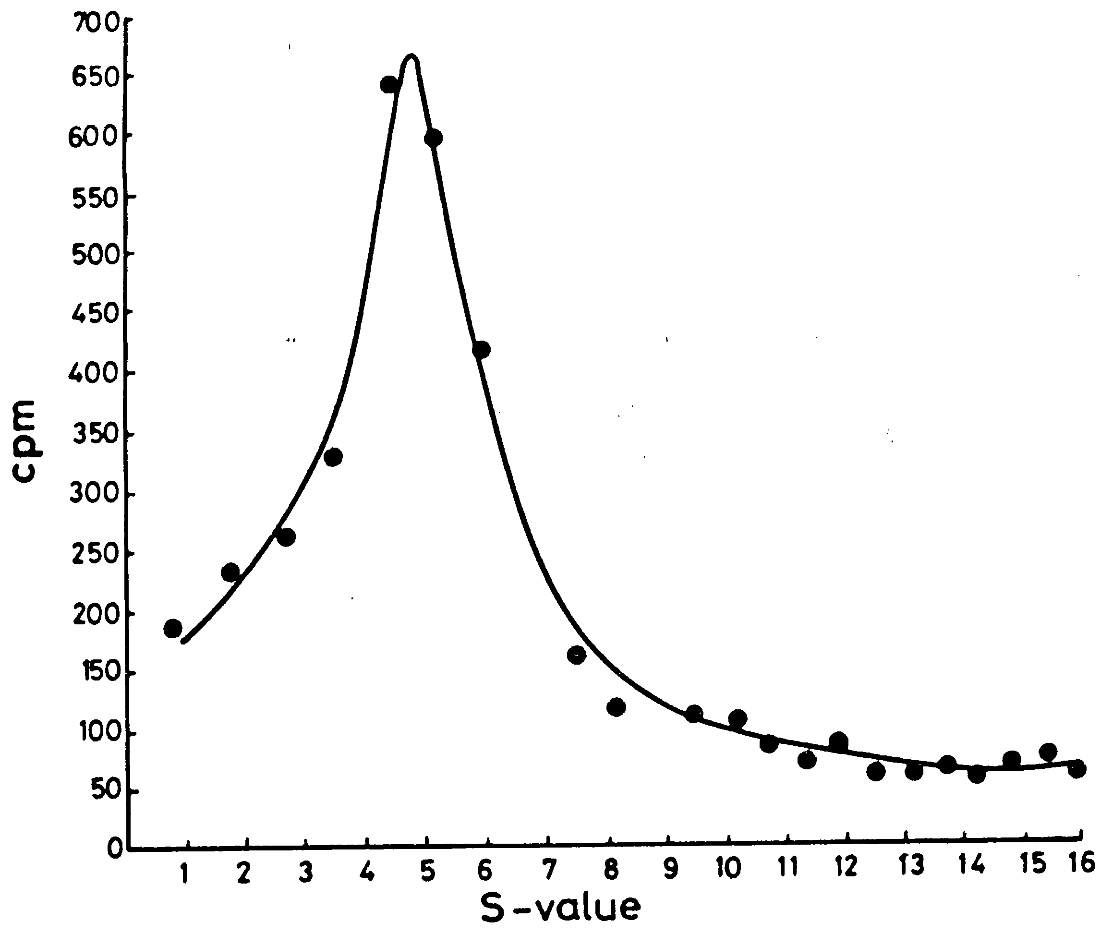

% DNA in hybrid
0
10
20
30
40
50
60
70
80
90
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0 1 2 3 4 5 6 7 8 9 10
RNA/DNA Ratio

It was thus decided next to prepare a "non-globin" c-DNA probe (see Fig. 6) and analyse the sequence content by kinetic methods. The resulting reassociation curve for the "non-globin" mRNA and its cDNA are shown in Fig. 7. The computer analysis of this curve resolved two components. The major component represents some 90% of the total cDNA (see Table 1) and has a true $R_ot_{1/2}$ very close to that previously observed for globin m-RNA under these conditions (29). The second and minor component is therefore the "non-globin" sequence content, but in fact seems to represent only 11% of the total cDNA. The discrepancy between this result and the titration data can be attributed to non messenger sequences in the non-globin RNA, presumably ribosomal RNA. Such a contamination is not totally unexpected as the preparation technique necessitated extraction of a very small m-RNA (poly A(+)) population from a high rRNA background.

Despite the "non-globin" m-RNA component only representing a small fraction of the total cDNA the data enables an estimate of the approximate non-globin sequence complexity albeit within the accuracy constraints imposed by analysing 10% components. The true $R_ot_{1/2}$ for this component of the cDNA is still extremely limited being just less than ten times greater than that for globin. Thus it would appear that the mouse reticulocyte expresses an abnormally small diversity of sequences; this analysis placing it between

FIGURE 6. Size Analysis of "Non-globin" c-DNA.

c-DNA (3000 cpm) was loaded onto 4%-11% alkaline sucrose gradients (0.9M NaCl, 0.1N NaOH) and centrifuged for 21 hrs at 30000 rpm at 20°C. Fractions (1 ml) were collected by upward displacement and used for liquid scintillation counting.



cpm
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
% Sucrose
2 4 6 8 10 12 14 16 18 20 22 24
Fraction No.

FIGURE 7. Kinetic Analysis of "Non-globin" m-RNA.

"Non-globin" c-DNA was hybridised with its template RNA under conditions described in Materials and Methods. The extent of hybridisation was estimated using S-1 nuclease; values are not adjusted for the resistant background.

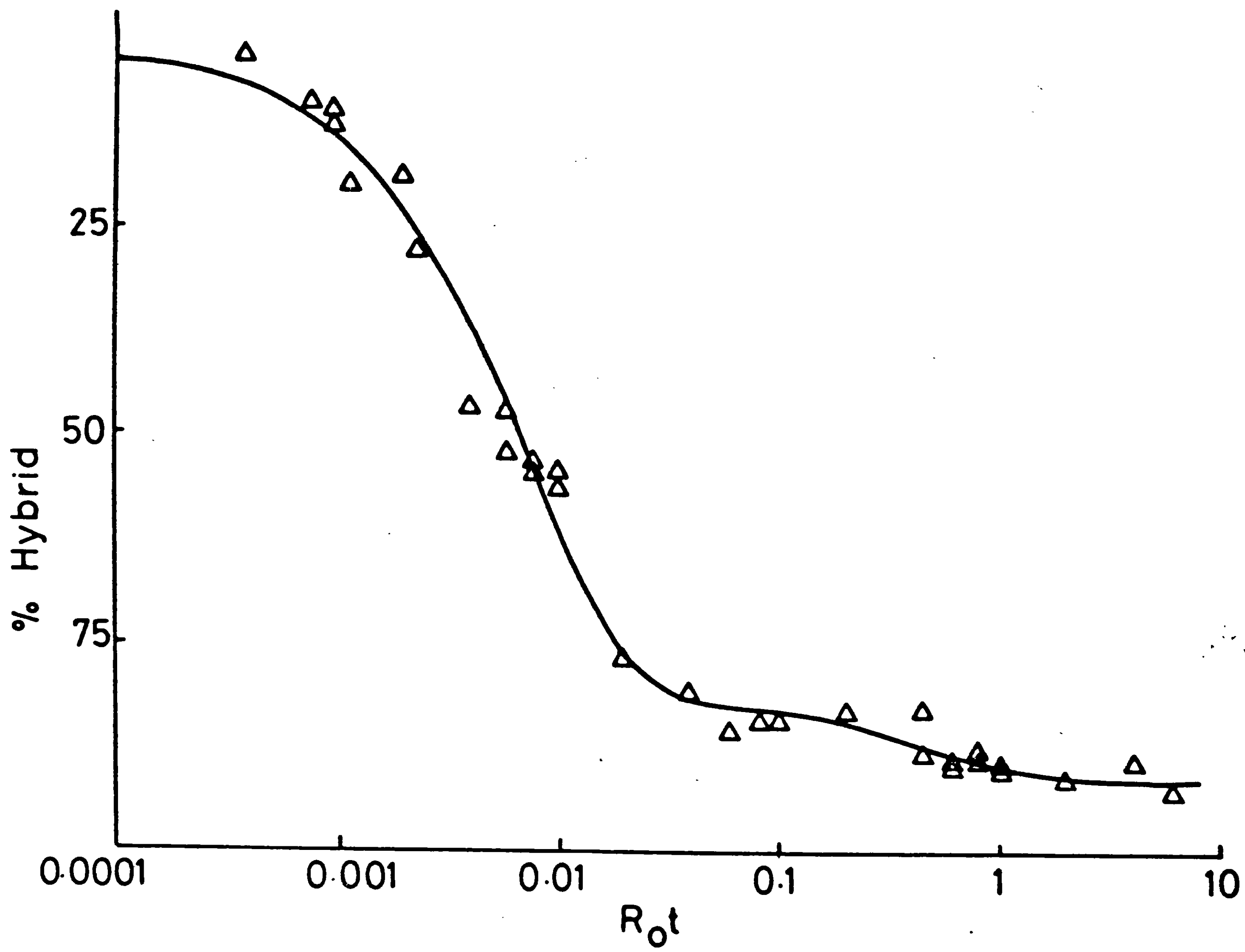

% Hybrid
25
50
75
0.0001
0.001
0.01
0.1
1
10
R_ot

TABLE 1. ANALYSIS OF NON-GLOBIN cDNA COMPLEXITY.

Component	cDNA mass fraction	$R_ot_{\frac{1}{2}}$obs.	$R_ot_{\frac{1}{2}}$corr[a]	Complexity[b]	No. of average sequences[c]
1	88.70%	5.14×10^{-3}	4.56×10^{-3}	4×10^{5}	2
2	11.30%	3.70×10^{-1}	4.19×10^{-2}	3.68×10^{6}	6

[a] The $R_ot_{\frac{1}{2}}$ obs. is corrected using the mass fraction of cDNA of that particular component to give $R_ot_{\frac{1}{2}}$ corr.

[b] Complexity was calculated using 4×10^{5} as a standard for globin.

[c] The complexity was converted to sequence number using the standard for total globin m-RNA and 6×10^{5} daltons as the average molecular weight for mouse messenger RNA.

4 and 9 different sized messengers of average size. Undoubtedly one cannot rule out the possibility that a further much more diverse sequence set is present but at an abundance too low to be observed under these conditions. Notwithstanding this proviso, the hybridisation data squares well with observations from other laboratories regarding the complexity of reasonably abundant non-globin products (39)(415).

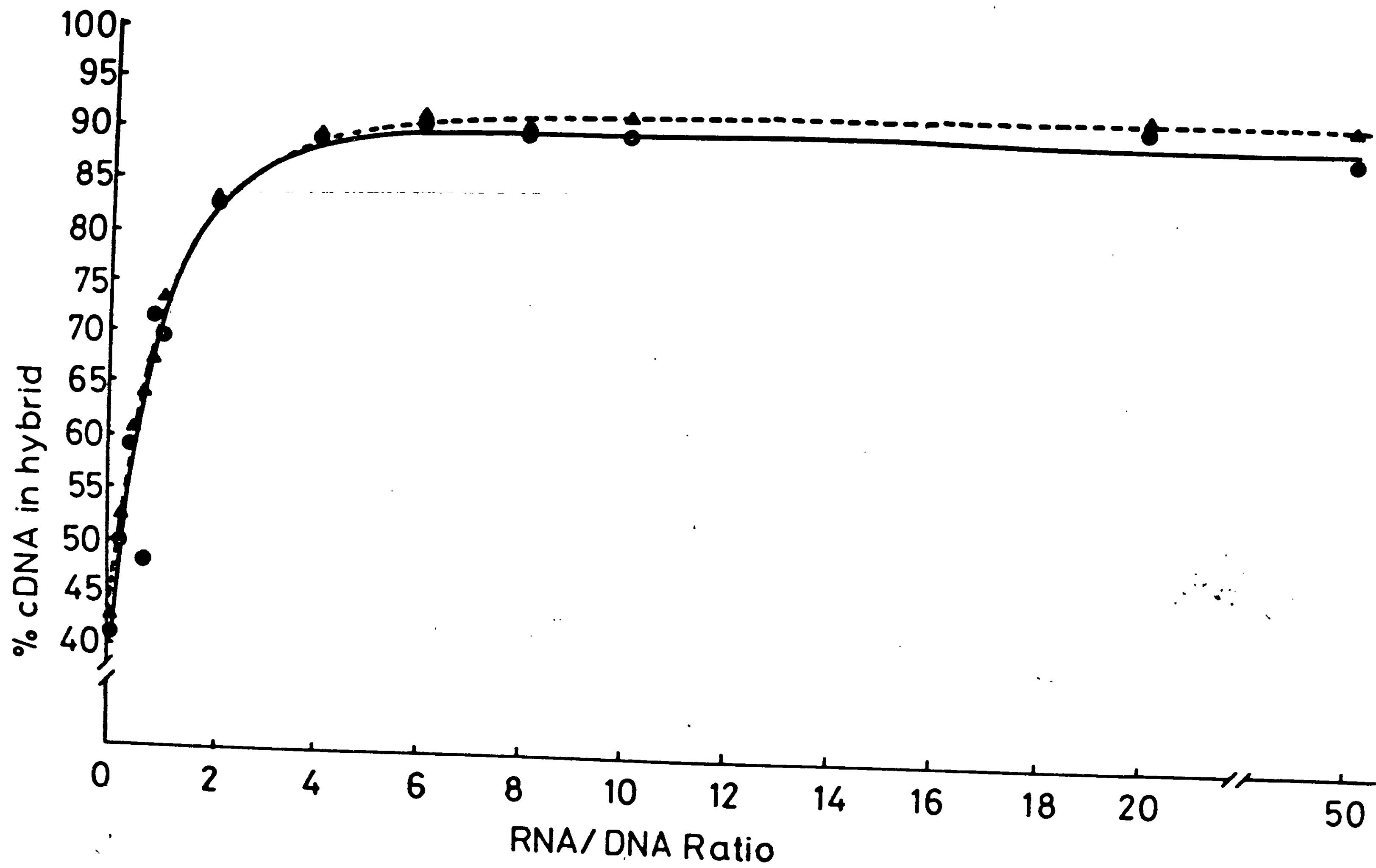
The implication of this kinetic investigation was that no enrichment of "non-globin" sequences had accrued from the gradient fractionation procedure. To examine this possibility, the "non-globin" cDNA was titrated with either unfractionated reticulocyte m-RNA or RNA taken from a 9S size cut (see Fig. 4) of m-RNA. Interestingly, the cDNA hybridises equally well to both RNAs, moreover the plateau is attained at a RNA/cDNA ratio very similar to that seen for the homologous reaction between unfractionated mRNA and cDNA. Whilst the 9S RNA may still contain some non-globin sequences the implication of this analysis is that it differs little from the unfractionated m-RNA and thus both contain all the sequences present in the cDNA, in roughly similar proportions. (Fig. 8).

Despite some evidence that the unfractionated m-RNA may in fact be markedly lower in "non-globin" sequence content than previously estimated, as in fact has been reported for mouse reticulocytes (62) by Aviv and his colleagues, the nature of the "non-globin" c-DNA population suggested that a greater enrichment may be obtained by other techniques. The first approach was to investigate the anaemia to ascertain if non-globin synthesis remained constant during this period.

FIGURE 8. Globin m-RNA Sequence Content of "non-globin" c-DNA.

"Non-globin" c-DNA was hybridised to reticulocyte m-RNA or 9S size fractionated RNA under conditions as described in Figure 5.

—▲— 9S RNA —●— reticulocyte m-RNA


% cDNA in hybrid
100
95
90
85
80
75
70
65
60
55
50
45
40
0
2
4
6
8
10
12
14
16
18
20
50
RNA/DNA Ratio

C. Analysis of Reticulocyte Proteins Labelled in Vitro

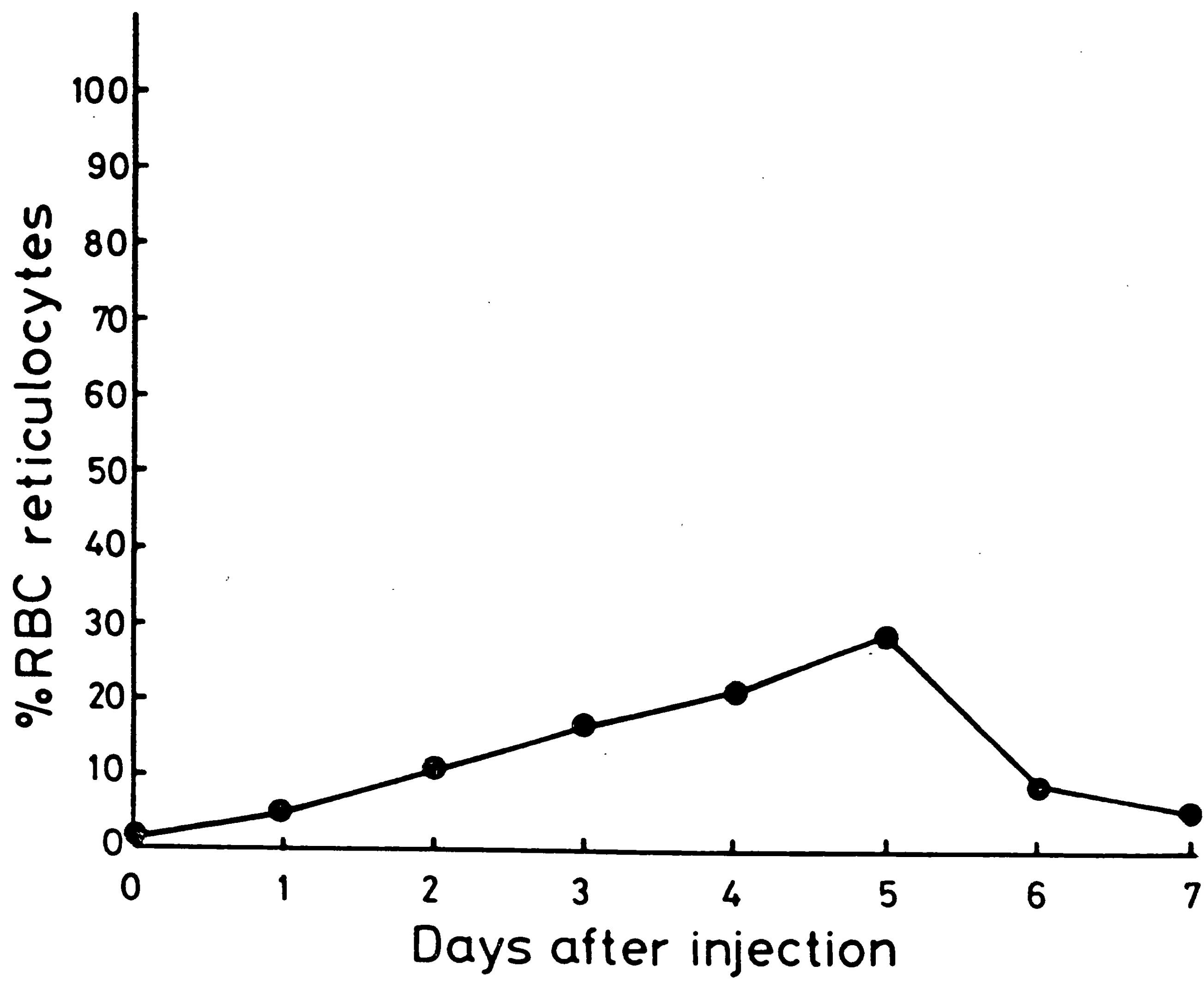
Evidence from a number of sources (68)(416)(417) has suggested that more immature cells may be synthesising greater quantities of "non-globin" proteins. A method was devised therefore to observe the in vivo protein synthesis patterns of reticulocytes at various stages of an induced anaemia to determine whether non-globin synthesis remains constant during reticulocyte maturation.

To assess non-globin protein synthesis mice were rendered anaemic by a single injection of acetyl phenyl hydrazine (0.1 ml 2.5%). On subsequent days after injection, blood from about 20 of the mice was collected and used to prepare ^{35}S labelled acid/acetone protein precipitates, using a haemolysate induced by osmotic shock. A small sample of the blood was subjected to Brilliant Cresyl Blue supra-vital staining to assess the severity of the anaemia. The progressive change to the circulating reticulocyte count is shown in Fig. 9.

A portion of the total acid/acetone protein was then subjected to fractionation on Sephadex G-100 under conditions similar to those described by Bulova and Burka (414). The separation was however performed using buffer containing 6 M Urea, 1% β-mercaptoethanol. The presence of a reducing agent is imperative when fractionating globins as aggregation due to the presence of cysteine bridges is common. The results of this analysis are shown in Table 2 and in Fig. 10.

FIGURE 9. Production of circulating reticulocytes during induced anaemia.

Mice were rendered anaemic by a single injection of acetylphenylhydrazine (0.1 ml, 2.5%). Blood was collected on subsequent days after injection and a small aliquot used for estimation of reticulocyte count, by Brilliant Cresyl Blue vital staining.


%RBC reticulocytes
0
10
20
30
40
50
60
70
80
90
100
0
1
2
3
4
5
6
7
Days after injection

The major peak of radioactivity was found always to correspond to the peak of ultra-violet absorbance. The major peak of radioactivity can also be seen to remain remarkably constant with regard to the fractions in which it was eluted (see particularly 10b and c), moreover the peak height displays a similar consistency. The fraction of the total protein contained within the pre-peak which eluted just behind the excluded material (dextran blue marker) however, varied markedly with the stage of the anaemia (see Table 2). In particular, whilst the peak reticulocytosis occurred at Day 5, when blood to synthesise messenger RNA was normally harvested, the maximum for "pre-peak" protein synthesis was quite clearly earlier, on the third day. As this pre-peak was obviously heterogeneous in nature and most likely non-globin proteins, all the pre-peak fractions (plus the first few of the globin tail) were pooled, concentrated and used for analysis on polyacrylamide/SDS slab-gels. Unfortunately, owing to heavy losses of material at the concentration step only two sets of samples produced meaningful results. The stained protein bands are shown in Fig. 11. Track 1 contains protein from Day 2 and Track 2 from Day 3. In both cases a good agreement between staining pattern and autoradiographic pattern was observed. For example both tracks show the presence of a major band, probably two different proteins, of approximately 30K mw (Band F).

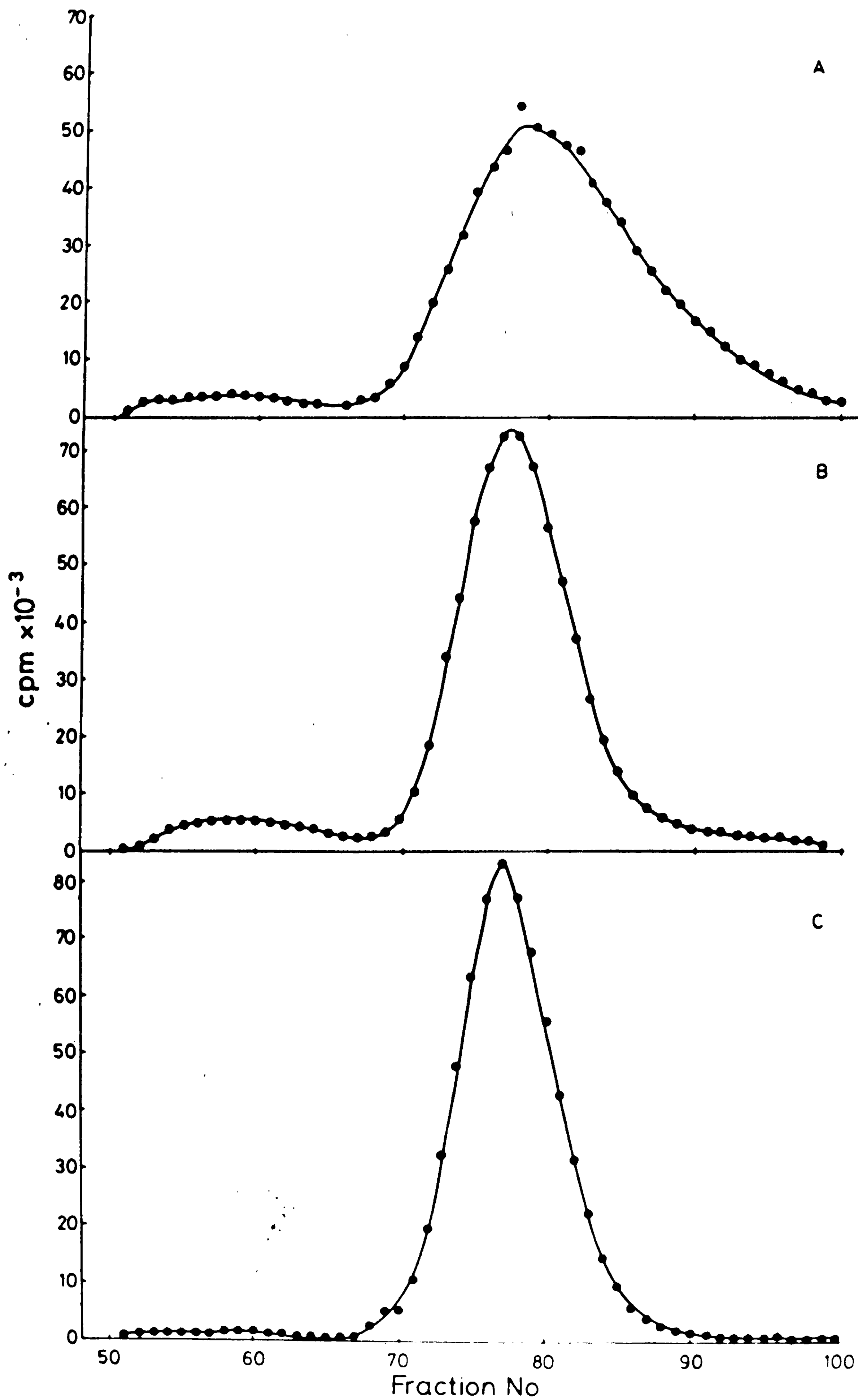
FIGURE 10. Fractionation of Acid Acetone Precipitated Proteins.

^{35}S labelled proteins from sequential days of an induced anaemia (see Fig. 9) were prepared by acid acetone precipitation. Fractionation into globin and "non-globin" components was performed on Sephadex G100 columns run in 6M urea; 1% β-mercapto-ethanol. 2.5 mls fractions were collected and 0.5 mls used for liquid scintillation counting.

a) 40 mg protein from Day 2.

b) 20 mg protein from Day 3.

c) 20 mg protein from Day 5.



A
B
C
cpm $\times 10^{-3}$
Fraction No
0
10
20
30
40
50
60
70
80
90
100

TABLE 2. ANALYSIS OF INDUCED ANAEMIA.

Days after injection[a]	Vol. of Packed Cells (mls)[b]	Yield of Protein (mg)[c]	Circulating Reticulocytes (%)[d]	Fraction Total Protein Non-globin[e]
1	2.0	69.2	5	-
2	2.0	112.0	11	5.0
3	2.0	34.7	17	7.9
4	1.7	67.0	21	3.7
5	1.0	40.0	31	2.6
6	1.5	107.7	9	-
7	-	-	6	-

[a] A single dose of acetylphenylhydrazine (0.1 ml, 2.5%) was administered on Day 1.

[b] Volume of packed cells obtained from 30 mice.

[c] Protein was extracted as an acid/acetone precipitate.

[d] Reticulocyte count was obtained from vital staining with Brilliant Cresyl Blue.

[e] Non-globin content of total protein was estimated from liquid scintillation counting of G-100 Sephadex column fractions (see Fig. 10).

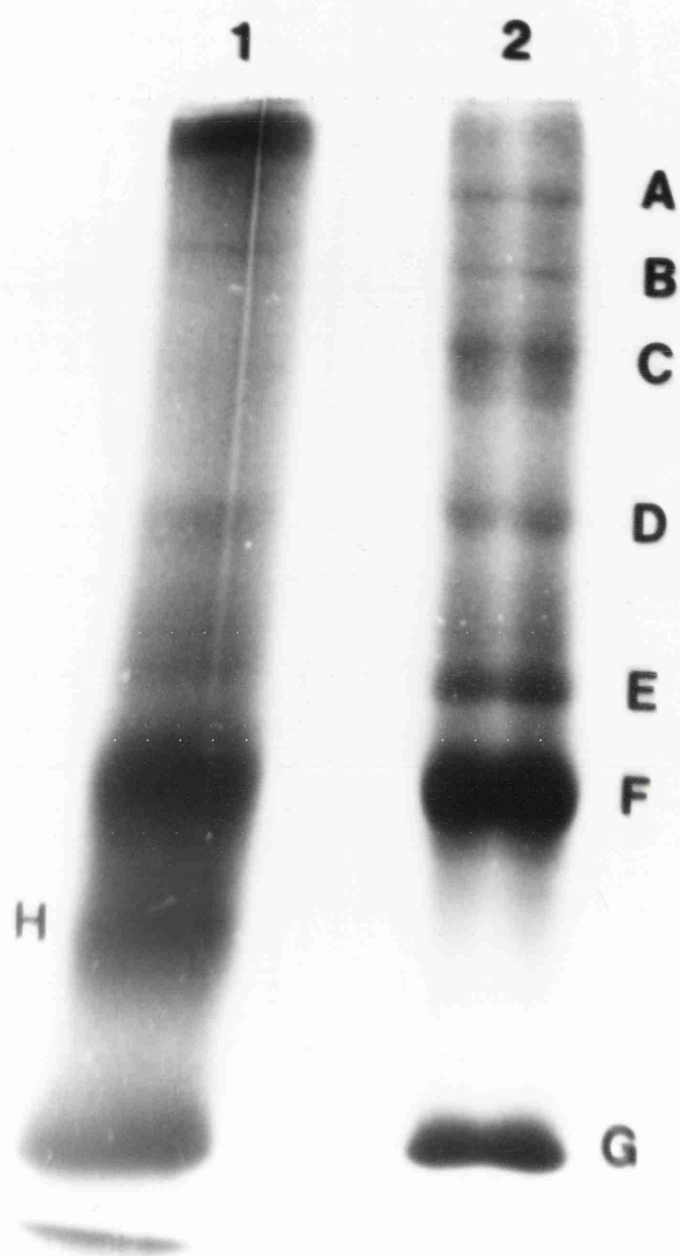
Band F was also the major radioactively labelled product. Band E however, although prominently staining, was found not to have incorporated radioactivity. A number of differences exist between tracks 1 and 2, most notably a low molecular weight band in Track 1 (Band H) which may be a bona fide protein or could conceivably be breakdown products of a higher molecular weight protein. Lodish has identified a protein of similar molecular weight in reticulocyte membranes, however, the protein in track 2 is not represented in track 3 and therefore may not be stable as are most membrane proteins. Some very high molecular weight species are lost between Day 2 and 3, this could possibly be some residual Spectrin synthesis occurring in less mature reticulocytes.

