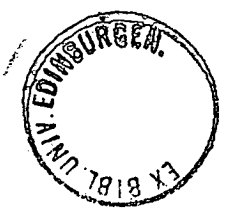
ISOLATION AND CHARACTERIZATION OF THE cDNA FOR CYSTIC FIBROSIS ANTIGEN

JULIA RUTH DORIN

Ph.D.
University of Edinburgh
1987



ACKNOWLEDGEMENTS

The past three years have been extremely enjoyable and I would like to thank the people responsible:-

Veronica,

the nicest supervisor possible, who always took time to guide, encourage, teach and cheer.

Nick,

for his unfailing enthusiasm guidance and demand for hot data.

Bob,

who taught me the A, B, cDNA, and helped me get the message.

Richard,

who helps when there RNA problems.

Anne,

the best benchmate anyone could have.

John,

who put up with me, made me oligonucleotides and corrected my spelling mistaks.

Judy,

for expert chromosome analysis and FACS.

Maureen,

for beautiful immunohistochemistry.

Richard B,

for hybrid cell DNA.

Sandy, Norman and Douglas,

for their superb photography.

I would also like to thank David Brock, Caroline Hayward and Michal Novak for helpful discussion and invaluable collaboration.

Lastly thanks to Mum and Dad who commiserated and celebrated throughout, and will be glad to know it's all over.

This work was carried out at MRC Clinical and Population Cytogenetics Unit, Edinburgh, with the encouragement of Professor H.J. Evans and was funded by an MRC training award.

DECLARATION

I declare that :-

a) this thesis has been composed by myself,

b) the work is either my own or the worker/author involved is clearly stated.

ABSTRACT

Cystic fibrosis antigen (CFAg) is a protein which is present in the serum of cystic fibrosis (CF) patients and clinically normal heterozygotes, but not in normal individuals. CFAg has been shown to be a major granulocyte protein in normals, and the gene mapped to chromosome 1. This thesis describes the molecular cloning and subsequent characterization of the cDNA for CFAg.

The availability of monoclonal antibodies to CFAg facilitated the identification of myeloid tissues which were actively synthesising the protein. A specific radiolabelled protein could be immunoprecipitated from ^{35}S-methionine labelled extracts of chronic myeloid leukaemia cells (CML), normal granulocytes and HL-60 cells differentiated towards granulocytes. In CML and granulocytes, CFAg appeared to be one of the most abundantly synthesised proteins.

Messenger RNA was isolated from CML cells and a cDNA library of ~70,000 members was constructed in the vector λgt10. The cDNA clone for CFAg was identified by screening the library with a highly degenerate oligonucleotide probe designed using available protein sequence data.

Verification of the identity of the largest cDNA clone isolated was obtained by comparing the amino acid sequence derived from the cDNA with that of the partially sequenced protein. The size of the protein from the deduced sequence combined with protein sequencing data is 10,938 Kd, which is in approximate agreement with the size observed on a reducing gel system. Using the cDNA clone in Northern blot analysis, a 550bp transcript was observed which was highly abundant in CML and differentiated HL-60 cells. Southern blotting of genomic DNA indicates that there is only a single

gene for CFAg in both human and rodent. The availability of a series of somatic cell hybrids containing fragments of human chromosome 1, enabled the assignment of the gene for CFAg to the region q12-q22. Comparing the predicted sequence of CFAg with protein sequences in the NBRF database, significant similarity was discovered to two calcium-binding proteins (S-100 and intestinal calcium-binding protein) found in various mammalian species. These proteins are of unknown function but may be involved in signal transduction. Homology was also observed to two sequences as yet not present in the database, but from the same family of calcium-binding proteins.

The CF locus is linked to DNA markers on chromosome 7, and therefore CFAg cannot be the product of the aberrant gene. However the function of this novel, granulocytic calcium-binding protein must now be pursued to explain its abnormal accumulation in CF serum and point the way to the biochemical pathway affected in CF.

Table of Contents

Chapter 1: Introduction 1
1.1 The genetics of cystic fibrosis 1
1.2 Prenatal diagnosis 2
1.2.a Microvillar enzymes 2
1.2.b RFLPs 2
1.3 Heterozygote detection 3
1.3.a Prospects for the cloned gene 4
1.3.b Biochemical detection of genetic diseases 5
1.4 The clinical picture 6
1.5 Functional changes in CF cells 9
1.6 Factors associated with the CF genotype 15
1.7 Cystic fibrosis antigen (CFAg) 18
Chapter 2: Materials and methods 23
2.1 Tissue culture 23
2.2 HL-60 induction 23
2.3 NBT reduction assay 23
2.4 Isolation of polymorphonuclear cells from whole blood 24
2.5 Preparation of cell lysates 24
2.6 Ouchterlony double diffusion 25
2.7 Internal cell labelling 25
2.8 Immunoprecipitations 26
2.9 SDS polyacrylamide gel electrophoresis 26
2.10 Autoradiography 28
2.11 Vectors 29
2.12 Bacterial strains 31
2.13 Media and buffers for bacterial and phage culture 31
2.14 RNA Extraction 33
2.15 Isolation of poly(A)$^+$ RNA 34
2.16 Gel electrophoresis of RNA 34

2.17 Northern blot analysis 35
2.18 Nick translation 35
2.19 Random priming of DNA 36
2.20 *In vitro* translation of mRNA 37
2.21 cDNA library construction 37
2.22 End labelling of DNA 42
2.23 Plating and screening a lambda gt10 library 43
2.24 Preparing phage DNA 44
2.25 Cloning into λgt11 45
2.26 Screening λgt11 libraries wih antibody probes 46
2.27 Differential screening 47
2.28 Restriction endonuclease digestion 47
2.29 Agarose gel electrophoresis 48
2.30 Ligation of DNA fragments 48
2.31 Transformation of bacterial cells 48
2.32 Small scale plasmid preparation 50
2.33 Large scale preparation of plasmid DNA 50
2.34 Isolation of insert DNA 52
2.35 Sequencing in M13 53
2.36 Extraction of DNA from cell pellets 55
2.37 Southern blotting 56
Chapter 3: CFAg expression in myeloid cells 58
3.1 Turnover of CFAg in various cell types 58
3.2 Method of analysis 62
3.3 Results 63
3.3.a CFAg in HL-60 cells 63
3.3.b CFAg protein turnover in granulocytes from normals, CML and HL-60 cells. 68
3.3.c Immunoprecipitation of CFAg using McAb, from normal, CML and induced HL-60 69
3.4 Conclusions 70
Chapter 4: cDNA library construction 77
4.1 Introduction 77
4.2 Synthesising and cloning double-stranded cDNA 78

4.2.a The general methodology used 78
4.2.b The practical problems encountered 78
4.3 Characterization of λgt10 CML cDNA library 89
4.4 Construction of a λgt11 cDNA library 92
Chapter 5: Attempt to isolate CML specific clones 97
5.1 General strategy 97
5.2 Differential screening 97
5.3 Application of the differential screening technique to the λgt10 CML cDNA library 99
5.4 Discussion 111
Chapter 6: Isolation of the cDNA for CFAg 115
6.1.a Design of oligonucleotide probes to screen the λgt10 CML cDNA library 115
6.1.b Screening the λgt10 CML cDNA library with oligonucleotide probes 117
6.1.c Insert size of the lambda clones 118
6.1.d Verification of clone 8-9 being cDNA for CFAg 119
6.2 Sequencing data 119
6.3 Tissue distribution, transcript size and abundance of CFAg mRNA 128
6.3.a Tissue distribution 128
6.3.b Transcript size 129
6.3.c Abundance of CFAg mRNA in myeloid tissues 129
6.4 Southern blot analysis of genomic DNA with CFAg cDNA clone 8-9 133
6.5 Conservation of CFAg between species 133
6.6 Chromosomal localisation of the gene for CFAg 134
6.7 Conclusion 139
Chapter 7: Comparison of CFAg deduced sequence to other protein sequences 140
7.1 Introduction 140
7.2 Proteins with sequence similarity to CFAg 141
7.3 Calcium-binding proteins 143
Chapter 8: Discussion 149
8.1 Calmodulin 150

8.2 Intestinal calcium-binding proteins 151
8.3 S-100 152
8.4 p11 154
8.5 2A9 154
8.6 CFAg as a calcium-binding protein 155
8.7 Signal transduction 155
8.8 Relationship of CFAg to the gene defect in CFAg 157
8.9 Defective chloride channel activity in CF 157
8.10 Evidence for the involvment of a calcium-binding protein in CF 159
8.11 CFAg 160
8.12 Use of CFAg in heterozygote detection 161
8.13 Granulocytes as an affected tissue 162
Abbreviations 165
References 167

CHAPTER 1

INTRODUCTION

This thesis describes the cloning and characterisation of the cDNA for cystic fibrosis antigen (CFAg). Cystic fibrosis (CF) is a life threatening disease, the biochemical basis of which remains unknown. CFAg is present at raised levels in the serum of individuals with CF and heterozygotes. Importantly, the amount of CFAg present in the serum appears to track the CF gene, i.e. heterozygotes have reduced levels of the protein compared to affected individuals. This fact, together with the fact that heterozygotes are clinically unaffected and yet still demonstrate the accumulation of CFAg in their serum, indicates that CFAg may be directly concerned with the basic gene defect in CF. The cloning of CFAg has revealed that it is a novel calcium-binding protein and allows speculation as to its involvement in CF.

1.1 The genetics of cystic fibrosis.

Cystic fibrosis is one of the commonest European genetic diseases. It is inhereited in an autosomal recessive fashion, i.e on average one in four of the offspring in affected families have CF. Romeo (1984) has shown that the number of first cousin marriages in Italian "CF parents" closely matches that expected from the hypothesis of a single locus disorder, as opposed to a model with mutations at two or more loci. Assuming Hardy-Weinberg equilibrium i.e random assortment of the gene in the gene pool, then approximately 1 in 23 individuals in Caucasian populations carry the recessive gene.

CF heterozygotes (parents of CF children are obligate heterozygotes) are clinically normal and therefore any abnormality that is present consistently in heterozygotes and homozygotes, must reflect the basic gene defect and is not a secondary affect of the disease. The high frequency of the CF mutation in the population means that prenatal diagnosis and heterozygote detection are of prime importance.

1.2 Prenatal diagnosis.

1.2.a Microvillar enzymes.

A prenatal diagnostic assay has been developed based on the observation that amniotic fluid supernatants from fetuses affected with CF, contain reduced levels of microvillar enzymes (Carbarns et al 1983). The deficiency of the enzymes are probably the consequence of abnormal passage of a thickened and viscid meconium from the fetal gut into the fluid. There are two main disadvantages of measuring the level of microvillar enzymes in amniotic fluid to predict the CF status of the fetus. The first is that the deficiency of these enzymes is due to a physiological abnormality which is secondary to the basic gene defect. This means that the frequencies of false negative and positive prognoses are too high to offer the test except to mothers with a 1 in 4 risk of recurrence of an affected child (Brock 1984). The second disadvantage is that optimally the test is carried out at 18 weeks of gestation after which time a termination of pregnancy can be very traumatic for the mother. The advantage of the test is that although a previous affected child is required to identify the family as "at risk", the test does not require any tissue from that individual.

1.2.b Restriction Fragment Length Polymorphisms (RFLPs).

The other method of prenatal diagnosis which can be used to determine whether a fetus has CF, uses RFLPs identified by DNA probes linked to the CF locus. Recently a number of groups (Wainwright et al 1985; White et al 1985; Knowlton et al 1985) have shown that the CF locus is tightly linked to DNA markers on chromosome 7. A

variety of DNA polymorphic markers have been characterized which are close enough to the CF gene to be used confidently in prenatal diagnosis. The test is based on the assumption that there will be no recombination between the CF gene and the closely linked DNA probe(s). Consequently, for a certain individual if the CF gene can be associated with a particular allele of a tightly linked DNA probe, then it is assumed that the association will remain in the next generation. One of the limitations of the test is that the prospective parents must be informative for at least one DNA probe, i.e it must be possible to distinguish allelic variants for at least one linked probe by restriction enzyme digestion of their genomic DNA. A further limitation is that there must be DNA available from a previous affected child. This is so that it can be determined which alleles of the linked probe are associated with the maternal and paternal CF genes. The fact that CF is a life reducing disease often means that the previous child is dead, and so DNA is not available unless some tissue has been stored at the time of death.

The advantages of this method of prenatal diagnosis are that :-

a) due to the tight linkage to the CF locus of the DNA probes available, the false prediction frequency is low;

b) the fetus DNA sampling can be performed at 9 weeks of gestation using chorionic villus sampling which reduces the trauma inflicted on the mother, should a termination of pregnancy prove advisable.

1.3 Heterozygote detection

Both the prenatal tests described above are only appropriate in families where a CF child has already been born. In the microvillar enzyme assay, the previous affected child identifies the 1 in 400 couples who on average have a 1 in 4 chance of producing a CF child. In the RFLP test, DNA from a previous affected child is intrinsic to the assay. Ideally a heterozygote detection system should be designed to identify "at risk"

couples before a CF child is born. If this could be achieved the microvillar enzyme test could be used on the relevant pregnancies.

1.3.a Prospects for the cloned gene.

The gene for CF has not been cloned as yet. However when this happens, it will be possible to determine the molecular basis of the CF mutation. If the mutation is shown to be identical in every case and not a heterogeneous set of intragenic mutations, then the presence of the mutant gene may be easy to diagnose by DNA analysis. This has been shown to be possible for sickle cell anaemia, where the same base change in the DNA is always responsible for the disease. Only the normal allele gene sequence will form stable duplexes with an oligonucleotide probe (Conner 1983). Thus the presence of the mutant allele can be determined for any individual or fetus directly, by hybridising the oligonucleotide probe to their DNA. Even if this were possible for CF, screening the whole population using DNA analysis to identify heterozygotes would be unthinkable even if only considering the expense involved. What is required is a reliable, economical and technically simple test which uses a sample of easily available tissue or body fluid to distinguish CF heterozygotes. At present, the biochemical basis of the CF phenotype is unknown. It maybe that cloning the CF gene will immediately enable identification of its product and the biochemical basis of CF will be obvious. The techniques of reverse genetics which have been used for cloning the gene for Duchenne muscular dystrophy (Kunkel et al 1985) and are being applied to the isolation of the CF gene (Scambler et al 1986), will not necessarily identify the gene product at fault in the disease. The putative genes for Duchenne muscular dystrophy (Monaco et al 1986) and for retinoblastoma (Friend et al 1986) have been cloned, but as yet no gene product has been identified although linked DNA probes to Duchenne muscular dystrophy have been available for the last five years. However, identification of the CF gene product surely must improve the possibility of designing a simple heterozygote test. Alternatively, it may be possible to design a

carrier detection test by examining in detail the physiological phenotype in CF heterozygotes and homozygotes.

1.3.b Biochemical detection of genetic diseases.

Two model systems which identify disease-affected individuals and heterozygotes are described below.

Phenylketonuria

The disease phenylketonuria (PKU), is due to impaired phenylalanine oxidation resulting in elevated tissue and serum phenylalanine (Tourain and Sidbury 1983). Because phenylalanine hydroxylase is absent and phenylalanine cannot be converted to tyrosine, homozygotes have hyperphenylalanemia which leads to neurotransmitter synthesis deficit and brain dysfunction. Neonatal diagnosis is performed by measuring the phenylalanine:tyrosine ratio and clinically normal heterozygotes can also be identified by this ratio (Paul et al 1978). Heterozygotes have reduced levels of phenylalanine hydroxylase, and consequently, raised levels of phenylalanine and reduced levels of tyrosine in plasma. However, these deviations from normal levels are not enough to affect the heterozygote individual clinically. It should be noted that the diagnosis of PKU does not involve measuring the amount of phenylalanine hydroxylase, the gene product of the mutant gene. It involves measuring the amount of a substrate and product directly affected by fluctuations in the level of the mutant gene product. A final point is that the understanding of the basic defect in PKU not only allows diagnosis of the disease, but enables the design of a complete therapy i.e a restricted diet (without phenylalanine) avoids hyperphenylalanemia in affected individuals and results in a normal phenotype.

I-cell disease

I-cell disease, is so called because skin fibroblasts, from patients with the disease, are filled with innumerable, phase-dense, acid phosphatase-positive inclusions (Neufeld

and McKusick 1983). The disease is caused by the complete or partial deficiency of the phosphorylating step of acid hydrolases which is the identifying signal for their inclusion into lysosomes. As a consequence, they are secreted instead. Serum levels of some acid hydrolases are raised 20 fold in homozygotes, and 2 fold in heterozygotes, compared to normals (van Elsen et al 1976). The measurement of serum levels of heat stable hexosaminidase, which is not the product of the defective gene, can be used to identify I-cell disease heterozygotes from normals (Vidgoff and Buist 1977) .

Clearly, identifying the basis of the CF disease is very important. Abnormalities present in heterozygotes merit investigation because they must reflect a defect related directly to the CF gene and their use in a heterozygote detection assay must be assessed. Before discussing factors present in CF heterozygotes, it is of value to examine the overall clinical picture of CF and the probable affected tissues in which to look for clues as to which process is perturbed in the disease.

1.4 The Clinical picture

Any theory which postulates the aberrant pathway in CF must take into consideration the clinical symptoms associated with the disease and account for them somehow. However, it must be acknowledged that some disease symptoms may be secondary to the basic disease process.

Cystic fibrosis is a generalized disease of the exocrine glands. The name is derived from the histological changes which are characteristically seen in the pancreas of diseased individuals (Talamo et al 1983). These include fibrosis of acini, dilation of ductules, plugs within acini and focal acinar calcifications. The main pathology of the disease involves the respiratory tract and pulmonary complications invariably arise. However autopsy studies of neonatal lung samples (from infants who die as a sequel to meconium ileus) show that they are normal at birth (Oppenheimer 1981). Once lung involvement sets in, obstruction of the small airways is evident and secondarily to this

there is bronchiolitis and mucual plugging of the airways with bronchial changes being more apparent than parenchymal ones (Bedrossian et al 1976). The lung epithelium shows stratification, metaplasia of squamous cells, and loss of cilia which contributes to the retention of secretions. Chronic infection of the respiratory tract is very common, affecting ~80% of CF patients, and bronchitis and airway obstruction eventually lead to respiratory or cardiac failure. Later pulmonary infection is mainly due to the bacterium *Pseudomonas aeruginosa.* Chemotherapy remains a problem with longterm antibiotic treatment being controversial (Zimarkoff et al 1983). It is not known why CF patients are so susceptible to infection by the very unusual mucoid *Pseudomonas* variants and although pulmonary alveolar macrophages are basically normal their phagocytic function can be impaired by CF serum (Thomassen et al 1980).

The upper respiratory tract is also affected secondarily to hyperactive mucus secreting glands. Nasal polyps give rise to nasal obstruction in 6-24% of CF patients (Schwachman et al 1962).

The classic fibrosis of the pancreas appears to follow loss of pancreatic enzyme activity and in all but a few cases this must be treated with enzyme replacement therapy to correct the malabsorption of fats, proteins and carbohydrates (Johansen et al 1968). The clinical repercussions of malabsorption are: poor weight gain in spite of large food intake, lack of muscle tissue and subcutaneous fat, the passage of bulky, oily, foul-smelling stools leading to rectal prolapse, and delayed puberty (Chase et al 1979). An increased fecal bile loss and a decreased bile acid pool is found as is bile malabsorption (Fondacaro et al 1980). Malabsorption of the fat soluble vitamins is also due to pancreatic insufficiency and reduced levels of vitamins A, D, E, K and B12 are common (Talamo et al 1983).

Carbohydrate intolerance, present in 40% of patients, develops into frank diabetes mellitus in 1-2%, although this is generally very mild due to the observed increase in the number of insulin receptor sites in peripheral tissues (Lippe et al 1980).

A blockage of the gut by meconium (meconium ileus) is present in 7% of newborn CF babies and is almost always an indication of the disease (Donnison et al 1966). It is thought that this is due to loss of pancreatic enzymes preceding morphological pancreatic damage (Oppenheimer et al 1973). Transient blockage of the fetal intestine appears to be present around 18-19 weeks of gestation in most CF fetuses, allowing the resulting reduced levels of intestinal microvillar enymes to be used for prenatal diagnosis of CF by amniocentesis (section 1.2.a).

The liver is also damaged in CF probably due to excessive accumulation of biliary mucus. The number of microvilli on bile canalicular epithelium has been shown to be reduced (Dominick et al 1976). However normal liver function is usual, and failure rare.

Infertility is observed in 98% of adult males with CF due to absence or reduction of vas deferens (Taussig et al 1972). The testes show active, but decreased spermatogenesis and neonatal autopsy reveals normal genital tract. CF women however can bear children, and infertility is mainly due to excessive extracellular cervical mucus (Oppenheimer et al 1970).

The salivary mucus secreting glands also show ductal dilation and eventual fibrosis. A significant rise in Na^+ concentration has been observed in small salivary gland saliva (Wiesman et al 1972).

The commonest clinical finding in CF is elevated sweat concentration of sodium, chloride, and potassium ions due to decreased transductal reabsorbtion of sodium and chloride (Schultz 1969). Mangos (1973b) showed that *in vitro* sweat gland ducts had defective reabsorption of sodium, but normal water permeability and Quinton et al (1983a, 1983b) more recently observed reduced chloride permeability. Diagnosis of CF is based on the quantitative pilocarpine iontophoresis sweat test (Gibson et al 1959) where localised sweating is stimulated by iontophoresis of pilocarpine into the skin and the electrolytic composition of collected sweat is measured. This defect was first

discovered in a New York heatwave in 1948. More recently the level of neonatal faecal proteases has been used as a diagnosis of CF as they are found to be highly reduced.

1.5. Functional changes in CF cells

CF appears to be a disease characterized by chronic lung disease, exocrine pancreatic insufficiency, and elevated sweat electrolytes (Di Sant' Agnese et al 1976). Consequently an examination of the cells closely involved in these functions may provide an indication of the pathophysiology of the disease.

Epithelial cells include the cells lining the respiratory and digestive tracts as well as the ducts of exocrine glands. In respiratory epithelia (cells which line the nasal cavity, trachea, lung, bronchus, and alveoli), it has been shown using electrophysiological techniques (Davis and Nadel 1983), that active transport of chloride ions occurs towards the airway lumen and sodium ions towards the submucosa. In CF tissue there appears to be an increased potential difference across airway epithelium (Knowles et al 1981). Knowles et al (1981) showed that the nasal tissue excised from subjects with CF exhibited higher voltage and lower conductance than control tissue.

Epithelial cells from normal human and CF nasal polyps were isolated, implanted into tracheal grafts, and injected into nude mice by Yankaskas et al (1985). The theory being that this technique, followed by analysis of the grafts, would separate secondary effects of the disease from primary cellular defects. Knowles et al (1981, 1983a, 1983b) previously demonstrated abnormalties of airway epithelial ion transport *in vivo* and *in vitro*. Amiloride induced a greater inhibition of the current generated by CF than normal cells. In normal cells amiloride abolishes Na^+ absorption but induces a secretion of Cl^-. In CF grafts, a diminished capacity to secrete Cl^- results in a smaller residual current than in normal cells. Normal tissues pretreated with amiloride can be stimulated by isoprenaline to secrete Cl^- more rapidly and hence increase the short circuit current. Freshly excised CF tissue does not respond quickly to the β-adrenergic

agonist isoprenaline, and nor do the CF cells which have been through immunosuppressed mice. In these grafted cells the major bioelectric abnormalities of CF respiratory epithelia observed by Knowles (1981) are retained.

Widdicombe et al (1985) used the fact that epithelial cells will grow easily in culture to establish confluent monolayers. These cells have tight junctions separating discrete apical and basolateral cell membranes and appear to retain their electrical properties (Handler et al 1979). Dog tracheal epithelium grown in culture retains differentiated transport function (Coleman et al 1984); active chloride and sodium ion secretion as well as absorption can be demonstrated. CF tracheal cells cultured 4 hours after the death of a 12 year old CF girl were compared with normal and dog tracheal cells. CF cultured cells' short circuit current was unaffected by bumetanide (an inhibitor of chloride secretion), and the responsiveness to isoproterenol and prostaglandin E_2 was one tenth that of normal cultured epithelial cells. This compounds the mounting evidence that chloride ion secretion and its stimulation by secretagogues is defective in CF tissue.

It has been subsequently shown (Knowles 1983a, Knowles 1983b) that amiloride-induced chloride secretion occurs only in normal cells and a chloride-free cell-bathing solution induces a smaller hyperpolarization in CF tissue than in normal tissue (Knowles et al 1983b). This implies that sodium ion absorption occurs across an epithelium relatively impermeable to chloride ions. Treatment with indomethacin, a chloride ion movement inhibitor, did not alter the potential difference in either control or disease groups (Knowles et al 1983a).

Stutts et al (1985) also cultured cells from CF and non CF individuals. Primary cultures of airway epithelial cells from CF and atopic (disease control) polyps and normal nasal airway epithelium were established and $^{36}Cl^-$ uptake was measured. The results showed that even in cultures where 88% of the cells had been produced by *in vitro* divisions there was still significantly reduced chloride uptake in the CF cells.

The work on chloride reabsorption has recently been clarified by Welsh (Welsh 1986a, Welsh and Liedtke 1986b) and Frizzell et al (1986). Welsh and Liedtke examined excised cell free patches of membrane from CF epithelial cells and normals and compared their chloride channel conductance. The results indicated that in excised cell free patches, chloride channel activity was identical in normal and CF samples. It appeared that the chloride channel was not acutely regulated by internal calcium ion concentration, indicated by changing the 1mM internal calcium to 1nM and observing no effect on the channel activity.

When chloride channels were examined in intact cells, only non-CF tracheal epithelial cells recorded any activity. The tissue from both CF and non CF was stimulated with β-agonists (isoprenaline and adrenaline), 8-bromo-cAMP and/or forskolin (an adenylate cyclase activator). Only cells from normal individuals responded to these stimuli. However a normal dose-dependent increase in cAMP was observed in both CF and normal tracheal epithelium when exposed to β-adrenergic stimulation. This implies that the defect in chloride channel regulation is at a point distal to cAMP accumulation and probably not in the actual channel protein itself.

Frizzell et al (1986) have essentially equivalent findings; although using the same patch clamp technique on primary cell cultures of airway epithelial cells, they find two different channel conductance levels. In excised patches both these channels exhibited identical chloride activity in normal and CF cells. However using cell attached assays once again no chloride channel activity was observed in the cells from CF airways even when stimulated with adrenaline, dibutryl cyclic AMP or forskolin (adenylate cyclase activator). The excised CF and normal patches had chloride channel activity which was inhibited by the addition of EGTA (a calcium chelator) to the bath. A free calcium ion concentration of 180 mM was required by inside out patches to activate the chloride channel activity. The calcium ionophore A23187 could stimulate chloride channel activity in cell attached CF cells, when added to a bath containing 1mM Ca^{2+},

in an equivalent manner to epinephrine or dibutryl cAMP stimulation in normal cells. The calcium activation of the chloride channel is not seen in the study by Welsh and Liedtke and may be due to differences in tissue culture technique (Case 1986).