As was indicated by the kinetic analysis, the protein labelling study suggested that reticulocytes were synthesising only a highly restricted set of protein products, comprising some six different species. Similar results have been obtained from work in other laboratories. Undoubtedly, factors such as detection level are again important here, as there may be a fairly large number of proteins of too low a specific activity to reveal themselves on an analysis of this nature. With so few observable products it is unlikely that much masking due to comigration was occurring but it may be possible to resolve Band F into its two presumed components by an analysis on 2-dimensional gels (418).

FIGURE 11. SDS/Polyacrylamide gel Electrophoresis of fractionated "non-globin" proteins.

Pooled "non-globin" protein fractions from Sephadex G-100 fractionations were concentrated and loaded onto a 10% polyacrylamide/SDS slab gel with a 4.5% stacking gel in buffer containing 0.01M Tris pH 8.0, 1 mM EDTA, 1% SDS, 5% β-mercaptoethanol. The gel was run in buffer containing 0.38 M glycine, 0.05 M Tris pH 8.3, 1% SDS,,until the Bromophenol Blue tracker dye had reached the bottom of the gel. (approx. 9 cms.) The gel was then stained with Coomassie brilliant blue in 50% TCA as descibed by O'Farrell (418) and prepared for fluorography according to the method of Bonner & Laskey (449).

1) "Non - globin" protein sample from Day 2 of induced anaemia.
2) As above but from Day 3.


1
2
A
B
C
D
E
F
G
H

D. Manipulation of the Anaemic Response

The previously described investigation of non-globin protein synthesis strongly indicated that cells harvested on the third day after injection should contain a higher proportion of non-globin m-RNA sequences than material obtained from the standard protocol, which collected blood on the fifth day. The problem to overcome was the low level of circulating reticulocytes at the early stages of the anaemia (only 17% on Day 3; see Fig. 9). Variations in dosage and administration time were adapted to boost the yield of reticulocytes at the early stages of the anaemic response. An extremely severe anaemia was found to be produced by the third day using an injection regime, which both doubled the original dose (Day 0) and then repeated this 48 hrs subsequently (Day 2). As judged by Brilliant Cresyl Blue staining, the anaemia was very severe; greater than 98% of circulating cells scoring as reticulocytes by this methodology. More importantly perhaps, a much greater proportion of cells than normal demonstrated a highly extensive staining pattern, indicative of the presence of immature cells, with a more highly developed endoplasmic reticulum. (Fig. 12). More mature cells show a localised more granular staining appearance (Fig. 12).

One further modification found fruitful was to prepare the acetyl phenyl hydrazine in PBS solution rather than citrate/NaOH as normally utilised. This had a quite marked

FIGURE 12. Photomicrograph of Brilliant Cresyl Blue stained blood smear.

Blood was obtained from anaemic mice as described in the text and stained to show the presence of reticulocytes as described previously. The arrow designates an immature cell showing an extensive open staining pattern. The triangle indicates a more mature cell with a more localised granular stained appearance.

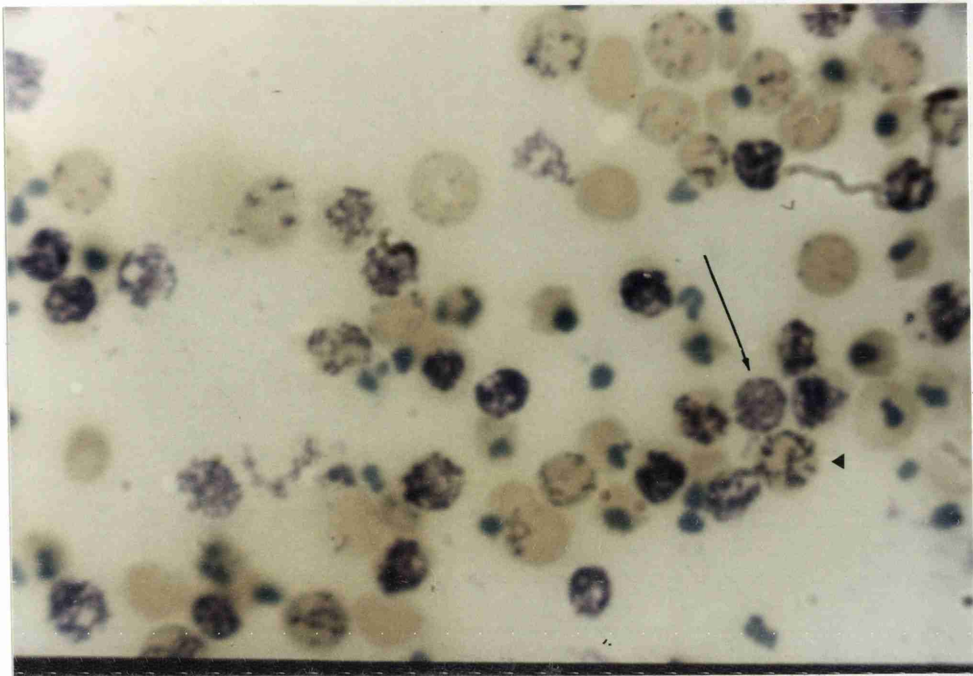
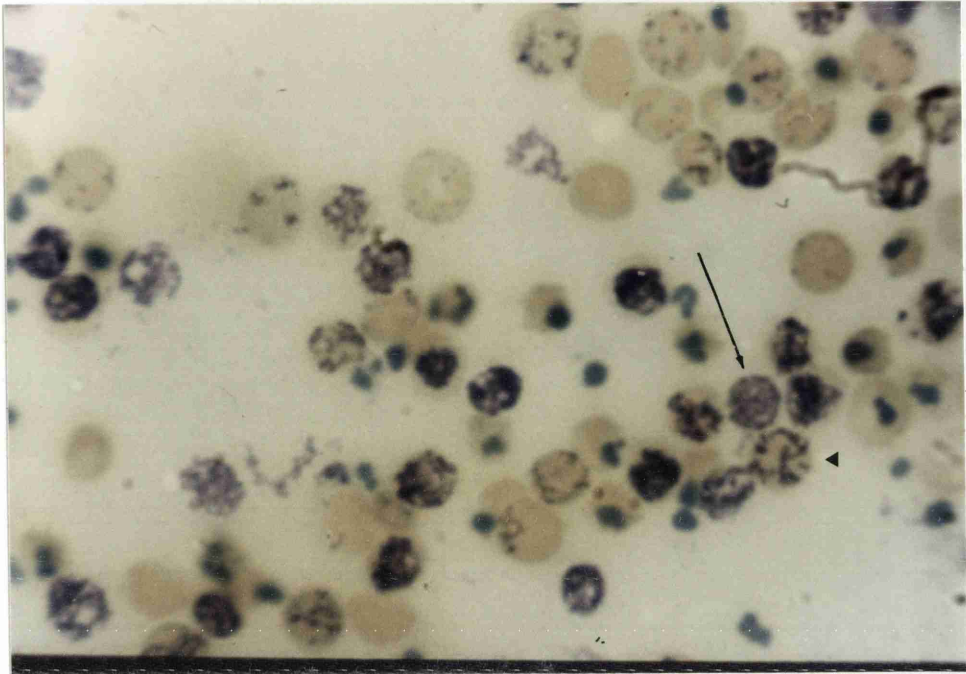


FIGURE 12. Photomicrograph of Brilliant Cresyl Blue stained blood smear.

Blood was obtained from anaemic mice as described in the text and stained to show the presence of reticulocytes as described previously. The arrow designates an immature cell showing an extensive open staining pattern. The triangle indicates a more mature cell with a more localised granular stained appearance.



effect on nonspecific mortality, that is death occurring before the onset of anaemia, reducing it to around 2% compared with a maximum of 15% when the standard phenyl hydrazine hydrochloride solution was used.

Such a gross anaemia was found to have some unfortunate repercussions however. Firstly, as may be expected there was a considerable drop in the haematocrit, the volume of packed cells obtained from a given number of mice being significantly reduced. Secondly, the presence of precipitated haemoglobin in the form of haemochromes, both in the cells and the plasma, caused a marked increase in the fragility of the reticulocytes, such that a considerable proportion of lysis occurred under centrifugation conditions necessary to pellet the cells and remove the buffy coat.

To attempt to overcome these problems an investigation was made of the possibility of using other media to separate the red and white blood cells. Initial results using "lymphoprep" (a ficoll/hypaque mixture) were promising. Lysis was virtually absent and the major white cell fraction, small lymphocytes were collected at the interphase between the lymphoprep and the overlayered whole blood. The use of a dense ficoll layer allowed centrifugation at lower than normal "g" forces but the density of lymphoprep is designed primarily for the preparation of lymphocytes, thus granulocytes may have been pelleting with the red cells. Therefore using the methodology of English and Andersen (419) a Ficoll/Triosil

mixture of greater density was prepared to ensure that granulocytes did not pellet. These cells did still not remain at the interphase but were suspended within the Ficoll layer. Examination of microscopical preparations confirmed that the white cell contamination of reticulocytes prepared in this was at least as low as that obtained by standard procedures, and was most probably somewhat lower.

Unfortunately, when the protocol was scaled up for a bulk preparation a fairly large proportion of cells did lyse, although this did not affect the purity of the obtained reticulocytes. It seems that despite care being taken to ensure the conditions for the bulk preparation were identical to those used for the pilot experiments, the precise height to volume ratio of the centrifugation vessel was critical in determining the extent of lysis. Naturally, the main effect of these problems was to limit the amounts of RNA finally obtained. However, as the RNA was to be used for cloning procedures it was felt that this was not too great a penalty to incur in order to obtain RNA from immature cells.

An example of a cell free translation of mRNA made by this protocol as compared with the previous protocol is shown in Fig. 13. A worthwhile enrichment in at least one non-globin sequence can be observed on this exposure.

FIGURE 13. Cell free translations of reticulocyte m-RNA.

m-RNA from mouse reticulocytes was translated in a message dependent lysate from rabbit reticulocytes prepared according to Pelham and Jackson (401). The translation mixes were applied to the slots of a 12.5% polyacrylamide/SDS gel with a 4.5% stacking gel. Sample and running buffers were as described for Figure 11 and the gel was also fluorographed as previously described. (The translations were kindly performed by Dr H. Jacobs).

Track A: m-RNA prepared from reticulocytes obtained by the standard conditions of induced anaemia.

Track B: m-RNA prepared from reticulocytes obtained by the modified protocol, which gives a higher proportion of immature cells.

mw $x10^{-3}$
A
B
43
30
20
14

II. CONSTRUCTION OF CLONED PROBES

A. Preparation of cDNA for Cloning

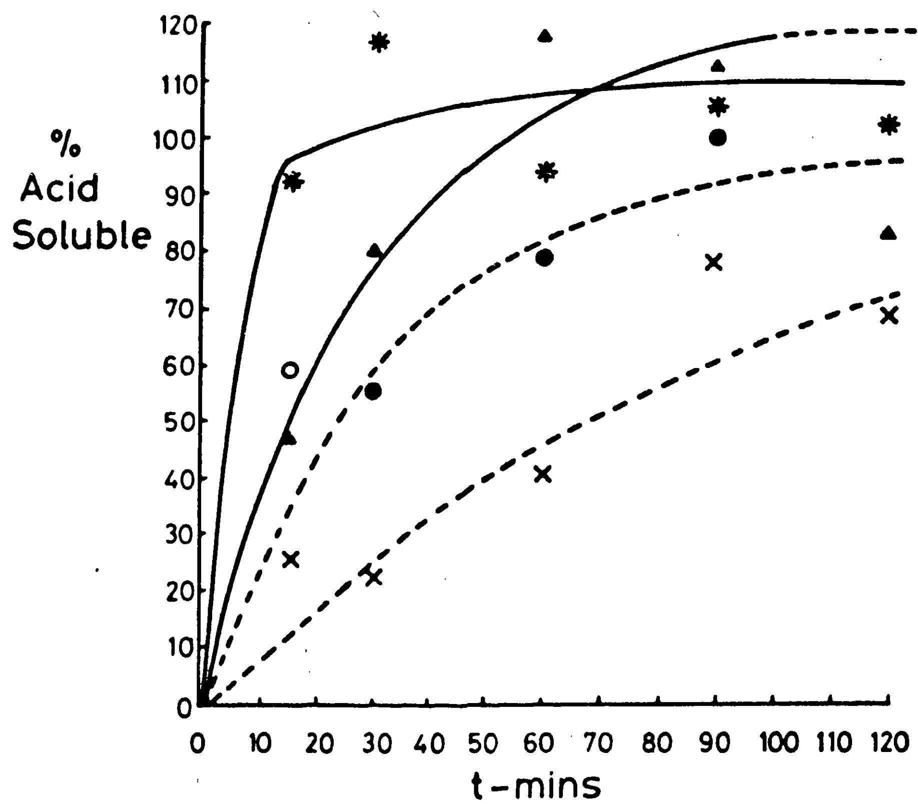
The cDNA used for establishing the cloning methodology was synthesised as detailed by Humphries *et al* (404); for its simplicity, later experiments utilised the protocol described by Schimke (408)(420). Both methods involve synthesis of the first strand by reverse transcriptase (gift of J.W. Beard) and use of DNA polymerase I for the second strand synthesis (see Methods). Similar yields were obtainable from both methodologies, approximately 50% of the single stranded cDNA being converted to duplex cDNA. The main limiting factor for either approach was the yield at first strand synthesis, which was mainly dependent on the reverse transcriptase preparation in use. This is in good agreement with data presented by Schimke.

The conditions required for S-1 digestion of the hairpin loop of the cDNA were investigated by monitoring the digestion of (^{3}H) labelled cDNA under conditions identical to those used for the d.s. cDNA reaction. The extent of digestion at various time intervals was assayed by estimating the percentage of cDNA rendered acid soluble. The results are shown in Fig. 14. The minimum quantity of enzyme commensurate with complete digestion in a reasonable time was chosen for the bulk reaction. The value chosen was 10 units in 2.5 mls; this accords well with the 5μ/ml reported in ref. (420).

FIGURE 14. Titration of S-1 Nuclease.

The time course of digestion of single stranded cDNA with varying amounts of enzyme was monitored as described in Materials and Methods, by assay of acid solubility, under the conditions used for hairpin scission of double stranded molecules.

2 units —X— , 5 units —●— ,
10 units —▲— , 20 units —*—



%
Acid
Soluble
120
110
100
90
80
70
60
50
40
30
20
10
0
0 10 20 30 40 50 60 70 80 90 100 110 120
t-mins

The products of the first and second strand syntheses were analysed by alkaline sucrose gradient centrifugation. Gradients of the d.s. cDNA before and after S-1 cleavage are shown in Fig. 15a-b respectively. A significant shoulder is visible after second strand synthesis indicating that the second strand synthesis is not complete; from the relation of the sizes of the fragments in the two peaks, it would appear that the mean s.s. cDNA size is approximately 230 nt. In support of this interpretation is the observation that the size of the d.s. molecules falls away sharply above 500 nt. This implies that the second strand synthesis is copying a proportion of the single strand DNA in its entirity rather than all molecules partially. Such a hypothesis is supported by the fact that after S-1 cleavage a marked peak at the original s.s. peak size is once again observed. Partial second strand synthesis should reveal itself by a reduction in the size of the products after S-1 treatment.

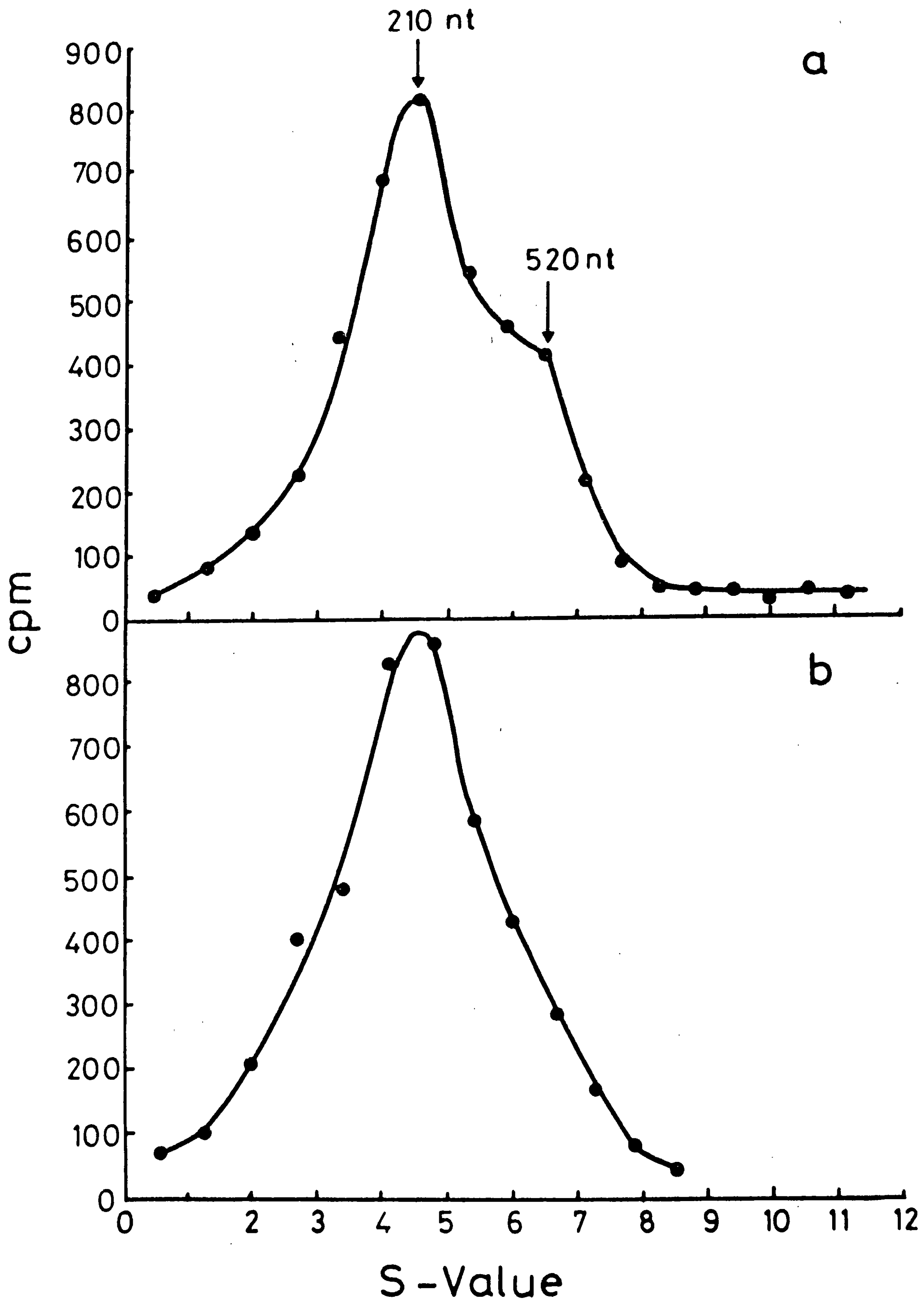
The cDNA was concentrated by ethanol precipitation of the S-1 reaction products (after phenol/chloroform extraction) in the presence of carrier E.coli t-RNA.

FIGURE 15. Analysis of the products of ds cDNA synthesis and S-1 cleavage.

The ds cDNA was sedimented in 12 mls. 4%-11% alkaline sucrose gradients using an IEC 6 x 14 rotor at 39,000 rpm for 13 hrs at 20°C. Fractions were collected by upward displacement and used for liquid scintillation counting, to detect the presence of cDNA.

a) ds cDNA products before S-1 scission of hairpin.

b) ds cDNA products after S-1 scission of hairpin.


210 nt
520nt
a
b
cpm
S-Value
0
100
200
300
400
500
600
700
800
900
0
1
2
3
4
5
6
7
8
9
10
11
12

B. Construction of Plasmid Vectors

1. Preliminary Characterisation of Plasmid DNA

Originally it was envisaged cloning in plasmid pBR322; however, the availability of its derivative pAT 153 enabled a change to the new plasmid and by use of the host E.coli HB101 a reduction of the original category 3 containment requirement to category 2.

However, the new plasmid was less well characterised than the parent molecule, so some preliminary investigation of the pAT 153 plasmid was made. The simplest analysis of the plasmids was to Hae II digest both of them as pAT was derived from a partial Hae II digestion of pBR 322. The digestion products were examined by ethidium bromide staining of a 6% polyacrylamide gel (see Fig. 16). All the expected bands were observed in pAT 153, the major missing fragment being the second largest, a 622 base pair fragment. Unexpectedly however, inspection of the lower molecular weight fragments revealed that a small 83 b.p. fragment was absent also. This fragment maps contiguously with the large missing fragment and thus establishes the total deletion of material from pBR 322 as 705 b.p. extending from map positions 1646 to 2351 on the map established by Sutcliffe (421)(422).

In preparation for cloning, plasmid DNA was linearised at the unique Eco RI site as shown in Fig. 17a. The schematic in Fig. 18 shows the steps involved in constructing recombinant plasmids by two common methodologies. The most frequently

FIGURE 16. Preliminary characterisation of plasmid pAT 153.

Plasmid DNA from pBR 322 or pAT 153 (1μg) was digested as specified by the manufacturer using Hae II. The products of the reaction were analysed by electrophoresis on 6% T.E.A. polyacrylamide gels and stained with ethidium bromide (1 μg/ml). The bands were visualised under u-v light.

Track A: pAT 153 digest.

Track B: pBR 322 digest.

The arrows indicate the bands missing from the corresponding pAT 153 digest, the sizes of which are as indicated.

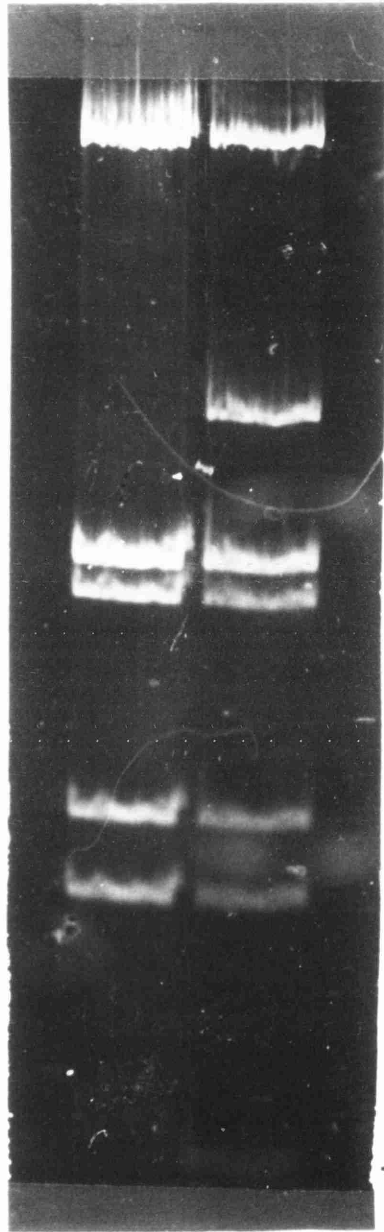

A
B
622 bp
83 bp

FIGURE 17. Linearisation of plasmid DNA.

Plasmids were linearised using enzymes cutting at unique sites within the plasmid, and the digests visualised on 1% agarose T.E.A. gels.

a) Track 1, unrestricted pAT 153 plasmid DNA;

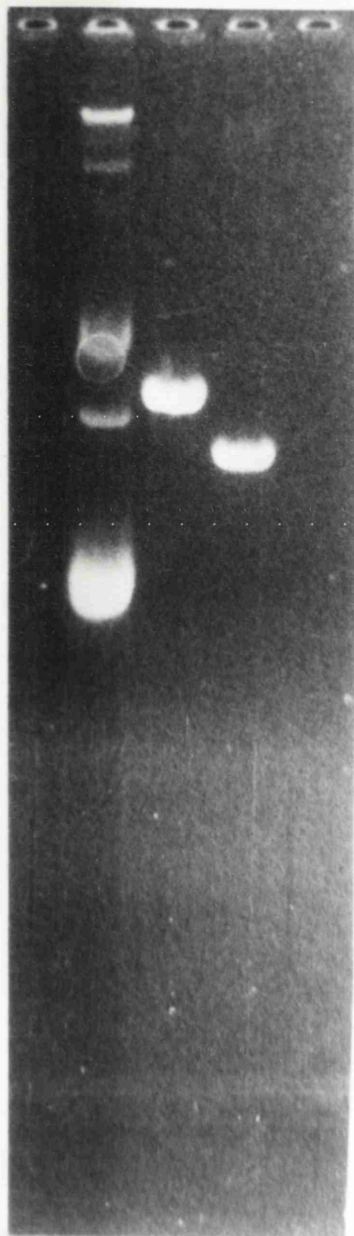
Track 2, pBR 322 linearised with EcoRI.
Track 3, pAT 153 treated as above.

b) Track 1, pAT 153 linearised with Sal I;

Track 2, pAT 153 linearised with BamHI.

a

1 2 3



b

1 2

utilised protocol is that shown on the left-hand side of the diagram, which utilises complimentary homopolymer extensions to the linearised plasmid and the double stranded cDNA.

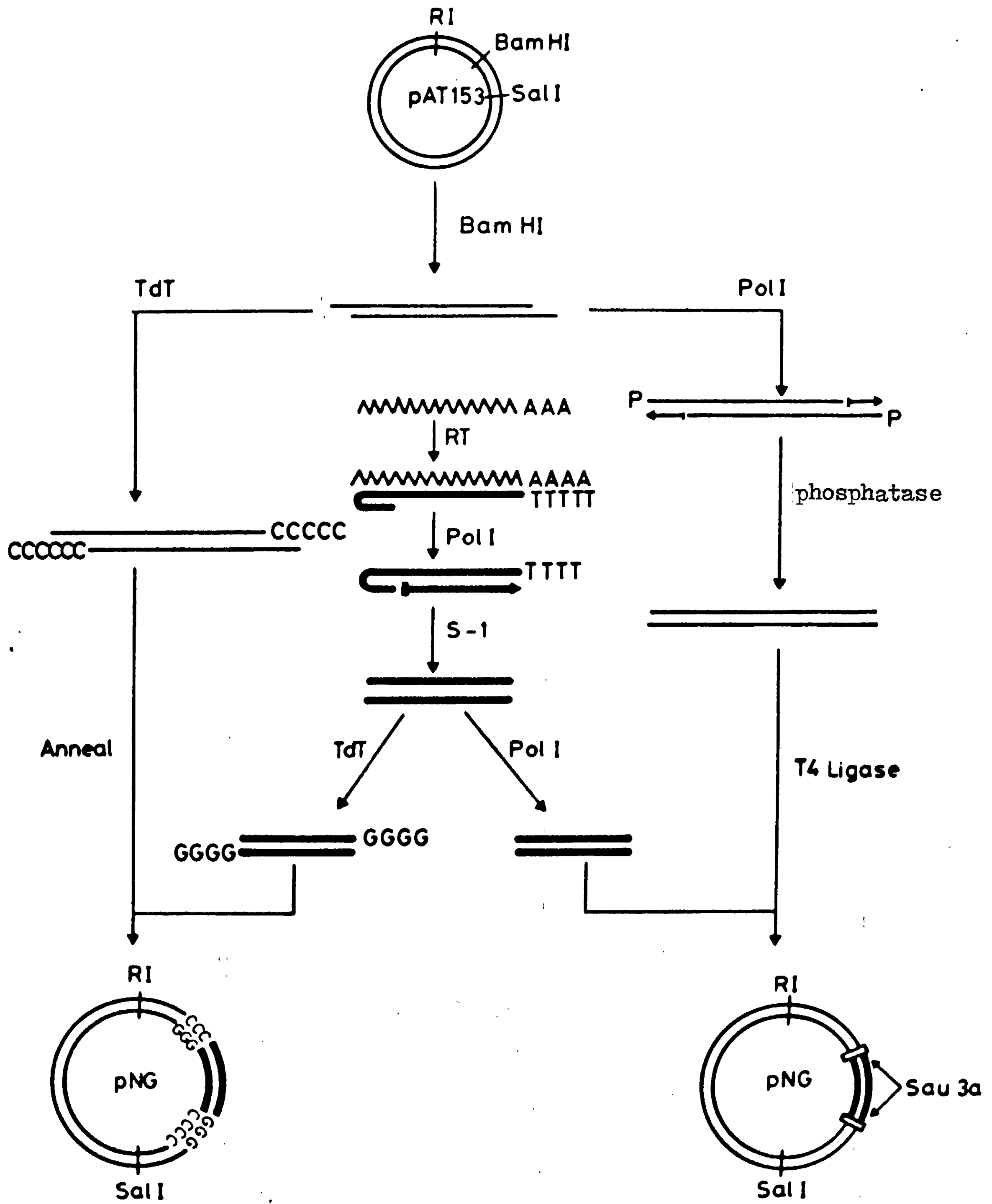
2. <u>Experiments Employing Tailed Plasmid</u>

The "tailing" reactions were all performed with terminal transferase prepared from calf thymus in this laboratory by the method of Bollum (420).

The strategy adopted was to add G-C "tails" as (owing to their high thermal stability) the optimal tail length is somewhat shorter than is necessary for A-T tails. This stability should also help to maximise the number of recombinant molecules formed. One disadvantage of this approach was that it does not lend itself to simple techniques for excising the cDNA insert; like the partial denaturation and S-1 cleavage possible for A-T tailed plasmid recombinants (see ref. 70 for details of this technique). The basic tailing procedures were as described by Humphries <u>et al</u>. (404) with some modifications. All reactions for example used cobalt as divalent cation as this helps to promote extension to recessed 3' termini like those generated by Eco RI cleavage (423). As addition of C residues is also somewhat more efficient, these were added to the plasmid; G residues were added to the essentially blunt ended cDNA.

The time course of C addition to plasmid DNA was monitored by addition of ^{3}H dCTP and assaying the number of

FIGURE 18. Schematic diagrams showing the steps involved in constructing recombinants by the "tailing" and "blunt end" ligation methodologies.



RI
Bam HI
pAT153
Sal I
Bam HI
TdT
Pol I
AAA
RT
P
P
AAAA
TTTTT
phosphatase
CCCCC
CCCCCC
Pol I
TTTT
S-1
Anneal
TdT
Pol I
T4 Ligase
GGGG
GGGG
RI
pNG
Sal I
RI
pNG
Sau 3a
Sal I

counts excluded on Sephadex G-50 at various incubation times. From the total number of counts eluting with the excluded material the number of residues added per 3' terminus was calculated. A similar procedure was adopted for addition of dGTP to the cDNA, except that owing to only limiting quantities being available, pilot experiments used blunt ended Hha I restriction fragments of phage λ DNA. The results are shown in Figure 19. Both reactions are characterised by a rapid phase of addition plus a slower one.

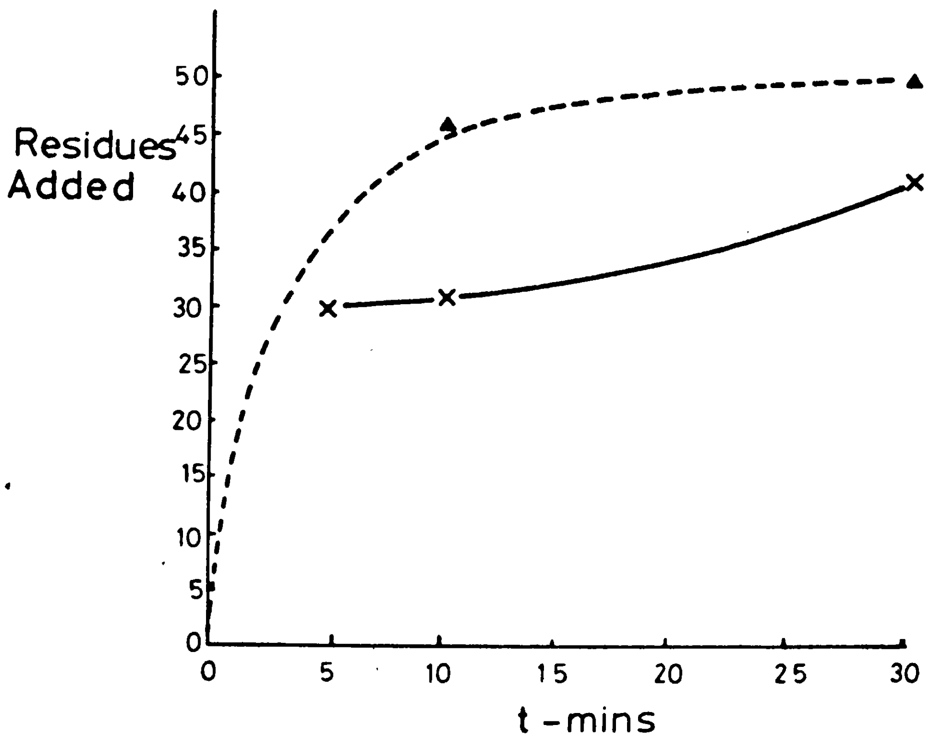
Having established conditions for adding an optimum number of 20 residues (see ref. 424), bulk reactions were performed and the homopolymer extended plasmid and cDNA dialysed against annealing buffer (10 mM Tris-HCl, pH 7.6, 100 mM $NaCl_2$, 1 mM EDTA).

Equimolar quantities of tailed plasmid and cDNA or tailed λ fragments were annealed under a variety of conditions and mixed with 0.1 M Tris before being used to transform competent E. coli cells and plated onto L-Agar plates containing 100 μg/ml Ampicillin and 10 μg/ml Tetracycline. The original annealing conditions were heating to 60°C for 1 hr followed by slow annealing, by allowing the water bath to cool to room temperature. A modification to this annealing protocol was to incubate at 60°C for 4 mins then for 1 hr at 43°C before allowing to cool to 30°C. The DNA mixture was then held at room temperature for 15 mins

FIGURE 19. Time course of homopolymer extension using terminal transferase.

Extension of homopolymer tails was monitored by estimation of radioactivity excluded on Sephadex G-50 using ^3H labelled dCTP- ▲ - or dGTP. -x-

A typical reaction mix contained 2 μg of plasmid DNA, 200 mM Hepes pH 7.0, 1mM $CoCl_2$, 1mM β-mercaptoethanol, 100 μCi of labelled dCTP (or dGTP), 11 μl T dT in a volume of 200 μl.



Residues Added
50
45
40
35
30
25
20
15
10
5
0
0
5
10
15
20
25
30
t - mins

before finally being kept on ice until ready for transformation.

The concentration of plasmid DNA was also increased from 1.4 μg/ml to 4 μg/ml (whilst maintaining equimolar ratio to cDNA) to try and influence annealing rate.
None of these protocols was able to produce recombinant clones containing cDNA or λ fragments however.

One conceivable cause of this was thought to be the possibility that Eco RI tailing was interfering with the promotor of the tet resistance gene which maps close by (425). A repeat experiment was tried plating out the bacteria onto L-agar containing only ampicillin as antibiotic. After 48 hrs, 32 colonies were observed and picked onto master plates for a further analysis.

The colonies continued to grow well in the presence of ampicillin both on plates and in liquid culture. However, no growth could be induced when tetracycline was also supplemented in the medium. This either indicated that the hypothesis was correct, or that the colonies did not contain plasmid DNA. Small scale plasmid isolation by the alkaline extraction technique (411) and Grunstein/Hogness type filter hybridisations (412) were both employed to test for the presence of plasmid DNA and were both found to give negative results. (Data not shown).

At this point it was decided to investigate the possibilities of an alternative strategy, especially as

E.M. investigations of similarly tailed plasmid within this laboratory indicated that the terminal transferase was extending at internal "nicked" sites within the plasmid, and was therefore most likely itself contaminated by exonuclease activity.

3. Experiments employing Blunt-End Ligation

One possible alternative strategy was to adopt techniques similar to those described by Fantoni (426) and by Chambon (409). This approach covalently bonds the cDNA and the plasmids before transformation using the enzyme T_4 ligase, which can perform this function for molecules having either cohesive or blunt-ended termini. (See Fig. 18 right hand pathway).

The first appraisal of the feasibility of this procedure was to linearise a sample of pBR 322 DNA at its unique Pvu II site which creates blunt ended molecules. The cleaved molecules were then incubated with T_4 DNA ligase and used for transformation. After overnight incubation, the antibiotic supplemented plates were found to contain a very large number of colonies, which was not observed on control plates.

The next step was to test the procedure with plasmid opened at the Eco RI site. To utilise such a restriction site it is first necessary to "repair" the cohesive termini and generate blunt-ends. This was performed using the Klenow fragment of DNA polymerase I, which lacks the $5' \rightarrow 3'$

exonuclease activity of the whole polymerase; this enzyme thus functions only as "gap-filler". After plasmid "repair" it was incubated with double stranded cDNA in a reaction mix containing 200 ng of plasmid and 10 ng of d.s. cDNA. One quarter of this material was then used for transformation and plated onto L-Agar plates containing ampicillin (100 μg/ml). After overnight incubation of the plates at 37°C a total of 102 colonies were found to be present, whereas control plates with a similar quantity of plasmid bore roughly one order of magnitude fewer colonies. The colonies derived from the ligated material were allowed to grow to a suitable size and then transferred using tooth picks to a reference agar plate and also streaked onto the surface of a nitrocellulose filter for Grunstein/Hogness screening. The plates were incubated overnight and the filter treated for screening, to lyse the colonies, as described in Materials and Methods. The filters were then hybridised to a ^{32}P cDNA probe prepared from the reticulocyte m-RNA used for the synthesis of the d.s. cDNA employed in the ligation. After washing and drying the filters they were then exposed to autoradiography. The results are shown in Fig. 20. Nine colonies were seen to have hybridised well to the probe and no hybridisation was observed to colonies containing only pAT 153 DNA. These nine positively hybridising colonies were repicked and replated onto both ampicillin and ampicillin plus tetracycline supplemented

L-agar plates. All the colonies grew on tetracycline supplemented medium although with somewhat differing rates of growth.

FIGURE 20. Grunstein/Hogness screening of cDNA recombinants.

Linear "repaired" plasmid (200 ng) was ligated to 10 ng of   s cDNA using T4 DNA ligase in a volume of 22 µl for 16 hrs at 12°C. One quarter of this mixture was used to transform competent E.Coli HB 101. After overnight incubation the colonies were streaked onto nitrocellulose filter overlaying L-agar plates (1.5%) and hybridised to ^{32}P labelled reticulocyte cDNA (2 x 10^5 cpm) for 16 hrs in 3 x SSC, 1 x Denhardt's solution at 68°C with 100 µg/ml salmon sperm DNA. The filters were washed as described and autoradiographed for 4 days.

C. Optimisation of the Technique of Blunt End Ligation

A major obstacle which must be overcome with the technqiue of blunt-end ligation is the problem of the background of drug resistant colonies which contain plasmid lacking a cDNA insert; having merely recircularised on ligation. Fortunately, overcoming this problem is not difficult. By treating the plasmid with alkaline phosphatase prior to ligation, the 5' terminal phosphate moieties can be removed. (409), (424). Under these circumstances the plasmid can then only recircularise if new 5' phosphates are donated by the cDNA. When so treated, the background of colonies containing plasmid alone is very low, representing only a small fraction of the recombinant clones (approximately 10-20%).

Another improvement to the original protocol was the use of alternative restriction sites within the plasmid as the site of cDNA insertion. As described in (409) the Sal I site or Bam HI site (see Fig. 18) can be used to advantage. This accrues from the "repair" reaction performed after the linearisation of the plasmid with either of the above enzymes. Because of the assymetric disposition of Sal or Bam cuts, after repair an extra nucleotide pair is generated at the end of each half of the original hexanucleotide recognition site. The new tetranucleotide sequences which will lie immediately adjoining the insert cDNA can thus be used to excise the cDNA by suitable choice of tetra-

nucleotide recognition enzyme (see Fig. 18). In the case of Bam HI, for example the recognition site is GGATCC. After repair, the cDNA will lie contiguously with the sequence GATC, which can subsequently be cut by the enzyme Sau 3a and thus release the insert DNA. The only disadvantages, of course, being that pAT 153 is cut into approximately eight major fragments itself by Sau 3a and one cannot guarantee that the insert will not itself contain a Sau 3a site.

Optimisation of the ligation reaction was approached in the following way: batches of T4 DNA ligase from various commercial outlets, Miles, BRL and N.E. Biolabs were titrated using a functional assay which measured ligation efficiency by estimating the number of transferred colonies generated by religation of pBR 322 previously cut at its unique Pvu II site. This assays directly the blunt end ligation activity as Pvu II cleavage generates base paired termini.

The results of these experiments are shown in Table 3, the best results being obtained with either of the latter two preparations.

Each subsequent step in the procedure was now monitored by assay of religation and transformation efficiency. Plasmid (pAT 153) was first linearised using Bam HI or Sal I (see Fig. 17b) and the cohesive termini "repaired" using POl I Klenow fragment, as described previously. The plasmids were then incubated with alkaline phosphatase and the transformation efficiency of religated plasmid, before and

TABLE 3. COMPARISONS OF LIGASE EFFICIENCY.

Enzyme	Units/ µg	Colonies/ µg	Efficiency rel. to wild type	Stimulation over linear
BRL	0.5	6×10^2	0.2%	1.25 X
"	1	1.9×10^3	0.65%	3.9 X
"	2	1.3×10^4	4.5%	27.0 X
"	16	6.9×10^4	23.8%	143.8 X
Miles	2.7	1.1×10^3	0.56%	2.2 X
"	5.4	0.9×10^3	0.45%	1.8 X
"	10.8	0.7×10^3	0.35%	1.4 X

Reactions were performed in a total volume of 6 µl in a mix containing 20 ng of plasmid DNA and varying quantities of ligase as shown above in the buffer described in Materials and Methods.

after phosphatase treatment was tested, as shown in Table 4. By comparison with the ligation efficiency of Pvu II cut plasmid, the "repaired" plasmids were able to achieve about 20% of that efficiency which indicates that roughly 1 in 5 molecules was completely base paired after repair. In contrast, the religation efficiency of phosphatased plasmid was very low, nearly two orders of magnitude below that observed for repaired non-phosphatased molecules.

A final test was now to mimic the addition of blunt-ended d.s. cDNA by using Hae III fragments of pAT 153 DNA. The results of this test are shown in Table 5. Whilst phosphatased plasmid displayed similar transformation efficiency as previously, when Hae III fragments were present to act as terminal phosphate donors a marked increase in transformation frequency was observed. This varied from nearly 20 fold increase for Sal cut plasmid to 60 fold for Bam cut. The latter stock of plasmid was therefore used for subsequent ligation reactions.

The reticulocyte cDNA library was established using the above protocol but two further modifications were found necessary to generate recombinants with these later cDNA preparations. Firstly it was found necessary to reincubate the cDNA with DNA polymerase to ensure that the ends were truly flush, secondly the ligation incubation was grossly extended from 20 hrs to 44 hrs with an addition of a fresh aliquot of enzyme after the first 20 hrs of incubation. These modifications resulted in final ligation efficiencies of about 1-1.5 x 10^4 colonies per microgram of input plasmid DNA. Simple variations from equimolar ratios of plasmid and cDNA were tried (eg. 2:1 and 1:2) but no advantage was found to accrue.

TABLE 4. COMPARISON OF LIGATION EFFICIENCY OF REPAIRED AND PHOSPHATASED PLASMIDS.

Condition of Plasmid	Bam HI	SAL I
Repaired–Ligase–Phosphatase	5.6×10^2	1.1×10^2
Repaired–Ligase+Phosphatase	6.4×10^2	4×10^1
Repaired+Ligase–Phosphatase	7.4×10^4	7.1×10^4
Repaired+Ligase+Phosphatase	1.1×10^3	8.4×10^2
Native	1.7×10^6	

Reactions were performed in a total volume of 6 µl in a reaction mix as described in Materials and Methods, using 50 ng of plasmid DNA.

TABLE 5. ABILITY OF EXOGENOUS HAE III PLASMID FRAGMENTS TO RESTORE LIGATION TO PHOSPHATASED PLASMID.

Conditions of Ligation	Bam HI	SAL I
-Ligase-Hae III fragments	-	-
-Ligase+Hae III fragments	-	6.2×10^{2}
+Ligase-Hae III fragments	1.5×10^{3}	2.8×10^{3}
+Ligase+Hae III fragments	9×10^{4}	5.2×10^{4}
Native plasmid	1×10^{6}	

The above reactions were performed using 20 ng of plasmid DNA, 1 ng of Hae III fragments (i.e. equimolar ratios) and 1 μl of Ligase in a total volume of 5 μl, in the ligation buffer described in Materials and Methods.

D. Screening Protocols and Development of Competition Hybridisation Assay

Having established a cDNA library, it was necessary to devise a methodology by which a reasonable number of clones could be rapidly screened with globin probes, to select out those of interest, which might carry non-globin cDNA inserts.

The first technical advance was the evolution of techniques for permanent storage of recombinants in an ordered array. The recombinant colonies were picked (with tooth picks) from the original plates, after spreading and growth of the transformation mix, into wells of 96-well microtitre plates, as used for tissue culture. Problems were experienced here with finding a suitable medium in which to grow the colonies, but the publication of Wensink and his colleagues (426) provided details of a highly suitable medium. This enabled good growth and by virtue of its glycerol content allowed permanent storage (at -20°C) by freezing. Replicas of such microtitre plate collections could be generated very simply by the use of transfer plates (Cooke Microtitre Ltd.).

Moreover this technique provided a simple way of making multiple copies of the plasmid collection on nitrocellulose filters, for screening with different probes.

The first approach to identifying the globin cDNA containing colonies was to label α and β globin cDNA clones (gift of C. Weissman) by nick translation. Hybridisation with such a probe however was found to be valueless owing to the homology which exists between the pCRI plasmid (in

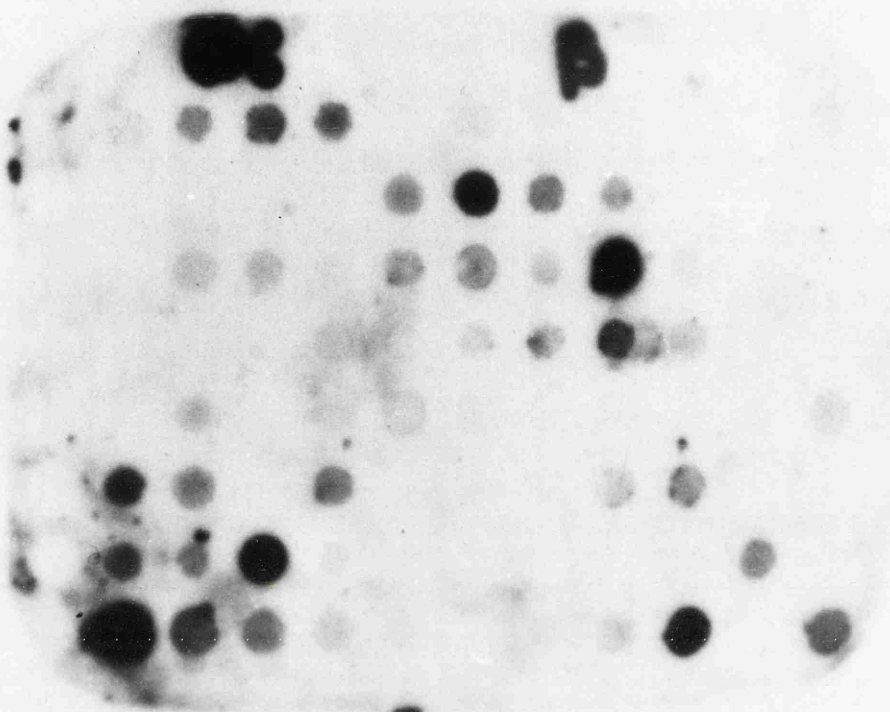
which the globin cDNA's were cloned) and the pAT 153 of the reticulocyte cDNA library. Attempts to remove this background by addition of gross excesses of competing pAT 153 DNA or E.coli DNA to either prehybridisation or hybridisation steps proved futile.

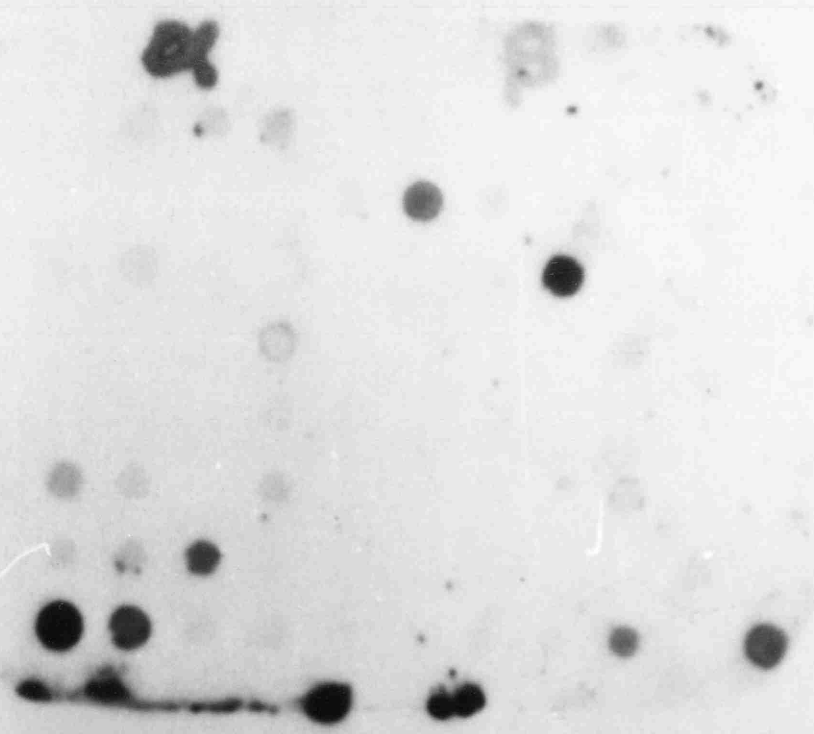
Eventually a competition hybridisation assay was developed using the α and β globin cDNA plasmids to compete cut globin sequences within the reticulocyte cDNA probe, used for screening the recombinants. A number of factors were found to be important. Again, initial attempts to effect competition, by adding the globin plasmids into the hybridisation mix had negative results. Additionally it was found obligatory to cut the pCRI globin plasmids, with an enzyme such as Hae III to reduce it to small pieces. Otherwise problems with non-specific background (which appeared to stem from concatameric associations of plasmid) were found to occur.

The final protocol adopted required a very large excess of the globin plasmids, restricted and denatured and then hybridised in solution to the ^{32}P cDNA probe (see legend to Fig. 21). The hybridisation of such a competed probe could then be compared to the results obtained for an identical amount of probe hybridised to a duplicate filter of recombinant colonies. Such a screening is shown in Fig. 21. Colonies which were reduced markedly in hybridisation signal intensity (like the control α/β globin cDNA clones) in the presence of competing cloned α and β cDNA sequences were

FIGURE 21. Competition hybridisation assay.

Duplicate nitrocellulose filters bearing reticulocyte cDNA recombinants were subjected to Grunstein/Hogness hybridisation as described for the previous figure. The uncompeted filter was hybridised to ^{32}P labelled reticulocyte cDNA (2×10^6 cpm). The competed filter was hybridised with an equal amount of ^{32}P reticulocyte cDNA, previously hybridised to a large excess of restricted and denatured pCR 1 α and β globin cDNA recombinants. After hybridisation the filters were washed and autoradiographed as described previously.


UNCOMPETED


COMPETED

scored as globin cDNA recombinants. Those colonies found to be unaffected by competition were considered putative non-globin cDNA recombinants.

From a screening of 96 colonies (all of which had scored positive on two independent screenings with uncompeted reticulocyte cDNA probe) ten were selected which appeared to be insensitive to competition on two successive independent screens. These ten colonies only were then subjected to competition hybridisation once more, in the presence of appropriate controls (see Fig. 22). With the possible exception of one clone (pNG-1) none showed any evidence of competition so all were designated putative non-globins and identified as pNG1-10.

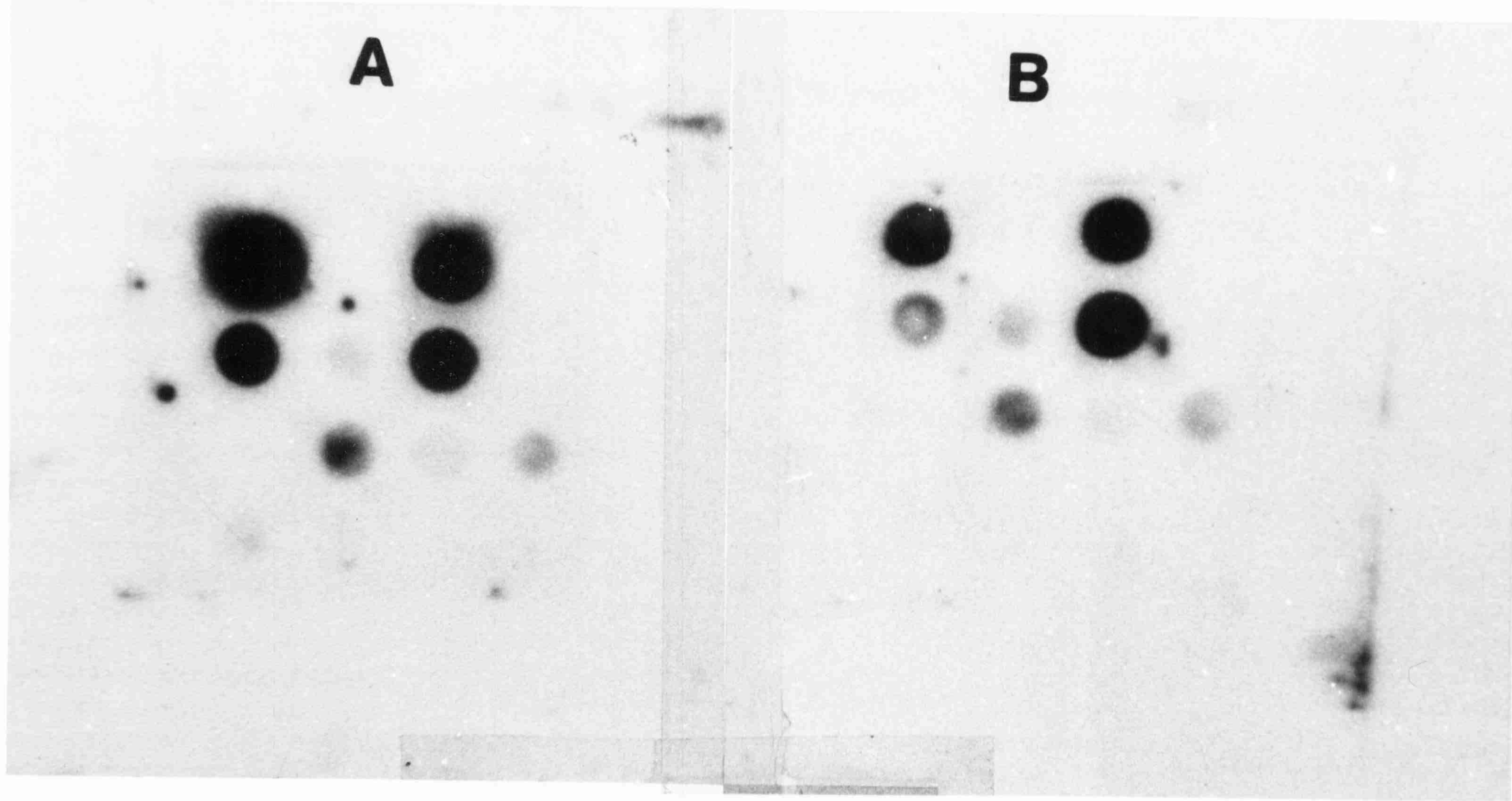
FIGURE 22. Competition hybridisation of putative "non-globin" recombinants.

Ten reticulocyte cDNA recombinants selected from the previous competition hybridisation were subjected to this procedure once again to confirm their assignment as "non globin" cDNA clones. The individual clones were designated pNG 1-10 as indicated in the schematic shown. The three controls are indicated by the letters a, p, b, representing the pCR1 α, pAT 153, pCR1 β clones respectively.

A) filter hybridised to uncompeted probe.

B) filter hybridised to competed probe.

Not all the "non globin" clones show visible hybridisation on this exposure.



a	p	b	
1	2	3	4
5	6	7	8
9	10		

III. CHARACTERISATION OF NON-GLOBIN CLONES

A. Use of Purified Globin Probes

As the non-globin cDNA clones had been selected on the basis of only a single criterion, namely the competition hybridisation assay, it was felt necessary therefore to prepare pure globin probes with which to challenge these clones.

Purified α and β specific probes were prepared in the following manner: plasmid containing α or β globin cDNA sequences was bound to nitrocellulose filters by the technique described in Materials and Methods. The filters were then hybridised to a total reticulocyte (^{32}P) labelled cDNA probe (see legend to Fig. 23). The filters were then washed to remove non-specifically bound counts, and the specifically hybridised globin sequences eluted from the filters by incubation in 70% formamide buffer at high temperature (68°C). The process was followed with a β-decay monitor, and a filter bearing pAT 153 DNA only was used as control. An α filter after washing, hybridised radioactivity to a level of approximately 500 c.p.s., whereas a pAT filter did not show retention of label above 2-3 c.p.s. The hybridisation was therefore shown to be sequence specific. Moreover, after elution no more than 10% of the bound counts remained on the filter.