Frizzell believes that the calcium signal that he observes does not involve phosphorylation as the inside out excised CF patches which demonstrated chloride channel activity when stimulated with calcium did not require adenosine 5' triphosphate in the medium.

Epithelial cells from the digestive tract are difficult to obtain, but buccal cells have been used. Mangos and Boyd (1984) report that CF patient buccal epithelium cell cytosol contains three times as much ionized free calcium as control cells (Mangos 1983).

Epithelial cells of the CF pancreatic duct are generally completely destroyed. A more accessible and more fully characterized gland is the sweat gland. This consists of a proximal secretory coil followed by an intermediate coiled duct segment and then a straight portion opening into the epidermis. Although there have been no morphological alterations observed in CF sweat glands (Spicer et al 1982), but based on the observation that sweat of CF patients contains increased levels of sodium and chloride, and on results that showed sodium reabsorbtion was defective (Mangos et al 1973a 1973b), Quinton (1983a) examined microdissected sweat glands from skin biopsies and examined their function.

Quinton (1983b) microdissected the reabsorptive duct from the secretory coil of sweat glands from skin biopsies from CF and normal patients. One end of the reabsoptive duct was cannulated with a micropipette which also served as a microelectrode in the lumen of the duct. During periods of perfusion when there were identical concentrations of NaCl in the bath and lumen, the average spontaneous potential across normal ducts was -6.8mV compared with -76.9mV in CF patients. Quinton performed a series of experiments on the CF and normal sweat duct which enabled him to

conclude that the chloride permeability of CF ducts was an order of magnitude lower than that of control ducts.

Sato and Sato (1984) also examined the secretory coils from normal and control subjects. They found a normal level of sweat secretion in both coil types. Sato (1983) had shown that in monkey palm eccrine sweat glands, maximal sweat rate was highest after cholinergic agonist stimulation (methacholine, acetylcholine, pilocarpine), reduced with β-adrenergic agonist (isoprotenol) and least of all with the α -adrenergic agonist phenylephrine. Removal of calcium ions from the medium abolished the sweat secretion induced by α-adrenergic or cholinergic agonists. Only the beta-agonists increased cAMP induction and this was not reduced by removal of calcium ions from the external medium. Isoproterenol, theophylline, and dibutryl cAMP all induced eccrine sweat secretion and Sato concluded that cAMP was the beta-adrenergic second messenger for sweat secretion from these glands. β-adrenergic stimulation of sweat secretion did not require external calcium, but this does not necessarily exclude the involvement of calcium released from internal stores.

Single sweat glands were isolated from skin biopsy specimens from normal and CF volunteers. Samples were either from the secretory coil or the end of the proximal tubule. CF sweat glands responded very poorly to either epinephrine or isoproterenol compared to normal sweat glands, although their response to methacholine was as avid as normals. The CF sweat glands accumulated tissue cAMP in response to isoproterenol alone or to the combination of isoprotenerol and theophylline or methacholine in a manner similar to the normal controls. The β-adrenergic response in normal sweat glands was not perturbed by the absence of calcium in the bathing medium. These results indicate that β-adrenergic stimulation of CF sweat glands is faulty and the defect is at some point distal to the adenylate cyclase step in the cAMP cascade.

In 1985 Heinz-Erian et al discovered that patients with CF were found to have a marked deficiency of vasoactive peptide (VIP) immunoactive nerves in both acini and

ducts of sweat glands in chest region skin biopsies. VIP promotes blood flow and movement of water and Cl- across epithelial surfaces and stimulates HCO_3^- secretion by pancreatic acini. It is an attractive hypothesis that deficient VIP innervation may be responsible for the relative impermeability of epithelium to Cl^-. The localisation of VIP immunoreactive nerves was determined using an anti-VIP antiserum. It was found that normal skin showed a rich network of VIP-immunoreactive nerves around secretory acini and a moderate innervation of the ducts. Individual VIP-positive nerve fibres were closely associated with the basement membrane of both acinar and duct cells. VIP innervation in CF samples was less pronounced in both cell types and absent around the ducts in 5 of 7 samples and minimal in the remaining 2. It is not known whether the depletion is a progressive symptom of CF or if it is present from birth and possibly a primary defect.

The secretory cells of exocrine glands are also of epithelial origin but have differentiated into complex structures. Mangos et al (1981) compared isolated CF parotid acinar cells with control cells and studied a variety of functional activities. CF cells had decreased potassium efflux in response to cholinergic receptor stimulation; reduced intracellular cGMP response to cholinergic receptor stimulation; reduced Ca^{2+} uptake by cells and reduced efflux of $^{22}Na^+$ across cell membranes. Intracellular cGMP response was increased upon β-adrenergic stimulation as was the secretion of amylase. The response of cellular ATP and oxygen consumption was unchanged relative to normal cells when stimulated with cholinergic agonists. McPherson et al (1985,1984) report a decreased stimulation of CF submandibular glands by β-adrenergic agonists resulting in a lack of mucin and amylase secretion. However a normal level of cAMP increase was observed when the agonist was applied.

Owing to the relative ease with which skin fibroblasts can be cultured, many studies have been carried out on this type of CF cell (Talamo et al 1983). The most interesting findings are that (Ansah et al 1980) CF skin fibroblasts have reduced mem-

brane Mg^{2+}/Ca^{2+}-ATPase activity, which may indicate a defect in the handling of calcium ions which would have far reaching functional implications. There is considerable evidence for altered calcium homeostasis in CF cells (Case 1984). Ca^{2+} concentration is increased in CF salivary secretions (Martinez 1982), isolated CF parotid glands (Mangos and Donnelly 1981) and cultured CF fibroblasts but there is decreased $^{45}Ca^{2+}$ efflux from preloaded cells (Shapiro et al 1982).

CF fibroblast mitochondria have increased accumulation of calcium ions (Shapiro et al 1982). The Ca^{2+} concentration goes up in normal fibroblasts when exposed to cultured medium from CF fibroblasts (Ceder et al 1983).

These physiological studies, indicate that the basic defect in CF lies in the control of chloride channel activity. However, there are no candidate proteins at present. It seems reasonable to assume that abnormalities associated with CF that are also observed in clinically normal heterozygotes must be relevant to the basic defect.

1.6 "Factors" associated with the CF genotype

Pursuing the identification of "factors" present in CF individuals, and more importantly clinically normal obligate heterozygotes, has been tempting as an indication of the basic gene defect and for the possibility of their use in a heterozygote detection system. Unfortunately research in this area has been dogged by findings that cannot be reproduced in other laboratories and promising carrier detection systems which fail in blind trials.

Breslow et al (1978) showed that cultured fibroblasts from patients with CF accumulate less Na^{+} in the presence of ouabain than do normal cells. There was a significant difference between CF homozygotes or heterozygotes and normals, but heterozygotes and CFs had a considerable degree of overlap. This was subsequently shown not to be reproducible (Breslow et al 1981).

Another effect employing cultured fibroblasts was demonstrated by Shapiro et al

(1982a). They found that when mitochondrial NADH dehydrogenase activity was measured, the pH optimum was related to genotype. CF patients gave optimal activity at pH 8.6, heterozygotes at pH 8.3 and normals at pH 8.0. Further, the Michaelis constant of the enzyme for NADH was also different for the three genotypes. This test appears to be limited to cultured fibroblasts and although the ability to distinguish the three classes is exciting, the results seem to be disappointingly not reproducible in other laboratories (Sanguinetti-Briceno and Brock 1982).

Lieberman et al (1981) using S-300 gel filtration, isolated from serum an IgM bound lectin cofactor. 95% of obligate heterozygotes had high lectin titres, controls had low titres and CF patients had elevated titres equal to, and exceeding, those of heterozygotes. Lectin cofactor was only found in the urine of CF individuals. However involvement of a lectin-like factor or an IgM bound cofactor in CF has not been determined and the test has not been reproduced in other laboratories as yet.

In a recent study, the erythrocyte superoxide dismutase and catalase activities in cells from CF sufferers and their parents were found to be significantly higher than in cells from age and sex matched controls, while the peroxidation of lipids was reduced (Markovics et al 1982).

Reduced β-adrenergic responses in lymphocytes and granulocytes of patients with CF and heterozygotes have been demonstrated by Davies et al (1983), although the absolute number of receptors is unaltered between CF and normal leucocytes.

Another system has been reported by Hosli and Vogt (1979) where gentle heat inactivation of plasma α-mannosidase and acid phosphatase allows differentiation of normals from both CF heterozygotes and patients. They reported that these two enzymes retained full activity (80-100%) in normals, (40-60%) in heterozygotes and lost all activity in homozygotes. There was no overlap between groups. The researchers claimed to have demonstrated that the genetic defect in CF affects the intracellular digestive system in a way that leads to a multiple leakage of hydrolytic enzymes into

the extracellular fluid and an accompanying partial deficiency of corresponding hydrolases. However the test has proved completely unreproducible in several cases (Patrick et al 1979, Harris et al 1979, Hultberget al 1981). Sack et al (1980) could only repeat the results in Hosli's laboratory. Ceder et al (1983a) repeated the results with α-mannosidase and acid phosphatase from plasma as well as from the medium from cultured fibroblasts. The non-reproducibility of the test outside the Pasteur Institute is puzzling. Ceder hypothesises that it may be due to the distilled water used at the Pasteur which is produced by pyrolysis and may contain a "factor" essential for the success of the test. A blind trial undertaken by the group at the Pasteur Institute (Katznelson et al 1983) correctly genotyped a whole set of 45 Israeli plasma samples. This remains an intriguing phenomenon and cannot be discounted in its relevance to CF.

The thickened mucus in CF was thought to be due to abnormal ciliary function. Sera from CF homozygotes and heterozygotes have been reported to contain one or more substances that inhibit the motility of rabbit tracheal cilia in tissue culture (Spock et al 1967). It has also been discovered that sera from CF genotypes will inhibit ciliary action in gills from freshwater oysters (Bowman et al 1969) or gills of freshwater mussels (Adshead et al 1975). Mouse sperm motility was also inhibited by sera from CF homozygotes and heterozygotes (Pivetta et al 1979). Fleming and Sturgess(1981) demonstrated that rat submandibular glands secrete more glycoprotein when submerged with CF sera and CF heterozygote sera than with normal sera. Conod et al (1977) found that the factor which caused ciliary dyskinesis in rabbit trachea was bound to IgG and had a molecular weight of 1 to 10 Kd. Wilson and Fundenberg (1977) described a factor similar to ciliary dyskinesis factor in the serum of asthmatic patients, and this was found to be C3a or a fragment thereof (Wilson and Fundenberg 1977b). The calcium ionophore A23187 causes ciliary dyskinesis and mucus secretion similar to CF sera when tested in the rabbit system (Bogart at al 1977). Bogart et al

(1979) showed that CF sera or CF IgG increase uptake of Ca^{2+} in rabbit tracheal explants relative to control sera or IgG. Three ciliary dyskinesis factors are secreted by CF lymphocytes and only one is specific for CF and T lymphocytes secrete this one (Wilson and Bahm 1980).

The oyster ciliary dyskinesis factor is nondialysable, heat labile, cationic and associated with IgG (Bowman et al 1973). It has also been detected in the media of fibroblasts from CF heterozygotes and patients (Bowman et al 1977). However, neither rabbit tracheal nor oyster gill assay systems could reliably identify the CF genotype of samples in a blind trial (Wood et al 1973,).

1.7 Cystic Fibrosis antigen (CFAg)

The factor associated with CF which can be most consistently demonstrated is the so called CFAg. Wilson et (1975) described a unique protein doublet at pI 8.4 when carefully collected serum samples were exposed to isoelectric focussing in the presence of urea. The doublet band, designated CF protein, was demonstrated in both CF patients and their parents (Wilson et al 1978). Other laboratories have also found this banding pattern on isoelectric focussing gels pH 2.5 - 10, in the pH8-9 range (Tully et al 1979, Jamieson et al 1984, Super et al 1984, Brock et al 1982, Scholey et al 1978, Nevin et al 1981). The bands are rarely seen in normals, and those that are positive could be undiagnosed carriers.

In an attempt to look specifically at CF protein (renamed CFAg), Manson and Brock (1980), excised portions of the acrylamide gel with the doublet band between pI values 8.2 and 8.5. The homogenized gel was then injected into guinea pigs. The absorption of the resulting crude antiserum with sera from individuals devoid of CFAg, produced a monospecific antiserum. The antiserum was then included in a rocket electrophoresis system with the serum from CF, heterozygotes, and normals. The pattern produced is shown below.

The observed pattern of expression with high levels in homozygotes, reduced levels in heterozygotes and very low levels in normals is a classic pattern for a single gene defect. The implication is that the antigen may be the product of the defective gene or involved in the same metabolic pathway. Immunoprecipitation and immunoradiometric assay (Bullock et al 1982) enabled the correct assignment of genotype on 94% of cases. However the collection of blood appears to very important for the success of this test. Blood must be collected into plain glass tubes and clotted overnight at $4^{o}C$. Sera collected into EDTA or lithium heparin tubes produce high levels of CFAg but give poor rockets. In the expectation of improving the characterization of CFAg by immunopurification, tissue localization, and accurate quantification, monoclonal antibodies were raised against CFAg. Repeated attempts to purify this antigen from serum by conventional fractionation techniques using guinea pig antiserum to follow purification, were unsuccessful. However identification of granulocytes as the tissue source of CFAg in normals (van Heyningen et al 1985) enabled the raising of monoclonal antibodies to this serum protein (Hayward et al 1986). The specificity of the monoclonal antibodies was established by their ability to remove CFAg, as detected by the antiserum, from CF serum. The monoclonal antibodies which were produced, defined at least two epitope regions on CFAg. This allowed a two site (sandwich) enzyme linked immunosorbent assay (ELISA) to be designed. The sandwich assay presents a sensitive reproducible quantitation of CFAg in serum and on this basis it is confirmed that there is a significant difference in CFAg between normals, CF heterozygotes and CF patients (Hayward et al 1987), although there is

considerable overlap between the genotypes.

The fact that CFAg is a granulocyte protein, and CF patients are not healthy individuals due to secondary infections, raises the point that increased levels of CFAg in the serum may be due to increased levels of granulocytes and therefore unrelated to the basic defect. Hayward et al (1987) have examined the serum levels of another granulocyte protein (lactoferrin), and an acute phase reactant (C-reactive protein) in CF, heterozygotes, normals and disease controls. The data show that CFAg levels are raised significantly in CF and heterozygotes, but lactoferrin and C-reactive protein are only significantly raised in the CF homozygotes and disease controls. This suggests that the elevated CFAg levels observed in heterozygotes are due primarily to a CF gene specific component. The raised level of CFAg in diseased controls is probably due to increased numbers of granulocytes as there is a concomitant increase in lactoferrin and C-reactive protein. In CF homozygotes the elevated CFAg levels may be ascribed to the compound effects of the presence of the CF gene and disease.

The tissue distribution of CFAg has been examined using the antiserum available. van Heyningen et al (1985) showed that peripheral leukocytes from chronic myeloid leukemia (CML) patients proved consistently to be strongly positive for CFAg and antigen from normal granulocytes was immunologically identical as demonstrated by Ouchterlony double diffusion. An exhaustive immunohistochemical tissue search, using the monoclonal antibodies raised against CFAg, has resulted in CFAg only being detected in polymorphonuclear granulocytes in blood or where these cells enter other tissues. The only exception has been squamous epithelial cells in tongue and oesophagus (M. Wilkinson personal communication). Both of these cell types are non-keratinised stratified epithelia. Exocrine tissue (apparently affected in CF) from normals were carefully studied and lung, pancreas, intestine and skin sweat glands from normals were all negative for CFAg. Adult CF tissues were not available, but CF fetal pancreas and lung samples were negative. The abnormal tissues were found to be

positive for CFAg by Wilkinson (personal communication) with the monoclonals from both epitopal groups. Squamous cell carcinoma of the skin and lung stained strongly in immunohistochemistry and so did the squamous epithelial cells from eczema and psoriasis.

van Heyningen et al (1985) also demonstrated, using a series of somatic cell hybrids, that CFAg expression was concordant with the presence of chromosome 1. This was done by fusing mouse myeloid stem cells (WEHI-TG) with CML cells and subcloning repeatedly for chromosome stability. Cell lysates were tested in Ouchterlony double diffusion analysis with antisera raised against the excised isoelectric focussing band, and also antisera raised against CML lysate partially purified material. The two antisera showed complete immunological identity against various antigen sources in this system.

Surveying the literature, there does not appear to be a granulocyte protein with the same properties as CFAg previously described, even though CFAg appears to be so abundant in polymorphonuclear granulocytes.

A large number of physiological and biochemical changes have been described to take place in CF. One of the most appealing observations as far as direct relevance to aetiology is concerned, is the elevation of serum levels of CFAg in CF homozygotes and heterozygotes. I set out to clone the gene for this protein when it was still a candidate product of the CF gene. The CFAg gene had been assigned to chromosome 1 by somatic cell genetics, so that when the segregation of CF with chromosome 7 markers put the CF gene on that chromosome, it was clear that CFAg is not the product of the CF gene.

Nevertheless the identity and function of CFAg remained of interest. The main justification for this research was that CFAg could be detected immunologically in sera, and the amount of CFAg which accumulated appeared to reflect the CF genotype of the individuals. Thus it could be postulated that although CFAg was not coded for

by the defective gene in CF, it may be affected directly by it or at least be involved in the aberrant pathway in CF. Hence characterization of the function of CFAg might elucidate the biochemical basis of CF.

A further aspect of CFAg research was the possibility that immunological quantitation of serum levels could be used in a heterozygote detection assay. The fact that antiserum was available and that serum was an accessible fluid made this an attractive possibility. Monoclonal antibodies raised against CFAg allowed purification and refinement of the heterozygote detection assay. However the test still contained spurious results although in general samples could be divided into three genotypes. It was felt that characterization of CFAg (which had only been identified immunologically) might help to improve the heterozygote test.

In this thesis I present the strategy used to clone the cDNA for CFAg. The sequence of the successfully cloned cDNA is given. The gene is sublocalized to a region on chromosome 1 and the abundance of mRNA in various tissues is examined. The predicted amino acid sequence is compared to other protein sequences in a computer database and its homology to calcium-binding proteins is discussed in relation to CF.

CHAPTER 2

MATERIALS AND METHODS

Chemicals - unless otherwise stated all chemicals were obtained from BDH Reagents Ltd.

SECTION 2.1

Tissue Culture

All cells were grown in RPMI 1640 medium (Gibco) with oxalocetate (1mM); pyruvate (0.45mM); porcine insulin (Novo) 0.2 International Units /ml; 0.03% glutamine, 3-(N-morpholino)propanesulphonic acid (MOPS) buffer 0.0125M plus 5% foetal calf serum (FCS) (Gibco) unless otherwise stated.

SECTION 2.2

HL-60 cell Induction

HL-60 cells uninduced, were grown in RPMI 1640 without serum but with insulin, sodium selenite, and transferrin, (ITS) (Sigma), and used according to the manufacturers instructions. Cells to be induced were grown in standard RPMI 1640 medium with the inducer present at the relevant concentration.

SECTION 2.3

Nitroblue tetrazolium (NBT) reduction assay

In order to test for differentiation, HL-60 cells were spun down and washed in phospate buffered saline (PBS). 10^5 cells were resuspended in 0.5ml PBS containing 1mg/ml NBT (Sigma) and 1μg/ml PMA (phorbol myrisate acetate) (Sigma) and incu-

bated for 20 minutes at 37°C. The cells were then examined under the microscope, and the percentage of blue (differentiated) cells was determined.

SECTION 2.4

Isolation of polymorphonuclear cells from whole blood

Polymorphonuclear cells were isolated from 10ml normal blood after collecting the sample into lithium heparin tubes. Plasma was removed after spinning the sample for 10 minutes at 1000 rpm in a Sorvall RT6000 centrifuge. The original volume of the sample was reconstituted with standard medium containing heparin (Evans Medical Ltd) at 10 units/ml. In order to remove the red blood cells, the cell suspension was mixed 1:3 (volume/volume), with 6% dextran 250 (Sigma) in 0.15M NaCl solution, and incubated at 37°C for 30 minutes. The straw coloured top layer was then removed and the remaining sample discarded. The polymorphonuclear cells were separated from the mononuclear cells by loading the suspension onto a cushion of lymphoprep (Gibco). The gradient was then spun at 2,000 rpm for 20 minutes. The mononuclear cells moved to the interface, and the granulocytes travelled with the remaining red blood cells to the bottom of the tube. Isolated cells were washed in PBS and counted in a haemocytometer by diluting 1 volume of cell suspension into 9 volumes of a solution of crystal violet in 2% acetic acid. 10ml of normal blood yielded ~10^7 polymorphonuclear cells.

Granulocytes were isolated from CML blood samples by allowing the blood collected into lithium heparin to settle. The buffy coat which formed was rich in granulocytes, and having removed this from the top of the blood sample, no more purification was necessary.

SECTION 2.5

Preparation of cell lysates

Cell pellets were lysed by the addition of 100µl phosphate buffered saline (PBS)

containing 10^{-6}M phenylmethylsulphonyl fluoride (Sigma) (a potent serine protease inhibitor) to 10^7 cells and subjecting them to 3 cycles of freezing and thawing. The cell lysate was then spun in a microfuge centrifuge for 5 minutes and the supernatant (lysate) was removed to a fresh tube.

SECTION 2.6

Ouchterlony double diffusion

10µl samples for Ouchterlony analysis (Ouchterlony 1967) were placed in wells formed in 1.5% noble agar (Difco) / 0.9% NaCl and allowed to diffuse overnight at room temperature in a moist, level chamber. The gel was poured into plastic plates and six peripheral and one central well were cut in the gel an equidistance apart. The gels were washed in 0.05% Tween 20 in 0.9% NaCl. The gels were then dried down onto a hydrophilic surface (Gelbond FMC Corporation) by covering the gel on the gelbond with Whatmans filter paper #1, paper towels and a glass plate and leaving at room temperature overnight. Gels were stained for protein in 0.25% Coomassie blue (Bio Rad) in methanol : water : acetic acid 5 : 5 : 1 for 30 minutes and destained in the same solution without Coomassie blue until the bands became clear.

SECTION 2.7

Internal Cell labelling

Cells to be labelled were spun down and counted. 10^8 cells were resuspended in approximately 5ml of F10 nutrient mixture (Gibco) without valine, methionine, lysine, isoleucine or L-glutamine. The missing amino acids were supplemented to a final concentration of 0.03% w/v except those for which radiolabelled analogues were to be used. The cells were incubated for 1 hour at 37°C in the medium to reduce the intracellular pool of the unlabelled amino acid of choice. Cells were resuspended in 1ml of the same medium with 250µCi of the radio-labelled methionine (800 Ci/mmole, Amersham) added together with 1% volume/volume dialysed FCS, and incubated at 37°C for 4 hours. In the case of HL-60 cells, the uninduced cells were incubated

without FCS and ITS supplement was used instead. Induced cells were labelled in the continuous presence of the inducer (10^{-6}M retinoic acid). After the 4 hour incubation, the cells were washed three times with PBS and lysed in 100µl PBS/10^7 cells by repeated freeze/thawing. The percentage of incorporation of radiolabel into protein was measured by trichloroacetic acid (TCA) precipitation on glass fibre filters and scintillation counting in Aquasol (Amersham) (section 2.15).

SECTION 2.8

Immunoprecipitations

Labelled cell lysates (~ 1x10^6 cpm) were incubated with 50µl monoclonal antibody culture supernatant for 1 hour followed by incubation with donkey anti-mouse antibody coupled to cellulose beads (Sac-cell; Wellcome Reagents Limited) for 1 hour. The cellulose-coupled antibodies were then spun down and washed with 0.05% Tween 20 in PBS, three times. The final wash contained 1% β-mercaptoethanol. The precipitate was resuspended in 30µl SDS PAGE (polyacrylamide gel electrophoresis) sample buffer (section 2.9) and boiled for 5 minutes, then analysed by SDS PAGE.

SECTION 2.9

SDS polyacrylamide gel electrophoresis

CHEMICALS

acrylamide, bisacrylamide, ammonium persulphate, N N' tetramethylethylenediamine (TEMED), glycine, Coomassie blue and molecular weight standards were obtained from Biorad Laboratories, Tris from Sigma Chemicals.

STOCK SOLUTIONS

30% acrylamide (30% acrylamide with 1% bisacrylamide) in distilled water.

Separating gel buffer: 1.5 M Tris pH 8.8, 0.5% SDS)

Stacking gel buffer: 0.5M Tris pH 6.8, 0.5% SDS)

10x Electrode buffer: 0.5M Tris, 0.5M glycine, 10% SDS)

Sample Buffer: 10% stacking gel buffer, 2% SDS, 5% β-mercaptoethanol, 10% glycerol, 0.005% bromophenol blue.

PROCEDURE

The glass plates were washed thoroughly and wiped with ethanol. 1.5mm spacers were sandwiched between two glass plates with LKB clamps and then placed on the gel casting apparatus. The acrylamide solutions were made up as follows-

15% acrylamide gel solution consisted of 18.5 ml 30% acrylamide stock solution, 9.25ml of separating gel buffer, 7.3ml of distilled water, 200µl freshly prepared 10% ammonium persulphate, 20µl TEMED. After the addition of the TEMED and ammonium persulphate, the gel was quickly poured between the plates to about 5cm below the wells. The gel was overlaid with isobutanol saturated water and allowed to set. After about 1 hour, the isobutanol and any unpolymerised acrylamide was washed away, a comb inserted and the stacking gel poured. The stacking gel consisted of 1.5ml of 30% acrylamide stock plus 2.5ml of stacking gel buffer, 5.7ml distilled water, 30µl 10% ammonium persulphate and 10µl TEMED. The gel was left for 30 minutes to set and then the comb was removed and the wells thoroughly washed out with electrophoresis buffer. At least one litre of electrophoresis buffer was added to the lower tank and 1 litre to the upper reservoir.

Samples were prepared by addition of up to 30µl of sample buffer and boiling for 5 minutes. Gels were run at 15mA overnight to ensure the blue dye only just ran off the bottom of the gel. The gel was then removed from the apparatus and soaked in destain (methanol : water : acetic acid, 5 : 5 : 1) solution for 30 minutes to fix the proteins in the gel. The gel was then soaked for half an hour in Amplify (Amersham) to enhance the signal. The gel was then placed on dampened filter paper and dried down on a gel drier (Bio Rad) following the manufacturer's instructions.