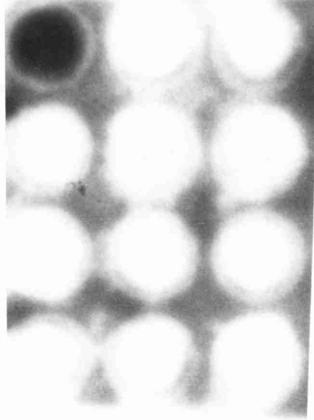
Following elution of the α and β specific cDNA sequences the probes were brought to 50% formamide and 3 x SSC suitable

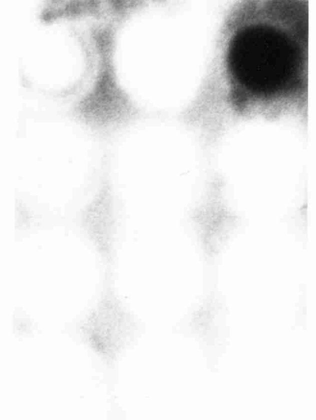
for hybridisation to two identical filters bearing the non-globin recombinant colonies, treated for screening as described previously. The results are shown in Fig. 23 and demonstrate the specificity of the probes. The α and β probes do not cross hybridise, nor do they hybridise to pAT 153 DNA, as would be expected. In addition none of the colonies previously designated non-globins show significant hybridisation which confirms the original designations.

FIGURE 23. Hybridisation of "non globin" recombinants with purified α and β globin cDNA probes.

Duplicate nitrocellulose filters bearing the ten "non globin" clones and suitable α and β globin and pAT 153 controls as indicated in the schematic were hybridised to α or β specific globin cDNA purified in the following manner: pCR1 α and pCR1 β plasmid DNA (10 μg) were bound to separate 1.3 cm. diameter nitrocellulose filter discs as described in Materials and Methods. The discs were prehybridised in a solution containing 3 x SSC, 1 x Denhardt's solution, 50% formamide for 4 hours at $42^{o}C$. The filters were then hybridised overnight at $42^{o}C$ in the same solution supplemented with 100 μg/ml salmon sperm DNA and reticulocyte cDNA probe (4×10^{6} cpm), (these reactions were performed in 5 ml glass bijoux). After hybridisation the filters were washed extensively at $50^{o}C$ in 0.1 x SSC, 0.1% SDS. Each filter was then transferred individually into 0.7 mls of a solution containing 70% formamide, 0.1 x SSC, 1 x Denhardt's solution, 100 μg/ml carrier DNA and incubated at $68^{o}C$ for 4 hours to melt off hybridised cDNA. This solution was then brought to the same concentrations as the original hybridisation mix and used to challenge the filters bearing recombinants described above.

The left-hand filter is probed with α cDNA; the right-hand filter with β cDNA.





A	pAT	B
1	2	3
5	6	7
8	9	10

B. Restriction Analysis of Non-Globin Recombinants

Restriction digests were employed to further characterise the recombinant plasmids. The simplest analysis is shown in Fig. 24 and shows the recombinants cut with Hinc II and EcoRI. This generates three fragments, a large fragment containing most of the plasmid DNA a small 452 bp fragment and a fragment containing the Bam HI site, normally 652 bp long, in the wild type pAT 153. In all cases it can be seen that the fragment containing the Bam site is shifted to a higher molecular weight than wild type indicating that the recombinants have material inserted at this restriction site. With the exception of clone pNG 5 none of the recombinant plasmids seem to have undergone gross modification during cloning. The plasmid DNA of clone pNG 5 however appears to have suffered extensive deletion. The insert sizes indicated by this gel are all moderate, lying in the region of 100-200 bp. with the exception of clone pNG 3 which seems to have an insert size of about 900 bp.

More precise size mapping of the inserts was performed by 6% polyacrylamide gel electrophoresis of digests using enzymes which cut more frequently. The insert sizes indicated by these digests are summarised in Table 6. In virtually all cases there is good agreements between insert size estimates obtained from different restriction digests.

Fig. 25 shows an extremely useful digest, that obtained from Hin f. This digest when coupled with Eco RI digestion provides fragments of 1000, 630, 520, 400, 300, 220 and 150 bp.

TABLE 6. RESTRICTION DIGEST SIZE ANALYSIS OF RECOMBINANT CLONES.

Clone	Enzyme Digest						Average Size Estimate (bp)
	Hinc II/ Eco RI	Hinf I/ Eco RI	Hae III	Taq 1	Dpn 1 (Sau 3a)	Ava II/ Eco RI	
pNG 1	155	150	160	160	165	140	155 ± 8
2	180	180	230	210	300[a]	200	200 ± 17
3	950	820	775	1000	1200[a]	950	900 ± 78
5[b]	-	-	-	-	-	-	-
6	175	160	170	175	195	180	176 ± 10
7	200	160	180	190	210	170	185 ± 17
8	90	600	1370	330	600	300	-[c]
9	200	180	200	195	200	170	191 ± 12
10	90	100	95	65	-	85	87 ± 11

a Estimated sizes only owing to the presence of deletions. Not included in size estimates.

b Sizes of inserted DNA cannot be determined owing to the large deletion in this plasmid.

c Owing to extreme variability in size, estimates, average was not estimated.

FIGURE 24. Restriction digest of "non-globin" recombinants using Hinc II/Eco RI double digest.

The results of this digest were visualised by ethidium bromide staining of a 2% agarose gel run in T.E.A. buffer. The numbers at the top of the tracks identify the "non-globin" clones, lane p contains a similar digest of pAT 153. The band indicated by an arrow is the band containing the Bam H1 site into which the reticulocyte cDNAs were cloned. The size markers in lane **H** are derived from a mixture of a Hin F and a Hin F/R1 double digest.

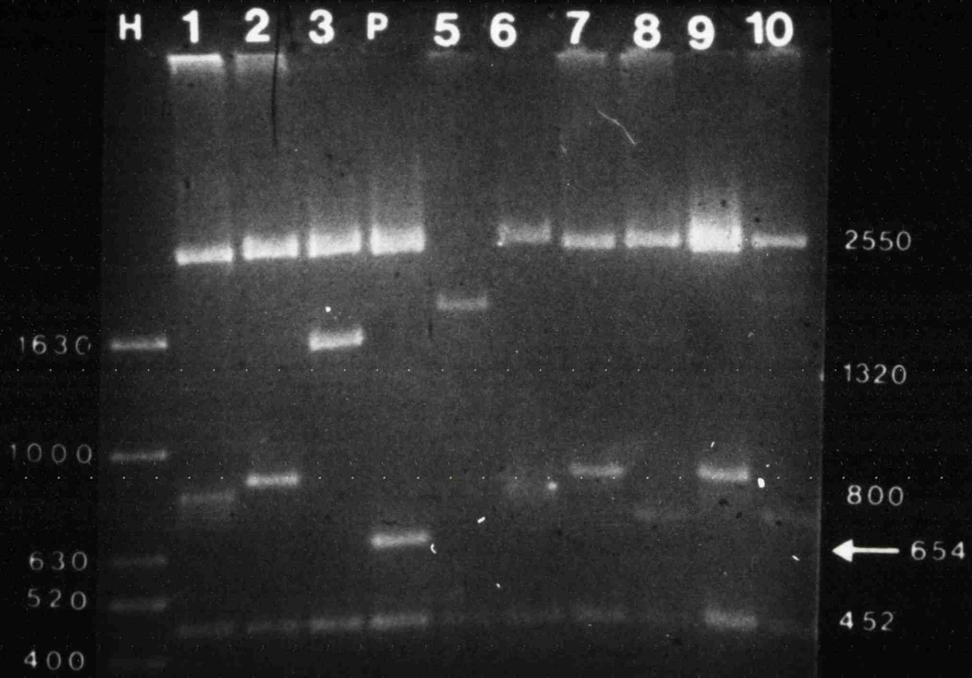
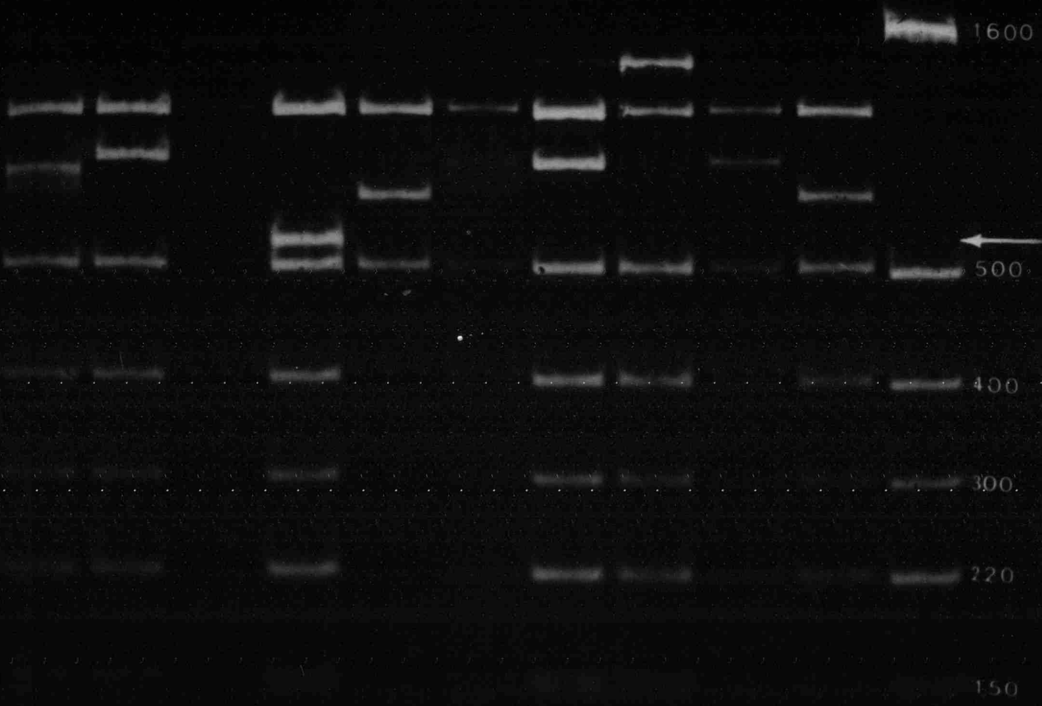

HINCII/RI
H
1
2
3
P
5
6
7
8
9
10
1630
1000
630
520
400
2550
1320
800
654
452

FIGURE 25. Restriction digest using a Hin F R1 double digest.

The clones are identified as described in the previous figure. The lanes marked p contain digests of pAT 153; the central lane being a double digest and the end lane a Hin F digest alone. The arrow again indicates the fragment containing the Bam H1 site. This analysis was performed on a 6% T.E.A. polyacrylamide gel.



HINF / R1
1
2
3
p
5
6
7
8
9
10
p
1600
500
400
300
220
150

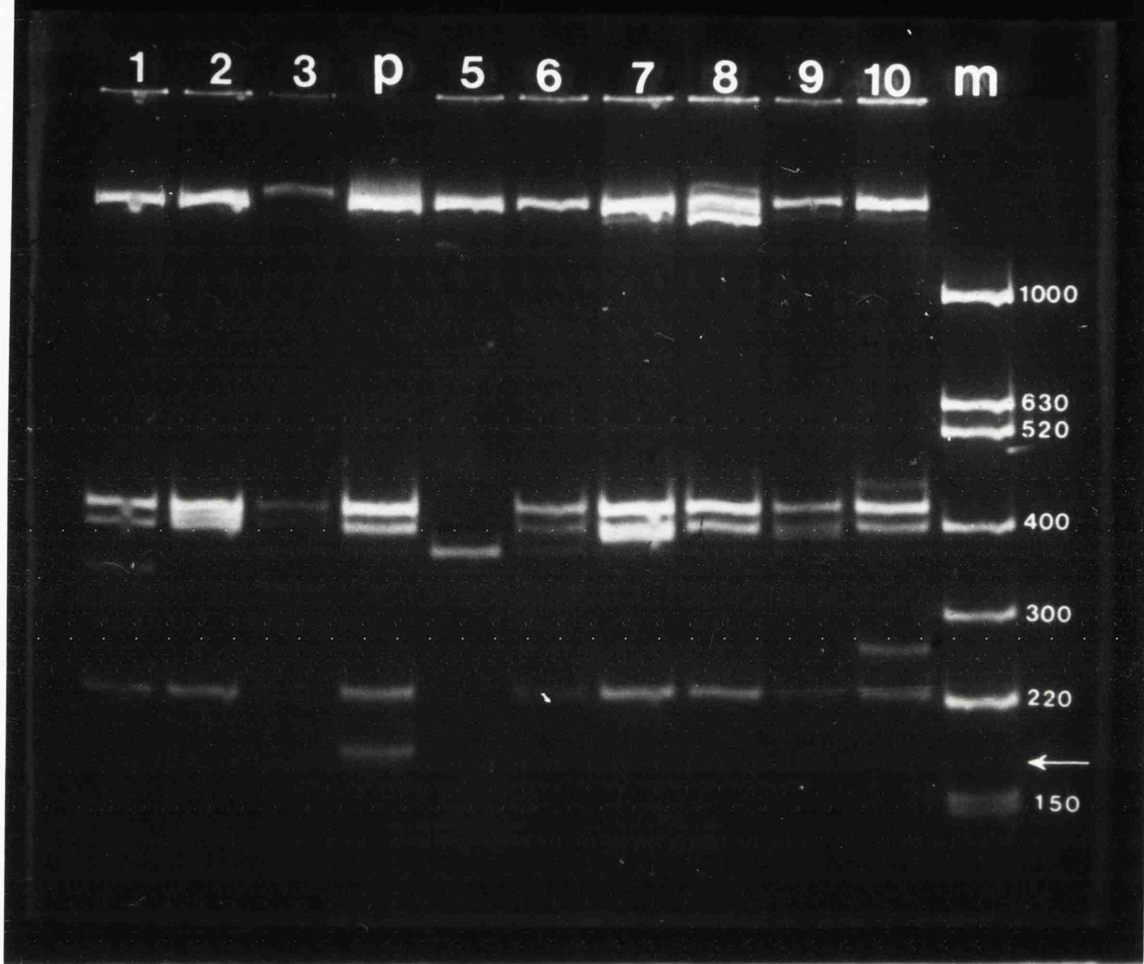
which thus makes an extremely useful set of markers for calibrating other digests. Sometimes these markers are a mixture of Hin f/RI and Hin f alone as this latter digest provides an additional marker at 1.6 Kb; this fragment being the one cleaved by Eco RI to generate the 1000 bp. and 630 bp. fragments.

Insert size estimates from this digest all accord well with the previous digest excepting clone pNG 8 which appears to have a very large insert, similar in size to that of pNG 3. (However see later blotting data in Fig. 30). The Hae II digest shown in Fig. 26 is most notable for the fact that clone pNG 3 is cut at sites within the insert to generate three new fragments. This is of some interest as Hae II sites are generally rare within the eukaryotic genome. This digest is also a fairly useful check on the presence of small deletions as the band containing the Bam site (arrowed) is only 181 bp. in length. Again pNG 8 is indicated as having a very large insert.

Interpretation of the Taq I digest in Fig. 27 is somewhat more complicated owing to the presence of a specific partial digestion product, which under conditions of enzyme excess used here has cleaved in some cases but not in others. The fragment containing the site which cleaves very poorly is further made cryptic by virtue of the fact that, as a composite of a 141 bp. and a 475 bp. fragment, it comigrates on this gel with the bona fide 622 bp. fragment. Only tracks for pAT 153 and pNG 2 show the presence of any 141 bp fragment

FIGURE 26. Restriction digest of recombinants with Hae II.

This analysis was performed as described in the preceding figure. The lane containing the size markers (lane m) shows a Hin F/RI double digest of pAT 153.


HAE II
1
2
3
p
5
6
7
8
9
10
m
1000
630
520
400
300
220
150

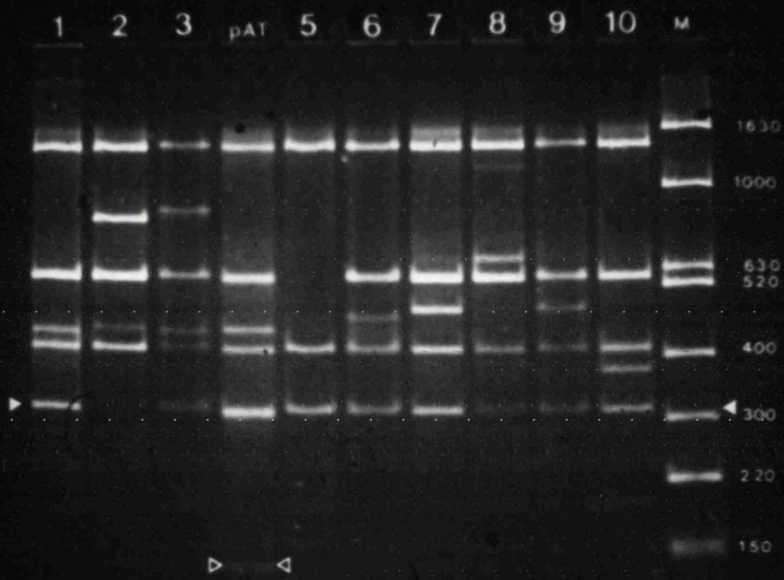
and thus some of the associated 475 bp. band. Clones pNG1 and 3 however show the presence of a fragment close in size to the 475 bp. fragment but in the absence of any 141 bp. band. Thus these fragments of close to 475 bp must originate from the insert itself plus sequences which normally runs as a 315 bp. band (part of the doublet which is arrowed) which contains the Bam site. The plasmid of clone pNG 3 shows the presence of an additional high molecular weight band thus establishing a Taq I site within the insert. Neither the Taq site nor the Hae II sites are found in the coding sequences of any of the globin genes sequenced to date. Taken with the insert size data for this clone (900 bp.), this is fairly strong corroborative evidence that this clone contains a non-globin insert. (See also tetranucleotide analysis in Fig. 29).

The presence of a deletion of sequences in the pNG 2 plasmid is revealed by the absence of the 312 bp. fragment which lies contiguously with the 315 bp. fragment that contains the Bam site. This deletion presumably extends in the anticlockwise direction (towards the RI site) and thus the junction of the two Taq fragments is missing.

Similar evidence for this deletion was obtained from Dpn I or Sau 3a digest of this clone. Evidence for a related type of deletion of sequence in clone pNG 3 was also revealed by Dpn I digestion. However this deletion cannot be quite so extensive as that seen in pNG 2 as the Taq site absent for the latter clone is still preserved in clone pNG 3 (see Fig. 27).

FIGURE 27. Restriction digest of recombinants with Taq 1.

This analysis was as described for Figure 25. The markers are as described for Figure 24. The fragment indicated by the solid black arrows carried the Bam site, the fragment indicated by the arrows shown in outline is the 141 bp fragment referred to in the text.


TAQ 1
1
2
3
pAT
5
6
7
8
9
10
M
1630
1000
630
520
400
300
220
150

As described earlier however the true significance of the Dpn I or Sau 3a digest is that these enzymes both recognise the sequence GATC which should lie in immediate juxtaposition with the inserted cDNA (see Fig. 18). Digestion with either of these enzymes should liberate all the plasmid fragments plus an addition fragment or fragments corresponding to the excised insert. Fig. 28 demonstrates the clearest example of this for clone pNG 7, which reveals the presence of a novel 230 bp. fragment specific to this clone. Three other clones pNG 1, 6 and 9 all behave in a similar fashion and presumably possess intact GATC sequences on either side of the inserted cDNA, although these seem more resistant to cleavage than pNG 7.

One other digest Ava II/RI was performed. This gave very similar results to the Hin f/RI digest so the results are shown only in Table 6. A feature of note in this digest was the absence of cleavage at site 1438 on the map according to Sutcliffe. (421),(427). The 303 bp. fragment was therefore replaced by a 345 bp. fragment, the contiguous 42 bp. piece not having been cleaved from it.

Interestingly, but perhaps not unsurprisingly, only pNG 3 showed any evidence for sites within the insert DNA, most likely due to the small size of the inserted material in the other cases. As clone pNG 3 did possess a fairly large insert, it was digested with a number of enzymes which cut more frequently than those so far discussed.

FIGURE 28. Dpnl digest of clone pNG 7.

Lane p) Dpn 1 digest of pAT 153.

Lane 7) Dpn 1 digest of pNG 7.

Lane M) size markers as indicated.

The arrows denote the novel insert fragment specific to the recombinant digest.

2. Tetranucleotide Mapping of Clone pNG 3

Enzymes which cut at tetranucleotide recognition sites were used to look for internal sites within pNG 3 insert DNA. The results are summarised in Table 7 and Fig. 29 The figure displays the data as pairs of tracks comparing pNG 3 with pAT 153. Novel fragments, specific to pNG 3 are distinguished by arrows, other features of note are indicated using appropriate symbols. One interesting anomaly was found in the Alu digest. No fragments at 226 or 136 bp. described by Sutcliffe, can be observed in the pAT 153 digest. Instead there is a new fragment seen at approximately 93 bp. plus what appears to be a doublet at 257 bp. (See legend to Fig. 29). A conceivable interpretation is that the 226 bp. plus the 136 bp. fragments have reassociated, possibly during the deletion of material from pBR 322 to generate pAT 153. The new disposition of sites then would create 93 bp. and 257 bp. fragments. However it should be pointed out that virtually all the composite fragments formed where they lie across the junction of the pBR 322 deletion are the sizes that would be predicted from the proposed Hae II coordinates (see Section II.B1).

A number of other small anomalies in the pAT 153 map have been located, for example there is a 303 bp. fragment seen in the Hha I digest (identified by a black square) which was most likely generated by failure to cleave at the junction of the 270 bp. and the 33 bp. fragments observed

in the pBR 322 sequence (421),(422). Another possible missing site is located at the junction of the largest Alu fragment, which should be 659 bp. in size. However it is unlikely that this would resolve from the 655 bp. fragment indicated below. Thus it seems that the contiguous 63 bp. fragment may still be attached, pushing this fragment to a size of 722 bp, which accords more closely with the observed distance of migration.

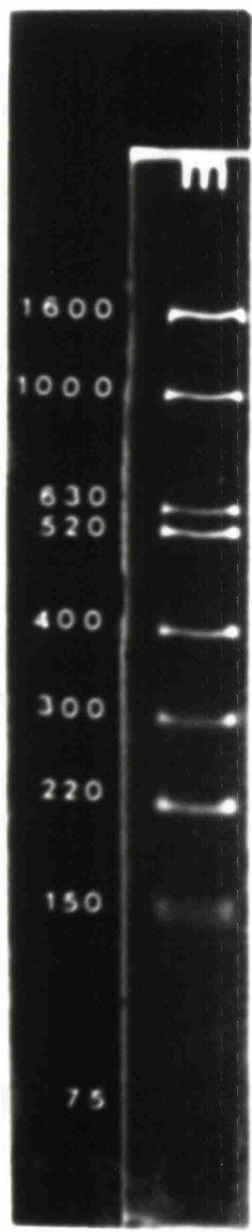
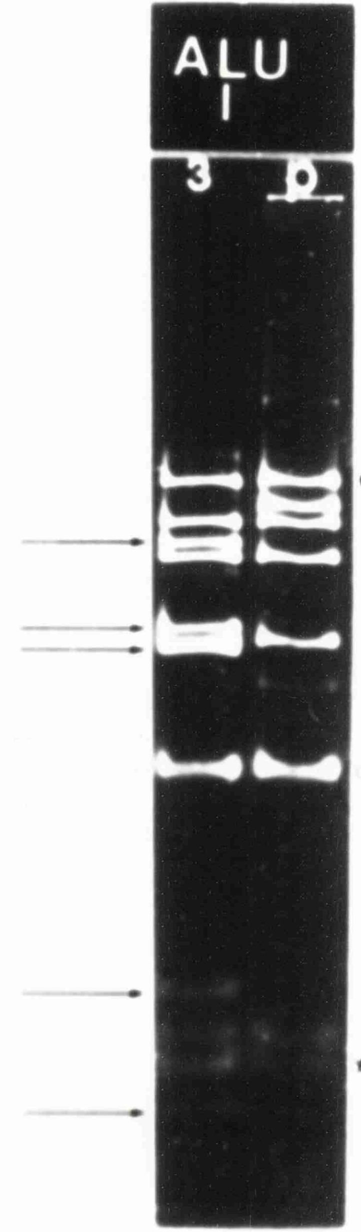
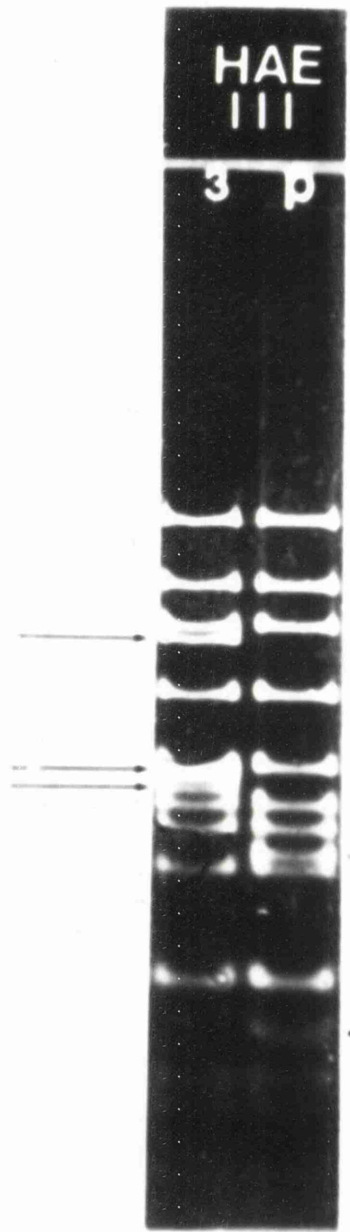
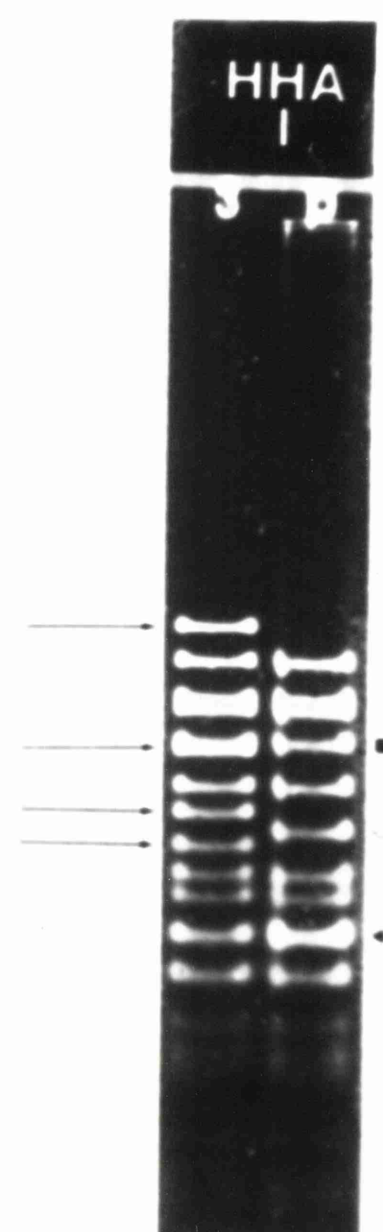
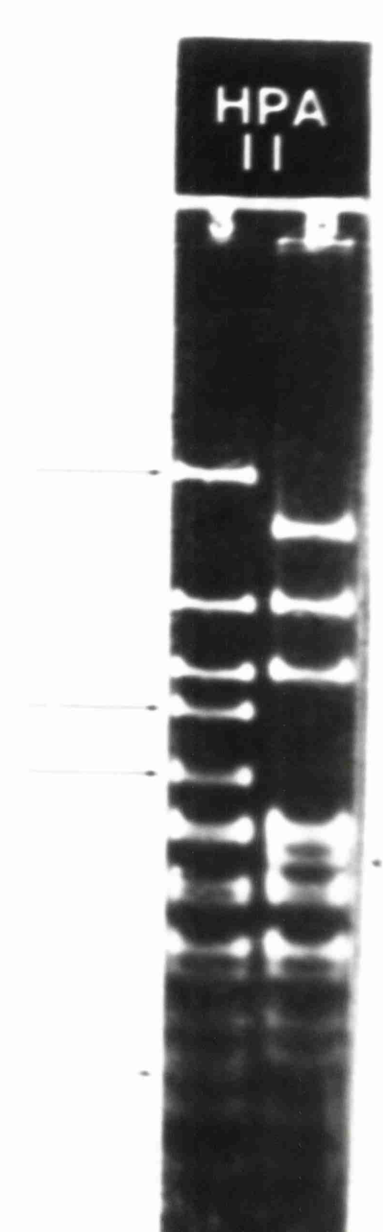
However, the 659 bp. Alu fragment also in fact maps over a region where a number of other fragments show somewhat anomalous migration for reasons not obviously explicable in the same terms. This applies to the top Hae III and Hpa II fragments and the 206 bp. Hha I fragment (see Figure legend) all of which migrate somewhat slower than predicted (i.e. apparently larger in size) from their reported sizes. This can be ascertained from the fact that a calibration made up from all the digests shown, plotted as size vs. migration distance as a semi-logarithmic relationship is in fact highly accurate and consistent.

The most perplexing observation has in fact been drawn from the Hin f/RI digests where, in digests analysed on 6% polyacrylamide the 520 bp. marker runs anomalously at approximately 580 bp, however if the digest is examined on 2% agarose (in the same buffer system) this fragment runs true to its expected size. (Compare for example its separation from the 630 bp. fragment above it in Figs. 24 and 29). It seems difficult to invoke secondary structure

FIGURE 29. Tetranucleotide cutting enzymes analysis of clone pNG 3.

The analysis was performed on 6% polyacrylamide gels as described previously. The tracks are arranged in pairs with the enzyme used for digestion as shown. Lanes 3 contain digests of clone pNG 3, lanes p digests of pAT 153. The markers are as described for Figure 24.