SECTION 2.10

Autoradiography

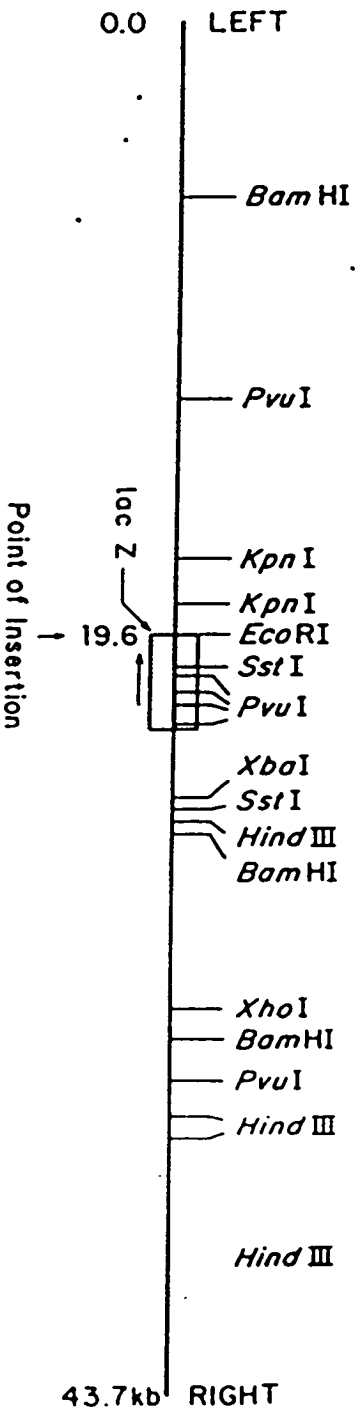
Gels to be autoradiographed were placed on filter paper and preflashed kodak XAR-5 X-ray film or Fuji X-ray film was placed over the gel. Fuji film gave clearer results but the Kodak film was more sensitive. Nitrocellulose blots to be autoradiographed were covered with clingfilm and preflashed Kodak X-ray film was routinely used. Both gels and blots were then placed in lead autoradiograph cassettes with intensifing screens and put at -70^{o}C for various periods of time.

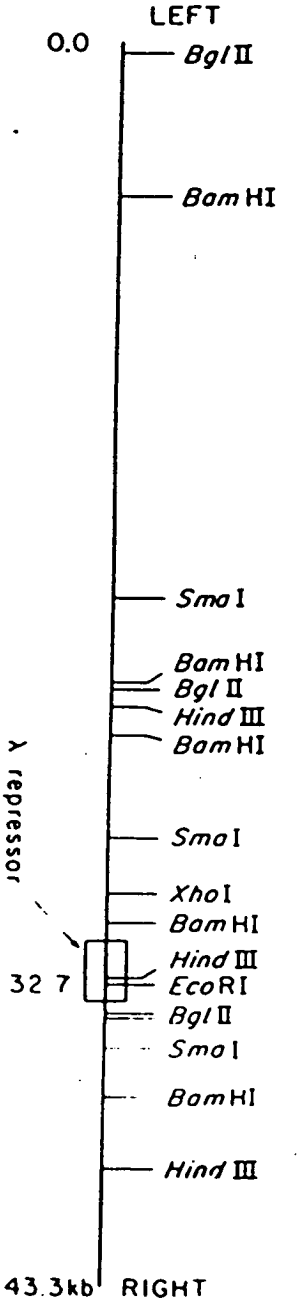
SECTION 2.11

Vectors

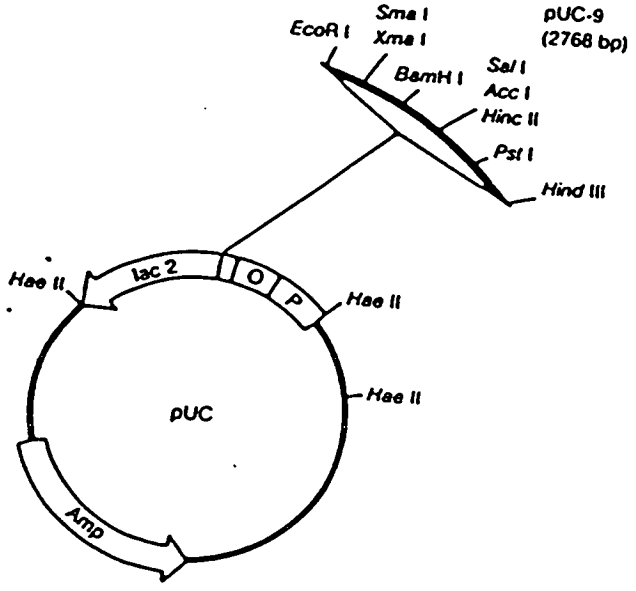
The following vectors were used in the experiments contained within the work described by this thesis.

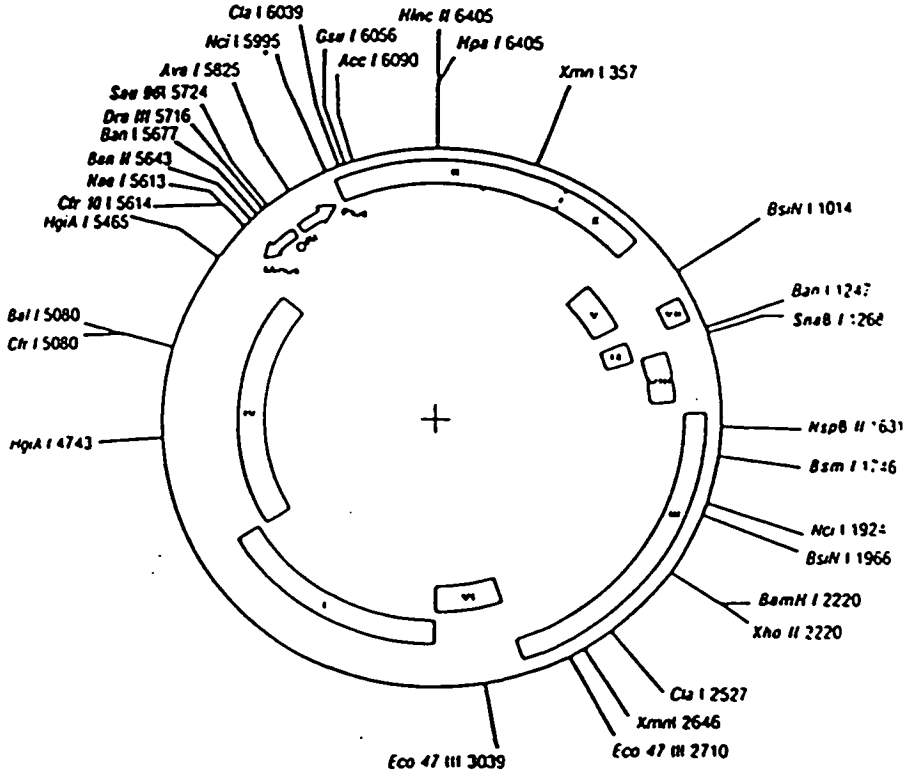
Restriction endonuclease maps of 1. λgt11 2. λgt10 (Huynh et al 1985).





Restriction Endonuclease map of pUC 9 (Vieira and Messing 1982) and M13mp8 (Messing and Vieira) showing the polylinker cloning region in detail.





DNA Synthesis

M13mp9/pUC9 ATG ACC ATG ATT ACG CCA AGC TTG GCT GCA GGT CGA CGG ATC CCC GGG AAT TCA CTG GCC

Hind III Pst I Sal I Acc I BamHI SmaI XmaI EcoRI Hae III

SECTION 2.12

Bacterial strains

E.coli C600 = *E.coli hsdR- hsdM+ sup E thr leu thi lacY1 tonA21.*

E.coli C600 *hfl* = *E.coli C600 hflA150(chr::Tn10).*

E.coli Y1088 (ATCC no 37195) = E.coli ^lacU169 supF supE hsdR- hsdM+ metB trpR tonA21 proC::Tn5 (pMC9). pMC9 =pBR322 lacI^{Q}.

E.coli Y1090 (ATCC no 37197) = *E.coli ^lacU169 proA+ ^lon araD139 strA supF(trpC22::Tn10) (pMC9).*

E.coli JM 83 = *E.coli* JM 83 *ara, ^lacpro, str A, thi, o80d, lac Z ^M15.*

E.coli =*E.coli* JM 101 ^lacpro, thi, supE, F'traD36, proAB, lac^{q}Z ^M15.

SECTION 2.13

Media and buffers for bacterial and phage culture

The *E.coli* bacterial strains C600, C600 *hfl*, Y1090 and Y1088 were all grown on LB (Luria-Bertani) agar plates or in LB medium. *E.coli* C600 *hfl* was grown in medium containing tetracycline (Sigma) (12.5µg/ml). *E.coli* Y1090 and Y1088 were grown in medium containing 50µg/ml ampicillin (Sigma) to maintain the plasmid pMC9.

LB Medium: per litre (pH 7.5) :

10 g bacto-tryptone (Difco)

5 g bacto-yeast extract (Difco)

10 g NaCl

LB agar: LB medium plus 1.5% agar (Difco).

LB top agar: LB medium plus 0.7% agarose (Sigma).

2x TY medium: per litre

16g Bacto tryptone

10g yeast extract

5g NaCl

H plates: per litre

10g Bacto tryptone

8g NaCl

12g agar

H top agar: as for H agar, but with 0.7% agarose instead of agar.

M9 medium: per litre

6g Na_2HPO_4

3g KH_2PO_4

0.5g NaCl

1g NH_4Cl

The pH was adjusted to 7.4, autoclaved and then the following reagents were added:

2 ml of 1 M $MgSO_4$; 10 ml of 20% glucose; 0.1 ml of 1 M $CaCl_2$.

Glucose/minimal medium plates:

15 g minimal agar (Difco) in 1 litre M9 medium plus 10 mM thiamine Cl.

Cells selected for pUC9 plasmid transformation, were grown on H plates containing 50μg/ml ampicillin. For indication of function of the lac Z gene in pUC plasmids, 0.02% X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside (2% in dimethylformamide stock solution)) was included in the agar media. For λgt11 and M13 both 1mM IPTG (isopropyl-β-D-thiogalactopyranoside) and 0.02% X-gal were included in the top agarose medium.

SECTION 2.14

RNA extraction

This was successfully carried out using the method of Chirgwin et al (1979) with some modifications. For extraction of RNA from cell lines, the cells were harvested and then washed in sterile phosphate buffered saline and pelleted in an Eppendorf microfuge. In the case of tissue (leukophoresis material), the cells were spun at 2000 rpm in a Sorvall RT6000 for 10 minutes . 4ml of 8M GHCl (8M Guanidine Hydrochloride 0.01M EDTA 0.05M Tris pH 7.5) per 10^8 cells was added to the cell pellets and either homogenised in a tight ball homogeniser, or else as in the case of tissue, a motor driven homogeniser was used for 30 seconds at the highest speed setting. The cells were then spun to remove debris (8,000 rpm for 10 minutes in an Sorvall superspeed RC 2-B) Half a volume of cold ethanol was then added to the supernatant, this was mixed and incubated at -20°C for 45 minutes. The RNA was then precipitated by centrifugation at 10,000 rpm for 15 minutes. The pellet was then taken up in 6M GHCl (6M Guanidine Hydrochloride 0.01M EDTA 0.05M Tris), homogenised, a half volume cold ethanol added, and the mixture incubated at -20°C for 45 minutes. The centrifugation step was then repeated and the pellet taken up in 6M GHCl as before. After spinning the precipitate down, the pellet was taken up in an appropriate volume of DEPC water (0.1% Diethylpyrocarbonate in double distilled water), homogenised and a half volume of filter sterile 7.5M ammonium acetate was added followed by two volumes of ethanol. This was then incubated at -70°C for 45 minutes or at -20°C overnight. The pellet was then collected by centrifugation as before, resuspended in DEPC water plus a one tenth volume of 2M sodium acetate pH 5.5 and 2.5 volumes of ethanol, and incubated at -70 for 45 minutes. This precipitation step was repeated and the amount of RNA determined by reading the optical density at 260 nanometers. At this wavelength, an absorbance of 1 is equivalent to an RNA concentration of 40 μg of RNA/ml.

SECTION 2.15

Isolation of poly $(A)^+$ RNA

This was carried out according to the method of Aviv and Leder (1972). Essentially, total RNA was dissolved in DEPC water and heated to 65°C for 5 minutes. An equal amount of 2x loading buffer (20mM Tris.Cl pH7.6; 0.5M NaCl; 1mM EDTA; 0.1% SDS) was then added to the RNA and applied to a 1ml oligo(dT)-cellulose column previously equilibrated with loading buffer and washed in turn with 3 column-volumes of a) sterile H_2O; b) 0.1M NaOH and 5mM EDTA; c) sterile H_2O. The column was checked to ensure that the effluent pH was less than 8, and was extensively washed with 5 volumes of loading buffer. After the RNA was loaded onto the column, the flow through was reheated to 65°C and reloaded. The column was washed with 5 volumes of loading buffer and then with loading buffer containing 0.1M NaCl. The poly $(A)^+$ RNA was then eluted with 3 column volumes of elution buffer (10mM Tris.Cl pH 7.5; 1mM EDTA; 0.05% SDS) and collected in 500µl fractions. Each fraction was checked for RNA by reading the OD at 260nm. The positive fractions were then pooled and the RNA precipitated with 0.3M sodium acetate (pH 5.2) and 2.2 volumes of ethanol.

SECTION 2.16

Gel electrophoresis of RNA

RNA was electrophoresed through formaldehyde denaturing gels (Maniatis et al 1982). 1.5% agarose; 10 mM $NaPO_4$ pH 6.5; 18% formaldehyde (from a 37% formalin solution, Fisher Scientific Ltd) in 100ml distilled water were run as horizontal submarine gels at 25 volts overnight. Samples were loaded in 3 volumes of 1.5x sample buffer (10mM $NaPO_4$; 18% formaldehyde; 50% formamide) and heated to 50°C for 10-15 minutes. The samples were then loaded in loading buffer (10x loading buffer is 10% bromophenol blue, 50%glycerol; 5mM EDTA; 0.4% xylene cyanol). If the RNA was not going to be transferred to nitrocellulose, the gel was stained in 0.5 µg/ml

ethidium bromide to visualise the 28s and 18s ribosomal RNAs.

SECTION 2.17

Northern blot analysis

After overnight electrophoresis the formaldehyde/agarose gel described above was soaked for 10 minutes in 10x SSC (1.5M NaCl; 0.15M sodium citrate) and laid inverted onto the transfer apparatus (Maniatis et al 1982). The nitrocellulose paper was soaked in water before being laid onto the gel, followed by one piece of Whatman filter paper number 17 soaked in water, and then one piece of dry filter paper and followed by paper towels. The RNA was then left to transfer overnight. The transfer buffer was 10x SSC. The nitrocellulose filter was then removed from the apparatus and baked at 80°C in a vacuum oven for 2 hours. The filter was then prehybridised in a solution of 10% dextran sulphate, 2x Denhardts (100x Denhardts = 2% (weight/volume) polyvinylpyrrolidone; 2% bovine serum albumin (BSA); 2% Ficoll 400; in distilled water), 6xSSC, 0.1% SDS, 0.1% sodium pyrophosphate and 150µg denatured salmon sperm for at least 1 hour at 68°C. The probe to be hybridised was then radiolabelled by nick translation (see section 2.4) and the boiled probe added to the hybridisation mix. Hybridisation was carried out at 68°C overnight. Filters were washed at 68°C in 2x SSC, 0.1% SDS and 0.1% sodium pyrophosphate unless otherwise stated.

SECTION 2.18

Nick translation

Double stranded DNA probes were labelled by nick translation (Rigby et al 1977). Typical reaction mix contained 100ng of DNA in a total volume of 20µl. This mix contained: 2µl 10x translation buffer (500mM Tris.Cl pH 7.5, 50mM $MgSO_4$, 0.1M β-mercaptoethanol, 10mg/ml BSA, and 1µl each of 0.1mM dATP, dCTP,dGTP (Sigma), 10µCi of dTTP 800Ci/mmol (Amersham), 1µl of DNase I (BRL) (0.1µg/ml freshly diluted from a 1mg/ml stock) and 1-3 units of *E.coli* DNA polymerase 1

(Boehringer Mannheim). The reaction was incubated at 16°C for a maximum of 2 hours.

The percentage incorperation of added labelled nucleotides into the DNA was estimated using trichloroacetic acid (TCA) precipitation (Furlong 1967). Briefly 0.5µl aliquots from the reactions were spotted onto glass fibre filters (Whatman GF/A) and the number of disintegrations per minute (dpm) recorded in a scintallation counter. The filter was then washed with 10% TCA and the dpm recounted. Normally the ^{32}P-labelled probes were prepared with specific activity of greater than 10^8 dpm/µg. The reaction was stopped by the addition of 20mM EDTA and the free counts were separated from the incorporated nucleotide over a Sephadex C-50 (Pharmacia) column. 5 ml of a suspension containing 30g/250ml Sephadex C-50 in column buffer (10x column buffer = 3M NaCl; 0.1M sodium acetate 0.5% SDS), was poured into a disposable 5ml plastic pipette, plugged with glass wool. The DNA was loaded onto the column in 100µl 10x column buffer and the progress of the DNA down the column was monitored with a geiger counter. As the separation of DNA from the unincorporated labelled nucleotides by gel filtration is on the basis of size, the larger labelled DNA molecules emerged from the column in the first peak. The DNA probe was then boiled for 5 minutes to separate the strands and put into the hybridisation bag.

SECTION 2.19

Random priming of DNA

When using small DNA molecules (~300bp), as probes, random priming (Feinberg and Vogelstein 1983) is a more efficient method of radiolabelling the DNA. Briefly, mixed sequence hexadeoxynucleotides are hybridised to the DNA, and DNA synthesis is primed from them. A typical 10µl reaction, contained 100ng of DNA (preboiled for 3 minutes and then on ice for 3 minutes), 2µl 10x nick translation buffer (see section 2.18), 1µl each 50mM dATP, dCTP, dGTP, 25µCi dTTP (800 Ci/mMol), and 2-3 units DNA polymerase I large fragment (Boehringer Mannheim). The reaction

was then incubated for 1 hour (or overnight) at 37°C in a lead pot. The reaction was then treated as for a nick translated piece of DNA (section 2.18).

SECTION 2.20

In vitro translation of mRNA

The translation of CML polyadenylated mRNA, was carried out using the rabbit reticulocyte lysate system (N90) supplied by Amersham. The lysate was stored in liquid nitrogen and when used, it was removed and thawed quickly. A reaction mix was made up of 40µl of reticulocyte lysate, 7µl L-^{35}S-methionine (Amersham code SJ.204, ~400 Ci/mMol) and 3µl (0.5µg) poly(A)$^+$ CML RNA. The mix was then left at 30°C and 1µl samples were withdrawn at 5 minute intervals over a 2 hour period. These aliquots were placed into 0.5ml of 1M sodium hydroxide/hydrogen peroxide and placed at 37°C for 10 minutes to hydrolyse the aminoacyl-tRNA complexes. The tubes were then removed to an ice bath and 2ml of ice cold 25% trichloroacetic acid solution containing casein hydrolysate was added. After 30 minutes the tube contents were filtered through Whatmann glass fibre discs and these were then washed with 8% TCA solution to remove unreacted, radiolabelled amino acid. The discs were then counted in scintillation fluid (Amersham) in a scintillation counter. The experiment was then repeated with an incubation time of 45 minutes as this was the time at which the maximal number of precipitable counts was recorded. The lysate was divided into 3 aliquots and each sample was immunoprecipitated with a different McAb (section 2.8).

SECTION 2.21

cDNA library construction

In constructing a cDNA library, the production of the first strand of cDNA from the RNA template is considered the most important and critical step. The pH of the reaction must be 8.3, and deviation of even 0.5 of a unit will reduce the number of full length transcripts. Potassium monovalent ions are considered give higher transcriptional efficiency than sodium-ions (Maniatis et al 1982). Retzel et al (1980) showed

that a high concentration of dNTPs is important for maximal production of cDNA. The optimal concentration of reverse transcriptase enzyme over template has been shown to be at 50-fold molar excess of enzyme (Freidman and Rosbash 1977), but this necessitates the use of highly purified reverse transcriptase and RNase inhibitors.

SECTION 2.21.1

First strand

A 200µl reaction mixture was made by adding the reagents in the order listed here. DEPC treated water and 10µg of polyadenylated RNA was added initially then 40 µl mix I (0.25M Tris.Cl pH8, 15mM $MgCl_2$, 1mM d(A,G,C)TP, 0.3mM dTTP), 5µl oligo(dT) at 100µg/ml (Collaborative Research), 4µl 1.5M KCl, 50 µCi (α-^{32}P) dTTP (~800Ci/mmol), 10µl 100mM DTT (Boehringer Mannheim) and 60 units RNAsin (Biotech). The reaction was put at 42°C for 10 minutes and then 200 units reverse transcriptase XL (NBL Enzymes) was added. The reaction was allowed to proceed for 30 minutes and was terminated by the addition of EDTA to a final concentration of 20mM. The reaction was then extracted twice with an equal volume of ultra pure phenol (BRL) saturated with TE (10mM Tris (pH 8.0); 1mM EDTA) followed by an ether extraction to remove the phenol. The cDNA was then precipitated twice with 0.2M sodium acetate and 2.2 volumes absolute alcohol at -70°C for 15-30 minutes.

SECTION 2.21.2

Second strand

The precipitated first strand cDNA was taken up in 50µl of DEPC water, and 5µl was retained for analysis. The remaining cDNA was made up to a reaction volume of 200µl with DEPC water and 40µl of a mix of 500µM each d(A,T,C,G)TPs , 10µl 0.15mM β-NAD, 20µl 10x single strand buffer (200mM Tris pH7.5, 50mM $MgCl_2$, 1M KCl, 100mM NaOAc 500µg/ml BSA), 10 units *E.coli* DNA ligase (NBL), 3 units RNase H (Boehringer Mannheim) and 45 units DNA polymerase I (New England Biolabs). The reaction was then incubated at 12°C for 1 hour followed by another hour

incubation at 22°C. The reaction was terminated by the addition of 1/10 volume 200mM EDTA and after 2 phenol extractions, the cDNA was precipitated with 0.2M sodium acetate and 2.5 volumes of ethanol.

SECTION 2.21.3

S1 nuclease digestion

The cDNA was resuspended in 50µl of DEPC water and 5µl were retained for analysis. 5µl of 10x S1 buffer (1mM $ZnCl_2$, 250mM NaCl, 30mM sodium acetate pH5.5) was added to the reaction and 100 units of S1 nuclease (Boehringer-Mannheim). The reaction was allowed to proceed for 20 minutes at 37°C. The cDNA was then extracted twice with phenol, once with chloroform and finally with ether. The cDNA was then precipitated twice with 0.2M sodium acetate and 2.5x volume of ethanol.

SECTION 2.21.4

Methylation

The precipitated cDNA was resuspended in a volume of 45µl, and 5µl were retained for analysis. This sample and the samples from the first and second strand synthesis were run on a 1.5% agarose gel to compare the size of the cDNA after each step. ϕx174 digested with Hae III was run as a size marker, and the position of the virus fragments marked on the gel with a scalpel blade. The gel was then dried down onto Whatman number 3 filter paper on a BioRad gel drier and autoradiographed (section 2.10).

The methylation reaction proceeded by the addition of 5µl of methylase buffer (100mM NaCl, 100mM Tris-Cl (pH 8.0), 1mM EDTA,) and 5µl S-adenosylmethionine (100 µM (Sigma)) and 40units of *E.coli* methylase (New England Biolabs) and was incubated for 20 minutes at 37°C. The enzyme was then inactivated by heat treatment at 70 °C for 10 minutes.

SECTION 2.21.5

Filling In

2µl of 300mM $MgCl_2$, 6µl d(A,C,G,T)TP (500µM), 1µl DEPC water and 1µl (10 units) DNA Polymerase I large fragment (Klenow fragment) were added to the inactivated methylase reaction and was allowed to proceed for 1 hour at room temperature.

SECTION 2.21.6

Addition of linkers

The cDNA was precipitated twice with 0.2M sodium acetate pH5.5 and 2.5 volumes of ethanol, and was then passed over a 200µl DE52 column (Pharmacia). The DE52 was equilibriated with 10mM Tris.Cl (pH 7.6), 1mM EDTA, 60mM NaCl. A slurry of the equilibriated DE52 (0.6ml will bind 20 µg DNA) was packed into a 1ml syringe containing a glass wool plug. The column was washed with TE (pH 7.6) containing 0.6M NaCl, followed by 3ml TE alone and finally 3ml 0.1M NaCl in TE. The cDNA was loaded onto the column in TE and the flowthrough was collected and reapplied to the column. The column was then washed twice with 1.5ml of TE containing 0.3M NaCl and the cDNA was eluted with three 0.5ml washes of TE containing 0.6M NaCl. The fractions containing radioactivity were pooled, phenol extracted, chloroform extracted, and finally ether extracted. The cDNA was then recovered by precipitation with ethanol and 10µg of yeast tRNA was used to promote precipitation.

In order to follow the addition of the *Eco RI* linkers to the cDNA the linkers were end-labelled (section 2.22).

The cDNA pellet was then taken up in 20µl of DEPC water and 5µl of the cDNA solution was added to 5µl of radiolabelled linker mix and 2µl 10x ligase buffer (500mM Tris-Cl (pH 7.8), 100mM $MgCl_2$, 200mM DTT, 10mM ATP, 500µg/ml BSA), 100 units T_4 DNA polynucleotide ligase (New England Biolabs). DEPC water was then added to a final volume of 20µl and the reaction was incubated at 16°C over-

night.

In order to test whether the ligation reaction had been successful, the cDNA and linkers were run on a 1.5% agarose gel which was then dried down, and autoradiographed.

SECTION 2.21.7

Generation of Eco R1 cohesive ends, and separation of excess linkers from cDNA

The ligation reaction was stopped by heat inactivation and the volume of reaction was increased to 50µl with DEPC water and 2µl of 10x *Eco R1* digestion buffer (500mM NaCl, 100mM Trs.Cl (pH 7.5), 100mM $MgCl_2$, 10mM DTT). 10 units of *Eco* R1 was added to the reaction and incubated for 1 hour at 37°C followed by a further addition of 10 units of *Eco* R1 and a further 45 minute incubation at 37°C.

The excess of linkers were separated from the cDNA by running the *Eco R1* cut sample on a 1.8% agarose gel. ϕx174 cut with *Hae III* was used as marker DNA. When the gel had been run and stained with ethidium bromide, the cDNA over 300bp was excised from the gel and put into dialysis tubing containing 2ml of 0.5x electrophoresis buffer (TBE) (0.5x TBE = 0.044 M Tris-borate; 0.044 M boric acid; 0.001 M EDTA), (Maniatis et al 1982). The sealed tubing was then laid across an electric field of 125v in 0.5x TBE for 1.5 hours. During this time the DNA is electroeluted from the gel onto the wall of the dialysis tubing. The field was then reversed for 2 minutes to release the DNA from the wall of the tubing. The tubing was then opened and the buffer into which the DNA had been eluted was collected, and the bag washed out with a small volume of 0.5x TBE. The cDNA was then made 0.2M with NaCl and passed through a primed Elutip-d mini-column (Schleicher and Schull) which binds up to 100µg of DNA of 50bp to 50,000bp in length. The column was then rinsed with low salt buffer (0.2M NaCl, 20mM Tris.Cl (pH 7.5), 1mM EDTA). The bound cDNA was eluted from the Elutip-d mini-column by passing 400µl of high salt buffer (1M NaCl, 20mM Tris.Cl (pH 7.5), 1mM EDTA) through the column and collecting the

elutant in a 1.5ml Eppendorf tube. The cDNA was then recovered by the addition of two volumes of ethanol and incubation at -70°C, spun down in a a microfuge, and the DNA pellet resuspended in an appropriate volume of TE buffer.

SECTION 2.21.8

Ligation of cDNA to lambda gt10

The cDNA isolated from the agarose gel was resuspended in a small volume, for ligation to the lambda vector. For efficient ligation, equimolar amounts of insert DNA to vector DNA should be present in the ligation. As it was not clear how much cDNA was remaining, half the remaining cDNA was included in the ligation. The reaction was in a volume of 5µl and contained 1µg lambda gt10 DNA (NBL), 0.5µl 10x ligation buffer (0.5M Tris pH7.4, 0.1M $MgCl_2$, 0.2M DDT, 10 mM spermidine, 10 mM ATP, 500µ/ml BSA), 100 units of T_4 ligase (New England Biolabs) and was incubated at 12°C overnight. To check the efficiency of selection for recombinant phage, a control ligation without any cDNA was also prepared. After overnight incubation, the phage was packaged using Gigapack (NBL) packaging mixture and packaging was carried out according to the manufacturers instructions. The packaged DNA was then plated out on 4 large (20x20cm) plastic plates.

SECTION 2.22

End labelling of DNA

10µg *Eco R1* linkers were diluted in 10µl 10x PNK buffer (0.7M Tris.Cl (pH 7.6), 0.1M $MgCl_2$, 50mM DTT), 100µCi ^{32}P-α ATP (8000Ci/mMol), 10 units T_4 polynucleotide kinase (PNK) (New England Biolabs) and DEPC water to a volume of 100µl. The enzyme catalyses the transfer of the γ phosphate of ATP to a 5' OH terminus in DNA or RNA (Richardson 1971). The labelling was allowed to proceed for 30 minutes at 37°C for and then 10µl 10mM ATP (neutralized) and 6 units polynucleotide kinase were added to the labelled linker mix and incubated at 37°C for a further 30 minutes. End labelling of oligonucleotides was carried out essentially in the

same way except that 5 pmole of oligonucleotide 5' ends was used in a 10µl reaction volume with 30µCi ^{32}P-α ATP, 10µl PNK buffer and 10 units T_4 PNK.

SECTION 2.23

Plating and screening a lambda gt10 library

The λgt10 phage were plated out on a lawn of the permissive host *E.coli* C600, and the non permissive host *E.coli* C600 *hfl.*

The following procedure was followed. An overnight culture of cells was grown up by innoculating LB broth /0.2% maltose with a single colony from a plate. Following incubation at 37°C overnight the cells were pelleted and resuspended in 1/2 the original culture volume of 10 mM $MgSO_4$. The cells were stored at 4°C until required for use. Cells up to 3 days old were routinely used.

The phage were then added at a suitable concentration to a 0.1ml aliquot of *E.coli*, if using 90mm petri dishes, and 1ml if using 200mm square petri dishes to plate out the phage. A suitable screening density is 5,000 - 10,000 pfu on 200mm square plates and 500-1000 pfu on 90mm petri dishes. After allowing the phage to adsorb for 15 minutes, 4ml of soft top agarose was added to the tube and poured onto an LB agar plate. The phage were then grown on inverted plates at 37°C overnight.

In order to screen the phage with a radiolabelled DNA probe, the procedure of Benton and Davis (1977) was used. Basically, plates were placed in the fridge for 1 hour and then overlaid with a sheet of nitrocellulose. The filter was laid onto the agar in such a way so as to avoid trapping any air bubbles. The filter position was marked on the plate by piercing the filter and agar with a needle filled with waterproof ink. The first filter remained on the agar for 30 seconds and subsequent lifts were left on an additional 30 seconds to the preceding filter time. The filter was then peeled from the plate and laid (plaque side up) in a shallow dish of denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30 seconds. The filter was then transferred for 5 minutes to a second dish containing neutralising solution (0.5 M Tris-Cl; pH 8.0, 1.5M NaCl). The filter

was then finally rinsed in 2x SSC (0.3M NaCl; 0.03 M sodium citrate) and then dried, and baked at 80°C for 2 hours. Filters prepared in this way were then hybridised with the appropriate probe under the same conditions as for Northern blots (section 2.14), and washed in 2xSSC, 0.1% SDS, 0.1% $Na_2H_2P_2O_7$). When the library was screened with oligonucleotides, the probes were endlabelled (section 2.22), hybridised at 37°C overnight in 6x SSC, 2x Denhardts, 0.1% SDS, 0.1% sodium pyrophosphate, and washed at 31°C or 37°C, in 2x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, with a final wash at 36°C or 42°C. The hybridisation and washing temperatures, depended on the oligonucleotide size and base composition (section 6.1.b). Filters were then autoradiographed, and any positive plaques were lined up to the plate and the plaques picked. The plaque was isolated from the agar with a toothpick and put into 1ml of SM buffer (50 mM Tris-Cl (pH 7.5); 100 mM NaCl; 10 mM $MgSO_4$; 2% gelatin). The phage were then plated out again, and rescreened. These phage should be enriched for positive phage and the process was repeated until all the plaques were positive, i.e a pure population of clonal phage was obtained. Phage were stored at 4°C over chloroform.

SECTION 2.24

Preparing phage DNA

Phage DNA was prepared by, producing a large scale plate lysate on a 200mm square LB agar plate. Bacteria were infected so that there was a confluent lawn of phage, and then following the method of Helm et al (1985), 25ml of SM buffer (50 mM Tris-Cl (pH 7.5); 100 mM NaCl; 10 mM $MgSO_4$; 2% gelatin) was put onto the plate and left shaking in the cold room overnight. The next day the phage had diffused into the buffer and the buffer was removed from the plate. Any agarose in the buffer was removed by spinning at 2,000 rpm in a Sorvall RT6000 centrifuge for 10 minutes. The phage were then pelleted from the supernatant by spinning at 35,000rpm for 90 minutes in a MSE Prepspin 63, for 90 minutes. Phage were resuspended in 1ml

phage buffer (10 mM Tris-CL (pH8.0); 2 mM $MgCl_2$) overnight at 4°C, and then spun 10 seconds in an microfuge centrifuge. The supernatant was then loaded onto a 2ml DEAE-cellulose column. The column run-through was discarded and the column was washed with 5ml of chase buffer (10mM Tris-Cl (pH8.0); 10 mM magnesium acetate; 60 mM sodium acetate). 1ml of elution buffer (10 mM Tris-Cl (pH8.0); 50 mM magnesium acetate) was then loaded onto the column and discarded. 0.6ml of of elution buffer was then loaded onto the column and the run-through was collected into a 1.5ml Eppendorf tube.

To the tube containing the 0.6ml of phage, 10µl of 0.1mg/ml proteinase K and 24µl of 10% SDS were added, and left at room temperature for 5 minutes. Then 100µl of 3 M potassium acetate was added and a precipitate formed. The tube was then heated to 88°C for 20 minutes and the precipitate disappeared. After cooling to 0°C for 5 minutes the precipitate reformed. The tube was then spun at ~12,000 x g for 15 minutes and the supernatant transferred to a fresh tube with 20µl of 1 mg/ml mussel glycogen being added to help precipitate the phage DNA. 700µl of isopropanol was then added and the sample was cooled to -20°C for 30 minutes, then precipitated by spinning at 12,000 x g for 20 minutes. The pellet was then washed in 100% ethanol and respun, dried and resuspended in 25-40µl of TE (pH8.0).

SECTION 2.25

Cloning into λgt11

The cDNA for cloning into λgt11 was prepared as described in section 2.21. The cDNA was ligated into alkaline phosphatase treated λgt11 arms (NBL) overnight at 12°C. The phage were then packaged with Gigapak (NBL) according to the manufacturers instructions. The phage were then plated out as with λgt10 but the bacterial strain used was Y1088 which are *hsdR- hsdM+*; this means that unmodified phage will not be restricted and will be modified to grow in *hsdR+* bacteria. The Y1088 also carry a lac repressor which will repress the production of the recombinant β-

galactosidase protein and so avoid creating a bias against poorly growing phage during library amplification. Packaged phage are plated out as for λgt10 (section 2.23) using *E.coli* Y1088. The percentage of recombinant phage was determined by including 40μl of 40mg/ml Xgal and 40μl of 1M IPTG in the top agarose for a 90mm petri dish. Bacteria infected with λgt11 are grown at 42°C because the phage has a temperature sensitive repressor (*CI857*) which is active at 37°C but inactive at 42°C.

SECTION 2.26

Screening λgt11 libraries with antibody probes.

Screening the λgt11 liver library (Clontech) with antibody probes, was carried out in the following manner. The phage were plated out on a lawn of *E.coli* Y1090 which contains lac repressor, *lon* protease deficiency and *supF* to suppress the phage mutation which causes defective lysis (S100). After the plaques had grown (by growing at 42°C for 3 hours), the plates were put at 37°C and overlaid with dry nitrocellulose filters, previously saturated with 10 mM IPTG. The filters were left on the plates for 2-3 hours, their positions marked, and then removed. The filters were rinsed briefly in TBS (Tris-Cl pH 8.0; 150 mM NaCl) and then incubated overnight in TBS + 3% bovine serum albumin (BSA) (Sigma). The filters were then incubated for 1 hour in TBS plus antiserum at a suitable dilution. Using a test system with purified albumin (Sigma) it was determined that a dilution of 1/500 Dako rabbit anti-human albumin antiserum would detect >1ng of albumin. It was also checked that the anti human albumin antiserum did not cross react with the BSA. After antibody incubation, the filters were washed in TBS 0.05% Tween 20 for 30 minutes and then 15 minutes in TBS plus 3% BSA followed by 15 minutes in TBS. The filters were then incubated in TBS containing ^{125}I-protein A (10^6 cpm/ 82 mm filter, specific activity 30 mCi/mg (Amersham)) for 1 hour at room temperature. The filters were then washed extensively in TBS plus 0.05% Tween 20 with 4 or 5 changes of wash. The filters were then dried and autoradiographed. In order to check that any positive clones were albumin clones, replica

filters were taken and probed with an albumin cDNA probe as described in section 2.23.

SECTION 2.27

Differential screening

This was carried out as follows. The library to be screened was plated out as previously described (section 2.23) and screened with first strand cDNA made to 1μg of poly adenylated RNA as described in section 2.21.1. Randomly primed cDNA was made by the method of Arrand (1986). Filters were hybridised and washed in the same conditions as for Northern and Southern blots (sections 2.17 and 2.37). Screening with the albumin probe was carried out as described in section 2.23.

SECTION 2.28

Restriction endonuclease digestion

Restriction enzymes were obtained from Boehringer Mannheim or New England Biolabs or NBL enzymes and used in high salt, low salt or medium salt (Maniatis et al (1982)) according to the manufacturers instructions.

High salt buffer

100 mM NaCl; 50 mM Tris (pH7.5); 10 mM $MgCl_2$; 1mM dithiothreitol.

Mid salt buffer

50 mM NaCl; 10 mM Tris (pH7.5); 10 mM $MgCl_2$; 1mM dithiothreitol

Low salt buffer

10 mM Tris (pH 7.5) 10 mM $MgCl_2$; 1 mM dithiothreitol.

Plasmid DNA was generally digested at a concentration of ~50μg /ml in buffer containing 3 to 10 fold enzyme units per μg of DNA, for several hours at the optimimum temperature for the enzyme. Digestion of genomic DNA was carried out at a concentration of 100-200μg DNA /ml in buffer containing 3 units of enzyme/ μg DNA, overnight at the optimum temperature for the enzyme.

SECTION 2.29

Agarose gel electrophoresis

DNA digestion and size were analysed by agarose gel electrophoresis. Small amounts of DNA (> 1μg) were analysed on 3mm thick 15 ml gels in tris-borate buffer (TBE) buffer (0.089M Tris; 0.089M boric acid; 0.002M EDTA (pH8.0)). Genomic DNA restriction enzyme digestion was checked by running a 1μg sample on a 3mm thick 100 ml agarose gel, buffered with tris-acetate buffer (TAE) (0.04M Tris-acetate; 0.002M EDTA (pH 8.0). Analysis of genomic DNA by Southern blotting was performed by initially running 10μg of digested DNA on a 3mm thick 300ml TAE buffered 0.8% agarose gel. DNA was loaded onto gels in "STOP mix" (30% Ficoll (Pharmacia Fine Chemicals); 0.25% Orange G; 0.5M EDTA and 10x electrophoresis buffer). The Ficoll increases the sample density and the Orange G enables the sample to be easily visualised and acts as a marker close to the front of the migrating DNA. After electrophoresis, gels were stained in electrophoresis buffer containing 1 drop of 10 mg/ml ethidium bromide for 15 minutes. The gels were then destained for 15 minutes in double distilled water. Ethidium bromide intercalates with DNA and its fluorescent properties allowed the DNA to be visualised when illuminated with UV light.

SECTION 2.30

Ligation of DNA fragments

Ligations were carried out in the following conditions :- linearised vector DNA (2-5μg/ml), insert DNA at 3-5 fold molar excess, 1/10th volume of ligation buffer (10x ligation buffer = 0.5M Tris pH 7.5; 0.1M $MgCl_2$), 1mM dithiothreitol, 1mM r-ATP, T_4 DNA ligase and sterile water to the required volume. Reactions were carried out at 12°C overnight in as small a volume as possible (~ 10μl).

SECTION 2.31

Transformation of bacterial cells

The *E.coli* K12 strain JM 83 was used for growing all plasmids. Competent cells were prepared as described by Maniatis et al (1982). Essentially, 100ml of LB medium was innoculated with 1ml of an overnight bacterial culture. The cells were then grown at 37°C with vigorous shaking to a 0.5 OD_{550}. The culture was then chilled on ice for 10 minutes and centrifuged at 4000 x g for 5 minutes at 4°C. After discarding the supernatant the cells were resuspended in half the original culture volume of 50mM $CaCl_2$. The cells were then left on ice for 20 minutes and recentrifuged. The resulting pellet was resuspended in 1/10 of the original volume of an ice cold, sterile solution of 50mM $CaCl_2$ and kept at 4°C for up to 48 hours. 300µl of cells were used in each transformation, and plasmid DNA was added in 100µl at a concentration of 1µg/ml. The tubes were kept on ice for at least 40 minutes and then heat shocked for 90 seconds at 42°C. The transformation mix was then made up to 1ml with LB broth and incubated for 1 hour at 37°C to allow expression of the plasmid encoded antibiotic resistance genes before growth on selective medium. Normally 50 to 200µl of the incubated mix was plated directly onto H-agar plates containing the appropriate antibiotics using a bent, sterile glass rod to distribute the mix. In the case of pUC 9, indication of insert DNA was the presence of white colonies when plated on X-gal plates.

The bacterial strain *E.coli* JM 101 was used for M13 transformations. The cells carry a plasmid which codes for the production of the pilus which is necessary for the M13 infection. This plasmid also codes for the proline gene while the bacterial DNA is deleted for the proline gene. Thus stock plates are glucose/minimal medium so only *E.coli* with the plasmid can grow. OD_{550} of ~0.4. At this time, 20ml of 2xTY was innoculated with a drop of the culture and allowed to grow. The bacteria with an OD of 0.4 were pelleted and resuspended in 1/2 volume of 50mM calcium chloride and left on ice for 20 minutes. The cells were then repelleted and resuspended in a volume 1/10th that of the original volume (2ml). The transformations were carried out by the

addition of the M13 DNA to 0.3ml of cells and left on ice for 40 minutes. The cells were then heat shocked as for the JM 83 cells but after the heat exposure 200µl of fresh cells were added and 40µl of 2% X-gal and 40µl of 100mM IPTG. The cells were then plated in 4ml of soft agarose on a 90mm LB agar plate.

SECTION 2.32

Small scale plasmid preparation

Small scale preparation of plasmid DNA was used to analyse large numbers of single colony bacteria to ensure they contained the correct plasmid before continuing with a large scale preparation. The technique used (a modification of Maniatis et al 1982) is as follows. 5ml of LB broth containing ampicillin at 50µg/ml was innoculated with a single bacterial colony. The culture was then grown overnight at 37^{o}C with vigorous agitation. 1.5ml of the culture was then poured into a 1.5 ml Eppendorf tube and the bacterial cells were pelleted. The pellet was then resuspended in 200µl of TELT (50 mM Tris-Cl (pH7.5); 62.5 mM EDTA; 2.5 M LiCl; 0.4% Triton X100) plus 1 mg/ml freshly added lysozyme, and boiled for 1 minute. The sample was then placed on ice for 8 minutes, spun at 12,000 x g for 15 minutes and the supernatant removed to a tube containing 0.6 volumes of isopropanol. After incubation on ice for 10 minutes, the tube was spun again for 15 minutes, the pellet dried and redissolved in 30µl of TE (pH 8.0). 10µl was then checked by digesting with a diagnostic restriction enzyme and run on a small (15ml, 3mm thick) gel.

SECTION 2.33

Large scale preparation of plasmid DNA

The fact that *E.coli* DNA is much larger than commonly used plasmid vector DNA, and that genomic DNA is isolated as broken linear molecules where plasmid DNA is on the whole, isolated as covalently, closed, circular DNA, allows plasmid DNA to be separated from genomic DNA. Exposure to heat or to alkali will irreversibly

denature the *E.coli* DNA whereas the covalently closed plasmid DNA will not be damaged in this way. When the two types of DNA are centrifuged to equilibrium in caesium chloride gradients, containing saturating amounts of the intercalating dye ethidium bromide, the plasmid DNA will bind much less of the dye and will band at a higher density in the gradient (Radloff et al 1967). The large scale preparation of plasmid DNA used was a modification of Maniatis et al (1982). Essentially, an overnight culture of transformed bacteria was grown in 500ml of LB Broth plus 0.1% glucose. The culture was spun down into two pellets and each dissolved in 10 ml of GTE (50mM glucose, 25mM Tris.Cl (pH8.0), 10mM EDTA) plus 20mg lysozyme and left on ice for 10 minutes. After this time, 40ml of alkaline SDS solution was added to the pellet (2ml 20% SDS, 2ml 5N NaOH, 36ml H_2O) and slowly stirred in. After 5 minutes on ice, 30 ml of high salt solution (3M potassium acetate, 2M acetic acid (pH4.8) was stirred into the mix . This was made by dissolving the potassium acetate in half the required volume of water and pH to 4.8 with glacial acetic acid. The full volume was then made up and this makes the acid molarity 2M.) The tubes remained on ice for 10 minutes and then were centrifuged at 10,000 rpm in a Sorval RC-5B for 15 minutes. 160ml of absolute ethanol was then added to the muslin filtered supernatant and left at -20^{o}C for 20 minutes. The solution was then resupun for 15 minutes at 10,000 rpm and the pellet was resuspended in 6ml TE plus 4ml 7.5M ammonium acetate and put on ice for 20 minutes after which time the tubes were spun at 14,000 rpm for 20 minutes. The supernatant was then treated with 20ml absolute alcohol to precipitate the DNA and spun at 10,000 rpm for 15 minutes. The pellet was then treated in one of two ways. Either the pellets * were resuspended in 1ml TE and half was frozen. The other half was treated with 10µl of a 10 mg/ml stock of RNase A for 30 minutes at 37^{o}C. The DNA was then phenol extracted and the aqueous phase re-extracted with chloroform followed by ether. 160µl of 7.5M ammonium acetate was then added to the tube and 350µl of isopropanol. The precipitated DNA was spun out in an Eppendorf

UNIV. EDINBURGH. EX BIBL.

centrifuge and then resuspended in 400µl DEPC treated water. This DNA solution was then cut with a suitable enzyme and the amount estimated by running the DNA on an agarose gel.

Alternatively the pellets * were resupended in 3.21ml TE plus 3.55g caesium chloride and 0.34ml of a 10mg/ml solution of ethidium bromide. The final density should be 1.55 g/ml. The solution was then transferred to a 5ml polycarbonate tube and spun in a TV 865 rotor in the Sorvall OTD-65 overnight at 40,000 rpm at 20°C. The following morning 2 bands should be visible, the lower band being the closed circular plasmid DNA. This lower band was collected into a tube and extracted with butan-2-ol to remove the ethidium bromide. As many extractions as were necessary to remove the dye plus one more, were carried out, and after dialysis against several changes of TE (pH 8.0) the DNA was precipitated and the concentration determined by reading on the spectraphotometer at 260nm. At 260nm, 1 OD = 50µg DNA.

SECTION 2.34

Isolation of insert DNA

The clone 8-9 cDNA insert was isolated from plasmid DNA by the following method (Weislander 1979). The plasmid was cut to completion with the enzyme *Eco RI* and 10µg was run slowly on a 1% low melting point agarose (FMC) gel. The gel was then stained with ethidium bromide and the relevant band excised with a scalpel. The agarose was then melted in 5 volumes of 20mM Tris-Cl (pH8.0) and 1 mM EDTA at 65°C for 15 minutes. The melted gel slice was then extracted by the addition of an equal volume of prewarmed phenol. The aqueous phase was recovered by centrifugation at 20°C and re-extracted with phenol/chloroform, chloroform and finally diethyl ether. The fragment DNA was then recovered by the addition of 1/10 volume of 2M sodium acetate (pH 5.5) and 2.2 volumes of ethanol.

SECTION 2.35

Sequencing in M13

The first stage in sequencing cDNA clone 8-9 was to subclone the DNA into M13. This was carried out by cutting the vector in its replicative form with *Eco R1* and cutting the plasmid containing the insert DNA with the same enzyme. The DNA was then ligated overnight (section 2.30) and JM 101 cells were transformed with the ligated DNA (section 2.31). The recombinant phage (single stranded) were identified from nonrecombinant phage by plating on X-gal plates. Recombinant phage gave colourless plaques. Recombinant plaques were picked with a toothpick and grown in 1.5ml of a 1/100 dilution of an overnight culture of JM 101 in 2xTY broth. The cultures were then grown at 37°C for 5 hours. The cultures were then transfered to Eppendorf tubes and centrifuged for 5 minutes. The supernatant (containing the single stranded phage) was then removed from the bacterial cells (containing the replicative double stranded form) and the supernatant was recentrifuged to ensure no bacterial cells were remaining. The "clean" supernatant was then added to 200µl of PEG/NaCl (20% polyethylene glycol 6000/2.5M NaCl), shaken, and allowed to stand for 15 minutes at room temperature. The tube was then centrifuged for 5 minutes and the supernatant was discarded. The tube was then recentrifuged for 2 minutes and any remaining PEG was carefully removed. 100µl of TE pH 8.0 was added to the pellet and 100µl of ultra pure phenol (BRL) was added. After vortexing, the tube was allowed to stand for 15 minutes and then was centrifuged for 3 minutes. The aqueous layer was transferred to a new Eppendorf tube and re-extracted with chloroform. The aqueous layer was then removed and the DNA precipitated with 10µl 2M sodium acetate (pH 5.5) and 250µl of ethanol, overnight at 20°C. The resulting pellet of single stranded template obtained after 10 minutes centrifugation, was resuspended in 50µl TE (pH8.0).

5µl of single stranded template was annealed to the 17 nucleotide primer supplied

with the Amersham sequencing kit. 1µl of primer (2.5ng) was added to 5µl of template with 1.5µl 10x buffer (100mM Tris pH 8.0; 50 mM $MgCl_2$) and 2.5µl double distilled water. The annealing reaction mix was then placed at 60°C for 1-2 hours after which the radiolabelled dTTP (1µl of >400Ci/mmol (Amersham)) was added to the tube together with 1µl of Klenow fragment (1U/µl (Boehringer Mannheim)). The reaction was then divided into four tubes numbered 1-4.

Stock 0.5mM deoxy NTP solutions and 0.5mM dideoxy ATP, 0.05mM dideoxy GTP, 0.02mM dideoxy CTP and 0.15M dideoxy TTP solutions were used to make the solutions for each of the four tubes. Ao consisted of 20µl dCTP, 20µl dGTP, 1µl dATP and 20µl TE pH8.0. Co consisted of 20µl dGTP, 1µl dCTP, 20ml dATP and 20µl TE . Go consisted of 20µl dATP, 20µl dCTP, 1µl dGTP and 20µl TE. To consisted of 20µl dATP, dCTP, dGTP and 20µl of TE.

A' contained 25µl Ao + 25µl ddATP, C' contained 25µl Co + 25µl ddCTP. G' contained 25µl Go + 25µl ddGTP. T' contained 25µl To + 25µl ddTTP.

Tube 1 contained 2µl A', tube 2 contained 2µl C', tube 3 contained 2µl G' and tube 4 contained 2µl T'. The reaction was divided between these four tubes as described above and left at room temperature for 15 minutes. After this time, 2µl of chase mix (0.5mM dTTP) was added. After a further 15 minutes 4µl of formamide dye mix (100ml formamide, 0.1g xylene cyanol FF, 0.1g bromophenol blue, 4ml 500mM EDTA) was added to the tubes and the samples were placed in boiling water bath for at least 3 minutes directly before loading 3µl aliquots onto a gel.