The arrows represent novel fragments generated by a cleavage within the inserted cDNA of clone pNG 3. The fragments indicated by black triangles are the fragments which contain the Bam site. The sizes of these fragments are also indicated. Fragments indicated by hollow triangles (size also indicated) delineate fragments deleted or absent from pNG 3. The fragments in the Alu 1 and Hha 1 designated by solid symbols (black spot and black square respectively) represent fragments migrating anomalously as referred to in the text. The remaining fragments marked in the Alu digest (white spot and black star) represent fragments not predicted from the mapping data of Sutcliffe (421).


m
1600
1000
630
520
400
300
220
150
75

ALU I
3
p
655

HAE III
3
p
192
104

HHA I
3
p
206
152

HPA II

in a fragment of this size being responsible for such a difference in behaviour. Perhaps some quirk of base composition is responsible; this interpretation would perhaps carry some weight with regard to the overlapping and anomalously migrating fragments pointed out above.

With regard to pNG 3 itself one or two other strange features are apparent also. However, the first point to note is that the pAT fragment containing the Bam site (indicated in the pAT digests by a small black triangle) is in all cases changed in molecular weight in the recombinant plasmid, confirming that the cDNA is indeed inserted at this position.

The disposition of internal sites is indicated in Table 7, all the enzymes used here, cut internally at least once. Again a tendency towards a high G-C content is implied by this data suggesting that this protein may be somewhat atypical in its base sequence composition compared to other eukaryotic cDNAs. However, the highly ubiquitous Alu sites are typical of other eukaryotic sequences.

Evidence for a small deletion of material at the junction of cDNA and the Bam site used originally to linearise the plasmid was apparent from a Tha I digest. This confirms the Sau 3a data referred to earlier. The extent of this deletion cannot be much greater than 30 bp. however, as the Taq I site only 10 bp. removed from the missing Tha and Sau 3a sites was seen to be present (see Fig. 27). Less explicable is the Hpa II digest of pNG 3 where the junction of the

TABLE 7. DISTRIBUTION OF RESTRICTION SITES IN INSERT OF CLONE PNG 3.

Number of sites	0	1	2	3	4
Enzymes	EcoRI	Taq I	Hae II		Alu I
	Hinc II	Hha I	Hae III	Hha I	
	Pst I			Hpa II	
	Sau 3a				
	Hinf I				
	Hind III				
	Ava II				

217 bp. fragment indicated and the top fragment must be missing. The 217 bp. fragment contains the Bam site and lies contiguously with the largest Hpa II fragment, which by virtue of material still attached is raised to a higher molecular weight in the pNG 3 digest. Accordingly, there is no fragment seen at 192 bp. in the pAT Hae III digest (indicated by a hollow triangle) which has a 3' junction very close to these Hpa II sites referred to. Similarly, the partially overlapping 206 bp. Hha I fragment is also missing from the pNG 3 digest. This data would indicate that at least 70 bp. were missing from this region of the plasmid (70 bp. is the distance from the 3' end of the largest Hpa II fragment to the 3' end of the 206 bp. Hha I fragment). Were this the case, then all this material would be missing from a single Taq I fragment normally 312 bp. in length, such a change in molecular weight would be clearly visible in the Taq digest shown in Fig. 27 . This fragment however is quite unaffected which poses some difficult problems of interpretation. As the two digests were made on different plasmid DNA preparations it may just be possible that this represents material deleted during the subsequent handling of pNG 3. However as plasmid preparations of pNG 3 were not prepared from material grown from a single colony this possibility seems highly unlikely. These rearrangements may alternatively stem from changes in sequence due to point mutation, or conceivably altered methylation patterns of

plasmid DNA. Notwithstanding such alternatives, no satisfactory explanation for these restriction patterns has been found to date. Data from plasmid sequencing would no doubt be able to resolve this question, but this was considered of little significance in relation to more important aspects regarding the nature of the inserted sequence.

C. <u>Southern Blot Analysis of Non-Globin Recombinant Clones</u>

The next approach was to utilise the restriction data to further characterise the non-globin clones by hybridisation. The first objective was to establish that the previously identified novel fragments seen in the recombinant digests did in fact contain the inserted cDNA. An additional advantage of this approach was that by choice of suitable digest the pCRI globin cDNA clones could be used as probes, as the homology between the pCRI and pAT 153 plasmids resides in regions of pAT separable from that region containing the Bam HI site.

The nine non-globin recombinants were restricted with Hin f and Eco RI to generate a pattern like that shown in Fig. 25. This digest was chosen as all the fragments still resolve clearly on a 2% Agarose gel. The Ethidium bromide stained pattern of this digest separated on agarose is shown in Fig. 30a. Fragments of pAT 153 DNA were also run as markers, these being a mixture of a Hin f digest and a Hin f/RI double digest. One of these tracks also contained radio-labelled fragments, labelled by use of reverse transcriptase as described by Stark (77). The globin plasmids were also restricted and run on this gel, but they were digested with the enzyme Hha I which generates a large fragment containing the globin cDNA sequences and many small fragments which do not resolve properly on a 2% gel.

The DNA fragments were then transferred to nitrocellulose by the Southern blotting technique (76) using 20 x SSC as

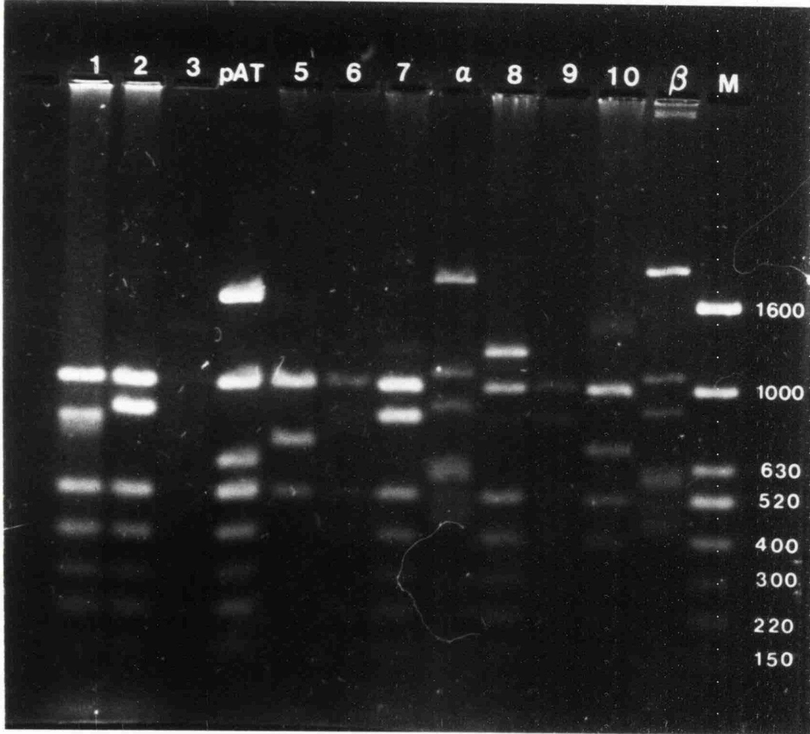
transfer buffer. After baking as usual to fix the DNA the filter was hybridised with ^{32}P labelled reticulocyte cDNA probe. The resultant autoradiograph is shown in Fig. 30b. Close comparison with the accompanying figure (Fig. 30a) demonstrated that in virtually all cases the fragment which hybridised with reticulocyte cDNA was the fragment specific to the recombinant digest, which contained the Bam HI site in the wild type plasmid. The only exception to this rule was the clone pNG 8 which seems to have hybridised material to a very minor fraction of the digest and not to the major new fragment, which can be clearly seen migrating at approximately 1.3 Kb in size. However, as reference to T ble 6 will show, this clone has given extremely variable size estimates for the inserted cDNA, for reasons which are still obscure. The hybridisation data presented here are consistent with an insert size of about 200 bp, which in fact is in good agreement with the original size estimate derived from the HincII/RI digest of Fig. 24. This recombinant, as well as clone pNG 10 also shows evidence of weak hybridisation to a second fragment which is not visible on the ethidium stained gel shown here, but which may be those seen at a similar position in the Hin f/RI digest of Fig. 25. This fragment could derive from some low level of RI* activity causing cleavage within the insert. If this is indeed the case then clones pNG 8 and 10 may well contain similar sequences inserted in them.

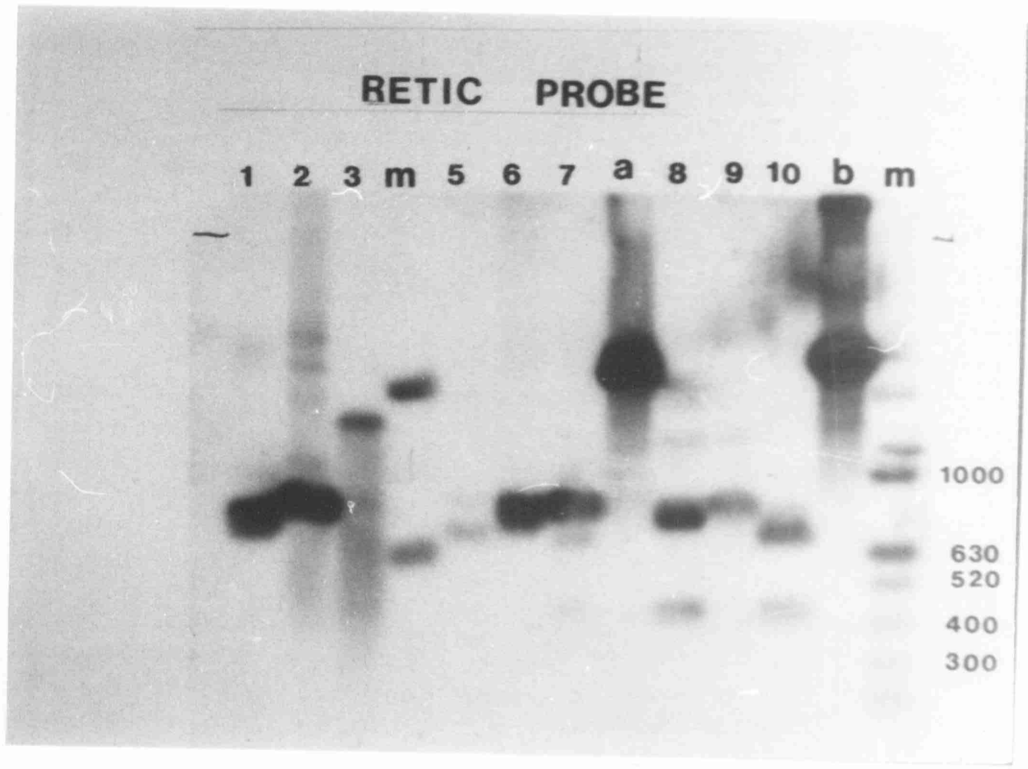
FIGURE 30. Southern blot analysis of "non-globin" recombinants.

Plasmid DNA from the "non globin" recombinants (2 µg) was digested using a HinF/R1 double digest (see Fig. 25) and run on a 2% agarose gel run in T.E.A. buffer (lanes 1-3, 5-7, 8-10). Hha 1 digests of the pCR 1 α and β globin cDNA plasmids were run in the lanes indicated. The remaining lanes contain pAT 153 marker fragments digested as for Figure 24. After electrophoresis the DNA fragments were transferred to nitrocellulose in 20 x SSC as described in Materials and Methods. The filter was then hybridised with reticulocyte ^{32}P cDNA and subjected to autoradiography for 16 hrs.

a) ethidium bromide staining of restriction digest fragments

b) autoradiograph of digest after transfer to nitrocellulose and hybridisation.


1 2 3 pAT 5 6 7 α 8 9 10 β M
1600
1000
630
520
400
300
220
150


RETIC PROBE
1 2 3 m 5 6 7 a 8 9 10 b m
1000
630
520
400
300

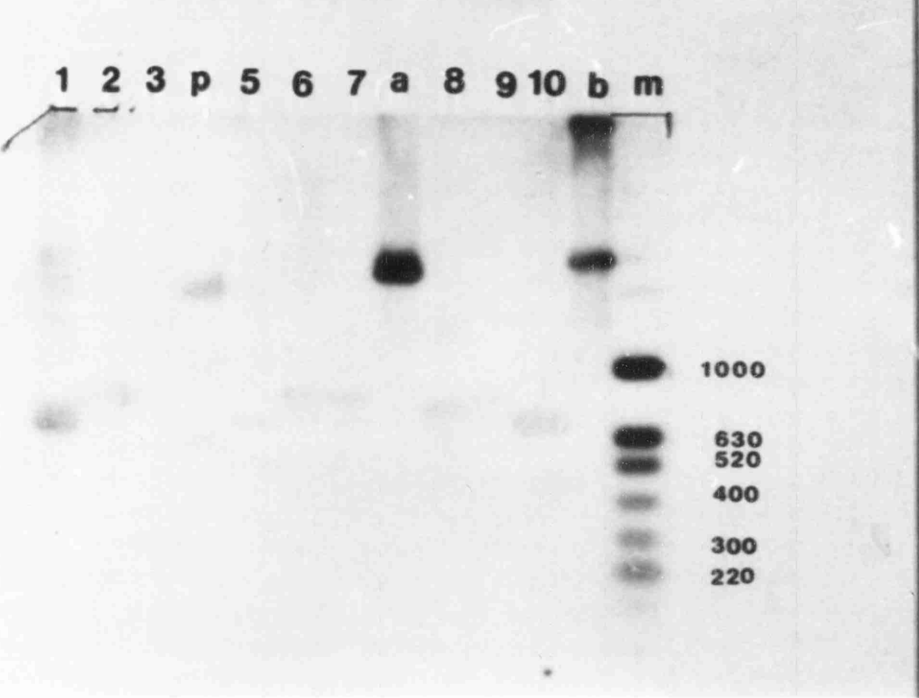
One somewhat unfortunate and surprising aspect of this hybridisation and of the subsequent ones, accrues from what can only be a spurious homology between globin cDNA and the pAT 153 vector, which is located within the particular fragment containing the Bam site. Virtually all the clones hybridise more strongly than the pAT however, in the case of clone pNG 3 it should be pointed out that the quantity of DNA is somewhat lower than for the other clones to the extent that it is barely visible on the ethidium stained pattern (Fig. 30a). The insert fragment can however be clearly seen by hybridisation. (Fig. 30b).

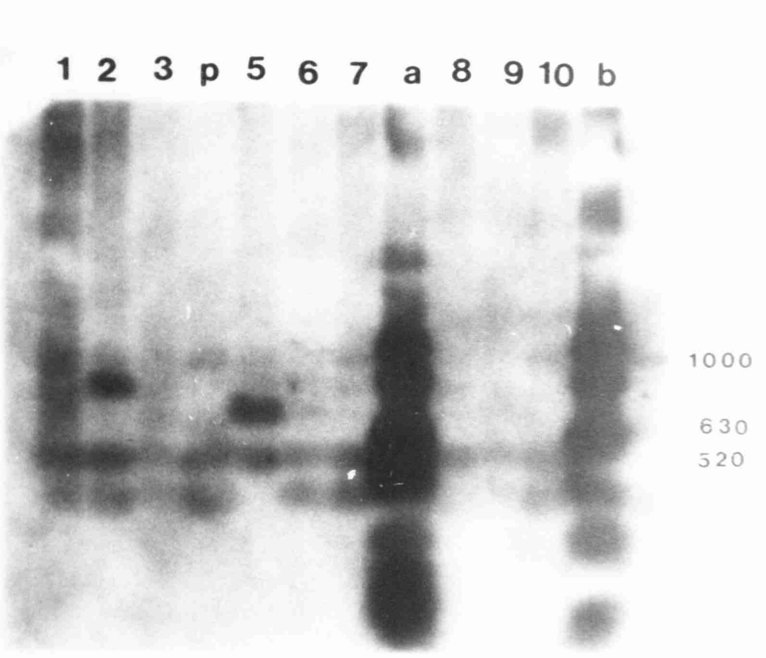
This blot analysis was extended by preparing purified α and β globin cDNA probes as described previously for the Grunstein/Hogness hybridisation (Fig. 23). The same filter was hybridised to this probe and the autoradiograph displayed in Fig. 23). The same filter was hybridised to this probe and the autoradiograph displayed in Fig. 31a. This probe shows good hybridisation to α and β insert DNA (tracks a and b) in the Hha digest of the pCRI globin cDNA recombinants. With the possible exception of clone pNG 1 none of the other clones displays hybridisation to these globin probes. Clone pNG 1 has previously shown evidence that it may be a globin cDNA insert (cf. Fig. 22). Finally, the filter was rehybridised using nick-translated α and β globin cDNA plasmids as probe (Fig. 31b). Homology between the pAT recombinants and the pCRI derived probe can be observed as hybridisation of the two fragments at 520 and 400 bp

FIGURE 31. Southern blot analysis.

a) nitrocellulose filter described in Figure 30 and hybridised with purified α and β globin cDNA. (See legend to Fig. 23).

b) as for 31 a) but hybridised with nick-translated pCR 1 α and β globin cDNA plasmids.


GLOBIN cDNA PROBE
1 2 3 p 5 6 7 a 8 9 10 b m
1000
630
520
400
300
220


GLOBIN PLASMIDS PROBE
1 2 3 p 5 6 7 a 8 9 10 b
1000
630
520

running across the whole filter. Hybridisation to the Hha digests of the α and β cDNA clones is of course observed to all the digest fragments and thus the α and β insert fragments are much less intense than previously, representing as they do a much smaller fraction of the probe than before. Additionally though two clones display specific hybridisation to the fragments containing their insert cDNA. In the case of pNG 2 this identifies the insert as globin sequences, for clone pNG 5 interpretation is far more difficult, owing to the large extent of material deleted from it. Thus the recombinant fragment of this cloned DNA also contains the region of homology between the pAT vectors and the pCRI probe sequences. More data on the identity of the sequences inserted into pNG 5 are required before it can be positively identified as a globin cDNA insert or not.

The hybridisation pattern of clone pNG 2 is revealing as it highlights a problem which takes on greater significance when the next series of experiments is discussed. Analyses of this cloned sequence previous to that shown in Fig. 31b gave no evidence that this contained a globin cDNA insert. Comparison for example of the result obtained using globin specific cDNA (Fig. 31a); the pertinent difference between the two probes being the representation of the entire β globin m-RNA sequence in the pCRI-β DNA, whereas the cDNA probes are derived from an oligo-dT primed synthesis of reticulocyte messenger RNA. The cDNA probe is thus much

greater enriched in sequences contiguous with the primer i.e. the 3' end of the messenger. Conversely such a cDNA probe will be rare in sequences derived from the 5' end of the messenger. As the cloned cDNAs are short they could contain sequences derived mainly from the 5' end of globin messenger RNA. When probed with a labelled oligo-dT primed cDNA, a clone of this type although containing globin sequences, will hybridise to a much lesser extent than a similar sized insert derived from the 3' end of the messenger. This presumably describes the nature of the sequences cloned in pNG 2 as only a "full-length" globin probe hybridises well to this sequence. Equally clone pNG 1 insert is probably disposed towards sequences derived from the 5' end of the α-globin messenger as this hybridises relatively poorly to reticulocyte cDNA. Unfortunately, this insert would not hybridise to the cloned α probe either (see Fig.31b) as this was found to represent sequences almost exclusively derived from the 3' end of the messenger. A cloned insert of the pNG 1 type would be difficult to sift out by the competition assay and thus could be expected to behave as a non-globin, by the previous hybridisation criteria.

D. Quantitation of Non-Globin Sequences.

1. Quantitative Estimation of Relative Abundance in Reticulocyte m-RNA of Sequences Complementary to Cloned cDNAs.

The relative abundances of the m-RNAs from which the non-globin cDNA clones were derived were estimated by a filter hybridisation assay. It was further hoped that quantitative data on abundances in reticulocyte m-RNA would help to confirm the assignment of non-globin status on these clones. Plasmid DNA was bound to small (1.3 cm) nitrocellulose filters as described in Materials and Methods (see also legend to Table 8). These were hybridised to reticulocyte cDNA in individual wells of a 24-well microtitre plate. Triplicates of each incubation were performed and the results shown are average values. The extent of hybridisation as estimated from liquid scintillation counting, after washing off non-specifically bound counts are shown in Table 8. This data is also shown expressed as a percentage of the extent of total hybridisation to the two globin sequences. The very high levels of hybridisation to the cloned sequences in pNG 1 and pNG 2 helps to confirm previous assignments that these do in fact contain globin cDNA sequences. The rather lower level of hybridisation to these clones than is observed for the pCRI α and β globin sequences can be explained(as described at the end of the previous section) if these clones represent sequences which come mainly from the 5' regions of the messenger RNA. The clone pNG 5 displays an extremely low level of hybridisation to reticulocyte cDNA and thus is probably not a globin sequence, despite the possible interpretation of the data shown in Fig. 31b.

TABLE 8. QUANTITATIVE ESTIMATES OF NON-GLOBIN m-RNA ABUNDANCE IN RETICULOCYTES.

Cloned Sequence[a]	% cDNA hybridised[b] above background	% Hybridisation relative to total globin
pCRIα	7.7	46.2
pCRIβ	5.2	31.2
pNG I	3.8	22.8
pNG 2	3.3	19.8
pNG 3	0.73	4.4
pNG 5	0.002	0.12
pNG 6	1.8	10.8
pNG 7	1.2	7.2
pNG 8	1.4	8.4
pNG 9	0.35	2.1
pNG 10	1.8	10.8

[a] 1 μg of each cloned sequence was bound to a 1.3 cm diameter nitrocellulose filter.

[b] Hybridisation values are corrected for hybridisation to a similar filter bearing pAT 153 DNA. Values shown are averages of triplicate filters.

The remaining clones all hybridise to somewhat greater extents than might have been expected, which places very strong constraints on the possible number of different sequences which have been cloned. However the observed abundances are not so great as to identify the clones as containing globin sequences.

Four clones for which the best evidence relating to non-globin status existed were chosen therefore for the tissue specificity studies described below.

2. <u>Analysis of Tissue Specificity of Non-Globin mRNAs from Reticulocytes</u>

A titration analysis, such as that described previously for estimating the abundances of non-globin sequences in reticulocyte m-RNA, was adopted to ascertain if the non-globins were expressed in cell types other than the reticulocyte.

Four clones were chosen for this study (for the reasons stated above) and grown in bulk culture to prepare suitable quantities of plasmid DNA. The DNA was prepared by the alkaline extraction technique with the additional caesium chloride gradient centrifugation as described in Materials and Methods. The four chosen clones were pNG 3, pNG 6, pNG 7 and pNG 9.

The tissues or cell types chosen for study were two erythroid cell types: induced and uninduced Friend cells (clone M2), one associated (by virtue of the location of foetal erythropoiesis) tissue type: adult liver (this was rat liver as mouse material was unobtainable), a non-erythroid

cell type: a fibroblast cell line (SC2), and a non dividing tissue: brain.

As previously described, plasmid DNA was denatured and bound to nitrocellulose filters then hybridised to ^{32}P labelled cDNA probe derived from one of the five m-RNAs described above, by oligo-dT primed synthesis. After extensive washing, the filters were subjected to liquid scintillation counting to estimate the extent of hybridisation. Triplicate filters for each incubation were used and the results are derived from hybridisations utilising two different levels of input counts (derived from two independent probe syntheses).

The results of this analysis are shown in Table 9, expressed as the fraction of the total input cDNA bound to the filter. Standards were β globin DNA (the pCRI cDNA clone) and pAT 153 DNA alone. The table shows mean values for the extent of hybridisation plus the standard error which was calculated from a full statistical analysis of variance. The 1% significance level for each experiment is also indicated in the table. This data is also represented histographically either tissue by tissue (Fig. 32) or clone by clone (Fig. 33). In the former case the standard errors are again indicated by error bars and the arrow at the ordinate represents the 1% statistical significance level.

Surprisingly perhaps, the cloned sequences seem to fall into at least three distinct patterns of hybridisation. Two clones pNG 3 and pNG 7 appear to be erythroid specific

TABLE 9. TISSUE SPECIFICITY OF RETICULOCYTE NON-GLOBIN CLONES.

CLONE	FLC^{+}	FLC^{-}	FIBROBLAST	ADULT LIVER	BRAIN
	Mean ± SEM	Mean ± SEM	Mean ± SEM	Mean ± SEM	Mean ± SEM
pAT 153	0.00 ± 0.30	0.00 ± 0.18	0.00 ± 0.13	0.00 ± 0.16	0.00 ± 0.23
pβG 1	57.10 ± 6.84	1.24 ± 0.35	0.33 ± 0.15	0.93 ± 0.24	1.05 ± 0.36
pNG 3	(00.60 ± 0.30)[a]	(0.39 ± 0.12)	0.53 ± 0.17	0.53 ± 0.21	0.66 ± 0.32
pNG 6	3.51 ± 0.71	1.72 ± 0.42	4.19 ± 0.46	2.47 ± 0.37	0.52 ± 0.30
pNG 7	14.90 ± 0.01	(0.18 ± 0.15)	0.31 ± 0.15	(0.28 ± 0.14)	0.51 ± 0.30
pNG 9	3.94 ± 0.75	0.42 ± 0.24	3.24 ± 0.38	1.03 ± 0.25	1.22 ± 0.48
1% SIG*	1.00	0.64	0.63	0.62	0.96

(Values are fraction of c-DNA hybridised x 10^4 and are not corrected for extent of reaction)

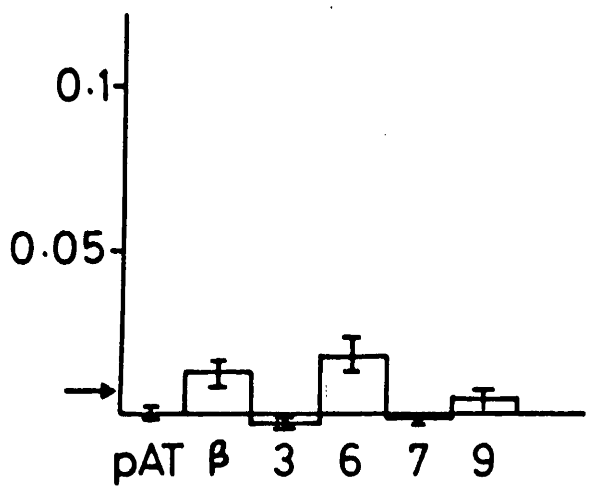
[a] values shown in brackets represent binding below control pAT 153 filters.

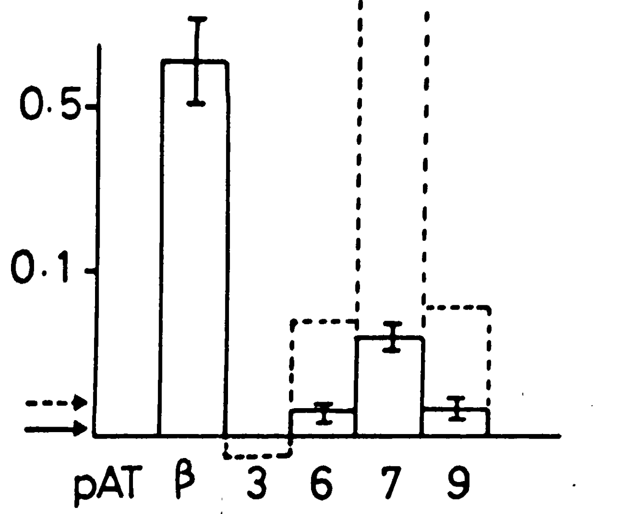
* Significance values based on analyses of variance.

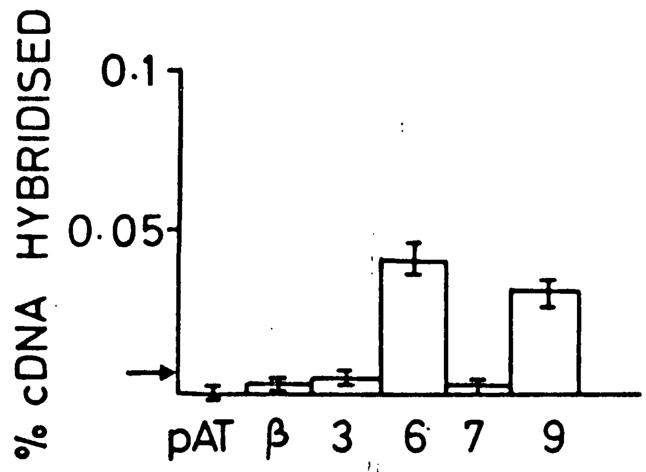
FIGURE 32. Tissue specificity of "non globin" mRNAs from mouse reticulocytes.

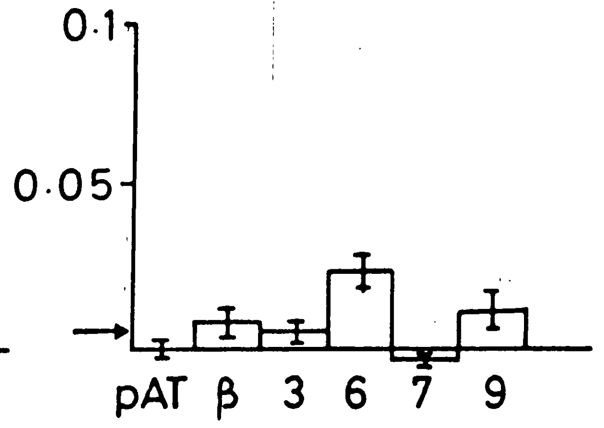
Filter bound plasmid DNA was hybridised to ^{32}P labelled cDNA probes from different tissues as indicated (for details see Materials and Methods). The data are represented histographically for all the cloned sequences hybridised with each cDNA probe; the height of the bar representing the fraction of cDNA bound.