The gels were prepared as follows. Two glass plates 20cm x 40cm, one with notched "rabbit" ears and one without, were cleaned with ethanol. The plate with the "rabbit" ears was coated with a siliconising solution (2% dimethyldichlorosilane in 1,1,1-trichloroethane (BDH)) and the unnotched plate was treated with γ-methacryloxypropyl-trimethoxy silane (Sigma). This silane compound was prepared by adding 30µl to 10ml ethanol and adding to this solution, 30µl of acetic acid in 300µl

of water. The gel side of the plate was then coated with this solution, allowed to dry and washed with ethanol. Plastic 0.35mm spacers were then placed down the sides of the plates and they were taped together with vinyl tape. 21g ultrapure urea (BDH), 7.5ml of acrylamide stock (38g acrylamide, 2g NN'-methylenebisacrylamide, in 100ml distilled water), 5ml 10x TBE buffer (section 2.29), in 50 ml with distilled water was prepared. 300µl freshly made 10% ammonium persulphate solution and 50µl TEMED was added just before the gel was poured. The gel was poured between the plates and a comb with 5mm teeth at 3mm intervals was inserted into the gel when it was set. The plates were held together more firmly with bull dog clips and the gel was then put into the vertical gel apparatus (BRL) with TBE as the running buffer. Gels were run at 27 watts until the blue dye front reached the bottom of the gel. Gels were then removed from the gel tank, the notched plate removed, and fixed in 10% acetic acid for 10 minutes. After extensive washing, the gel was dried onto the glass plate by placing it in an 80°C oven for 1 hour. The gel was then autoradiographed overnight at room temperature and the sequence could be read the following morning on developing the autoradiograph. The smallest fragment is the first band at the bottom of the gel. The next fragment up in the next smallest and so on. Gels read up towards the comb are being read in a direction 5'-3' from the primer. If the smallest fragment for example was in the T track, then the M13 template DNA is an A as the bases are complementary to the template. The gels could be read more easily by remembering that the upper C in a doublet is always more intense than a lower C and an upper G is often more intense than a lower G in a doublet.

SECTION 2.36

Extraction of DNA from cell pellets

Cell pellets of approximately 10^7 cells were used to extract DNA for analysis by Southern blotting. Pellets of ~5 x 10^7 were stored at -20°C and when DNA was extracted, they were thawed into 3 ml of lysis buffer (0.5% SDS; 150 mM NaCl; 10

mM Tris (pH 8.0) 10 mM EDTA). 10µl boiled, DNase free RNase A was added per ml of lysate, and the cells were incubated at 37°C for 1 hour after which time 50µl/ml of proteinase K (5mg/ml in lysis buffer) was added. The tube was then placed at 37°C for 4hr-overnight. The solution was extracted with an equal volume of phenol (water saturated) followed by reextraction of the aqueous phase with chloroform:phenol 1:1, and finally chloroform:isoamyl alcohol 24:1. After each organic addition the tube was shaken vigorously and spun at 12,000 x g to clarify the upper aqueous phase. After the final extraction, 0.5 volumes of 7.5M ammonium acetate was added to the solution and 2.5 volumes of absolute alcohol. The DNA was precipitated by the alcohol and was removed from the tube on a flame-sealed glass pipette. The DNA was then air dried, washed in 66% ethanol in 0.8M ammonium acetate, redried, and redissolved overnight in Tris-EDTA (10 mM Tris (pH7.5); 1mM EDTA). The amount of DNA was estimated by reading the optical density at 240, 260 and 280nm. At 260nm, 1 mg of DNA gives an OD of 20. Alternatively 1 OD at 260nm = 50µg of DNA/ml. The purity of the sample can be roughly estimated by the ratio of the 260nm:280nm reading. If the ratio is ~2, then the absorption is probably due to nucleic acids. If it is below 1.6 there may be protein or other UV absorbing molecules in the sample.

SECTION 2.37

Southern blotting

The technique of Southern blotting (Southern 1975) was used to transfer DNA from agarose gels to nitrocellulose filters. The DNA is denatured, neutralised and transferred in a high salt buffer by capillary action. The denatured, single stranded DNA binds to the filter, is then permanently bonded by baking the filter and is then hybridised to a radiolabelled probe to detect homologous sequences. In detail, a large (20 cm x 40cm) agarose gel was run to electrophoretically separate DNA which had been previously digested with restriction enzymes. The gel was then stained with ethidium bromide and a photograph taken. After this, the gel was placed in a plastic tray

containing 500ml of denaturing solution (1.5 M NaCl; 0.5M NaOH) for 1 hour at room temperature. The gel then was neutralised in 1 M Tris-Cl, 1.5 M NaCl (pH 5.5) for 1-2 hours at room temperature. The was gel transfered to a nitrocellulose filter on a transfer apparatus (Southern 1975) containing 20x SSC. The gel, covered with a nitrocellulose filter (presoaked in 2x SSC), 2 sheets of Whatmann 3mm filter paper and a wad of paper towels, was left at room temperature overnight. The following day, the nitrocellulose filter was removed and baked for 2 hours in a vacuum oven at 80^{o}C for 2 hours. The filter was then ready to be hybridised (as for Northern blotting, section 2.17) to a suitable radiolabelled probe (section 2.18), and autoradiographed (2.10).

CHAPTER 3

CFAg EXPRESSION IN MYELOID CELLS

3.1 Turnover of CFAg in various cell types

In order to isolate the cDNA for CFAg, it was first necessary to determine which of the tissues which express CFAg are actively synthesising the protein. It was assumed that if active protein turnover could be demonstrated then the mRNA for the gene would be present.

Granulocytes

Using the guinea pig antiserum (Manson and Brock 1980) raised against the excised isoelectric focussing band cut out of the gel (anti-IEF), van Heyningen et al (1985) demonstrated demonstrated that CFAg was present in normal and CML granulocytes. However granulocytes (polymorphonuclear leukocytes) are short lived end cells and the majority of their tissue specific products are synthesised in the bone marrow when granule formation occurs. The variable lobulated shape of the nucleus (which gives polymorphs their name) contains mainly clumped chromatin; there is almost no disperse chromatin and no nucleolus implying that little of the genome is being transcribed, and there is very little formation of new ribosomes (Rogers 1983). Neutrophils have very little rough endoplasmic reticulum (RER) and only a small centrally placed Golgi apparatus. Secretory granules and primary lysosomes are manufactured on the RER and processed through the Golgi apparatus in the course of differentiation. The lack of these organelles suggests that synthesis of protein and its packaging in external membrane have ceased.

The granulocytes' main function is to protect the body from infection and they are the first cell type at the site of inflammation. The bone marrow stage of the neutrophil life cycle commences with proliferation of commited stem cells and ends with the release of the neutrophil into the blood stream. Proliferation lasts about 5 days, maturation is completed after approximately 10 days and the mature cells are released into the circulation 1.5 days later (Athens 1970). The mature neutrophil stays in the circulation for less than 1/2 a day on average and then migrates into the tissue. 10^{11} neutrophils are released every day even in healthy adults with no inflammation.

The term granulocyte population refers to all polymorphonuclear (PMN) cells in the blood and they comprise of neutrophils, eosinophils and basophils. Neutrophils however account for 90% of these cells.

The neutrophil consists of primary and secondary or azurophilic and specific granules. The primary granules are peroxidase positive and uniquely contain myeloperoxidase, elastase, cathepsin G, proteinase 3, beta-glucuronidase, alpha-mannosidase, cathepsin B, cathepsin D (Baggiolini et al 1978) and all these components are produced in the promyelocyte cell. The next cell stage the myelocyte produces the specific granules which contain collagenase, lactoferrin, vitamin B12 binding proteins and in common with the primary granules, they also contain lysozyme. The developing neutrophil then continues through metamyelocyte and band cell stages to mature neutrophil characterised by its lobed nucleus. There also appears to be another granule type named a C particle which contains some proteases and the bulk of plasminogen activator (Bretz 1978). Although at the time of these experiments it was not known whether CFAg is associated with the cytoplasm or particulate fraction of granulocytes, there is now considerable evidence to suggest that CFAg is cytoplamic.

CML cells

CML cells also contain CFAg. It was hoped that if normal granulocytes did not actively synthesise CFAg then it might be synthesised by these more immature cells.

CML is characterised by massive expansion of pools of committed myeloid progenitors (Champlin et al 1985). In the chronic stage of the disease mature segmented cells, are found alongside myeloblasts, through promyelocytes to myelocytes of variable maturation. Subtle abnormalities in granulocytic function are observed (Olofsson et al 1976). The number of granulocytes is highly elevated 1-4 x 10^{11}/litre compared to 1-2x 10^{9} /litre in normals (Chikkappa et al 1976). In the aggresive leukemia stage the proportion of promyelocytes and blast cells increases and in blast crisis the cells no longer differentiate to mature granulocytes. Blast crisis can be "lymphoid" or "myeloid" (Rosenthal et al 1977) and at this stage other karyotypic abnormalities are superimposed on the Phildelphia chromosome phenotype which is present 90% of CMLs (Rowley 1975).

HL-60 cells

A third possibility for isolating mRNA coding for CFAg was to examine established myeloid cell lines. If a cell line growing in culture produced immunologically detectable antigen then it must contain mRNA for the protein. The promyelocytic cell line HL-60 (Collins et al 1977) was chosen as a possible candidate cell line. HL-60 was established from a patient with acute promyelocytic leukemia. These cells can be induced to differentiate towards macrophage like cells (Lotem et al 1979; Huberman and Callahan 1979; Rovera et al 1979; Rovera et al 1979a ; Koeffler et al 1981) when treated with tumour promoting phorbol diesters. This is characterised by arrested division and DNA synthesis and adherence of the cells to plastic substrates. The cells acquire complement receptors and phagocytic ability (Chaio et al 1981).

HL-60 can also be induced to differentiate towards granulocytic cells when treated with a variety of chemicals. These chemicals do not include glucocorticoids but this unresponsiveness is not due to a lack of steroid receptors (Freidman et al 1976). High concentrations of agents which induce a rise in intracellular levels of cAMP and 500 mM N,O dibutryl cAMP also induce HL-60 to mature to the myelocyte and

metamyelocyte stage (Chaplinski et al 1982). Low levels of retinoic acid (RA), 10nM or agents known to increse levels of cAMP will not induce differentiation, but priming for 24hrs by treatment with RA followed by treatment with cAMP inducing agents will stimulate the cells to mature towards granulocytes (Olsson et al 1982).

A large variety of compounds will induce differentiation of HL-60 along the granulocyte pathway (Koeffler 1983). These include dimethyl sulphoxide, hypoxanthine, hexamethylene bisacetamide, thymidine, butyric acid, retinoic acid, vitamin D_3, actinomycin. The degree of maturation is determined on the basis of morphology and the ability of the cells to reduce nitroblue tetrazolium (NBT) due to the presence of superoxide anion.

RA appears to be one of the most powerful inducers of HL-60 as relatively low concentrations are necessary to produce a high percentage of mature cells. The 13-cis form of the chemical is reported to be equally as effective as the all trans form (Breitman et al 1981). Ferrero et al (1983) examined the surface changes HL-60 produced on stimulation with DMSO and RA. The cells did not produce all the biochemical markers symptomatic of differentiation but treatment with RA produced significantly more cells which react with a monoclonal antibody (McAb) specific for mature myeloid cells. The induced maturation of HL-60 to granulocytes is not therefore a perfect model of myeloid differentiation and the absence of alkaline phosphatase and lactoferrin from induced cells indicates that the secondary granules are absent (Gallagher et al 1979). However they do not possess HLA-DR antigen which is present normally on myeloblasts and not on mature granulocytes (Zebrowski et al 1977).

On the basis of the apparent effectiveness of RA in inducing HL-60 to mature towards granulocytic cells, it was decided to use this compound to induce HL-60 and look for CFAg expression.

3.2 Method of analysis

The antibody used to detect CFAg in the above mentioned tissues was kindly supplied by Professor D. Brock. It was raised in guinea pigs against the band peculiar to CF and heterozygote individuals visible in isoelectric focussing gels of these individuals' serum (Manson and Brock 1980). This antiserum is termed anti-IEF antiserum. Another antiserum which was used was raised in guinea pigs against partially purified CFAg-positive fraction of myeloid leukaemia cells and is termed anti-CML (van Heyningen et al 1985). The method used to establish immunoreactive identity between these two antisera was by Ouchterlony double diffusion. In this technique a buffered agar gel is prepared and antigen and antibody are loaded in wells bored in to the agar. After incubation for 12-48 hrs antibody antigen recognition is detected by an immunoprecipitin line between the wells. The formation of a fused immunoprecipitin line between two different antibody sources to the same antigen well infers immunoreactive identity i.e the antisera recognise the same antigen in the protein mix in the antigen well. Equally, a fused precipitin line without spurs between an antibody and two different antigen sources suggests antigen identity.

3.3 Results

3.3.a CFAg in HL-60 cells

HL-60 cells were grown in serum-containing and serum-free medium and cell lysates were tested against the guinea pig antiserum raised against the band cut out of the isoelectric focussing gel (Manson and Brock 1980), in Ouchterlony double diffusion analysis. No reaction was observed when cells were grown in serum-free medium but a very faint immunoprecipitation was seen on cell lysates grown in 5% fetal calf serum, suggesting that fetal calf serum acts as a mild inductive agent. Cells were then induced to differentiate by the continuous presence of 10^{-6}M RA and cells were harvested on each day from day 0 to day 6. Differentiation was assesssed by the ability of cells to reduce NBT and figure 1 shows a typical view of dark blue differentiated cells and colourless undifferentiated cells.

Figure 2 shows the number of viable cells on each day and the number of cells capable of reducing NBT when grown in the continuous presence of RA. The results agree with the findings of Breitman et al (1980, 1981). By day 2 over 40% of the cells were morphologically mature i.e metamyelocytes and beyond. After 5 days continual exposure to 10^{-6}M RA, only 70% of the cells were viable as determined by trypan blue exclusion and 80% of the viable cells were differentiated as determined by NBT reduction.

Cells were harvested and lysed in PBS (10^7 cells per 100ul) and tested for antigen. The results show that there is no CFAg detected in cells grown in medium with 5% FCS when undiluted (anti-IEF) antiserum is used. However when the antiserum is diluted to 1/8, a precipitin line is visible. In this technique, antigen and antibody diffuse towards each other through the agar. Where a precipitin line forms is dependant on the molecular weight of the antigen and upon equivalence concentration being obtained at some point in the overlapping antigen /antibody gradients. When antigen concentration is very low precipitation may only occur if the antibody is

diluted. When HL-60 was grown in serum free medium no precipitation was seen at any antigen or antiserum dilution.

Figure 3a-c shows antigen levels as detected by Ouchterlony double diffusion on days 1-5.

Day 0 - equivalence is at 1 in 8 dilution of anti-IEF antiserum.

Day 1 - equivalence is at 1 in 2 dilution of antiserum

Day 2 - equivalence is at 1 in 4 dilution of cell lysate

Day 3 - equivalence is at 1 in 2 dilution of antiserum

Day 4 - equivalence is at 1 in 2 dilution of antiserum

Day 5 - equivalence is at 1 in 2 dilution of antiserum

This implies that CFAg expression is increased 32 fold from day 0 to day 2 and HL-60 grown on serum free medium has no detectable levels of CFAg. CFAg levels in HL-60 cell (grown for 2 days in the presence of RA), and CML cell lysates, were compared. Both cell lysates were prepared from equal cell numbers. Equivalence against anti-IEF was achieved at a 1 in 32 to 1 in 64 dilution of CML lysate and at a 1 in 4 dilution of HL-60 lysate. Thus CML appears to have at least 16 fold greater steady state CFAg compared to HL-60 induced with RA for two days. No CFAg could be detected when HL-60 was treated with the tumour promoting agent 12-0-tetradecanoylphorbol-13-acetate.

FIGURE 1

Photograph to show the appearance of a differentiated and undifferentiated HL-60 cell as determined by NBT reduction. The dark cells are differentiated while the light cells are not (x315).

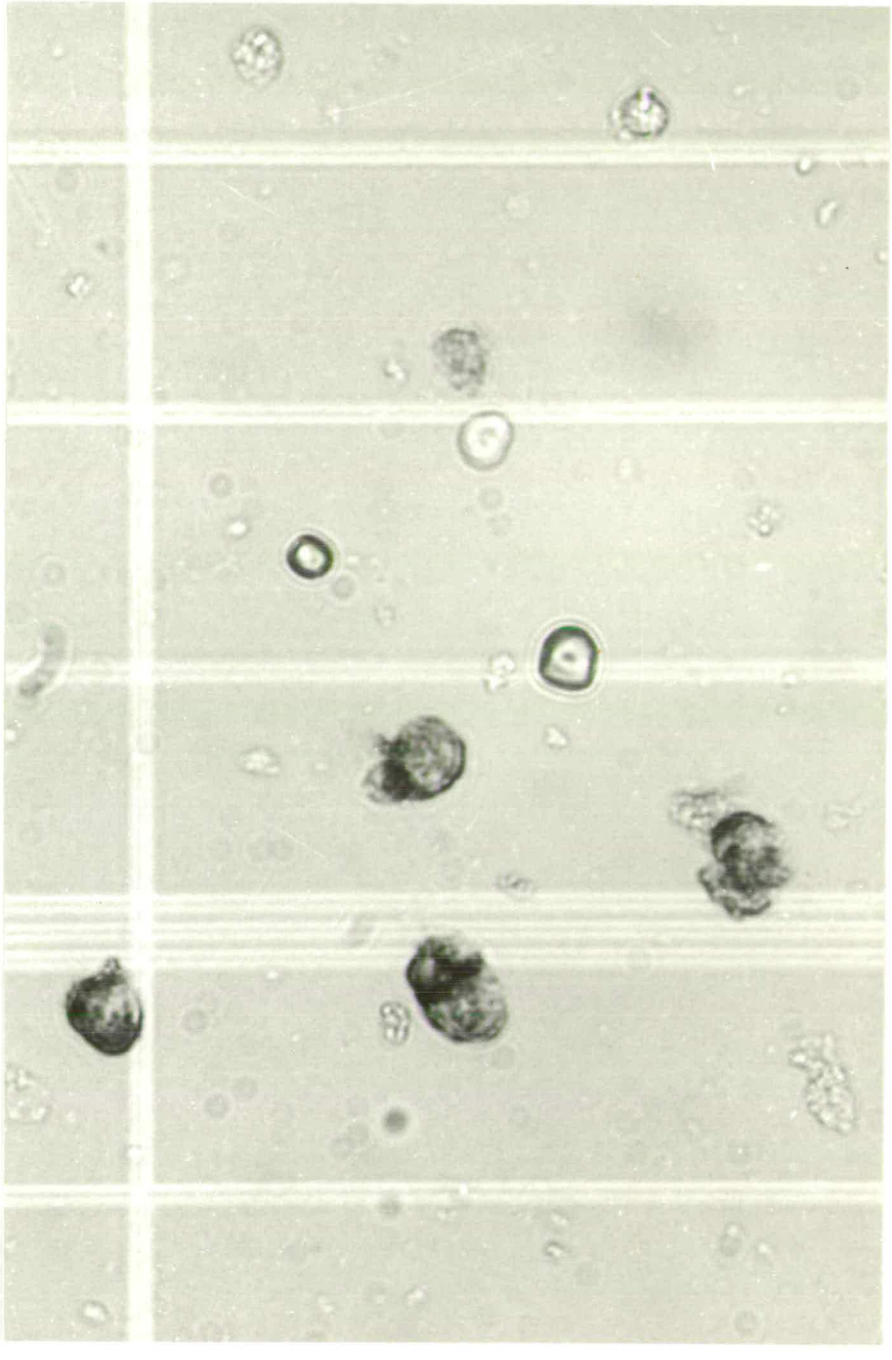
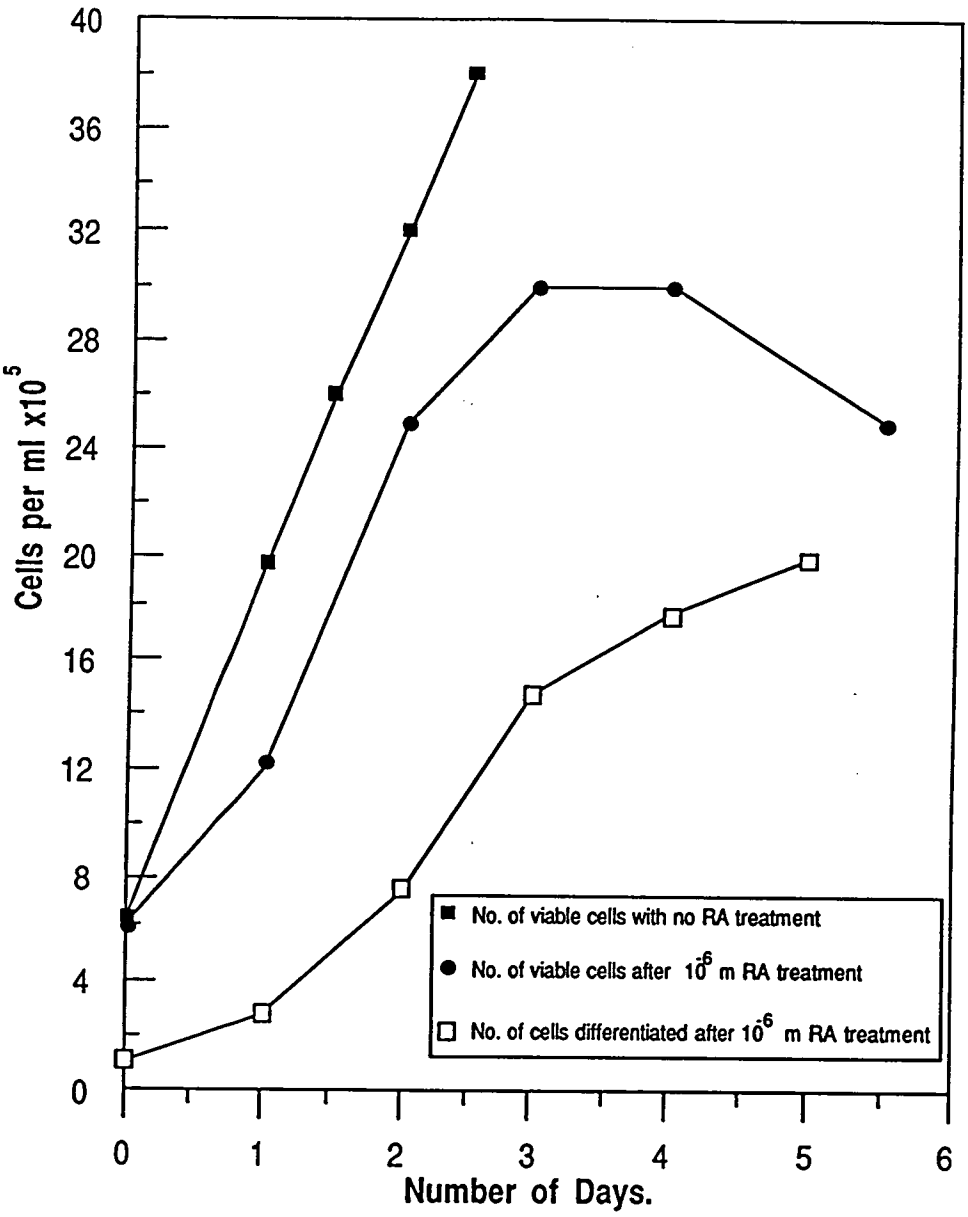
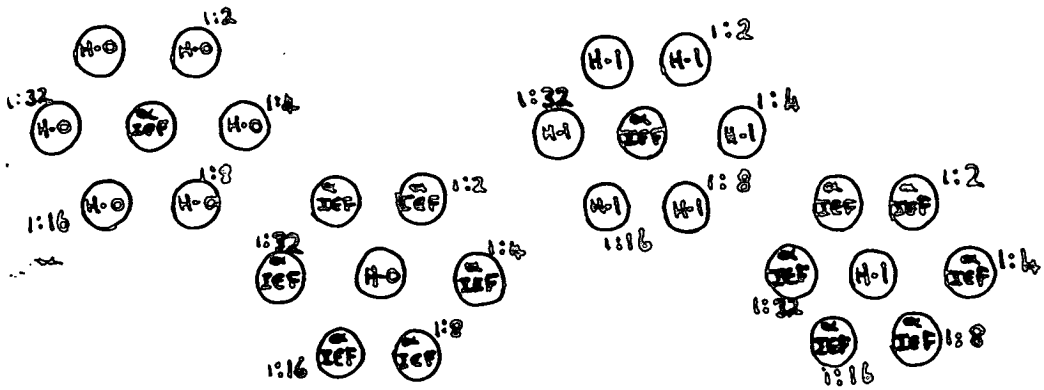


FIGURE 2

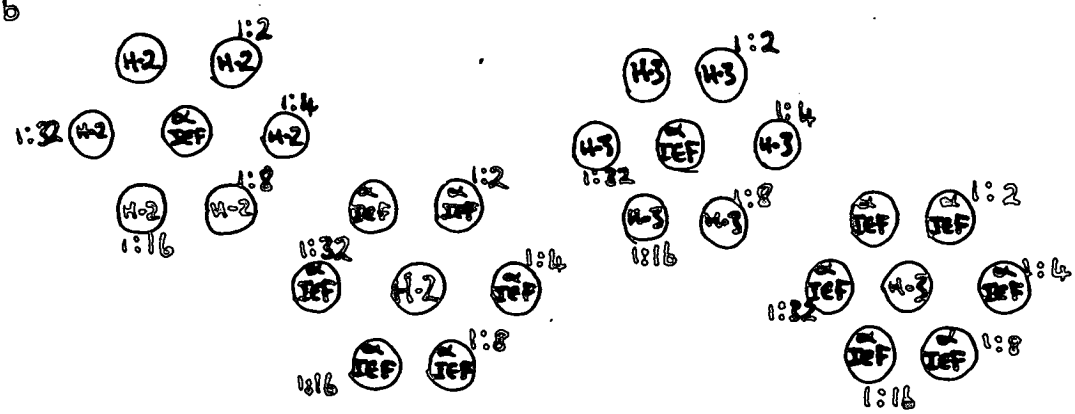
Graph to show the number of viable HL-60 cells and the number of differentiated cells on each day of a five day stimulation with 10^{-6}M RA. Differentiation was determined by NBT reduction, and cells were assessed as to their viability by exclusion of the dye nigrosin.



3a


H-0
H-1
1:2
1:4
1:8
1:16
1:32
REF

3b


H-2
H-3
1:2
1:4
1:8
1:16
1:32
REF

3c

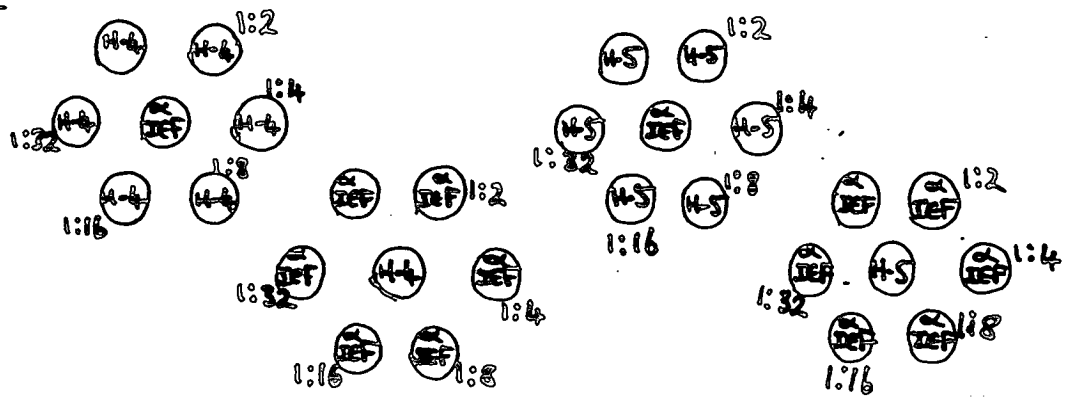
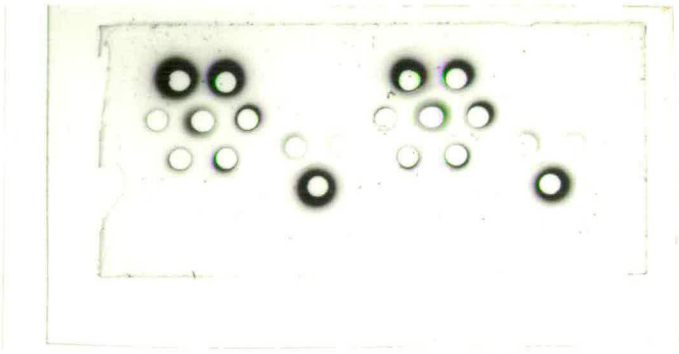

H-4
H-5
1:2
1:4
1:8
1:16
1:32
REF

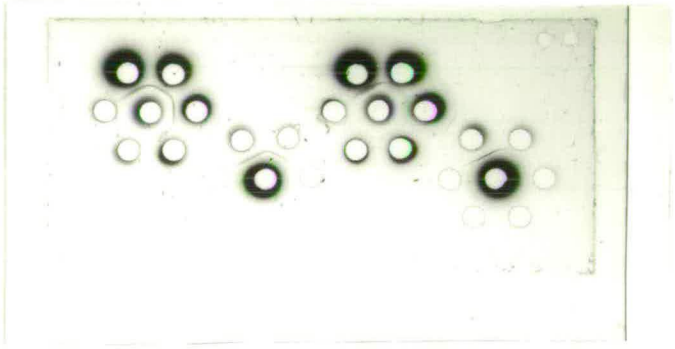
FIGURE 3 a,b,c

Figure 3 shows Ouchterlony double diffusion analysis of HL-60 cell lysates against anti-IEF antisera at various dilutions. The cell lysates are from cells grown in the presence of RA for variable lengths of time. The well contents and the dilution are shown adjacent to the wells. H stands for HL-60 and the number following indicates the length of time exposed to RA, and the lysate dilution. For example "H-0 1:2" is lysate from HL-60 (grown without RA) diluted 1 in 2. Equivalent numbers of cells are used to make each lysate. Gels are stained with Coomasie blue.

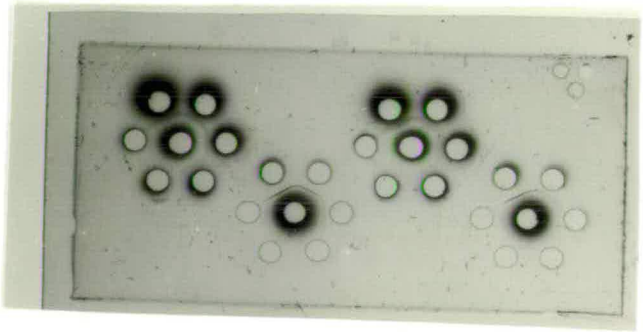
3a



3b



3c



3.3.b CFAg protein turnover in granulocytes from normals, CML patients and HL-60 cells.

In order to determine if CFAg was actively synthesised in CML and normal granulocytes, cell labelling with ^{35}S-methionine was carried out. ^{35}S-methionine will only be incorporated into proteins that are being synthesised i.e the mRNA for the protein is being translated and the protein assembled. In the case of CML samples buffy coats from 10mls of blood were used as >90% of the cells were myeloid. Lymphocytes were separated from red blood cells in normal samples over dextran 250 and PMN were separated from mononuclear cells by separation over Ficoll-hypaque. Cells were incubated for 1 hour without methionine and then for 4 hours with ^{35}S-methionine.

Ouchterlony double diffusion analysis

Cell lysates were tested after labelling, by Ouchterlony analysis and stained initially with Coomassie blue and then autoradiographed to see if any label had been incorporated into the precipitin band. Figure 4-a shows the immunodiffusion of anti-CML antiserum against internally ^{35}S-methionine labelled cell lysates of CML and normal granulocytes. Both cell sources appear to have incorporated the radiolabel into the precipitin line, indicating that there is active CFAg synthesis in both CML and normal granulocytes.

SDS polyacrylamide gel electrophoresis of bands excised from Ouchterlony gels.

A second step was to cut out the labelled precipitin band from the Ouchterlony gel and run the sample on a polyacrylamide gel to indicate subunit size. Figure 4-b shows the gel which had the bands in figure 4-a cut out and electrophoresed. The gel has CML total lysate and VvH normal PMN total lysate, the band patterns of which look similar but more complex in the CML lysate. Only the tracks which have CML lysate immunoprecipitates give detectable bands in the autoradiograph of the gel. In this case two bands are seen of ~14 kDa and 12.5 kDa. Both these bands appear to be

abundant proteins in the total labelled lysate samples of both CML and normal PMN samples.

The immunoprecipitated antigens are contaminated.

As a negative control lactoferrin and catalase antisera were used to in a parallel experiment. CML internally labelled lysate was immunoprecipitated in an Ouchterlony with lactoferrin, catalase and anti-CML antisera. The precipitin bands for all three antisera were cut out and run on a 12% SDS polyacrylamide gel. The results are shown in figure 5. Each antiserum has the same two radiolabelled bands. This means that it is not possible to identify a specific protein band for CFAg. The two protein species which are present in the precipitin bands irrelevant of the antisera used are major bands in the total lysate. It is thus due to their abundance and perhaps the nature of the proteins that they are trapped non specifically in the antigen/antibody lattice. These gels were washed for 48 hours in saline containing 0.05% Tween 20, but this was obviously insufficient.

3.3.c Immunoprecipitation of CFAg using McAb, from normal, CML granulocytes and induced HL-60.

HL-60 cells induced for 2 days with RA, uninduced HL-60, CML and normal granulocytes were all internally labelled for 4 hours with ^{35}S-methionine. Using monoclonal antibodies to alphafetoprotein (AFP) (van Heyningen et al 1983) as a negative control and lactoferrin (Chung et al 1985), as a positive control in the case of CML, a series of immunoprecipitations was carried out. The precipitations were washed extensively with buffer containing 0.1% SDS and β-mercaptoethanol. The results are shown in figure 6a, 6b and 6c. Lactoferrin monoclonal precipitates a protein of 79 kDa only from the CML lysate. CFAg monoclonal antibody CF46 precipitates two specific molecules of ~14 and 12.5 KDa from CML, normal PMN and induced HL-60. There is no specific molecule precipitated by CF46 from uninduced HL-60. These results are convincing because in the total HL-60 2 day RA lysate the

precipitated band is not a major species whereas it is in the granulocyte lysates. It is therefore less likely to be a contaminant purely because the protein is very abundant.

3.4 Conclusions

1. CFAg is synthesised in the promyelocytic cell line HL-60 when it is induced to differentiate towards granulocytes. With 10^{-6}M RA maximal antigen production is achieved after 2 days of continual exposure.

2. CFAg is actively synthesised in normal and CML granulocytes.

3. Monoclonal antibodies directed against CFAg, precipitate two molecules from labelled cell lysates of approximately 12 and 14 kDa.

4. Both these proteins appear to be abundant in the total cell lysates of ^{35}S-methionine labelled CML and normal granulocytes but poorly represented in induced HL-60 and not present in HL-60 grown in serum free medium. This agrees with the steady state CFAg protein data.

On the basis of these results it was decided that CML tissue contains the highest level of protein synthesis and therefore was used as the source of mRNA in the construction of a cDNA library.

FIGURE 4-a

Autoradiograph of an Ouchterlony double diffusion gel, where ^{35}S-methionine internally labelled CML and normal granulocytes, are allowed to diffuse with undiluted anti-IEF antiserum. Well 1 contains anti-IEF antiserum. Well 2 contains ^{35}S-methionine labelled CML lysate. Well 3 contains ^{35}S-methionine labelled normal granulocyte lysate.

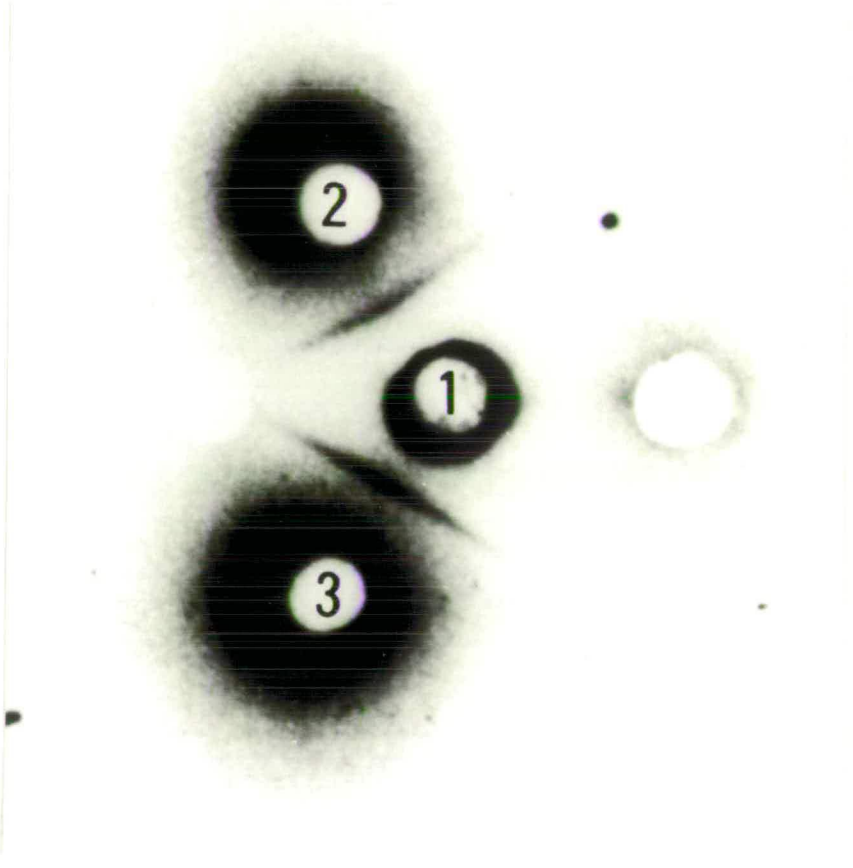


FIGURE 4-b

Autoradiograph of a 12% SDS PAGE separating labelled proteins present in Ouchterlony precipitin bands. lane 1 CML total lysate; lane 2 granulocyte total lysate; lane 3 excised Ouchterlony precipitin band between normal granulocytes and anti-CML antiserum; lane 4 excised Ouchterlony precipitin band between CML and anti-CML antiserum; lane 5 normal granulocyte labelled lysate; lane 6 excised Ouchterlony precipitin band between normal granulocytes and anti-IEF antiserum; lane 7 CML total lysate; lane 8 excised Ouchterlony precipitin band between CML lysate and anti-IEF antiserum.

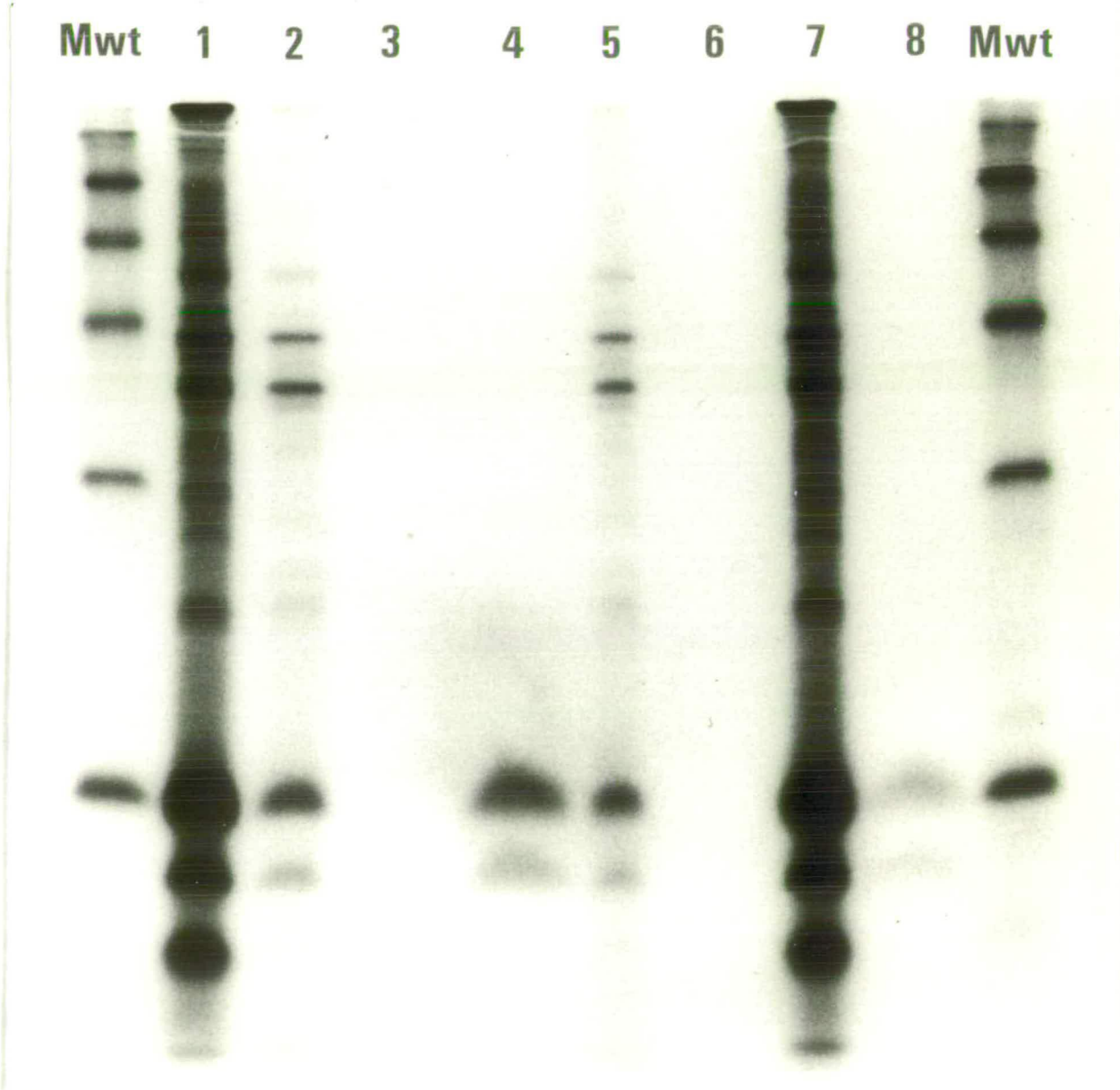
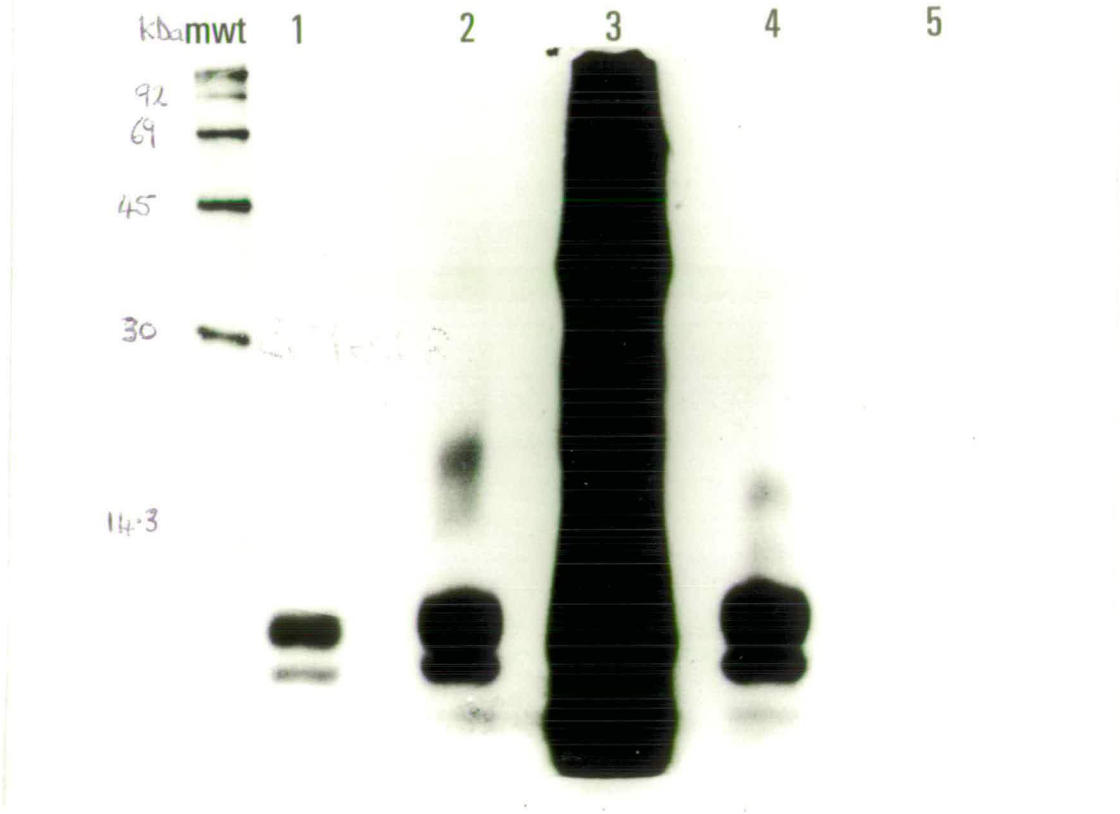


FIGURE 5

Autoradiograph of 12% SDS PAGE to show labelled proteins present in Ouchterlony precipitin bands between CML lysate and a variety of antisera.

lane 1- Excised Ouchterlony precipitin band between CML labelled lysate and catalase antiserum. lane 2-Excised Ouchterlony precipitin band between CML labelled lysate and anti-CML antiserum. lane 3-Excised Ouchterlony precipitin band between CML labelled lysate and lactoferrin antiserum. lane 4-Excised Ouchterlony precipitin band between CML labelled lysate and anti-IEF antiserum. lane 5- Excised Ouchterlony precipitin band between CML labelled lysate and anti-IEF antiserum.



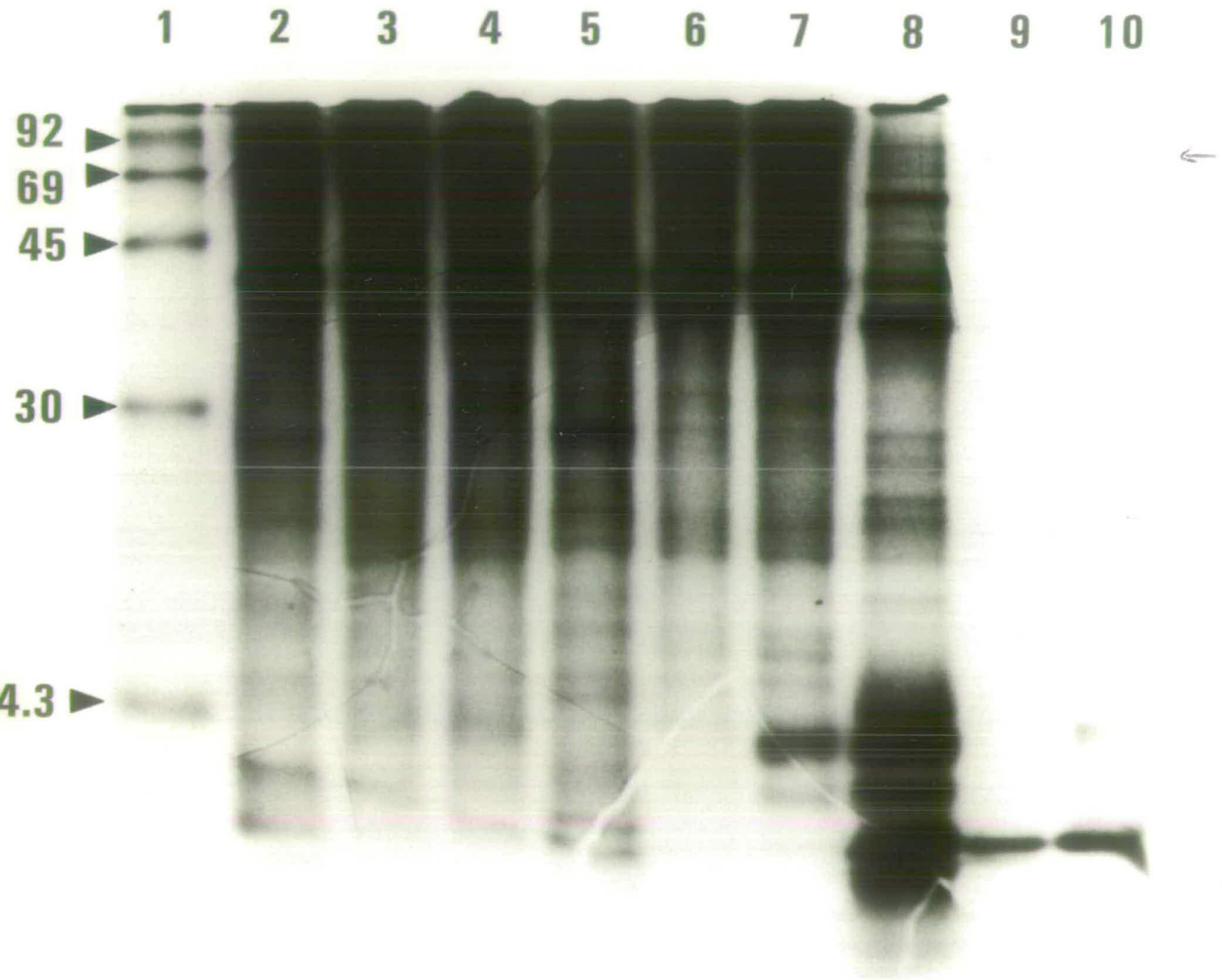

1
2
3
4
5
6
7
8
9
10
92
69
45
30
4.3

FIGURE 6 a-c

Autoradiograph of 15% SDS PAGE of immunoprecipitations from ^{35}S-methione internally labelled cell lysates. In each case 10^6 cpm were used as starting material in the immunoprecipitations, and 2 x 10^4 cpm were loaded straight onto the gel in the case of total lysates. Numbers next to arrows down the side of the gels indicate the position of the molecular weight markers and their size in kDa.

FIGURE 6-a

lane-1 molecular weight markers, with their size in kDa indicated; lane-2 total labelled lysate of HL-60 cells grown without serum; lane-3 as above immunoprecipitated with a pool of anti- lactoferrin monoclonal antibodies; lane-4 as in lane 2 immunoprecipitated with a pool of monoclonal antibodies against CFAg; lane-5 total labelled lysate of HL-60 cells grown 2 days in the presence of RA; lane-6 as lane 5 immunoprecipitated with a pool of anti-lactoferrin monoclonal antibodies; lane-7 as lane 5 immunoprecipitated with a pool of CFAg monoclonal antibodies; lane-8 total cell lysate from internally labelled CML cells; lane-9 as lane 8 immunoprecipitated with a pool of anti-lactoferrin monoclonal antibodies, the position of lactoferrin indicated with an arrow; lane-10 as lane 8 immunoprecipitated with a pool of CFAg monoclonal antibodies.

FIGURE 6-b

lane-1 total labelled lysate from normal peripheral blood granulocytes; lane-2 as lane 1 immunoprecipitated with anti-AFP monoclonal antibody; lane-3 as lane 1 immunoprecipitated with CF145 monoclonal antibody; lane-4 total labelled lysate from HL-60 cells grown 2 days RA; lane-5 as lane 4 immunoprecipitated with anti-AFP monoclonal antibody; lane-6 as lane 4 immunoprecipitated with CF145; lane-7 internally labelled uninduced HL-60 cell lysate; lane-8 as lane 7 immunoprecipitated with anti-AFP monoclonal antibody; lane-9 as lane 7 immunoprecipitated with CF145.