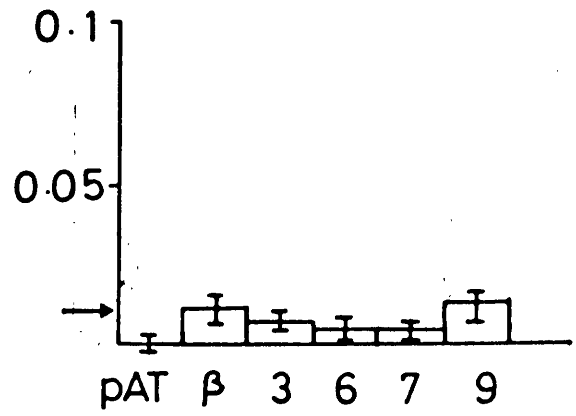
Standard errors on these values are shown as error bars and the arrow at the ordinate indicates the 1% statistical significance level. (See also Table 9).


FLC⁻
0.1
0.05
pAT
β
3
6
7
9


FLC⁺
X5
0.5
0.1
pAT
β
3
6
7
9


Fibroblast
% cDNA HYBRIDISED
0.1
0.05
pAT
β
3
6
7
9


Adult Rat Liver
0.1
0.05
pAT
β
3
6
7
9


Brain
0.1
0.05
pAT
β
3
6
7
9

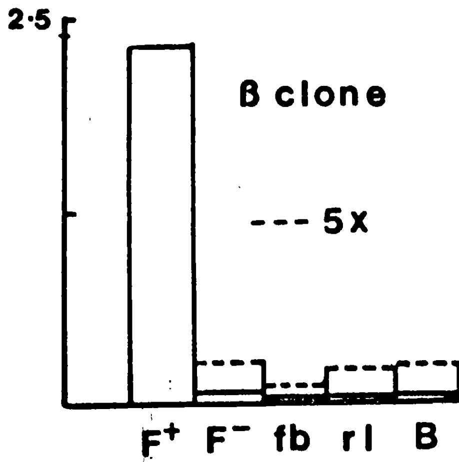
but the remaining two are represented much more widely. Clone pNG 7 is represented only in erythroid cell types and demonstrated a marked induction during Friend cell differentiation. Indeed this sequence reaches 30% of the level of β globin in these cells, after five days of growth in D.M.S.O. The clone pNG3 whilst also erythroid specific, seems to be expressed only in reticulocytes at detectable levels, for it was undetectable even in induced Friend cells. (See Table 8 for comparison). Analysis of material from foetal liver or anaemic spleen might demonstrate at which stage of erythropoiesis this particular sequence is induced. Availability of Friend cell cultures able to undergo terminal differentiation to the reticulocyte stage could be particularly illuminating.

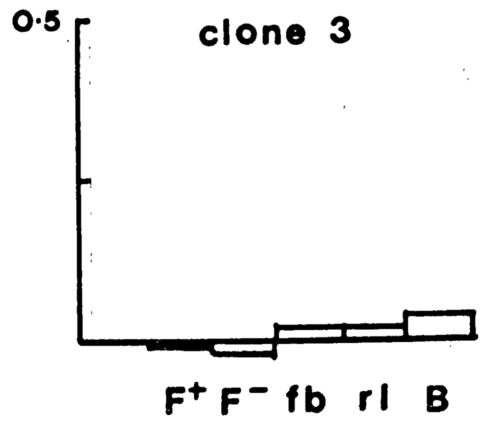
A totally different behaviour is displayed by the clones pNG 6 and 9 (Fig. 33). The former in particular being more prevalent than β-globin in the uninduced Friend cell (Fig. 32) upon induction though only a moderate change in abundance occurs, a factor of about four fold. For the β-globin sequence a greater than 50 fold change in steady state level was observed. Clone pNG 9 rises to similar levels as pNG 6 upon Friend cell induction but the precise increase in relative abundance between the induced and uninduced states cannot be accurately determined, as pNG 9 is on the limits of detection in these latter cells; below the 1% significance level in fact (Fig. 32).

FIGURE 33. Tissue specificity of "non globin" mRNAs from mouse reticulocytes.

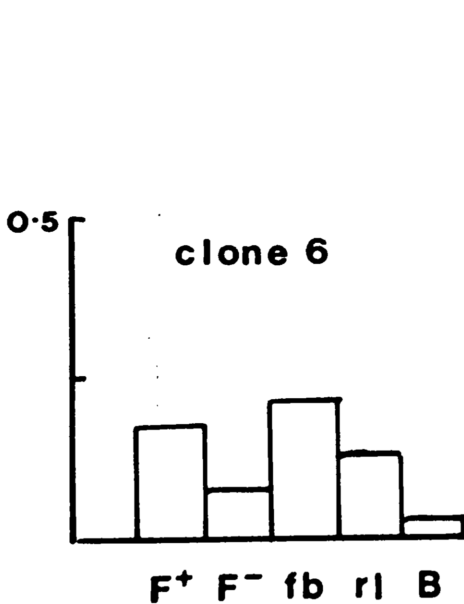
This figure presents the same data as displayed in the previous one but shows the relative levels of hybridisation for individual clones in all the different tissues. In this instance the fraction of cDNA hybridised has been corrected for the extent of reaction (approx. 20%). In all cases the abscissa represents the level of hybridisation to filter bound pAT 153 DNA.

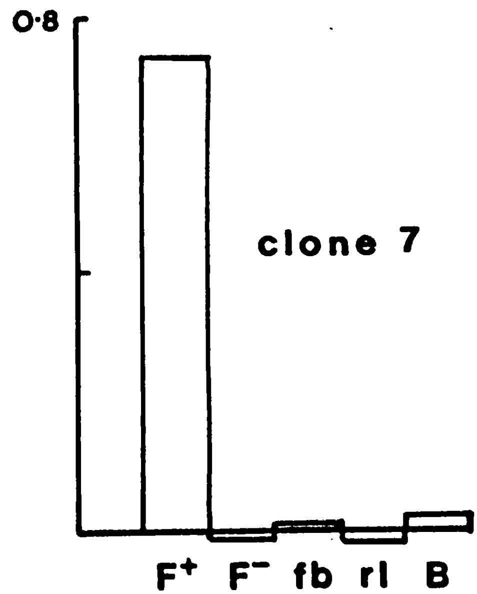
Retic cDNA recombinants
Tissue specificity of mRNA expression

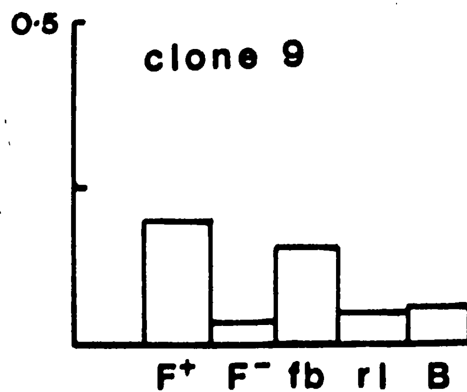




Percent cDNA hybridized at saturation







Furthermore both these clones reveal a broad spectrum of representation in other cell types (Fig. 33) hybridising significantly above background in adult liver and most noteworthy, in fibroblasts. Indeed clone pNG 6 hybridises to an extent greater than one order of magnitude above the β-globin control. The results in the fibroblasts are revealing as, whilst the levels observed in liver and brain could possibly be explained by residual contamination of these tissues, with blood cells for example, this cannot be true of cells grown in tissue culture. Additionally despite the fact that β-globin hybridises significantly to both adult liver and brain and thus may well indicate contamination of these tissues with blood, this can in no way account for the higher levels of hybridisation observed for clones pNG 6 and 9. Were this the case then the contribution to the hybridisation value could not be more than one third the level displayed for β-globin (cf. Table 8) as this is the maximum level seen for these clones in reticulocytes. The hybridisation to liver cDNA is most likely a reflection of a true expression of these sequences within this tissue.

Interestingly, although these clones (pNG 6 and 9) seem to represent sequences of a relatively ubiquitous nature, they occur at much higher abundance in reticulocytes than in other cell types. Additionally both sequence increase in abundance during D.M.S.O. induction of Friend cells. This strongly implies that despite being widely distributed,

they are also sequences intimately connected with erythroid maturation, and presumably therefore with red cell functions. The possibility must therefore be considered that the hybrising sequences in non-erythroid cells may not be identical to those cloned but are the products of closely related genes which are normally expressed in other cells or tissues. Possible candidates for such genes will be discussed further below.

DISCUSSION

I.A.1. Gradient fractionation of RNA

Evidence has been presented above for the successful cloning of non-globin messenger RNAs from mouse reticulocytes. The procedure employed was a ligation technique, joining blunt-ended cDNA and plasmid molecules. The total cDNA population was cloned and the desired sequences selected, by different screening approaches, from the total plasmid collection.

As other workers (39),(56) have reported successful fractionation of reticulocyte messenger RNA sequence population, to enrich for non-globin sequences, it may be pertinent to analyse whether such an approach could have been adopted here. The first point of relevance is that, for the cloning procedure outlined above it was considered a priority to have good quality (high proportion of intact molecules) m-RNA from which to begin synthesis of d.s. cDNA. This meant that gentle handling of the RNA was critical at all times. The gradient fractionation seemed an attractive possibility as it does not involve extremes of temperature or ionic conditions. Moreover, for the conditions of centrifugation described, optical density monitoring and radioactive counting suggested that a worthwhile separation should have occurred. With such a large globin m-RNA content the problem of contamination through aggregation is a very real one and this is considered the most likely cause for poor enrichment. Alternatively it is not absolutely

certain that no change in message structure occurred by centrifugational fractionation, as subsequent investigation on *in vivo* synthesised proteins at least revealed that the mouse reticulocyte may well possess a lower non-globin content than the rabbit. Certainly, the evidence of Aviv and his colleagues (62),(114) is in agreement with such an interpretation. Possibly this stems from an earlier release from marrow and thus a higher proportion of immature circulating cells in the latter species. Certainly, different strains of mice show differential responses to induction of reticulocytosis. Moreover genetic variation at the FV-2 locus is thought to regulate certain cycling parameters of precursors within the marrow (428), which could certainly affect the rate of release of reticulocytes into the circulation. Nevertheless it could have been revealing to try further gradient purification of the material from early reticulocytes; however, owing to the lysis occurring during preparation of this material it was felt that further possible exposure to nucleases should be avoided so this was not considered the best course of action for a maximum purification of fivefold. Alternative purification methodologies do exist however as it was clear that no fractionation procedure could obviate the necessity to further select out globin cDNA containing recombinants at the screening stage. These were not followed up exhaustively.

Both the sets of authors referred to above used preparative hybridisation using HAP chromatography but for the reasons previously stated, the use of high temperature and exposure to calcium phosphate was considered likely to significantly degrade the messenger or the cDNA derived from it; the latter of course relies on the preservation of its hairpin loop for second strand synthesis.

The possibility of using such a methodology for preparation of a non-globin enriched probe was considered strongly however. Naturally, the best purification should be obtainable from hybridisation of pure globin sequences, so attempts were made on this basis to fractionate reticulocyte cDNA by hybridisation to the pCRI α and β globin cDNA plasmids. For reasons not totally clear the hybridisation between plasmid sequences seemed more rapid than between plasmid and cDNA complementary sequences. Thus it was found impossible to drive the majority of sequences into hybrid even after sonication or restriction of plasmid sequences prior to hybridisation. The results obtained from the competition hybridisation make it probable that this failure was due to an insufficient plasmid excess being used. For efficient competition a mass excess of about 50,000 was employed.

The fractionation approach was therefore rejected in favour of the competition hybridisation, which naturally had the advantage of simplicity. Although, it may subsequently have been thought worthwhile further separating the hybridised

and unhybridised sequences by HAP chromatography. This could also possibly have provided a probe enriched in non-globin sequences; however the ability to purify the globin sequences from a cDNA probe looked a more viable proposition so was adopted instead.

2. Non-globin RNA and protein synthesis

The analysis of the reticulocyte messenger population and the protein labelling study presented here are both consistent in indicating a very low complexity of non-globin messenger and protein synthesis. Such figures would suggest that mammalian reticulocytes possess an abnormally low complexity, orders of magnitude less than observed in other cell types, for example. One precedent for this type of organisation in reticulocytes has come from the study made by Tobin and his colleagues (56),(57) in avian reticulocytes. Their study indicated that approximately 80 non-globin sequences are found in the avian cells which would suggest that either a second non-globin component should be found, at lower abundance perhaps, in the mouse reticulocytes, or that the avian reticulocyte still expresses a wider range of functions. This latter possibility may not be too unexpected owing to the ongoing transcription within the nucleated reticulocytes of the chick. The absence of nuclei in mammalian reticulocytes would therefore be instrumental in determining the lower complexity. In the chick, no evidence for other than a single non-globin sequence

abundance class was obtained but again (as in this report) in the absence of truly well fractionated probes the presence of additional sequences at really low abundance cannot be ruled out.

In this context though it is pertinent to observe that in mouse anaemic spleens, consisting mainly of orthochromatic erythroblasts, the sequence complexity of the non-globin fraction was also found to be abnormally limited, being in the region of 500 sequences (429). Ikawa and his colleagues have, also in the above report, shown evidence for a somewhat wider sequence diversity in mouse reticulocytes; they interpret their data as indicating that about half of the spleen erythroblast m-RNA complexity was still present in reticulocytes. Their interpretation, however, is based on a heterologous hybridisation curve between erythroblast non-globin cDNA and reticulocyte m-RNA. The $R_ot_{1/2}$ for such a reassociation curve is only of significance if the mass fraction of the RNA driving the reaction is known or can be estimated. In this context these authors, like others (68)(430) have estimated that about 10% of the reticulocyte m-RNA is non-globin and this accords well with the kinetic data presented in this report. If this is taken as a guideline, then at a maximum (if all the reticulocyte sequences are represented in the spleen population) only eight averaged size sequences are indicated by their data as present in reticulocytes. Such a value is in good agreement with data presented here and also with

data presented previously by Ikawa (39).

Other evidence relating to the non-globin messenger complexity of reticulocytes is sparse. Bishop and others (9) have shown the presence of RNA species other than globin in cDNA excess reactions hybridising to genomic DNA. The number of sequences involved was not ascertained from this study though.

It has been work performed in rabbit reticulocyte lysates through labelling of translation products which has provided most of the information regarding messenger content. Alternatively *in vivo* labelling of proteins, often of the reticulocyte membrane can provide some idea of the numbers and nature of the non-globin proteins to expect would be encountered. Lodish and Small for instance (68) have identified a 64 K molecular weight non-globin protein synthesised in rabbit reticulocyte lysates, which appears to be the major non-globin protein product, when assayed by *in vitro* translation in the wheat germ cell-free system. The level of synthesis of this protein drops significantly during reticulocyte maturation. A phenomenon which is equally well described for general non-globin protein content in rabbit reticulocytes (416). Lodish and Desalu have further identified six major non-globin proteins synthesised in reticulocyte lysates and in whole cells (415). Two of these proteins seem to be membrane proteins (417),(431),(432) but again the predominant synthetic product is a 64 K protein.

I.A.2. Non-globin RNA and protein synthesis.

The results described by Lodish and his colleagues show a good concordance with the non-globin protein analysis in this study. In particular the number of protein bands observed was similar in number and also of limited complexity which is also in accordance with the kinetic analysis of m-RNA described above. Other similarities are notable. Band H (Fig. 11 track 1) is of similar molecular weight to the Band F of Lodish, corresponding to a protein of about 20 K. The Band E (Fig. 11 track 2) shown here is very similar to the migration expected of actin (43 K) which would agree with evidence of Lodish and which is a known component of reticulocyte membranes. In contrast though, this band was not labelled in this study although a prominent staining component. Another particular difference is that in rabbits at least no major 30 K protein species have been observed, although some membrane proteins of 33 K and 36 K have been identified (431), (433). In this analysis (Fig. 11 track 2) the Band F is by far the major non-globin product. Chang et al. (432) though, in a study of mouse erythroid membrane protein synthesis revealed that the qualitative distribution of protein synthesis in these two species may well be different. No 64 K protein was associated with the mouse or rabbit membrane. This is therefore most likely cytoplasmic in location. Mouse reticulocytes appear to synthesise two membrane proteins, which these authors equate with Bands 4.1 and 4.2 of the human erythrocyte membrane; these are in the range 70 - 80 K molecular weight (432). Band C (Fig. 11

track 2) runs in this region of the polyacrylamide gel analysis of the G-100 column fractionated proteins. However another protein of similar molecular weight (approximately 70 K) has been identified by Thiele et al. (434). This is is a lipoxygenase; an enzyme capable of mitochondrial dissolution. What distinguishes this enzyme is its increased synthesis during reticulocyte maturation, which is in complete contrast to other reticulocyte non-globins. This change in synthetic rate seems to be governed by translational control and not post-translational modification of the protein. Band C of this report also seems not to be present in the earliest material (Fig. 11 track 1) but is clearly visible in proteins from later during the induced anaemia (Fig. 11 track 2).

Other possible candidates for non-globin synthesis in reticulocytes would be: the major ion channel protein, known as Band 3 in human membranes, which is about 96 K molecular weight, another prominent membrane protein is the enzyme glyceraldehyde-3-phosphate dehydrogenase, an enzyme crucial to glycolysis. Reticulocytes, as they destroy their mitochondria, are dependent on glycolysis for their energy. Thus, glucose uptake is of major importance as well. Another cytosolic protein known to be present in erythrocytes and which is active in reticulocytes is carbonic anhydrase. The activity of this enzyme is elevated during the later stages of erythropoiesis (435) and its synthesis is regulated like globin via the haemin control

of translation which operates in reticulocytes (436). Moreover this protein has a monomer molecular weight of about 30,000 which would make at least one of the components of Band F a likely candidate. Indeed red cells contain two isozymic forms of carbonic anhydrase so this could account for a doublet at approximately this molecular weight which is what was observed. A further point worthy of note is that only one of these forms, although they are very similar is erythroid specific, the other is widely represented. Additionally reticulocytes have been found to have increased levels of catalase (437) an enzyme which in monomeric form is approximately 60,000 molecular weight. This could well be the major 64,000 dalton protein of rabbit reticulocytes. The molecular weight difference can be accounted for by post translational modification, certainly in liver m-RNA where catalase is also a relatively abundant protein product, the de novo synthesised protein as determined by cell free translation is about 4000 daltons larger than the <u>in vivo</u> synthesised protein. (438)

Bulova and Burka (414) have used G-100 sephadex separation, as used here, to analyse non-globin synthesis in rabbit reticulocytes. Only very few products, about five in all, were identified as non-globins but no information on their identity was available. Some bad contamination between fractions was observed there and this may stem from the lack of a reducing agent (such as mercaptoethanol) in their separating buffer.

I.A.3. Cloning Methodology and analysis of clones

The protocol described here is naturally not the sole technique which could have been exploited. It should be said however, that many of the strategy decisions were directed by choice of restriction site which could be utilised for cloning. This was generally dictated by the containment requirement (as determined by the G.M.A.G.). Thus the Pst I site, which is considered optimal for tailing reactions (owing to the overhanging 3' termini) could not be utilised except under Category 3 conditions; this was due to the strong promotor which lies near this restriction site. Alternatively, the Pvu II site in pBR 322 would have been, in many ways, ideal for blunt end ligation as it produces flush ended termini. However this site lies within the fragment which was excised in forming the pAT 153 plasmid derivative.

It is undoubtedly true though, that some method for convenient excision of the inserted cDNA is highly advantageous. Here despite the fact that in a number of cases the expected San 3a sites were recreated the problems of separating the insert fragment from the eight plasmid fragments also generated is a major obstacle.

Restriction is the most reliable method for performing this function despite the attractiveness of S-1 cleavage of A-T tailed recombinants. In this context the use of linkers to insert the cDNA could be beneficial for future protocols.

The use of different screening strategies has already been discussed but one slightly contentious point should perhaps be pointed out. The success of the competition assay, depends on the ability of Grunstein/Hogness hybridisation, to respond in at least, a semi-quantitative fashion. If some form of quantitative relationship exists between hybridisation signal and representation of sequence within the probe then the recombinants should group into two categories; those showing high level hybridisation, which should be globin recombinants and low level hybridisation, which should be to non-globin recombinants. It is clear however, that the recombinants do not group this way. In particular, many globin recombinants must hybridise with much lower efficiency than would be expected. The variability in quantity of DNA bound to the filter may be one source of error. In some cases additional variation may have stemmed from colonies detaching from the filter. Globin cDNA clones derived from 5' message sequences may well have hybridised poorly to oligo-dT primed cDNA probes, as was highlighted earlier. Recombinants with this structure should however be fairly rare in the plasmid collection, unless first strand synthesis of cDNA was full-length and second strand synthesis was markedly poorer. In fact evidence for the precise reverse of this situation was presented here. One other explanation for variability in the hybridisation could also stem from differences in the size of the insert. This could be especially true for a situation where the hybridisation is performed in filter bound DNA excess, and the

individual colonies are competing for the same sequences within the probe. Under these conditions longer inserts may well be able to sequester greater amounts of probe.

I.A.4. Restriction Data.

The detailed restriction analysis has revealed a number of interesting features of somewhat unexpected nature. The Taq I and Ava II digests have drawn attention to the problems that can be encountered when digesting methylated DNA substrates. This problem has also been encountered elsewhere (439) and has been reported as stemming from methylated sequences lying contiguous with the recognition site but not within it. For example the Taq I recognition sequence is TCGA, however if the recognition sequence is TCGA^{me}TC as in pAT 153 then no scission occurs, unless very large enzyme excesses are used. This phenomenon is much easier to identify for the pBR 322 digest as this contains no other fragment running in the region of the composite fragment. This situation is in contrast with the behaviour displayed by Sau 3a and Mbo I for example where the methylation lies centrally within the restriction site and will prevent scission by Mbo I but does not interfere with Sau 3a (or Dpn I).

Not so straightforward to reconcile however is the anomalous appearance of the Alu I map of pAT 153. Some true rearrangement of sequences relative to pBR 322 seem to have occurred. As no other regions of the plasmid seem to have been affected, indeed all the expected sized fragments,derived from the

removal of the Hae II pieces, which straddle the join were observed. This anomaly may actually stem from a strain variation in the plasmid stock held at this laboratory or may have occurred during passage. That this is a sequence rearrangement and not a point mutation is evinced from the fact that a composite fragment was not observed, however two novel sized pieces were found strongly suggesting the former interpretation. No sequences of easily detectable size appear to be missing but undoubtedly detailed sequence information such as exists for pBR 322, especially around this region (near the replication origin) would be of particular benefit.

Some of the recombinants pose interesting problems. Clone pNG 8 for example which has cDNA complementary sequences in a fragment which makes up only a very minor fraction of the digest. This could have indicated that the clone was not purified or was really a mixture of two plasmids within a single clone but different digests have implied different sizes for the inserted sequences. The Hinc II/RI digest was in complete agreement with the hybridisation data for example; the Hae II digest quite markedly wrong. Uncut plasmid shows no evidence for the presence of more than one plasmid type (data not shown) but indicates a size commensurate with the Hinc II/RI data.

More interestingly perhaps, clone PNG 3, on tetra-nucleotide sequence analysis, has revealed a large insert size and moreover a high proportion of rare eukaryotic cutting

sites. The preponderance of GC containing restriction sites could indicate that this sequence codes for an arginine rich and therefore basic protein. This must be considered a somewhat speculative appraisal however. An alternative hypothesis could be that pNG 3 really is an amalgam with two inserted sequences ligated together, this would account for its unusual length. If this is the case, then one of these inserts could be of prokaryotic origin (plasmid fragment or E.coli genomic DNA) which could account for the prevalence of sites rare within the eukaryotic genome. One possible test of this interpretation would be by blot hybridisation to a restriction digest which makes an internal cut in the insert DNA. Unfortunately none of those identified to date are amenable to separation on agarose.

I.A.5. Quantitative estimates of abundance and tissue specificity.

The titration data, in particular the analysis of tissue specificity and developmental variation of the four clones studied in detail, raises some fascinating aspects.

In the type of quantitative analysis performed here, under conditions of filter bound DNA excess, the rate of reaction has been shown to depend on relative abundance of the hybridising sequence within the probe and is directly proportional to it (440). Thus at any given time the extent of hybridisation should be proportional to the abundance of the reacting sequence. One source of potential error has been identified - this relates to the possibility that the abundance of cloned sequences restricted to sequences derived

from the 5' region of the messenger RNA will be underestimated owing to its poor representation in a probe synthesised by priming from the 3' end of the messenger RNA. The main relevance of this lies in sorting out globin specific clones which hybridise poorly and thus appear as non-globins. Thus two clones pNG 1-2 were found to hybridise to a lesser extent than the pCRI globin cDNA plasmids. One possible solution to this problem could be to use a randomly primed probe, although for quantitative analysis that approach has other problems. Alternatively, labelled RNA not cDNA could be employed as a probe especially if kinase labelled at the 5' end. This would provide a good correlate relative to hybridisation values obtained by cDNA methodology. However, two reasons make it unlikely that such factors will have interfered with the analysis presented here. Firstly, recombinants which are non-globins, as they are comparatively rare in the m-RNA population, are very unlikely to have been selected out at all if they showed a reduced level of hybridisation to an oligo-dT primed probe. Equally, 5' end specific clones should be rare in the original cDNA; if they represent already rare sequences, this type of clone will doubly be underpresented in the plasmid collection. Furthermore this effect, like any other variation due for example to insert size, could only reduce the level of observed hybridisation there is no possibility of overestimate. Thus, comparison of the hybridisation behaviour of any given clone in different tissues would at all times be valid. Under certain specific circumstances, comparison of the

absolute quantitative levels of hybridisation between cloned sequences within a given tissue could be erroneous.

Examination of the data presented here argues against that possibility. With regard to length of cloned sequence it can be seen that despite its extra size clone pNG 3 at no time showed a greater abundance than other non-globin sequences. More importantly perhaps, the relative abundances for α and β globin in the reticulocyte RNA as assessed here were in very close accordance with previously obtained values (99),(108),(148); the α sequence being found in excess although somewhat shorter than the β insert. The actual ratio of 1.5:1 is in very close agreement with Phillips et al. (108)(1.25:1), Mezl et al. (99)(1.4:1) and a 1.4:1 ratio predicted from analysis of initiation rates by Lodish (148). Only in abnormal erythropoiesis or during Friend cell induction have α to β ratios indicating excess β m-RNA been reported. Much variation in these parameters has been experienced however, depending on the clone analysed and the inducer employed.

The results obtained with the β-globin filters also agree well with other analyses of Friend cells. A background level of 0.062% hybridisation was observed here (Fig. 33) (corrected for approximately 20% level of hybridisation relative to saturation). Values of 0.06% for α and β m-RNA have been shown by Gilmour et al. (37), 0.09% by Nudel for β alone (clone 745A) or <0.05% for α + β by Minty et al. (29) (M2 cells). Some clones appear to have

higher basal levels, in particular clone 707, but variation due to growth medium factors has also been noted (29). The induction ratio observed here is also in close agreement with those seen elsewhere. For example by this technique globin is induced 50 fold in its steady state level, on five days induction in D.M.S.O. Compared with other recorded values of 50 times (101), 32 times (29), 50-100 times (37), 25 times (38). The first three of these estimates are for clones 745A, M2 and M2 respectively, the fourth for 707.

The other factor of relevance here is the observed tissue specificity of β-globin sequence expression. Here, as for some other workers, significant levels of β-globin m-RNA have been detected in both brain tissue and adult liver. One possible source of this β-globin sequence representation could stem from contamination of these tissues, as referred to earlier. However with regard to the liver results one factor may be particularly pertinent, as this cDNA was derived from rat liver, and in the rat, even in quite mature adults, a proportion of erythropoiesis still takes place in the liver (W. Ostertag, pers. commun.). No similar explanation can be invoked for brain tissue of course, but this material is difficult to wash clean of blood. On the other hand the level of β-globin sequences observed in brain are very low, on the limits of statistical significance and, as less than optimal amounts of probe were available here (brain m-RNA is also a relatively poor template for reverse transcriptase), coupled with the high complexity of

this tissue's m-RNA population it is possible this result is spurious. It is notable that these results stand in stark contrast to these also obtained here in fibroblasts (S2) which despite being a transformed line (29) show no evidence for expression of the β-globin gene, a situation at odds with that observed by Humphries et al. (84) but similar to that seen by other authors (85),(86). This is revealing as extensive sequence crossover had been shown for uninduced Friend cells and this fibroblast line (29).

The spectrum of tissue specificity displayed by the four non-globin clones is interesting. The clone pNG 3 for example would appear only to be expressed at the reticulocyte stage during erythroid maturation and also is specific to that cell type, as far as those tested can indicate. One possible candidate for the m-RNA coded by this clone could be the lipoxygenase alluded to before (434), which was found not to be expressed until late maturation of the reticulocyte. Its presence as messenger might hope to be detected at a later stage than the Friend cell equivalent, after the final cell division for example. Certainly, the mitochondrial destructive function would be of little value to cells other than reticulocytes, except possibly other cells which are destined to die shortly, epidermal cells perhaps, or which have high levels of glycolytic activity.

The behaviour of clone pNG 7 is well in line with it being one of the classic erythroid markers, as it appears

specific to the erythroid lineage and induces markedly in Friend cells. Of possible candidates for the identity of pNG 7, catalase though would seem unlikely as this should be seen in liver as well as erythroid cells (437),(438). Moreover the rise in catalase activity does not usually occur until the orthochromatic stage (435); a somewhat later stage than the Friend cell appears to represent. Carbonic anhydrase levels begin to rise in polychromatic erythroblasts which would fit with the observed data for pNG 7. In addition as pointed out earlier carbonic anhydrase is the most likely major product synthesised in mouse reticulocytes. However, Kabat and his colleagues have described only a moderate induction of this enzyme activity during Friend cell induction (3 fold) (102). Again carbonic anhydrase activity is detectable in liver and particularly in kidney. No evidence here relates to the kidney levels of any of the cloned sequences but this could be a worthwhile addition to the tissues examined. The general behaviour of carbonic anhydrase perhaps fits the pattern described for the clones pNG 6-9 which show evidence for a wider distribution and more moderate (2 fold and 9 fold respectively) levels of induction in Friend cells. However both these clones show a higher level of representation in liver than would be expected for carbonic anhydrase, more in line with catalase type behaviour.