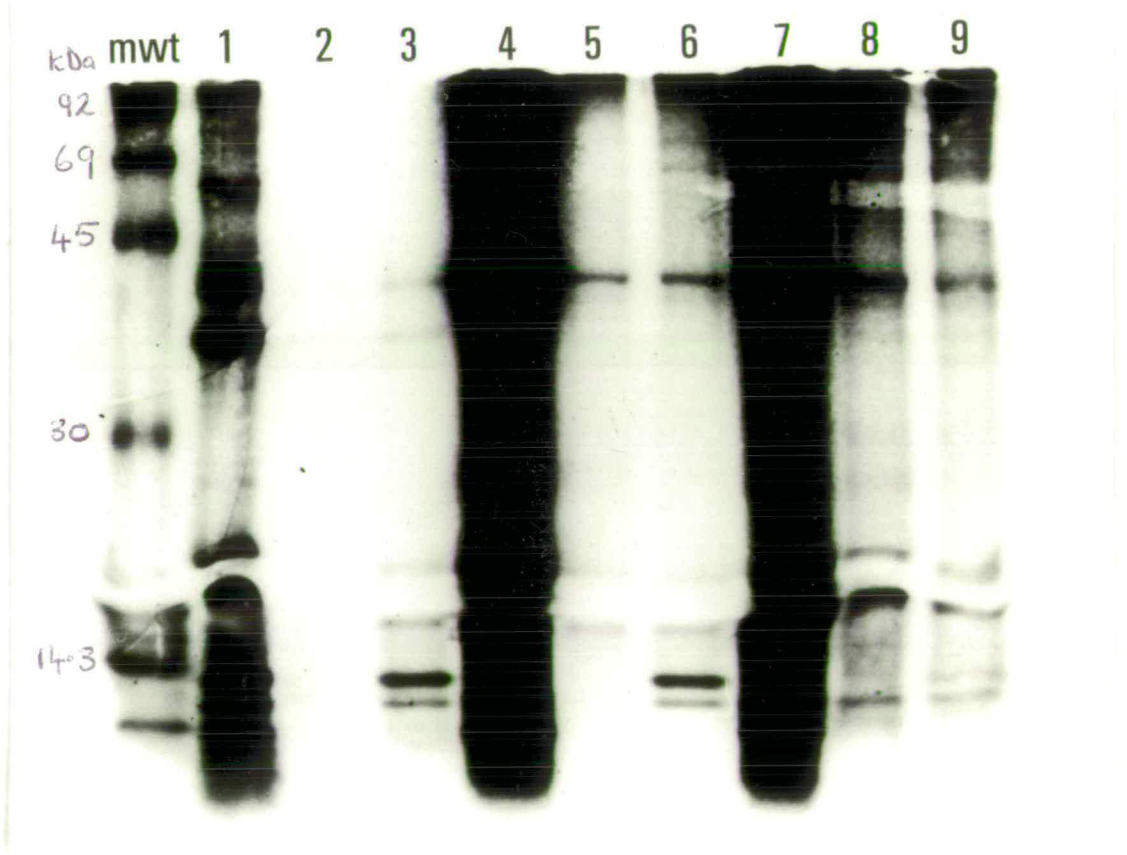
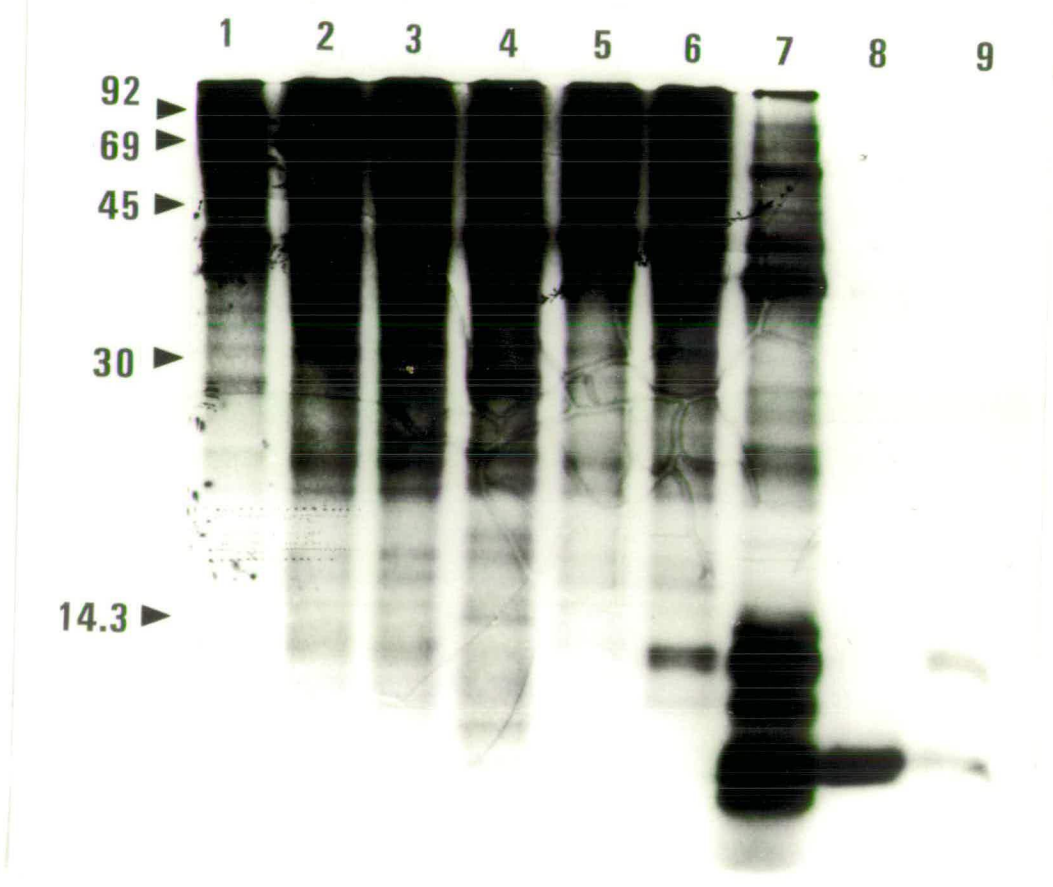


FIGURE 6-c

lane-1 lysate from intermally labelled uninduced HL-60 cells; lane-2 as lane 1 immunoprecipitated with anti-AFP monoclonal antibody; lane-3 as lane 1 immunoprecipitated with CF145; lane-4 internally labelled HL-60 induced with RA for 2 days; lane-5 as lane 4 immunoprecipitated with anti-AFP monoclonal antibody; lane-6 as lane 4 immunoprecipitated with CF 145; lane-7 labelled lysate from CML cells; lane-8 as lane 7 immunoprecipitated with anti-AFP monoclonal antibody; lane-9 as lane 7 immunoprecipitated with CF145.



CHAPTER 4

cDNA LIBRARY CONSTRUCTION

4.1 Introduction

In order to isolate the cDNA for CFAg it was decided to construct a cDNA library in λgt10 using polyadenylated RNA from CML cells as a starting material. The frequency at which cDNA clones of a particular mRNA species appear in a cDNA library is generally proportional to the abundance of that species in the mRNA population. It was assumed that as CFAg appeared to be an abundantly synthesised protein in CML cells then the mRNA would be abundant. This means that CFAg should be highly represented in a library made from CML mRNA and candidate clones could be isolated by differential screening -see Chapter 5.

A lambda vector was chosen instead of a plasmid because of the high efficiency and reproducibility of *in vitro* packaging as a method of transfecting *E.coli*. λgt10 was chosen because it has a selection system for recombinant phage which remain healthy and grow well to give large plaques. Murray et al (1977) describe a series of vectors which utilize the repressor CI gene to select for recombinants. In wild type lambda the CI gene product binds to the operator genes to suppress transcription of the genes necessary for lytic growth of the phage. Thus insertion into this gene yielding CI- phage will give clear plaques as all phage will undergo lytic growth. CI+ phage give turbid plaques as some phage will express the CI gene product and some will not. λgt10 does not require an insert fragment to yield packagable DNA and so without selection 70-90% of the cDNA library will be parent phage. When λgt10 is grown on

an *E.coli* strain carrying the high frequency lysogen mutation Hfl A150 (Hoyt et al 1982) the CI+ phage are repressed so efficiently that plaque formation is supressed. CI- phage however will plaque with normal efficiency on the Hfl A150 strain. λgt10 has a unique Eco R1 site in the CI gene, into which inserts are normally cloned. The phage can accept up to 7.6kb of insert assuming a packaging capacity of 105% wild type length (Huynh et al 1985)

4.2 Synthesising and cloning double-stranded cDNA

4.2.a The general methodology used is:

A. Extraction of RNA from CML cells.

B. Isolation of polyadenylated RNA

C. cDNA synthesis catalysed by reverse transcriptase using poly(A)$^+$ RNA as template and oligo $(dT)_{12-18}$ as a primer.(Efstratiadis et al 1976)

D. mRNA-cDNA hybrid is used as a substrate for RNase H, E.coli DNA polymerase I and E.coli DNA ligase in the presence of deoxyribonucleoside triphosphates. RNase H introduces nicks in the RNA to produce Okazaki fragment like primers which provides substrates for the replacement of RNA by DNA (mediated by Polymerase I). E.coli DNA ligase will seal the nicks but unlike T4 ligase, will not ligate double-stranded molecules to each other. This technique for second strand synthesis was introduced by Okayama and Berg(1982) and modified by Gubler and Hoffman(1983). Classical methods of cDNA manufacture, hydrolyse the RNA with alkali and prime the second strand from a hairpin structure found at the 3' end of cDNA molecules (Efstratiadis et al 1976). This is disadvantageous because the looping results in loss of sequence at this region.

E. The double stranded cDNA is blunt-ended by treatment with S1 nuclease. This is a poorly controlled step and the enzyme must be titrated before use to prevent degradation beyond the overlapping ends.

F. The cDNA internal Eco R1 sites are protected from subsequent cleavage by treatment with Eco R1 methylase.

G. Synthetic linkers with Eco R1 restriction sites are ligated onto the cDNA.

H. Cleaving with Eco R1 to remove all but the terminal unit of the linker and thus generating Eco R1 termini on the molecule.

I. Separation of cDNA from excess of linkers

J. Cloning of cDNA into Eco R1 cut vector.

4.2.b The practical problems encountered.

Isolation of mRNA

The initial problem encountered was isolation of mRNA from CML cells-the tissue of choice. From 10mls of leukaemic blood, with 100 times normal levels of

polymorphonuclear leukocytes it was possible to separate ~4 x 10^9 myeloid cells, however extraction of RNA using buffer saturated phenol with proteinase K digestion (Frazier et al 1983) or guanidium-HCl (Chirgwin et al 1979) proved very difficult and yields of RNA were very low. No RNA could be successfully extracted from normal polymorphonuclear leukocytes. If the same number of liver cells were used (~ 10^9), the yield would be at least 10mg of RNA. However the liver has several abundant mRNA species (Galau et al 1977). Granulocytes are terminally differentiated end cells designed to destroy invading organisms and therefore do not synthesise many proteins and are probably rich in endogenous ribonuclease. Finally, I received a leukophoresis sample from a CML patient and from 10^{11} cells using the guanidine-HCl method it was possible to extract 10mg of RNA. More than 90% of the cells were of myeloid morphology and so no cell separation was undertaken. The patient (CD) from whom the leukophoresis sample was taken, had been admitted to hospital with symptoms of fatigue and anaemia. The haemoglobin was 7.9g/dl, the white cells 55.9 x 10^9/l and platelets 54 x 10^9/l. She was regarded by the consultant haematologist as having a chronic myeloid leukaemia distinct from typical chronic granulocytic leukaemia and chromosome analysis of the bone marrow was normal i.e no Philadelphia chromosome was seen. She had been treated with cytotoxic drugs with minor haematological improvement. The patient has since died of an intracerebral haemorrhage. The technique used to isolate RNA from this patient (CD) was used subsequently to isolate RNA from another leukophoresis sample (BS) who was Philadelphia chromosome positive.

Enrichment for polyadenylated RNA

The isolated RNA was then enriched for poly $(A)^+$ RNA over an oligo(dT) column and yielded 400 μg from 4mgs.

The mRNA was electrophoresed on a 1% agarose/2.2Mformaldehyde gel and probed with a catalase cDNA probe (van Heyningen et al 1985a) to show that the

mRNA was intact (Figure 7). This Northern blot also includes poly $(A)^+$ samples from normal human liver, a liver hepatoma cell line (Hep G2), CD CML, and induced and uninduced HL-60 cells. The HL-60 cell mRNA is somewhat degraded, but the CML mRNA appears to be intact.

The CML poly $(A)^+$ RNA was then translated in a cell free rabbit reticulocyte lysate translation system and immunoprecipitated with monoclonal antibodies (McAb) to lactoferrin, AFP or CFAg. mRNA for lactoferrin is known to be present in the cells of CML patients (Rado et al 1984) but AFP mRNA will not be present. Figure 8 shows that lactoferrin McAb immunoprecipitates a specific band from the translated, ^{35}S-methionine labelled products. The AFP McAb does not precipitate a specific radiolabelled band, and CF46 immunoprecipitates a band assumed to be CFAg.

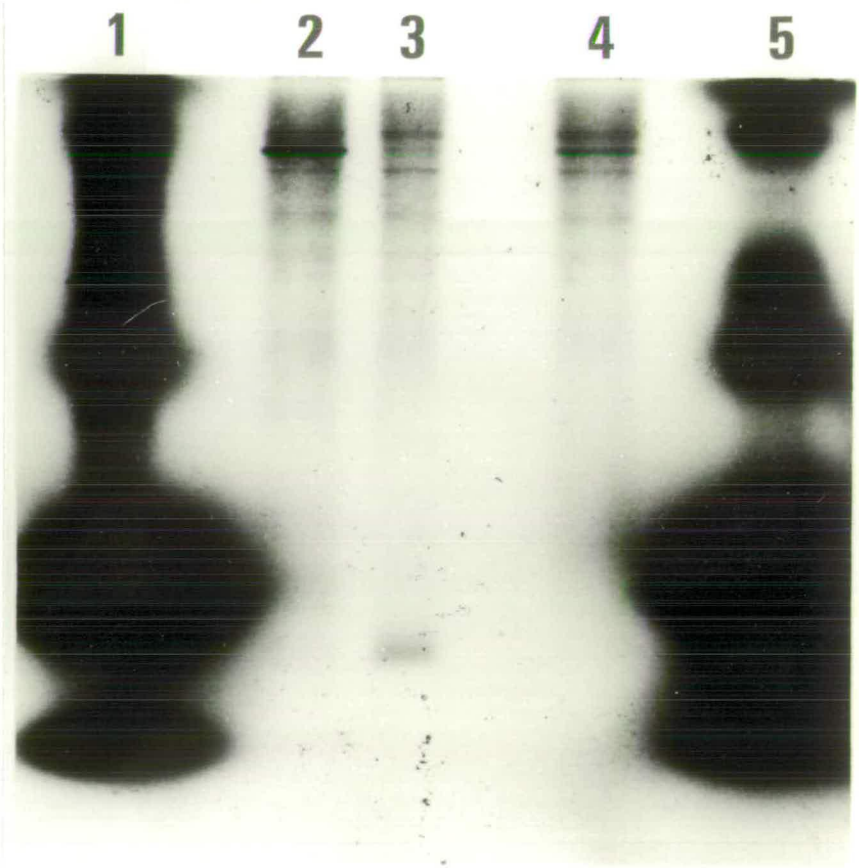
Thus, the mRNA prepared from CML cells is intact and contains mRNA which can be translated in a cell free system and from the translation products a specific protein can be immunoprecipitated by McAb directed against CFAg.

FIGURE 7

Autoradiograph of northern blot probed with human catalase cDNA clone to show that CML CD mRNA is intact. Each lane contains 2μg poly(A)$^+$ from various sources. Lane-1 uninduced HL-60; lane-2 HL-60 induced 2 days RA; lane-3 CML; lane-4 HEPG-2 (hepatoma cell line); lane-5 human brain lane-6 human liver.

FIGURE 8

Autoradiograph of 12% SDS polyacrylamide gel to demonstrate immunoprecipitation of ^{35}S-methionine labelled CFAg and lactoferrin, but not AFP from rabbit reticulocyte translation products of CML poly(A)$^+$ RNA. 10^6 cpm were used as starting material in each immunoprecipitation and 50,000 cpm were loaded in lanes 1 and 5. lane-1 total translated products from CML RNA; lane-2 as lane 1 immunoprecipitated with anti-lactoferrin McAb; lane-3 as lane 1 immunoprecipitaed with anti-CFAg McAb 145 ; lane-4 as lane 1 immunoprecipitated with anti-AFP McAb; lane-5 Rabbit reticulocyte translation system, no mRNA added. Lanes 1 and 5 are overexposed to reveal the bands in lanes 2, 3 and 4.



Double stranded cDNA synthesis, and S1 nuclease treatment

The cDNA library was constructed as described in the methods section 2.21. The synthesis of cDNA was traced by the incorporation of (α^{32}P)TTP.

The average size of first strand products was determined by running 1/10 of the sample on a 1.5% agarose gel and autoradiographing the dried down gel. In initial experiments the first strand synthesis showed cDNA of low average size - less than 1Kb. The reverse transcriptase (RT) used in this case was from Boehringer Mannheim and so in an effort to improve the size of cDNA molecules, reverse transcriptase from three sources were compared, in an experiment where 2 µg of CML poly $(A)^+$ RNA was used as a template for each of the three enzymes. The NBL Enzymes Ltd reverse transcriptase and Amersham reverse transcriptase required 60 units of enzyme where the NBL XL (a higher grade) needed only 20 units as it has 3 fold greater specific activity. The percentage of radiolabel incorporated in each case was the same, (~8%). It was found that Amersham reverse transcriptase gave poor cDNA size compared to NBL Enzymes' regular and XL RTs. The XL preparation was very concentrated compared to the other two samples and gave cDNA molecules of high average size as shown in figure 9.

It was noticed, that in some instances within the first strand cDNA products discrete bands could be visualised. These were probably due to self primed cDNA synthesis from DNA and/or ribosomal RNA contaminating the preparation. This phenomenon was particular to certain poly $(A)^+$ RNA preparations.

The results below are for the cDNA used to make the library. Using 10µg CML poly$(A)^+$RNA as starting material.

1st strand- % of label incorporated = 5%

Amount of unlabelled dTTP = 53nmoles

Amount of unlabelled dTTP incorporated = 5% of 53 nmoles

= 2.65nmoles

assume average Mwt of dNTP = 350

Weight of cDNA = 927ng

Therefore % of mRNA transcribed = (927/10,000)x100 = 9.27%

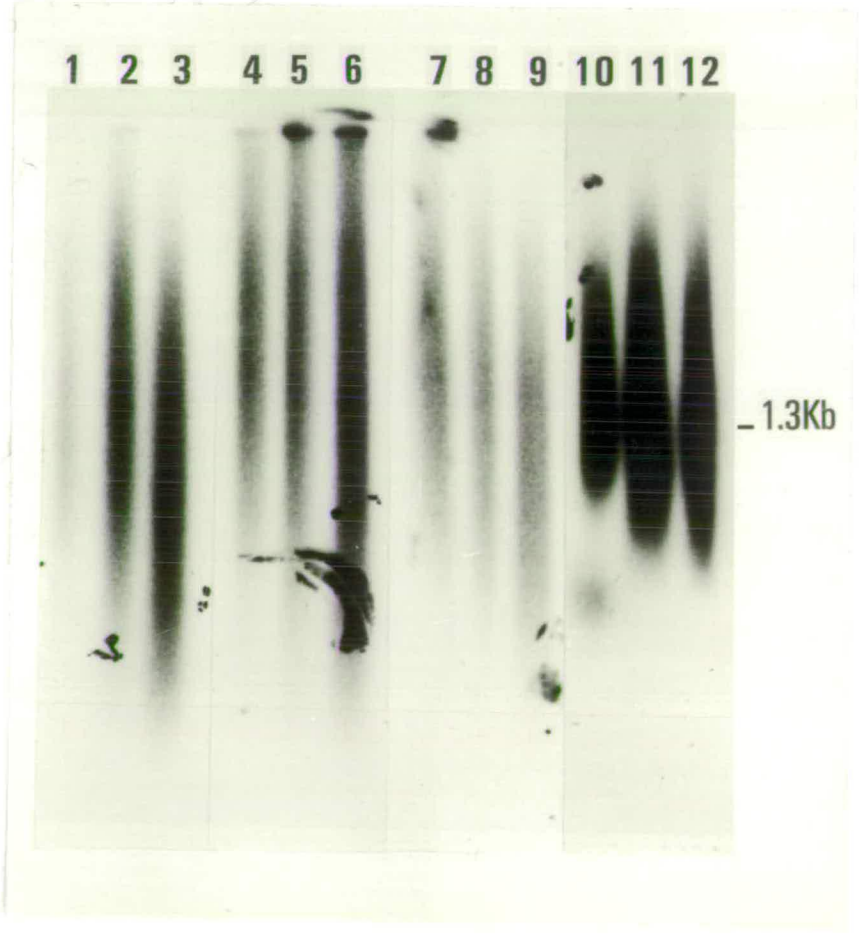
Assuming that at least 90% of 1st strand becomes second strand, total amount of cDNA = 1.8μg.

Figure 9 also shows the cDNA used to make the cDNA library. First strand has average sized species of ~ 1.5kb. The second strand is of the same size average as first strand and S1 nuclease treatment reduces the average size very slightly. An alternative method to using S1 nuclease to blunt end the cDNA, is to use T_4 Polymerase. T_4 polymerase will "fill in" the ends of the cDNA and so the problem of over digestion with S1 nuclease can be obviated. However as the S1 nuclease appeared to cause no difficulties, this step was not altered.

FIGURE 9

Autoradiograph of dried down non denaturing 1.5% agarose gel, containing samples of cDNA synthesised with three different reverse transcriptases. In each case, 2µg of CML poly(A)$^+$ was used as starting material. 50,000 cpm from each reaction were saved from the first strand, second strand and S1 stage and run on the gel. The size markers were Hind III cut lambda DNA and Hae III cut ϕx174. Lanes 1 to 3 uses NBL regular reverse transcriptase, lanes 4 to 6 uses NBL XL reverse transcriptase and lanes 7 to 9 uses Amersham RT. lanes 1,4,7 are first strands, 2, 5, 8 are second strand and lanes 3, 6, 9 are after the S1 nuclease treatment.

Lanes 10, 11, 12, first strand, second strand and S1 nuclease treated double stranded cDNA used to construct the CML λgt10 library. 10µg of poly(A)$^+$ RNA was used as starting material and NBL XL RT was used to synthesise the cDNA. The size markers are Hae III cut ϕx174.



Methylation and addition of Eco R1 linkers

The cDNA methylation step caused no problems but the addition of *Eco R1* linkers proved extremely difficult. The methylation of the cDNA was not checked directly but the *E.Coli methylase* was shown to be active before being used in the reaction. This was done by treating purified lambda DNA with the enzyme and then demonstrating that the it could no longer be restricted with *Eco R1*. In order to follow the ligation of the linkers onto the cDNA, the linkers were end-labelled with (γ^{32}P) ATP and polynucleotide kinase. It was found that polynucleotide kinase was extremely unstable and after being used once or twice it would be inactive. To ensure that the linkers could ligate to themselves, a control ligation reaction was carried out without the addition of the cDNA. It was possible to see quite easily when the linkers had ligated to themselves because the unlinked linkers at the bottom of a 1.5% agarose gel disappeared and a smear up the gel to about 600bp size was seen. Because double insertions are undesirable, a molar excess of linkers to cDNA was used. This means that there should still be concatamers of linkers seen in a smear at the bottom of the gel and an increase of label in the cDNA when ligation of linkers and cDNA has successfully occurred. This would indicate that some linkers have ligated to themselves and some have ligated to the cDNA. It was difficult to detect appreciable increase in label in the cDNA but the smear of self ligated linkers and the diminished amount of free linkers should have been easily visualised.

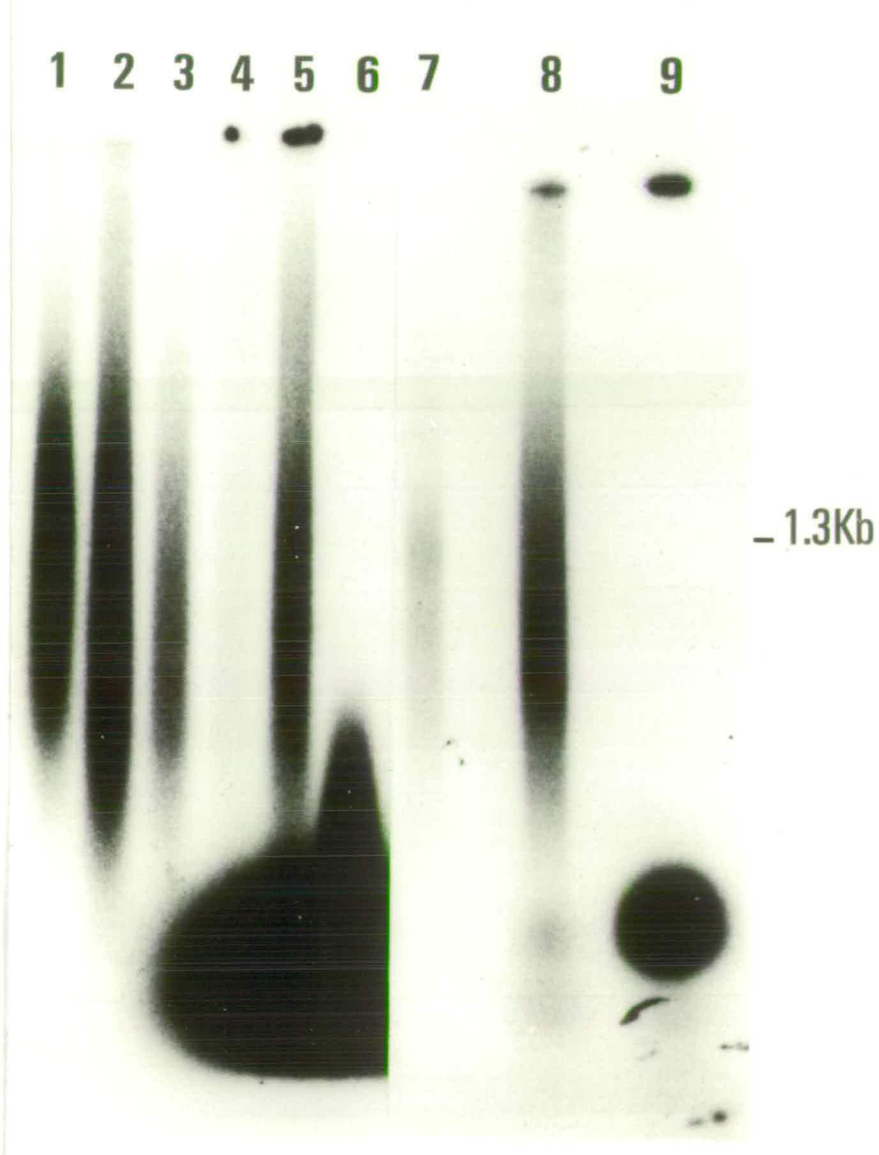
It was found that the ligation of the linkers would procede satisfactorily until cDNA was added to the reaction, when no reduction in the unligated linkers was seen and no smear of self ligated linkers. Sodium acetate was used to precipitate the cDNA after each step in case the ammonium ions from the ammonium acetate were inhibiting ligation. This appeared to make no difference. 1μg of lambda DNA cut with *Eco R1* and included in the ligation reaction without cDNA did not inhibit the ligation of the linkers to themselves. Precipitating the lambda DNA from each of the cDNA reaction

mixes and then including the DNA in the ligation reaction did not inhibit the reaction of the linkers ligating to themselves. 1mM spermidine was included in the reaction mix and appeared to reduce ligation of the linkers even in the control reaction without the cDNA. According to Pheiffer and Zimmerman (1983), polymers such as polyethelyene glycol (PEG 6000), Ficoll, bovine serum albumin or glycogen can stimulate blunt-end ligation by a factor of up to ten. A reaction buffer including PEG 6000 proved to give extremely efficient ligation of the linkers to themselves without the cDNA, but did not allow the reaction to proceed with the cDNA present. If there is some sort of inhibitor present then perhaps one would expect the polymer effectively to concentrate it in the reaction volume. It had to be concluded that there was something contaminating the cDNA, which was inhibiting the ligation reaction. In an effort to remove this inhibitor, a 200µl DE 52 column was made and used in an attempt to clean up the cDNA preparation. The column is anionic and so it will bind negatively charged molecules. In this way the cDNA can be bound to the column and washed free of any contaminants. The cDNA was applied to the column in 0.1M NaCl in TE (pH 7.6), washed with 0.3M NaCl in TE and eluted in 0.6M NaCl/TE. The cDNA was then precipitated twice with ethanol and added to the ligation reaction. In fact the first time this step was used to try and eliminate any inhibitors it coincided with the making of completly fresh buffers and the linkers appeared to link very efficiently to the cDNA. However on subsequent attempts at least half of the counts of the cDNA stayed on the DE52 column and could not be eluted even with 3M NaCl. Even with the inclusion of the column step in the method, the ligation reaction still did not work routinely. Figure 10 shows a typical example of experiments where the ligation reaction was unsuccessful and then finally successful.