Neither of these assignments squares too well with the fairly high levels observed in fibroblasts. No catalase has

been detected in fibroblasts, nor has acetylcholine esterase which also induces during Friend cell inhibition. Of other inducible Friend cell products some of the most critical are in the haem pathway, δ-amino laevulinic acid synthetase is one and is rate limiting for haem synthesis (441). Haem synthetase is another enzyme whose function is crucial to maturation in Friend cells (118),(119), a mutant line deficient in this enzyme is unable to differentiate. Cytidine deaminase has also been identified by Friend herself (442) as an enzyme activity raised during both normal erythropoiesis and during Friend cell induction. This enzyme is also reported to be specifically associated with stress erythropoiesis in the mouse. Arnstein and his colleagues have also identified a late rise in the level of adenylate kinase (435).

A number of these functions could be attributed to pNG 7, especially these of more specialised erythroid nature, but it should be said that at this stage there is no categorical evidence ruling out the possibility that pNG 7 is a portion of a globin sequence, although its somewhat greater representation in Friend cells than in reticulocytes, relative to the globins, somewhat argues against this. It may possibly be a minor globin, of β or α related type (the mouse has a number of uncharacterised β-like genes) which is affected by stress erythropoiesis, like the β^{c} chain of the sheep (116). It may be of foetal-like function; a foetal globin β-chain has been reported in Friend cells and foetal liver (443), (444) though its existence is somewhat

contentious. No evidence for its existence in reticulocytes has been produced however.

Enzymes of a wider representation which might be expected to be more ubiquitous, are glycolytic enzymes, as reticulocytes obtain all their energy from this form of respiration, and the associated pentose-phosphate pathway enzymes such as phosphogluconate dehydrogenase which supplies NADPH (445). Perhaps another likely candidate is super oxide dismutase, which works hand in hand with catalase, and incidentally also has a molecular weight in the 30 K region.

An alternative explanation for the existence of these products and or some of the more erythroid specific enzymes, in fibroblasts could be due to the transformed nature of these cells. This applies for two reasons. Firstly, tumour cells have a link in common with reticulocytes in that they often also undergo aerobic glycolysis for their energy requirement. This phenomenon could account for elevated levels of glycolytic enzymes in fibroblasts in common with reticulocytes. Secondly, and more prosaically perhaps, the type of transcriptional leakiness associated with transformed cells could be responsible for the cross hybridisation between the reticulocyte non-globins and the fibroblast c-DNA. Such virally induced derepression can be quite specific for different sequences, as witnessed by the effect of RSV transformation on avian erythrocytes (85). This could account for the possible hybridisation to the non-globins pNG 6 and 9 and the negative result for the

β-globin. As already stated, this fibroblast line does show evidence for holding most of its sequences in common with Friend cells.

Further investigations are naturally required to make the above discussion anything more than speculation. One major step to identifying the cloned non-globin sequences would be to obtain data on the proteins they code for by *in vitro* translation after hybridisation selection to filter bound plasmid. Preliminary evidence suggests this is a fairly good method of corroborating the non-globin status of the cloned sequences, by analysis of the translation products on acid-urea gels (446). That at least is not troubled by problems relating to the location of the cloned sequence within the messenger. Although the recent availability of both genomic cloned sequences for α and β globin could reinforce the original assignments in perhaps a more unequivocal fashion. More positive evidence on the nature of the cloned sequences could be informative. To date analysis by Southern hybridisation and to filter bound RNA have been unsuccessful in determining the messenger size of the cloned sequence or its genomic organisation. It may be found necessary to excise the inserted DNA for an analysis of this sort.

Other extensions to the approach so far adopted would be to further the analysis of how these sequences vary in abundance during erythropoiesis. The stem cell line (416) of Dexter could prove informative in this context (447). It

is also interesting to note that a direct link in sequence content between reticulocyte m-RNA and the Friend cell has been established here, for sequences other than globin. This stands diametrically opposed to the work of Obinata et al. (39) who have stated that reticulocyte non-globin cDNA does not hybridise to Friend cell RNA. However that was not the only contentious result of that report and the advantage of cloned probes could be considered decisive.

In analysing both tissue specificity and erythropoiesis, some information on the level at which expression is mediated could be obtained. For example, can the non-globin sequences be found present in nuclear RNA prior to their appearance in the cytoplasm? Are they regulated coordinately with globin m-RNA? And can they be detected in the nuclear RNA of cells of tissues in which they are not expressed? If they appear to be transcriptionally regulated, then do changes in the DNAse I sensitivity of the chromatin occur and if so, at what developmental stages? Perhaps most tantalisingly, does the regulation of those sequences specific to the erythroid lineage vary from the mechanisms of regulation for sequences of a more ubiquitous nature? By creating a battery of cloned probes derived from a single cell type some of these questions can be approached which could reveal some of the mechanisms of erythroid differentiation.

More immediately perhaps the cloned non-globin sequences could be used to attack two smaller but by no means trivial questions. Firstly, the aspect of differential stabilities

for the globin and non-globin compartments could be addressed using specific non-globin sequences to see if they become destabilised in reticulocytes as Aviv has predicted (114).

Secondly, some debate has occurred as to whether non-globin reticulocyte proteins are synthesised exclusively on membrane bound polyribosomes as suggested by Burka and colleagues (414),(448), but denied by Lodish (431).

The possession of cloned reticulocyte non-globin sequences could provide one way of resolving that debate at least.

REFERENCES

1. HASTIE ND, BISHOP JO. Cell 1976; 9: 761-774.

2. YOUNG BD, BIRNIE GD, PAUL J. Biochemistry 1976; 15:2823-2829.

3. AXEL R, FEIGELSON P, SCHULTZ G. Cell 1976; 7: 247-254.

4. GALAU GA, KLEIN WH, DAVIS MM, WOLD BJ, BRITTEN RJ, DAVIDSON EH. Cell 1976; 7: 487-505.

5. DAVIDSON EH, KLEIN WH, BRITTEN RJ. Dev. Biol. 1977; 55: 69-84.

6. PATERSON BM, BISHOP JO. Cell 1977; 12: 751-765.

7. MAURON A, SPHOR G. Nucleic Acids Res. 1978; 5: 3013-3032.

8. WILLIAMS JG, PENMAN S. Cell 1975; 6: 197-206.

9. BISHOP JO, BECKMANN J, CAMPO MS. Phil. Trans. R. Soc. Lond. B. 1975; 272: 147-157.

10. AFFARA NA, JACQUET M, JAKOB H, JACOB F, GROS F. J. Biol. Chem. 1977; 12:509-520.

11. KLEIMAN L, BIRNIE GD, YOUNG BD, PAUL J. Biochem. 1977; 16: 1218-1223.

12. MONAHAN JJ, HARRIS SE, O'MALLEY BW. J. Biol. Chem. 1976; 251:3738-3748.

13. HARDING JD, MACDONALD RJ, PRZYBYLA AE, CHIRGWIN JM, PICTET RL, RUTTER WJ. J. Biol. Chem. 1977; 252: 7391-7397.

14. AFFARA NA, DAUBAS P. Dev. Biol. 1979; 72: 110-125.

15. AZZIZ S, BALMAIN A, KNOWLER JT. Eur. J. Biochem. 1979; 100:85-94

16. JAQUET M, GROS F. Nucleic Acids Res. 1979; 6: 1639-1655.

17. SOPOWIT SC, ROSEN JM. Biochem. 1980; 19: 3452-3460.

18. TIMBERLAKE WE. Dev. Biol. 1980; 78: 497-510.

19. FIRTEL RA. J. Mol. Biol. 1972; 66: 363-377.

20. HOUGH BR, SMITH M, BRITTEN RJ, DAVIDSON EH. Cell 1975; 5: 291-299.

21. WOLD BJ, KLEIN WH, HOUGH-EVANS BR, BRITTEN RJ, DAVIDSON EH. Cell 1978; 14: 941-950.

22. FIRTEL RA, BONNER JT. J. Mol. Biol. 1972; 66: 339-361.

23. LODISH HF. Miami Winter Symposium 1978; 15: 169-183.

24. BONNER JT. Signalling Systems in Dictyostelium. in: Graham CF, Wareing PF, Eds. The Developmental Biology of Plants and Animals. London: Blackwell Scientific Publications, 1976; 204-215.

25. TIMBERLAKE WE. Dev. Biol. 1976; 51: 202-214.

26. CLUTTERBUCK AJ. Genetics 1969; 63: 317-327.

27. RUTTER WJ. Miami Winter Symposium 1978; 15: 67-90.

28. AZZIZ S, BALMAIN A, KNOWLER JT. Eur. J. Biochem. 1979; 100: 95-100.

29. MINTY AJ, BIRNIE GD, PAUL J. Exp. Cell Res. 1978; 115: 1-14.

30. HARRISON PR.

31. GREEN H. Miami Winter Symposium 1978; 15: 13-16.

32. STRICKLAND S, SMITH KK, MARROTT KR. Cell 1980; 21: 347-355.

33. COX RF. Biochem. 1977; 16: 3433-3443.

34. PARKER MG, MAINWARING WIP. Cell 1977; 12: 401-407.

35. HIGGINS SJ, PARKER MG, FULLER FM, JACKSON PJ. Eur. J. Biochem. 1979; 102: 431-440.

36. HEREFORD LM, ROSBASH LM. Cell 1977; 10: 463-467.

37. GILMOUR RS, HARRISON PR, WINDASS JD, AFFARA NA, PAUL J. Cell Diffn. 1974; 3: 9-22.

38. HARRISON PR, GILMOUR RS, AFFARA NA, CONKIE D, PAUL J. Cell Diffn. 1974; 3: 23-30.

39. OBINATA M, KAMEJI R, UCHIYAMA Y, IKAWA Y. In: Differentiation of normal and neoplastic hematopoietic cells. Cold Spring Harbor Laboratory, N.Y., 1978; 319-334.

40. LEIBOVITCH MP, LEIBOVITCH SA, HAREL J, KRUH J. Eur. J. Biochem. 1979; 97: 321-326.

41. JAQUET M. AFFARA NA, ROBERT B, JAKOB H, JACOB F, GROS F. Biochem. 1978; 17: 69-79.

42. MARTIN GA. Cell 1975; 5: 229-243.

43. GETZ M, BIRNIE GD, YOUNG BD, McPHAIL E, PAUL J. Cell 1975; 4: 121-129.

44. HERMAN RC, WILLIAMS JG, PENMAN S. Cell 1976; 7: 429-437.

45. CHIKARAISHI DM, DEEBS S, SUEOKA N. Cell 1978; 13: 111-120.

46. LEWIN B. Cell 1975; 4: 77-93.

47. TIMBERLAKE WE, SHUMARO DS, GOLDBERG RB. Cell 1977; 10: 623-632.

48. TILGHMAN SM, CURTIS PJ, TIEMEIER DC, LEDER P, WEISSMANN C. Proc. Natl. Acad. Sci. USA 1978; 75: 1309-1313.

49. SMITH K, ROSTECK P, LINGREL JB. Nucleic Acids Res. 1978; 5: 105-115.

50. KINNIBURGH AJ, MERTZ JE, ROSS J. Cell 1978; 14: 681-693.

51. DERMAN E, GOLDBERG S, DARNELL JE. Cell 1976; 9: 465-472.

52. NAKAZATO H, EDMONDS M. Methods Enzymol. 1974; 29E: 431-443.

53. BANTLE JA, MAXWELL IH, HAHN WE. Anal. Biochem. 1976; 72:413-427.

54. SMITH MJ, HOUGH BR, CHAMBERLIN ME, DAVIDSON EH. J. Mol. Biol. 1974; 85: 103-126.

55. ERNST SG, BRITTEN RJ, DAVIDSON EH. Proc. Natl. Acad. Sci. USA 1979; 76: 2209-2212.

56. LASKY L, NOZICK ND, TOBIN AJ. Dev. Biol. 1978; 67:23-29.

57. LASKY L, TOBIN AJ. Biochem. 1979; 18: 1594-1598.

58. AFFARA NA, J. Mol. Biol. 1980; 141: 441-458.

59. AFFARA NA, J. Mol. Biol. 1980; 141: 459-

60. WEINTRAUB H, GROUDINE M. Science 1976; 193: 848-856.

61. WILLIAMSON R, MORRISON M, LANYON G, EASON R, PAUL J. Biochemistry 1971; 10: 3014-3023

62. BASTOS RN, VOLLOCH Z, AVIV H. J. Mol. Biol. 1977; 110: 191-203.

63. ROSEN JM, WOO SLC, MEANS AR, O'MALLEY BW. Purification and characterisation of ovalbumin mRNA in the chick oviduct. In: Methods in Molecular Biology 8 Eukaryotes at the Subcellular Level. J. Last ed. New York and Basel: Marcel & Dekker, 1976; 369-434.

64. TONEGAWA S, BRACK C, HOZUMI N, MATTHYSSENS G, SCHULLER R. Immunological Rev. 1977; 36: 73-94.

65. SUZUKI Y, TSUDA M, JSUJIMOTO Y, OHSHIMA Y, GRIZA PE. Structural and functional studies of the fibroin gene gene in Carnegie Inst. Washington Yearbook 1978; 78: 84-100.

66. STRAIR RK, YAP SH, SHAFRITZ DA. Proc. Natl. Acad. Sci. USA 1977; 74: 4346-4350.

67. MORRIS GE, BUZASH BA, ROURKE AW, TEPPERMAN K, THOMPSON WC, HEYWOOD SM. Cold Spring Harbour Symp. Quant. Biol. 1972; 37: 535-541.

68. LODISH HF, SMALL B. Cell 1976; 7: 59-65.

69. LACY E, HARDISON RC, QUON D, MANIATIS T. Cell 1979; 18: 1273-;283.

70. SIM GK, KAFATOS FC, JONES CW, KOEHLER MD, EFSTRATIADIS A, MANIATIS T. Cell 1979; 18: 1303-1316.

71. KING CR, UDELL DS, DEELEY RG. J. Biol. Chem. 1979; 254: 6768-6786.

72. ZIMMERMAN CR, ORR WC, LECLERC RF, BARNARD EC, TIMBERLAKE CR. Cell 1980; 21: 709-715.

73. FYRBERG EA, KINDLE KL, DAVIDSON N, SODJA A. Cell 1980; 19: 365-378.

74. MANIATIS T, HARDISON RC, LACY E. et al. Cell 1978; 15: 687-701.

75. BENTON WD, DAVIS RW. Science 1977; 196: 180-182.

76. SOUTHERN EM. J. Mol. Biol. 1975; 98: 503-517.

77. ALWINE JC, KEMP DJ, PARKER BA. et al. In Methods in Enzymol. Wu R. ed. New York Acad. Press 1979; 68: 220-242.

78. MAXAM AM, GILBERT W. Proc. Natl. Acad. Sci. USA 1977; 74: 560-564.

79. SMITH A. In Methods in Enzymol. Wu R. ed. 1980; 69:

80. HARRISON PR, BIRNIE GD, HELL A, HUMPHRIES S, YOUNG BD, PAUL JP. J. Mol. Biol. 1974; 84: 539-554.

81. YOUNG BD, HARRISON PR, GILMOUR RS et al. J. Mol. Biol. 1974; 84: 555-568.

82. KEDES LH. In Ann. Rev. Biochem. Snell EE. ed. California Annual Reviews Inc. 1979; 48: 837-870.

83. HEREFORD L, FAHRNER K, WOOLFORD J.Jr., ROSBASH M. Cell 1979; 18: 1261-1271.

84. HUMPHRIES S, WINDASS J, WILLIAMSON R. Cell 1976; 7: 267-277.

85. GROUDINE M, WEINTRAUB H. Proc. Natl. Acad. Sci. USA 1975; 72: 4464-4468.

86. SPECTOR DH, SMITH K, PADGETT T. et al. Cell 1978; 13: 371-379.

87. RUTHERFORD TR, CLEGG JB, WEATHERALL DJ. Nature 1979; 280: 164-165.

88. BENZ EJ, MURNANE MJ, TONOKOW BL. et al. Proc. Natl. Acad. Sci. USA 1980; 77: 3509-3513.

89. LOCKHARD RE, RAJ B, HANDARY UL. Cell 1976; 9: 747-760.

90. PROUDFOOT NJ, LONGLEY JI. Cell 1976; 9: 733-746.

91. BARALLE FE. Cell 1977; 10: 549-558.

92. HEINDELL HC, LIU A, PADDOCK GV, STUDNICKA GM, SALSER WA. Cell 1978; 15: 43-54.

93. PROUDFOOT NJ. Cell 1977; 10: 559-570.

94. BASTOS RN, AVIV H. Cell 1977; 11: 641-650.

95. COURTNEY M, WILLIAMSON R. Nucleic Acids Res. 1979; 7: 1121-1130.

96. ROSS J. J. Mol. Biol. 1976; 106: 403-420

97. WEAVER R, BOLL W, WEISSMAN C. Experimentia 1979; 35: 983.

98. WEAVER RF, WEISSMAN C. Nucleic Acids Res. 1979; 7: 1175-1193.

99. METZL VA, KAWASAKI ES, HUNT JA. Biochem. J. 1979; 179:

100. CONKIE D, KLEIMAN L, HARRISON PR, PAUL J. Exp. Cell Res. 1975; 93: 315-324.

101. NUDEL U, SALMON J, FIBACH E. *et al*. Cell 1977; 12: 463-469.

102. KABAT D, SHERTON CC, EVANS LH, BIGLEY R, KOBE R. Cell 1975; 5: 331-338.

103. CHUL HK, DETTI DM, MARKS PA, RIFKIND RA. J. Cell Biol. 1971; 51: 585-595.

104. HARRISON PR. Nature 1976; 262: 353-356.

105. MARKS PA, RIFKIND RA. Science 1972; 1975: 955-961.

106. ORKIN SH, SWAN D, LEDER P. J. Biol. Chem. 1975; 250: 8753-8760.

107. CHENG TC, KAZAZIAN HH. Proc. Natl. Acad. Sci. USA 1976; 76: 1811-1815.

108. PHILLIPS JA, SNYDER PG, KAZAZIAN HH. Nature 1977; 269:442-445.

109. HUNT JA. Biochem. J. 1974; 138: 487-498.

110. HUNT JA. Biochem. J. 1974; 138: 499-510.

111. LOWENHAUPT K, LINGREL JB. Cell 1978; 14: 337-344.

112. LOWENHAUPT K, LINGREL JB. Proc. Natl. Acad. Sci. USA 1979; 76: 5173-5177.

113. AVIV H, VOLLOCH Z, BASTOS R, LEVY S. Cell 1976; 8: 495-503.

114. BASTOS RN, AVIV H. J. Mol. Biol. 1977; 110: 205-218.

115. VOLLOCH Z, HOUSEMAN D. Cell 1981; 23 in press.

116. BENZ EJ, BARKER JE, PIERCE JE, TURNER PA & NIENHUIS AW. Cell 1978; 14: 733-740.

117. HARRISON PR, RUTHERFORD T. CONKIE D. _et al_. Cell 1978; 14: 61-70.

118. RUTHERFORD TR, WEATHERALL DJ. Cell 1979; 16: 415-423.

119. RUTHERFORD TR, HARRISON PR. Proc. Natl. Acad. Sci. USA 1979; 76: 5660-5664.

120. TSAI MJ, TING AC, NORDSTROM JL, ZIMMER W, O'MALLEY BW. Cell 1980; 22: 219-230.

121. ROOP DR, TSAI MJ, O'MALLEY BW. Cell 1980; 19: 63-68.

122. RYFFEL GU, WYLER T, MUELLENER DB, WEBER R. Cell 1980; 19: 53-61.

123. CHAMBON P, BENOIST CR, BREATNACH R. In: Miami Winter Symposium 1979; 16: 55-82.

124. HARPOLD MM, DOBNER PR, EVANS R, BANCROFT FC, DARNELL JE, Jr. Nucleic Acids Res. 1979; 6: 3133-3144.

125. McKNIGHT GS, PALMITER RD. J. Biol. Chem. 1979; 254: 9050-9058.

126. ROBINS DM, SCHIMKE RT. J. Biol. Chem. 1978; 253: 8925-8934.

127. McKNIGHT GS, LEE DC, PALMITER RD. J. Biol. Chem. 1980; 255: 148-153.

128. McKNIGHT GS, LEE DC, HEMMAPLARDH D, FINCH CA, PALMITER RD. J. Biol. Chem. 1980; 255: 144-147.

129. GANGULY R, MEHTA NM, GANGULY N, BANERJEE MR. Proc. Natl. Acad. Sci. USA 1979; 76: 6466-6470.

130. TOOLE JJ, HASTIE ND, HELD WA. Cell 1979; 17: 441-458.

131. FEIGELSON P, DE LAP L, CHEN CC, CHAN K, KURTZ D. Glucocorticoidal and developmental control of specific hepatic messenger RNA species *in vivo* and in hepatocytes *in vitro*. In: Busch H. ed. The Cell Nucleus Vol. VII Chromatin Part D. New York: Acad. Press, 1979; 229-257.

132. HASTIE ND, HELD WA TOOLE JJ. Cell 1979; 17: 449-457.

133. INNIS MA, MILLER DL. J. Biol. Chem. 1979; 254: 9148-9154.

134. SALA-TREPAT JM, DEVER J, SARGENT TD, THOMAS K, SELL S, BONNER J. Biochem. 1979; 18: 2167-2178.

135. EVANS GA, ROSENFELD MG. J. Biol. Chem. 1979; 254: 8023-8030.

136. KURTZ DT, CHAN K-M, FEIGELSON P. Cell 1978; 15: 743-750.

137. MOEN RC, ROWE DW, PALMITER RA. J. Biol. Chem. 1979; 254: 3526-3530.

138. KREAM BE, ROWE DW, GWOREK SC, RAISZ LG. Proc. Natl. Acad. Sci. USA 1980; 77: 5654-5658.

139. HIETER PA, HENDRICKS MB, HEMMINKI K, WEINBERG ES. Biochem. 1979; 18: 2707-2716.

140. GRUNSTEIN M. Proc. Natl. Acad. Sci. USA 1978; 75: 4135-4139.

141. DOWBENKO DJ, ENNIS HL. Proc. Natl. Acad. Sci. USA 1980; 77: 1791-1795.

142. CHILDS G, MAXON R, KEDES LH. Dev. Biol. 1979; : 153-174.

143. MARGOLSKEE JP, LODISH HF. Dev. Biol. 1980; : 50-63.

144. LODISH HF, MARGOLSKEE JP, BLUMBERG DD. In: Miami Winter Symposium 1978; 15: 169-183.

145. ALTON TA, LODISH HF. Cell 1977; 12: 301-310.

146. WILLIAMS JG, LLOYD MM, DEVINE JM. Cell 1979; 17: 903-913.

147. MACLEOD C, FIRTEL RA, PAPKOFF J. Dev. Biol. 1979; : 263-274.

148. LODISH HF. Ann. Rev. Biochem. 1976; 45: 39-72.

149. THIREOS G, SHEA RG, KAFATOS FC. Proc. Natl. Acad. Sci. USA 1979; 76: 6279-6283.

150. BARNETT T, PACHL C, GERGEN JP, WENSINK PC. Cell 1980; 21: 729-738.

151. ZITOMER RS, MONTGOMERY DL, NICHOLS DL, HALL BD. Proc. Natl. Acad. Sci. USA 1979; 76: 3627-3631.

152. RUDERMAN JV, WOODLAND HR, STURGESS EA. Dev. Biol. 1979; 71: 71-82.

153. ORDAHL CP, KIOUSSIS D, TILGHMAN SM, OUITT CE, FORNWALD J. Proc. Natl. Acad. Sci. USA 1980; 77: 4519-4523.

154. LASKY LA, LEV Z, XIN J-H, BRITTEN RJ, DAVIDSON EH. Proc. Natl. Acad. Sci. USA 1980; 77: 5317-5321.

155. HARPOLD MM, EVANS RM, SALDITT-GEORGIEFF M, DARNELL JE jr. Cell 1979; 17: 1025-1035.

156. BRITTEN RJ, DAVIDSON EH. Science 1969; 165: 349-357.

157. DAVIDSON EH, BRITTEN RJ. Quart.Rev. Biol. 1973; 48: 565-613.

158. DAVIDSON EH, BRITTEN RJ. Science 1979; 204: 1052-

159. KLEIN WH, THOMAS TL, LAI C, SCHELLER RH, BRITTEN RJ, DAVIDSON EH. Cell 1978; 14: 889-900.

160. SCHELLER RH, CONSTANTINI FD, KOZLOWSKI MR, BRITTEN RJ, DAVIDSON EH. Cell 1978; 15: 184-203.

161. CONSTANTINI FD, DAVIDSON EH, BRITTEN RJ. Nature 1980; 287: 111-116.

162. KINDLE KL, FIRTEL RA. Nucleic Acids Res. 1979; 6: 2403-2422.

163. KIMMEL AR, FIRTEL RA. Cell 1979; 16: 787-796.

164. ROBERTS R. In: News and Views. Nature 1980; 283: 132-133.

165. FIRTEL RA, TIMM R, KIMMEL AR, McKEOWN M. Proc. Natl. Acad. Sci. USA 1979; 76: 6202-6210.

166. JELINEK W, EVANS R, WILSON M. SALDITT-GEORGIEFF M, DARNELL JE. Biochemistry 1978; 17: 2776-2783

167. JELINEK W, LEINWAND L. Proc. Natl. Acad. Sci. USA 1980; 77: 1348-1402.

168. BRAUDE P, PELHAM H, FLACH G, LOBATTO R. Nature 1979; 282: 102-105.

169. JELINEK W. Proc. Natl. Acad. Sci. USA 1978; 75: 2679-2683.

170. RUBIN CM, HOUCK CM, DEININGER PL, FRIEDMANN T, SCHMID CW. Nature 1980; 284: 372-374.

171. HOUCK CM, RINEHART FP, SCHMID CW. J. Mol. Biol. 1979; 132: 289-306.

172. DUNCAN C, BIRO PA, CHOUDARY PV. et al. Proc. Natl. Acad. Sci. USA 1979; 76: 5095-5099.

173. COWAN K, TEGTMEYER P, ANTHONY DD. Proc. Natl. Acad. Sci. USA 1973; 70: 1927-1930.

174. JELINEK W, LEINWAND L. Cell 1978; 15: 205-214.

175. WEINER AM. Cell 1980; 22: 209-218.

176. BRANLANT C, KROL A, EBEL J-P. et al. Nucleic Acids Res. 1980; 8: 4149-

177. LERNER MR, STEITZ JA. Proc. Nat. Acad. Sci. USA 1979; 76: 5495-5499.

178. LERNER MR, BOYLE JA, MOUNT SM, WOLIN SL, STEITZ JA. Nature 1980; 283: 220-224.

179. WISE JA, WEINER AM. Cell 1980; 22: 109-118.

180. HICKS JB, HERSKOWITZ I. Genetics 1977; 85: 373-393.

181. KUSHNER PJ, BLAIR LC, HERSKOWITZ I. Proc. Natl. Acad. Sci. USA 1979; 76: 5264-5268.

182. SPRAGUE GF, RINE J, HERSKOWITZ I. Nature 1981; 289: 250-252.

183. KLAR AJS, STRATHERN JN, BROACH JR, HICKS JB. Nature 1981; 289: 234-244.

184. NASMYTH KA, TATCHELL K, HALL BD. *et al*. Nature 1981; 289: 244-250.

185. PAIGEN K. Ann. Rev. Genet. 1979; 13: 417-466.

186. PAIGEN K, MEISLER M, FELTON J, CHAPMAN V. Cell 1976; 9: 533-539.

187. PAIGEN K, SWANK RT, TOMINO S, GANSCHOW RE. J. Cell Physiol. 1975; 85: 379-392.

188. PAIGEN K, LABARCA C, WATSON G. Science 1979; 203: 554-556.

189. BERGER FG, PAIGEN K. Nature 1979; 282: 316-

190. LUSIS AJ, PAIGEN K. Cell 1975; 6: 371-378.

191. COLEMAN DL. J. Biol. Chem. 1969; 241: 5511-

192. COLEMAN DL. Science 1971; 173: 1245-1246.

193. BERNSTINE EG. J. Biol. Chem. 1979; 254: 83-87.

194. BERNSTINE EG, KOH C, LOVELACE CS. Proc. Natl. Acad. Sci. USA 1979; 76: 6539-6541.

195. BERNSTINE EG, KOH C. Proc. Natl. Acad. Sci. USA 1980; 77: 4193-4195.

196. SZOKA PR, PAIGEN K. Genetics 1978; 90: 597-612.

197. BOUBELIK M, LENGEROVA A, BAILEY DW, MATOUSEK V. Dev. Biol. 1976; 47: 206-214.

198. ABRAHAM I, DOANE WW. Proc. Natl. Acad. Sci. USA 1978; 75: 4446-4450.

199. CHOVNICK A, GELBART W, McCARRON WM. Cell 1977; 11: 1-10.

200. SPRADLING AC, MAHOWALD AP. Cell 1979; 16: 589-598.

201. WARING GL, MAHOWALD AP. Cell 1979; 16: 599-607.

202. SPRADLING AC, WARING GL, MAHOWALD AP. Cell 1979; 16: 609-616.

203. WELDON-JONES C, ROSENTHAL N, RODAKIS GC, KAFATOS FC. Cell 1979; 18: 1317-1332.

204. BLAU HM, KAFATOS FC. Dev. Biol. 1979; 72: 211-225.

205. NADEL MR, THIREOS G, KAFATOS FC. Cell 1980; 20: 649-658.

206. IATROU K, TSITILOU SG, GOLDSMITH MR, KAFATOS FC. Cell 1980; 20: 649-658.

207. WOOD WG, OLD JM, ROBERTS AVS, CLEGG JB, WEATHERALL DJ. Cell 1978; 15: 437-446.

208. FRITSCH EF, LAWN RM, MANIATIS T. Nature 1979; 279: 598-603.

209. BENZ EJ, Jr., FORGET BG, HILLMAN DG, COHEN-SOLAL M, PRITCHARD J, CAVALLESCO C. Cell 1978; 14: 299-312.

210. OLD JM, PROUDFOOT NJ, WOOD WG, LONGLEY JI, CLEGG JB, WEATHERALL DJ. Cell 1978; 14: 289-299.

211. FLAVELL RA, BERNARDS R, DE BOER E. Nucleic Acids Res. 1979; 6: 2749-2759.

212. CHANG JC, TEMPLE GF, TRECARTIN RF, KAN YW. Nature 1979; 281: 602-603.

213. BERNARDS R, FLAVELL RA. Nucleic Acids Res. 1980; 8: 1521-1534.

214. VAN DER PLOEG LHT, KONINGS A, OORT M, ROOS D, BERNINI L, FLAVELL RA. Nature 1980; 283: 637-642.

215. BERNARDS R, KOOTER JM, FLAVELL RA. Gene 1979; 6: 265-280.

216. TILGHMAN SM, TIEMEIER DC, SEIDMAN JG, <u>et al</u>. Proc. Natl. Acad. Sci. USA 1978; 75: 725-729.

217. JEFFREYS AJ, FLAVELL RA. Cell 1977; 12: 1097-1108.

218. BREATHNACH R, MANDEL JC, CHAMBON P. Nature 1977; 270: 314-319.

219. DOEL MJ, HOUGHTON M, COOK EA, CAREY NH. Nucleic Acids Res. 1977; 4: 3701-3713.

220. WEINSTOCK R, SWEET R, WEISS M, CEDAR H, AXEL R. Proc. Natl. Acad. Sci. USA 1978; 75: 1299-1303.

221. GILBERT W. Nature 1978; 271: 501.

222. LACY E, MANIATIS T. Cell 1980; 21: 541-553.

223. JAHN CL, HUTCHISON CA, PHILLIPS SJ. Cell 1980; 21: 159-168.

224. PROUDFOOT NJ, MANIATIS T. Cell 1980; 21: 537-544.

225. CLEARY MC, HAYNES JR, SCHON EA, LINGREL JB. Nucleic Acids Res. 1980; 8: 4791-4801.

226. FRITSCH EF, LAWN RM, MANIATIS T. Cell 1980; 19: 959-972.

227. BERGET SM, MOORE C, SHARP PA. Proc. Natl. Acad. Sci. USA 1977; 74: 3171-3175.

228. ALONI Y, DHAR R, LAUB O, HOROWITZ M, KHOURY G. Proc. Natl. Acad. Sci. USA 1977; 74: 3686-3690.

229. BERK AJ, SHARP PA. Proc. Natl. Acad. Sci. USA 1978; 75: 1274-1278.

230. THOMAS M, WHITE RL, DAVIS RW. Proc. Natl. Acad. Sci. USA 1976; 73: 2294-2298.

231. HOROWITZ M, LAUB O, BRATOSIN S, ALONI Y. Nature 1978; 275: 558-559.

232. SARGENT TD, WU J-R, SALA-TREPAT JM, WALLACE RB, REYES AA, BONNER J. Proc. Natl. Acad. Sci. USA 1979; 76: 3256-3260.

233. CHIEN Y-H, THOMPSON EB. Proc. Natl. Acad. Sci. USA 1980; 77: 4583-4587.

234. LOMEDICO P, ROSENTHAL N, EFSTRATIADIS A, GILBERT W, KOLODNER R, TIZARD R. Cell 1979; 18: 545-558.

235. BELL GI, PICTET RL, RUTTER WJ, CORDELL B, TISCHER E, GOODMAN HM. Nature 1980; 284: 26-32.

236. BELL GI, PICTET RL, RUTTER WJ. — Nucleic Acids Res. 1980; 8: 4091-4109.

237. MEANS AR, O'MALLEY BW. — Cell 1980; 21: 681-687.

238. LINDENMAIER W, NGUYEN-HUU MC, LURZ R, _et al_. — Proc. Natl. Acad. Sci. USA 1979; 76: 6196-6200.

239. NAGATA S, MANTEI N, WEISSMANN C. — Nature 1980; 287: 401-410.

240. GOEDDEL DV, SHEPARD HM, YELVERTON E. _et al_. — Nucleic Acids Res. 1980; 8: 4057-4073.

241. DAWID IB, WAHLI W. — Dev. Biol. 1979; 69: 305-328.

242. GALLWITZ O, SURES I. — Proc. Natl. Acad. Sci. USA 1980; 77: 2541-2550.

243. NG, R, ABELSON J. — Proc. Natl. Acad. Sci. USA 1980; 77: 3912-3916.

244. MONTGOMERY DC, HALL BD. — Cell 1979; 16: 753-761.

245. HOLLAND JP, HOLLAND MJ. — J.Biol.Chem. 1979; 254: 9839-9854.

246. EFSTRATIADIS A, POSAKONY JW, MANIATIS T, _et al_. — Cell 1980; 21: 653-668.

247. PATIENT RK, ELKINGTON JA, KAY RM, WILLIAMS JG. — Cell 1980; 21: 565-573.

248. FRYBERG EA, KINDLE KL, DAVIDSON N, SODJA A. — Cell 1980; 19: 365-378.

249. KIMMEL AR, FIRTEL RA. — Nucleic Acids Res. 1980; 8: 5599-5609.

250. GILBERT W. — Nature 1978; 271: 501.

251. DARNELL JE, Jr. — Science 1978; 202: 1257-1260.

252. JUNG A, SIPPEL AE, GREZ M, SCHUTZ G. — Proc. Natl. Acad. Sci. USA 1980; 77: 5759-5763.

253. EATON WA. — Nature 1980; 284: 183-185.

254. CRAIK CS, BUCHANAN SR, BEYCHOK S. — Proc. Natl. Acad. Sci. USA 1980; 77: 1384-1388.

255. COCHET M, GANNON F, HEN R, MAROTEAUX L, PERRIN F, CHAMBON P. — Nature 1979; 282: 567-

256. CORDELL B, BELL G, TISCHER E, _et al_. Cell 1979; 18: 533-543.

257. NISHIOKA Y, LEDER A, LEDER P. Proc. Natl. Acad. Sci. USA 1980; 77: 2806-2809.

258. VAMIN EF, GOLDBERG GI, TUCKER PW, SMITHIES O. Nature 1980; 286: 222-

259. BREATHNACH R, BENOIST C, O'HARE K, GANNON F, CHAMBON P. Proc. Natl. Acad. Sci. USA 1978; 75: 4853-4857.

260. CATTERALL JF, O'MALLEY BW, ROBERTSON MA, STADEN R, TANAKA Y, BROWNLEE GG. Nature 1978; 275: 510-513.

261. LEDER P, HANSEN JN, KONKEL D, LEDER A, NISHIOKA Y, TALKINGTON C. Science 1980; 209: 1336-1342.

262. PROUDFOOT NJ, SHANDER MHM, MANLEY JL, GEFTER ML, MANIATIS T. Science 1980; 209: 1329-1336.

263. MARTIN SL, ZIMMER EA, KAN YW, WILSON AC. Proc. Natl. Acad. Sci. USA 1980; 77: 3563-3566.

264. FLAVELL RA, KOOTER JM, De BOER E, LITTLE PFR, WILLIAMSON R. Cell 1978; 15: 25-41.

265. MACDONALD RJ, CRERAR MM, SWAIN WF, PICTET RL, THOMAS G, RUTTER WJ. Nature 1980; 287: 117-122.

266. HELLIG R, PERRIN F, GANNON F, MANDEL JL, CHAMBON P. Cell 1980; 20: 625-637.

267. DODGSON JB, STROMMER J, ENGEL JD. Cell 1979; 17: 879-887.

268. DEISSEROTH A, NIENHUIS A, LAWRENCE J, GILES R, TURNER P, RUDDLE FH. Proc. Natl. Acad. Sci. USA 1978; 75: 1456-1460.

269. KAY RM, WILLIAMS JG. Cell 1980; 21: 555-564.

270. WELDON JONES C, KAFATOS FC. Nature 1980; 284: 535-638.

271. GOLDSCHMIDT-CLERMONT M. Nucleic Acids Res. 1980; 8: 235-252.

272. BERK AJ, SHARP PA. Cell 1977; 12: 721-732.

273. SHATKIN AJ. Cell 1976; 9: 645-653.

274. BUSSLINGER M, PORTMANN R, IRMINGER JC, BIRNSTIEL ML. Nucleic Acids Res. 1980; 8: 957-977.

275. HENTSCHEL C, IRMINGER J-C, BUCHER P, BIRNSTIEL ML. Nature 1980; 285: 147-151.

276. BENOIST C, O'HARE K, BREATHNACH R, CHAMBON P. Nucleic Acids Res. 1980; 8: 127-141.

277. INGOLIA TD, CRAIG EA, McCARTHY BJ. Cell 1980; 21: 669-679.

278. TSUJIMOTO Y, SUZUKI Y. Cell 1979; 16: 425-436.

279. BUSSLINGER M, PORTMANN R, BIRNSTIEL ML. Nucleic Acids Res. 1979; 6: 2977-3008.

280. PROUDFOOT NJ, BROWNLEE GG. Nature 1976; 263: 211-214.

281. NUNBERG J, KAUFMAN R, CHANG A. et al. Cell 1980; 19: 355-364.

282. SMITH HO, BIRNSTIEL ML. Cell 1978; 14: 655-671.

283. WEIL PA, LUSE DS, SEGALL J, ROEDER RG. Cell 1979; 18: 469-484.

284. MANLEY JL, FIRE A, CANO A, SHARP PA, GEFTER ML. Proc. Natl. Acad. Sci. USA 1980; 77: 3855-3859.

285. LUSE DS, ROEDER RG. Cell 1980; 20: 691-699.

286. CORDEN J, WASYLYK B, BUGHWALDER A, SASSONE-CORSI P, KEDINGER C, CHAMBON P. Science 1980; 209: 1406-1414.

287. WASYLYK B, KEDINGER C, CORDEN J, BRISON O, CHAMBON P. Nature 1980; 285: 367-373.

288. WASYLYK B, DERBYSHIRE R, GUY A, et al. Proc. Natl. Acad. Sci. USA 1980; 77: 7204-7208.

289. BENOIST C, CHAMBON P. Proc. Natl. Acad. Sci. USA 1980; 77: 3865-3869.

290. GROSSCHEDL R, BIRNSTEIL ML. Proc. Natl. Acad. Sci. USA 1980; 77: 1432-1436.

291. MINTY A, NEWMARK P. In: News and Views. Nature 1980; 288: 210-211.

292. BAKER C, HERISSE J, COURTOIS G, GALIBERT F, ZIFF E. Cell 1979; 18: 569-580.

293. HAEGEMAN G, FIERS W. Nucleic Acids Res. 1978; 7: 2359-2371.

294. PELLICER A, ROBINS D, WOLD B, et al. Science 1980; 209: 1414-1422.

295. DEISSEROTH A, HENRICK D. Cell 1978; 15: 55-63.

296. WOLD B, WIGLER M, LACY E, MANIATIS T, SILVERSTEIN S, AXEL R. Proc. Natl. Acad. Sci. USA 1979; 76: 5684-5688.

297. MANTEI N, BOLL W, WEISSMANN C. Nature 1979; 281: 40-46.

298. BREATHNACH R, MANTEI N, CHAMBON P. Proc. Natl. Acad. Sci. USA 1980; 77: 740-744.

299. HUTTNER KM, SCANGOS GA, RUDDLE FH. Proc. Natl. Acad. Sci. USA 1979; 76: 5820-5824.

300. HAMER DH, LEDER P. Nature 1979; 281: 35-40.

301. MULLIGAN RC, HOWARD BH, BERG P. Nature 1979; 277: 108-114.

302. HAMER DH, KAEHLER M, LEDER P. Cell 1980; 21: 697-708.

303. WEINMAN R, ROEDER RE. Proc. Natl. Acad. SCi. USA 1974; 71: 1790-1794.

304. BIRKENMEIER EH, BROWN DD, JORDAN E. Cell 1978; 15: 1977-1058.

305. NG S-Y, PARKER CS, ROEDER RG. Proc. Natl. Acad. Sci. USA 1979; 76: 136-140.

306. WU GJ. Proc. Natl. Acad. Sci. USA 1978; 75: 2175-2179.

307. SAKONJU S, BOGENHAGEN DF, BROWN DD. Cell 1980; 19: 13-25.

308. BOGENHAGEN DF, SAKONJU S, BROWN DD. Cell 1980; 19: 27-35.

309. PELHAM HRB, BROWN DD. Proc. Natl. Acad. Sci. USA 1980; 77: 4170-4174.

310. ENGELKE DR, NG S-Y, SHASTRY BS, ROEDER RG. Cell 1980; 19: 717-728.

311. HONDA BM, ROEDER RG. Cell 1980; 22: 119-126.

312. CLEVER U. Ann. Rev. Genet. 1968; 2: 11-30.

313. ASHBURNER M, BONNER JJ. Cell 1979; 17: 241-254.

314. COMPTON JL, McCARTHY BJ. Cell 1978; 14: 191-201.

315. YAMAMOTO KR, ALBERTOS BM. Ann. Rev. Biochem. 1976; 45: 721-746.

316. HIGGINS SJ, GEHRING U. Adv. Cancer Res. 1978; 28: 313-97

317. OHNO S, CHRISTIAN L, ATTARDI BJ, KAN J. Nature New Biol. 1973; 245: 92-93.

318. LINN S, RIGGS AD. Cell 1975; 4: 107-111.

319. TSAI M, TSAI S, O'MALLEY B. In Vitro Chromatin Transcription in: Busch H. ed. The Cell Nucleus Vol. VII Chromatin Part D. New York: Acad. Press, 1979; 163-197.

320. GOTTESFELD J, MURPHY RF, BONNER J. Proc. Natl. Acad. Sci. USA 1975; 72: 4404-4408.

321. NOLL M. Nucleic Acids Res. 1974; 1:

322. GROUDINE M, DAS S, NEIMAN P, WEINTRAUB H. Cell 1978; 14: 865-878.

323. STALDER J, GROUDINE M, DODGSON JB, ENGEL JD, WEINTRAUB H. Cell 1980; 19: 973-980.

324. STALDER J, LARSEN A, ENGEL JD, DOLAN M, GROUDINE M, WEINTRAUB H. Cell 1980; 20: 451-460.

325. WEISBROD S, GROUDINE M, WEINTRAUB H. Cell 1980; 19: 289-301.

326. WEISBROD S, WEINTRAUB H. Cell 1981; 23: in press.

327. GAREL A, AXEL R. Proc. Natl. Acad. Sci. USA 1976; 73: 3966-3970.

328. PANET A, CEDAR H. Cell 1977; 11: 933-940.

329. BELLARD M, GANNON F, CHAMBON P. Cold Spring Harb. Symp. Quant. Biol. 1978; 42:

330. GAREL A, ZOLAN A, AXEL R. Proc. Natl. Acad. Sci. USA 1977; 74: 4867-4871.

331. SHEPHERD JH, MULVIHILL ER, PALMITER RD. J. Cell Biol. 1977; 75: 353a.

332. MILLER DM, TURNER P, NIENHUIS AW, AXELROD DE, GOPALAKRISHNAN TV. Cell 1978; 14: 511-521.

333. GREAVES M.,JANASSY G. Biochem.Biophys.Acta 1978;516: 193-230.

334. WU C, BINGHAM PM, LIVAK K, HOLMGREN R, ELGIN SCR. Cell 1979; 16: 797-806.

335. WU C, WONG Y-C, ELGIN SCR. Cell 1979; 16: 807-814.

336. KUO MT, MANDEL JL, CHAMBON P. Nucleic Acids Res. 1979; 7: 2081-2103.

337. WU C. Nature 1980; 286: 854-860.

338. SCOTT WA, WIGMORE DJ. Cell 1978; 15: 1511-1518.

339. VARSHAVSKY A, SUNDIN O, BAHN M. Cell 1979; 16: 453-466.

340. OUDET P, GROSS-BELLARD M, CHAMBON P. Cell 1975; 4: 281-300.

341. OLINS AL, OLINS KE. Science 1974; 183: 330-332.

342. SCHEER U. Cell 1978; 13: 535-549.

343. LOHR D, TATCHELL K, VAN HOLDE KE. Cell 1977; 12: 829-836.

344. LEVY-WILSON B, KUEHL L, DIXON GH. Nucleic Acids Res. 1980; 8: 2859-2869.

345. MAYFIELD JE, SERUNIAN LA, SILVER LM, ELGIN SCR. Cell 1978; 14: 539-544.

346. JOHNS EW, GOODWIN GH, WALKER JM, SAUNDERS C. Ciba Foundation Symposium 1975; 28: 95-108.

347. MATHIS D, OUDET P, CHAMBON P. C.R.C. Crit. Rev. Biochem. 1980

348. BIRD AP. Nucleic Acids Res. 1980; 8: 1499-1504.

349. MANDEL JL, CHAMBON P. Nucleic Acids Res. 1979; 7: 2081-2103.

350. VAN DER PLOEG LHT, FLAVELL RA. Cell 1980; 19: 947-958.

351. VAN DER PLOEG LHT, GROFFEN J, FLAVELL RA. Nucleic Acids Res. 1980; 8, 4563-4574.

352. WEINTRAUB H. (Unpublished observations).

353. GROUDINE M. (Unpublished observations).

354. THOMAS GP, MATTHEWS MB. Cell 1980; 22: 523-539.

356. PERLE MA, NEWMAN SA. Proc. Natl. Acad. Sci. USA 1980; 77: 3773-3777.

357. SARAGOST IS, MOYNE G, YANIV M. Cell 1980; 20: 65-73.

358. GOTTESFELD JM, BLOOMER LS. Cell 1980; 21: 751-760.

359. WEINTRAUB H. Nucleic Acids Res. 1978; 5: 1179-1188.

360. WEINTRAUB H. Nucleic Acids Res. 1980; 8: 4745-4753.

361. BRINSTER RL, CHEN HY, TRUMBAUER ME, AVARBOCK MR. Nature 1980; 283: 499-501.

362. FARMER SR, BEN-ZEEV A, BENECKE B-J, PENMAN S. Cell 1978; 15: 627-637.

363. BENECKE B-J, BEN-ZEEV A, PENMAN S. Cell 1978; 14: 931-939.

364. HEYWOOD SM, KENNEDY DS, BESTER AJ. Proc. Natl. Acad. Sci. USA 1974; 71: 2428-2431.

365. BESTER AJ, KENNEDY DS, HEYWOOD SM. Proc. Natl. Acad. Sci. USA 1975; 72: 1523-1527.

366. LENK R, HERMAN R, PENMAN S. Nucleic Acids Res. 1978; 8: 3057-3070.

367. PERRY RP, SCHIBLER U, MEYUHAS O. Miami Winter Symposium 1979; 16: 187-205.

368. MEYUHAS O, PERRY RP. Cell 1979; 16: 139-148.

369. WOODLAND HR, FLINT FH. Dev. Biol. 1980; 75: 199-213.

370. IATROU K, DIXON GH. Cell 1977; 10: 433-441.

371. WILSON MC, SAWICKI SG, WHITE PA, DARNELL JE. J. Mol. Biol. 1978; : 23-36.

372. MERKEL CG, WOOD TG, LINGREL JB. J. Biol. Chem. 1976; 251: 5512-5515.

373. PALATNIK CM, STORTI RV, CAPONE AK, JACOBSON A. J. Mol. Biol. 1980; : 99-118.

374. MARBAIX G, HUEZ G, BURNY A, *et al*. Proc. Natl. Acad. Sci. USA 1975; 72: 3065-3067.

375. HUEZ G, MARBAIX G, BURNY A. *et al*. Nature 1977; 266: 473-474.

376. HUEZ G, MARBAIX G, GALL-WITZ D, *et al*. Nature 1978; 271: 572-573.

377. SEGHAL PB, SOREQ H, TAMM I. Proc. Natl. Acad. Sci. USA 1978; 75: 5030-5033.

378. NEVINS JR, DARNELL JE, Jr. Cell 1978; 15: 1477-1493.

379. FRASER NW, NEVINS JR, ZIFF E, DARNELL JE. Jr. J. Mol. Biol. 1979; 129: 643-656.

380. BERK AJ, SHARP PA. Proc. Natl. Acad. Sci. USA 1978; 75: 1274-1278.

381. REDDY VB, THIMMAPPAYA B, DHAR R, *et al*. Science 1978; 200: 494-502.

382. KNAPP G, BECKMAN JS, JOHNSON PF, FUHRMAN SA, ABELSON J. Cell 1978; 14: 221-236.

383. HOPPER AK, BANKS F, EVANGELIDIS V. Cell 1978; 14: 211-219.

384. REDDY V, GHOSH P, LEBO-WITZ P, PIATAK M, WEISSMANN S. J. Virol. 1979; 30: 279-296.

385. TRAPNELL BC, TOLSTOSHEV P, CRYSTAL RG. Nucleic Acids Res. 1980; 8: 3659-3672.

386. JACOB M. (Unpublished observation).

387. AVVEDIMENTO VE, VOGELI G, YAMADA Y, MAIZEL JV, Jr., PASTAN I, CROMBRUGGHE B. Cell 1980; 21: 689-696.

388. KINNIBURGH AJ, ROSS J. Cell 1979; 17: 915-921.

389. HAMER DH, LEDER P. Cell 1979; 18: 1299-1302.

390. HAMER DH, LEDER P. Cell 1979; 17: 737-747.

391. GRUSS P, LAI C-J, DHAR R, KHOURY G. Proc. Natl. Acad. Sci. USA 1979; 76: 4317-4321.

392. LEE A-S, THOMAS TL, LEV Z, BRITTEN RJ, DAVIDSON EH. Proc. Natl. Acad. Sci. USA 1980; 77: 3259-3263.

393. ROGERS J, EARLY P, CARTER K, *et al*. Cell 1980; 20: 303-312.

394. EARLY P, ROGERS J, DAVIS M, *et al*. Cell 1980; 20: 313-319.

395. LIU C-P, TUCKER PW, MUSHINSKI F, BLATTNER FR. Science 1980; 209: 1348-1358.

396. TUCKER PW, LIU C-P, MUSHINSKI JF, BLATTNER FR. Science 1980; 209: 1353-1360.

397. KANTOR JA, TURNER PH, NIENHUIS AW. Cell 1980; 21: 149-157.

398. LAZOWSKA J, JACQ C, SLONIMISKI PP. Cell 1980; 22: 333-348.

399. DUJON B. Nature 1979; 282: 777-778.

400. ROBERTS BE, PATERSON BM. Proc. Natl. Acad. Sci. USA 1973; 70: 2330-2334.

401. PELHAM HRB, JACKSON RJ. Eur. J. Biochem. 1976; 67: 247-256.

402. CORDELL B, WEISS SR, VARMUS HE, BISHOP JM Cell 1978; 15: 79-91.

403. AFFARA NA, YOUNG BD. 1976 MSE Application Information Sheet A 12.6.76.

404. HUMPHRIES P, OLD R, COGGINS LW, McSHANE T, WATSON C, PAUL J. Nucleic Acids Res. 1978; 5: 905-924.

405. ADESNIK M, SALDITT M, THOMAS W, DARNELL JE, Jr. J. Mol. Biol. 1972; 71:21-30.

406. AVIV H, LEDER P. Proc. Natl. Acad. Sci. USA 1972; 69: 1408-1414.

407. ROTHENBERG E, BALTIMORE D. J. Virol. 1976; 21: 168-178.

408. WICKENS MP, BUELL GN, SCHIMKE RT. J. Biol. Chem. 1978; 253: 2483-2495.

409. COCHET M, PERRIN F, GANNON F. et al. — Nucleic Acids Res. 1979; 6: 2435-2452.

410. BIRNIE GD. — Isopycnic Centrifugation in ionic media in Birnie GD & Rickwood D. eds. Centrifugal Separations in Molecular and Cell Biology. London: Butterworths. 1978. 170-215.

411. BIRNBOIM HC, DOLY J. — Nucleic Acids Res. 1979; 7: 1513-1522.

412. GERGEN JP, STERN RH, WENSINK PC. — Nucleic Acids Res. 1979; 7: 2115-2136.

413. RIGBY P. — J. Mol. Biol. 1977; 113: 237-

414. BULOVA SF, BURKA ER. — J. Biol. Chem. 1970; 245: 4907-4912.

415. LODISH HF, DESALU O. — J. Biol. Chem. 1973; 248: 3520-3527.

416. LODISH HF, SMALL B, CHANG H. — Dev. Biol. 1975; 47: 59-67.

417. LODISH HF, SMALL B. — J. Cell Biol. 1975; 65: 51-64.

418. O'FARREL PH. — J. Biol. Chem. 1975; 250: 4007-4021.

419. ENGLISH D, ANDERSON BR. — J. Immunol. Methods 1974; 5:249-252.

420. BOLLUM FJ, CHANG LMS, TSIAPULIS CM, DORSON JW. — Methods Enzymol. 1973; 29E: 150-173.

421. SUTCLIFFE JG. — Nucleic Acids Res. 1978; 5: 2721-2728.

422. SUTCLIFFE JG. — Cold Spring Harbor Symposium

423. ROYCHOWDRY R, JAY E, WU R. — Nucleic Acids Res. 1976; 3: 863-877.

424. HUMPHRIES P, COCHET M, KRUST A, GERLINGER P, KOURILSKY P, CHAMBON P. — Nucleic Acids Res. 1977; 4: 2389-2406.

425. BOLIVAR F, RODRIGUEZ RL, GREENE PJ, et al. Gene 1977; 2: 95-113.

426. FANTONI A, BOZZONI I, ULLU E, FARACE MG. Nucleic Acids Res. 1979; 6: 3505-3517.

427. SUTCLIFFE JG, CHURCH GM. Nucleic Acids Res. 1978; 5: 2313-2319.

428. SUZUKI S, ALEXRAD A. Cell 1980; 19: 225-236.

429. OBINATA M, IKAWA Y. Nucleic Acids Res. 1980; 8: 4271-4282.

430. STEWART AG, CLISSOLD PM, ARNSTEIN HRV. Eur. J. Biochem. 1976; 65: 349-355.

431. LODISH HF. Proc. Natl. Acad. Sci. USA 1973; 70: 1526-1530.

432. CHANG H, LANGER PJ, LODISH HF. Proc. Natl. Acad. Sci. USA 1976; 73: 3206-3210.

433. KOCH PA, GARDNER FH, CARTER JR, Jr. Biochem. Biophys. Res. Commun. 1973; 54: 1296-1299.

434. THIELE BJ, BELKNER J, ANDREE H, RAPOPORT TA, RAPOPORT SM. Eur. J. Biochem. 1979; 96: 563-569.

435. DENTON MJ, SPENCER N, ARNSTEIN HRV. Biochem. J. 1975; 146: 205-211.

436. BEUZARD Y, RODVIEN R, LONDON IM. Proc. Natl. Acad. Sci. USA 1973; 70: 1022-1026.

437. CONSCIENCE JF, MILLER RA, HENRY J, RUDDLE H. Exp. Cell Res. 1977; 105: 401-

438. ROBBI M, LAZAROW PB. Proc. Natl. Acad. Sci. USA 1978; 75: 4344-4348.

439. BACKMAN K. Gene 1980; 11: 169-171.

440. KAFATOS FC, JONES CW, EFSTRATIADIS A. Nucleic Acids Res. 1979; 7: 1541-1552.

441. EBERT PS, IKAWA Y. Proc. Soc. Exp. Biol. Med. 1974; 146: 601-

442. REEM GH, FRIEND C. Proc. Natl. Acad. Sci. USA 1975; 72: 1630-1634.

443. WU NC, SIKKEMA DA, ZUCKER RM. Biochem. Biophys. Acta 1978; 536: 306-311.

444. WU NC, ZUCKER RM. Biochem. Biophys. Acta 1979; 537: 299-302.

445. LICHTMAN MA. In Haematology & Oncology; The Science and Practice of Clinical Medicine Vol. 6. Ed. Lichtman MA, New York: Grove and Stratton 1980; 152.

446. HENTSCHEL CC, KAY RM, WILLIAMS JG. Dev. Biol. 1979; 72: 350-362.

447. ELIASON JF, TESTA NG, DEXTER TM. Nature 1979; 281: 382-384.

448. SCHREML W, BURKA ER. J. Biol. Chem. 1968; 243: 3573-3580.

449. BONNER J, LASKY R. Eur. J. Biochem. 1974; 46: 83-88.

Acknowledgements

A long list of thank yous: to the M.R.C. for a Ph.D. Studentship, to Lorna Peedle for her extremely efficient and speedy typing of this manuscript, to the whole staff of the Beatson Institute for providing an amiable and helpful atmosphere in which to work, to David Tallach for his help with the artwork and particularly to my supervisor, Paul Harrison, for his guidance and criticisms during the evolution of this work. Finally, one special thank you to Lena, without whose kindness and understanding it would not have been written.