FIGURE 10

Autoradiograph of dried down 1.5% agarose gel with an example of an unsuccessful ligation. Using 2 day RA induced HL-60 poly(A)$^+$ as the template. lane-1 first strand cDNA; lane-2 double stranded cDNA; lane-3 S1 nuclease treated double stranded cDNA; lane-4 cDNA plus linker ligation reaction mix with 1mM spermidine included in the reaction. lane-5 cDNA plus linker ligation reaction; lane-6 linker reaction no cDNA;

CML mRNA successfully linked to *Eco R1* linkers. lane-7 blunt ended double stranded cDNA; lane-8 cDNA plus linker ligation reaction; lane-9 unligated linkers.



Cleavage of the synthetic linkers

The cDNA was then cut with *Eco R1* and there was a slight reduction of size of the cDNA indicating that the enzyme had successfully cut.

Separation of the linker DNA from cDNA

The unlinked linkers were partially removed from the cDNA by ethanol precipitation which favours the larger cDNA molecules over the oligomers. The cDNA was then run on a 1.5% low melting point agarose gel and the DNA molecules above ~300bp were excised. The cDNA was extracted from the gel by excising the gel fragment and removing the cDNA by electroelution.

Ligation of cDNA to lambda DNA

As it was unclear how much cDNA was remaining after all the sampling, the whole amount was ligated to 1μg of lambda DNA which was obtained from Promega already cut with *Eco R1*. The lambda DNA was also ligated in parallel with the control insert provided by the company in parallel and then both reactions were packaged using Gigapak (Vector Cloning Systems). The test insert ligated to λgt10 produced ~2x 10^6 plaques per μg λ gt10 on the bacterial strain selecting for inserts (C600 hfl) and ~10^7 plaques on the permissive strain C600.

The cDNA ligated to λgt 10, when plated out gave a library on the C600Hfl strain, of 40,000 recombinants, and subsequently a second library of 70,000 recombinants was obtained by the same techniques. Phage from the primary plating on C600Hfl were collected in SM buffer and 4ml aliquots were stored at 4^oC over chloroform.

4.3 Characterization of λgt10 CML cDNA library.

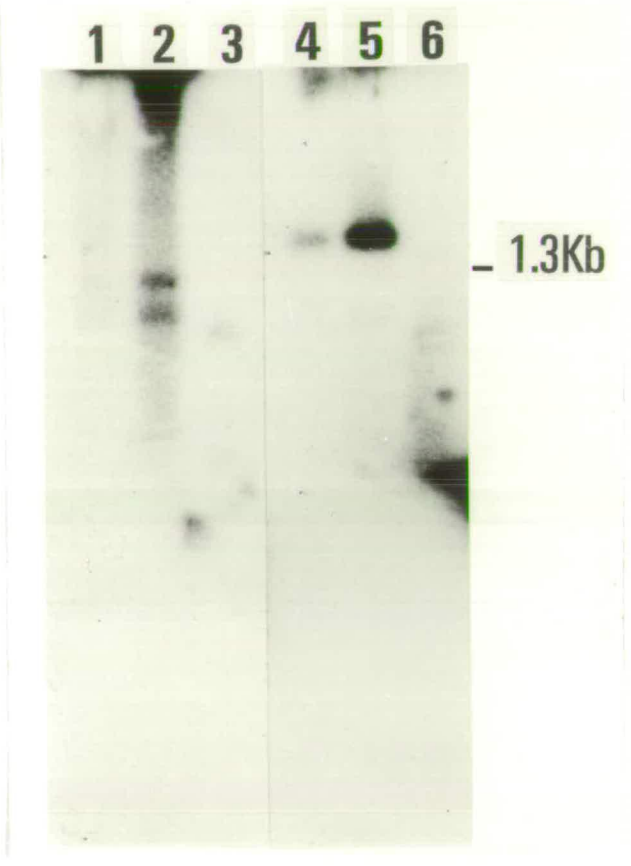
The library was plated out (~5,000 pfu) and screened with the plasmid pBR322 in order to check that there was no plasmid contamination. The autoradiograph of the filter probed with the nick translated plasmid was negative, indicating that the library

was not contaminated with plasmid DNA.

In order to ascertain whether the library contained cDNA inserts corresponding to expressed genes , lambda DNA was isolated from a clear plate lysate representing the whole library. The DNA was then analysed by Southern blotting using a human catalase cDNA probe (van Heyningen et al 1985a), or a mouse class I cDNA probe (gift of P.Brown). Figure 11 shows DNA from a human liver cDNA library (van Heyningen et al 1985a) and from λgt10 CML library side by side, probed with the two cDNA probes. The class I probe hybridises to two major bands in the CML library DNA and to one slightly smaller band in the liver library DNA. The catalase cDNA probe hybridises to one band only in both libraries. This indicates that the CML cDNA library has cDNA inserts corresponding to mRNA transcripts for catalase and class I genes. Analysing the DNA from the CML cDNA library and the human liver cDNA library with a human albumin probe did not reveal a hybridising band in the first case but did in the latter (data not shown).

FIGURE 11

Autoradiograph of DNA from CML cDNA and human liver cDNA libraries probed with human catalase cDNA probe (lanes 1,2,3) or human HLA class I cDNA probe (lanes 4,5,6). lanes 1 and 2 ~ 1µg CML cDNA library DNA cut with Eco R1 to release insert DNA; lane 3 human liver cDNA library DNA cut with Pst I to release insert DNA. lanes 4 and 5 as 1 and 2; lane 6 as lane 3.



4.4 Construction of a λgt11 cDNA library

As antiserum and McAb to CFAg were available, it was decided to make a cDNA library from CML cells in an expression vector.

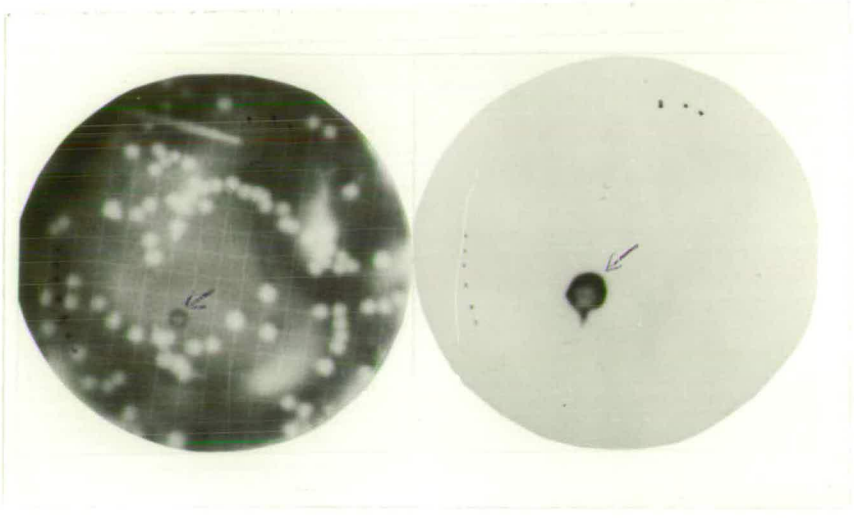
The vector of choice was λgt11. This vector can produce a hybrid polypeptide consisting of the β-galactosidase gene product and the insert gene coded polypeptide (Young et al 1983). This enables the library to be screened with antibodies directed against the protein and positive plaques can be identified using a radiolabelled or enzyme linked second antibody (Helfman et al 1983, Young and Davis 1983). Before the library was constructed the screening procedure was tested. Using an antiserum raised in rabbits against human albumin, a human liver library in λgt 11 was screened according to section 2.26. Positive plaques were identified using I^{125} radiolabelled protein A. Positive plaques were found at a frequency of ~5%, and duplicate filters screened with a human cDNA albumin probe, showed that all the antibody positive plaques hybridised to the albumin probe, but not all the plaques hybridising to the probe were positive with the antibody (figure 12). This is as expected because some of the recombinant phage may have inserts which code for some of the albumin protein but not enough to enable the antibody to recognise the protein. However a high percentage (80%) of the plaques hybridising to the probe also gave positive signal with the antiserum. However only 1 in 6 of the albumin cDNA inserts will by chance be in the correct frame and orientation to produce albumin protein. The high number of clones which react both with the antibody and the probe is probably a statistical artefact due to the fact that only 6 albumin clones were present in the screen.

It is to be expected that antiserum would be efficient at recognising native protein which has not necessarily been folded into its correct configuration as antiserum is a mix of antibodies directed to different immunogenic sites of the antigen. If McAb were to be used the success of the experiment would depend on whether the epitope on the protein to which the McAb was directed could still be recognised in the fusion protein.

However antiserum to CFAg was available as well as McAb to two separate epitopes (Hayward et al 1986) so success seemed probable if a suitable expression library was made.

FIGURE 12

Autoradiograph of a set of replica filters from λgt11 human liver cDNA library probed with (on the left) rabbit anti-human albumin serum followed by ^{125}I-protein A, or (on the right) with a human albumin cDNA probe labelled by nick translation. The filters are aligned so that their orientations are the same, and positive colonies are arrowed.



Lambda gt11 contains a unique Eco R1 site within the lacZ gene. This gene codes normally for β-galactosidase and the *Eco R1* site is 53bp upstream from the translation termination codon. Foreign DNA sequences inserted into this site can then be expressed as hybrid proteins with β-galactosidase. The phage have no selection for recombinants but there is an indication system. The β-galactosidase will produce blue coloured plaques when plated on the chromogenic substrate 5-Bromo-4-chloro-3-indoly-b-D-galactopyranoside (xgal). When the gene has an insertion the β-galactosidase can no longer function and so white plaques are produced. So that the proteins coded for by the foreign DNA do not harm the growth of the *E.coli,* the host bacteria contain a plasmid with the lac repressor gene (lacI) which prevents transcription of the lacz and hence the inserted gene until induced by the inactivation of the repressor by isopropyl -D-thiogalactopyranoside (IPTG). In order to prevent the bacteria destroying the antigen, the phage are grown in protease deficient bacteria (lon-) (Charnay et al 1980). The phage also carry a temperature sensitive CI gene which facilitates the growth of phage as lysogens at 32°C and as lytic phage at 42° C. The recombinant library is initially plated out on Y1089 which is hsdR- and hsdM-, and carries the plasmid borne lac repressor.

The construction of λgt11 library was carried out at the same time as the second λgt10 library was made. The protocol is identical for the two phage except that there is no selection for recombinants in the λgt11 system so a molar excess of insert DNA is used in the ligation of the phage to the cDNA. Alkaline phosphatase treated λgt11 arms were purchased from Promega Biotech and the ligation was done in 2μl volume. The control ligation with no insert DNA gave 80% of plaques (all blue) obtained with non alkaline phosphatase treated, *Eco R1* cut lambda gt11. When the library was plated out, 50% were white plaques on x-gal and IPTG and assumed to be recombinant phage. I then amplified the library as for λgt10 and replated out on Y1089 with x-gal and IPTG. The new ratio of recombinants to non-recombinants was 1:300. The non-

recombinants had obviously outgrown the recombinants in the course of amplification and hence the drop in number. A better strategy would have been to screen the library straight away, as this first plating out is the most representative. It was decided not to attempt to make another expression library, but to isolate possible myeloid specific clones from the λgt10 library and then reclone those inserts into λgt11 and see if any were positive with the antiserum raised against the band cut out of the IEF gel.

CHAPTER 5

Attempt to Isolate CML Specific Clones.

5.1 General Strategy

Having constructed a library which should contain the cDNA for CFAg the next step was to isolate those clones. At this time no protein sequence data was available. The strategy to isolate putative clones was as follows:

a) Isolate CML specific clones by differential screening. Confirm their specificity by Northern analysis of mRNA from CML and other tissues.

b) Of these clones identify which ones will hybridise to mRNA isolated from HL-60 induced with RA but not to mRNA from HL-60 grown on serum free medium.

c) Of these clones, it should be relatively easy to recognise which ones map to chromosome 1 due to the availability of chromosome 1 containing somatic cell hybrids.

d) Any clones which are positive by all these criteria can be further tested to see if they will hybridise to mRNA extracted from CFAg expressing somatic cell hybrids (van Heyningen et al 1985) and not to mRNA from a non-CFAg expressing hybrid.

e) As a definitive test, the positive clones could be recloned into a suitable expression vector so that the proteins for which the cDNAs code could be tested for reactivity with the anti-IEF antiserum.

5.2 Differential Screening

The technique of differential screening has been used successfully in a number of cases. Dworkin and Dawid (1980), found that 20% of the clones from a Xenopus laevis cDNA library contained inserts that could give detectable signal when screened with ^{32}P cDNA synthesised from homologous DNA. Clones which gave no detectable

autoradiographic signal contained sequences comprising less than 0.06% of the radioactive cDNA while positive clones contained sequences homologous to increasingly larger fractions of of the probe, correlating to increasing autoradiographic signal intensity. Thus, for a specific mRNA population, only abundant or moderately abundant sequences will be represented highly enough in the cDNA probe to carry enough signal to be effective in hybridisation. From the relative amount of CFAg radiolabelled in internally labelled cells it was thought that CFAg would be of at least moderate abundance in CML mRNA.

In theory the cDNA library is plated out and replica filters are screened with a cDNA probe homologous to the mRNA used to construct the library and a cDNA probe synthesised from mRNA of a different tissue or stage in development. Plaques which hybridize only to the cDNA from one tissue type must be either expressed only in that tissue or else more abundantlly expressed in that tissue compared with the other. There is some evidence that the more abundant sequences may represent tissue specific sequences (Hastie and Bishop 1976). Dworkin and Dawid (1980) and Williams and Lloyd (1979) agree that a 5 fold difference in expression can be detected by this technique. Williams and Lloyd (1979), successfully used this technique to isolate developmentally regulated mRNAs from the slime mould Dictyostelium discoideum and Rebagliati et al (1985) localized maternal mRNAs from xenopus eggs.

Axel and Lemke (1985) isolated the cDNA for the rat Schwann cell glycoprotein Po using the technique of differential screening and hybrid selection. They constructed a λgt10 cDNA library from poly(A)$^+$ RNA isolated from the sciatic nerve of 8-10 day rats. The library was then screened in duplicate with ^{32}P-labelled cDNAs from sciatic nerve and brain. They report that 2% of the cDNA clones hybridised only to the sciatic nerve cDNA, but 70% of these were found subsequently to cross hybridise. However one clone was completely absent in brain tissue and highly expressed in Schwann cells of the sciatic nerves. Thus the number of differentially expressed clones which

were detected indicates that a less abundantly expressed clone would be much more elusive.

5.3 Application of the differential screening technique to the λgt10 CML cDNA library

1. Screening CML cDNA library with cDNA from various tissues in an attempt to isolate CML specific clones.

It was decided to screen replica filters from the CML cDNA library with labelled cDNA complementary to mRNA from a B lymphoblastoid cell line (Daudi) (Klein et al 1968); mRNA isolated from uninduced HL-60 and to the CML mRNA which was used for the construction of the library. Both the former tissues should have had some mRNA species for housekeeping genes in common with CML, and the HL-60 cDNA should contain some myeloid specific sequences. The CML cDNA however should contain sequences which are specific to that tissue type and so it was expected to hybridise uniquely to some plaques.

RESULTS

The library was plated out on large 20x20cm plates at approximately 5,000 plaques per plate. 3 replica filters were taken as described by Benton and Davis (1977). Poly-adenylated RNA was isolated from the tissues to be used in each case. The cDNA used to screen the filters was made essentially in the same way that 1st strand cDNA was made for constructing the library. The developing strand of complementary DNA was labelled by the incorporation of (α^{32}P) dTTP. The RNA-DNA hybrid was then cleaved with alkali treatment and approximately equal number of counts was used for the cDNAs being compared.

Three plates were screened and no differences between any of the filters could be identified. The CML cDNA hybridised to 5% of the clones giving a very strong signal, 2% of the clones gave a fainter signal. Every clone which hybridised to the CML

cDNA also hybridised to both the Daudi cDNA and the HL-60 cDNA. Furthermore those clones which were less intense with the CML cDNA were also less intense with the other two cDNAs. At this point it was thought that perhaps the similarity of the tissue types was resulting in all the highly expressed transcripts being present in all three tissues. Consequently the next step was to try to isolate sequences that were present in CML and not present in liver mRNA.

2. Screening with liver cDNA versus CML cDNA

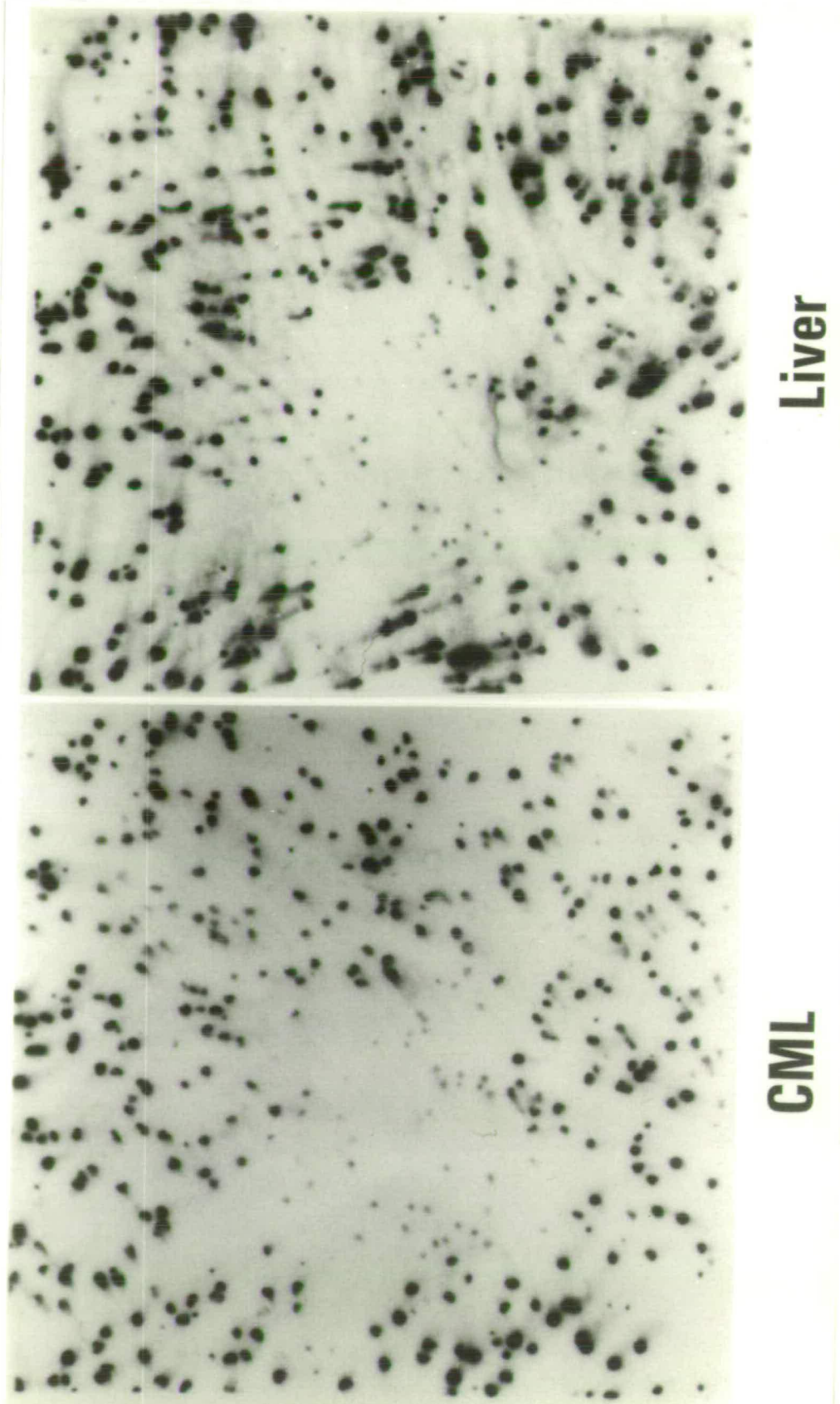
Liver poly(A)$^+$ RNA contains highly expressed tissue specific mRNAs, (Galiu et al 1977) and therefore liver cDNA should contain a different pattern of abundant molecules from other tissues. The CML cDNA was rescreened with cDNA from CML poly(A)$^+$ and cDNA made to normal liver poly(A)$^+$.

RESULTS

The results were the same as before in that each plaque that was positive with the CML cDNA was also positive with the liver cDNA (see figure 13).

FIGURE 13

Overnight autoradiographs of replica nitrocellulose filters taken from CML cDNA library, probed with CML cDNA or liver cDNA. Each 20x20cm plate contained ~5,000 plaque forming units (pfu). No differences could be observed between the two autoradiographs even after 1 week exposure. The filters are arranged in the same orientation.



3. Screening fewer plaques per plate

In the previous experiments large numbers of phage were being screened to detect differences in expression i.e 50,000 recombinant phage were screened distributed over 10 plates and no differences were detected.

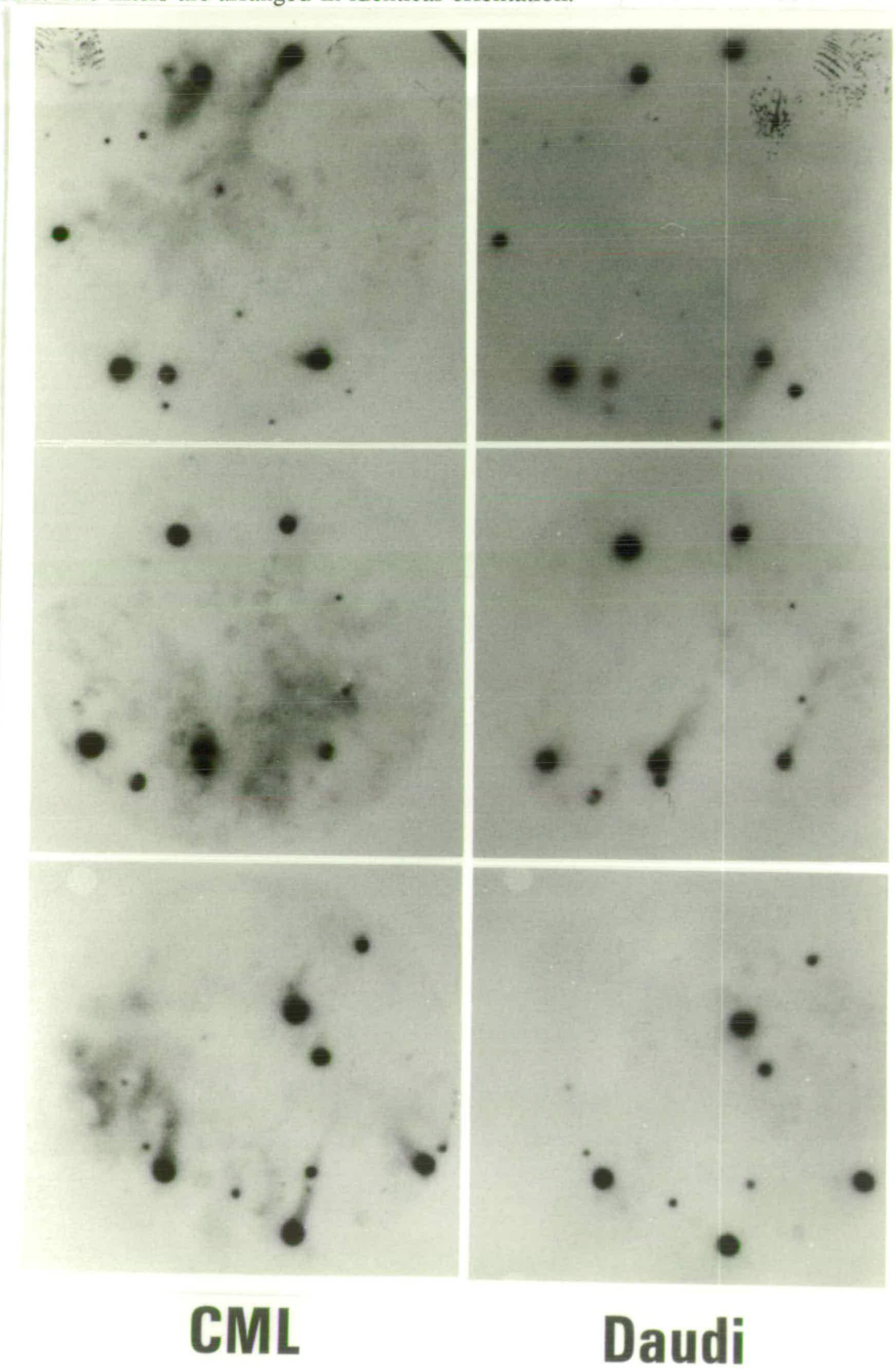
As smaller filters are easier to handle and compare, it was decided to plate out smaller numbers of less dense phage. As cDNA reflects the mRNA population then genes which are represented as ~0.5% of the mRNA should be present at 1 in 200 clones. Replica filters from 20 small plates with ~200 pfu per plate were screened with the liver, Daudi and CML cDNAs.

RESULTS

The results (figure 14) show ~5% of the screened clones are positive in each case and no hybridisation differences can be detected between any of the cDNAs. Two plaques only appeared to be very faintly positive with Daudi and liver cDNA and strongly positive with CML cDNA (figure 15; top row). This plaque (1 out of 2000 plaques) however appeared to have no difference in signal when tested in secondary screening (figure 15 bottom row). The possibility that repeated sequences within the untranslated region of the mRNA could be masking tissue specific mRNA species was investigated next.

FIGURE 14

Overnight autoradiographs of nitrocellulose filters taken from CML cDNA library probed with CML cDNA or Daudi cDNA. Plates contained 200 plaque forming units. No differences could be observed between the two autoradiographs after 10 days exposure. The filters are arranged in identical orientation.



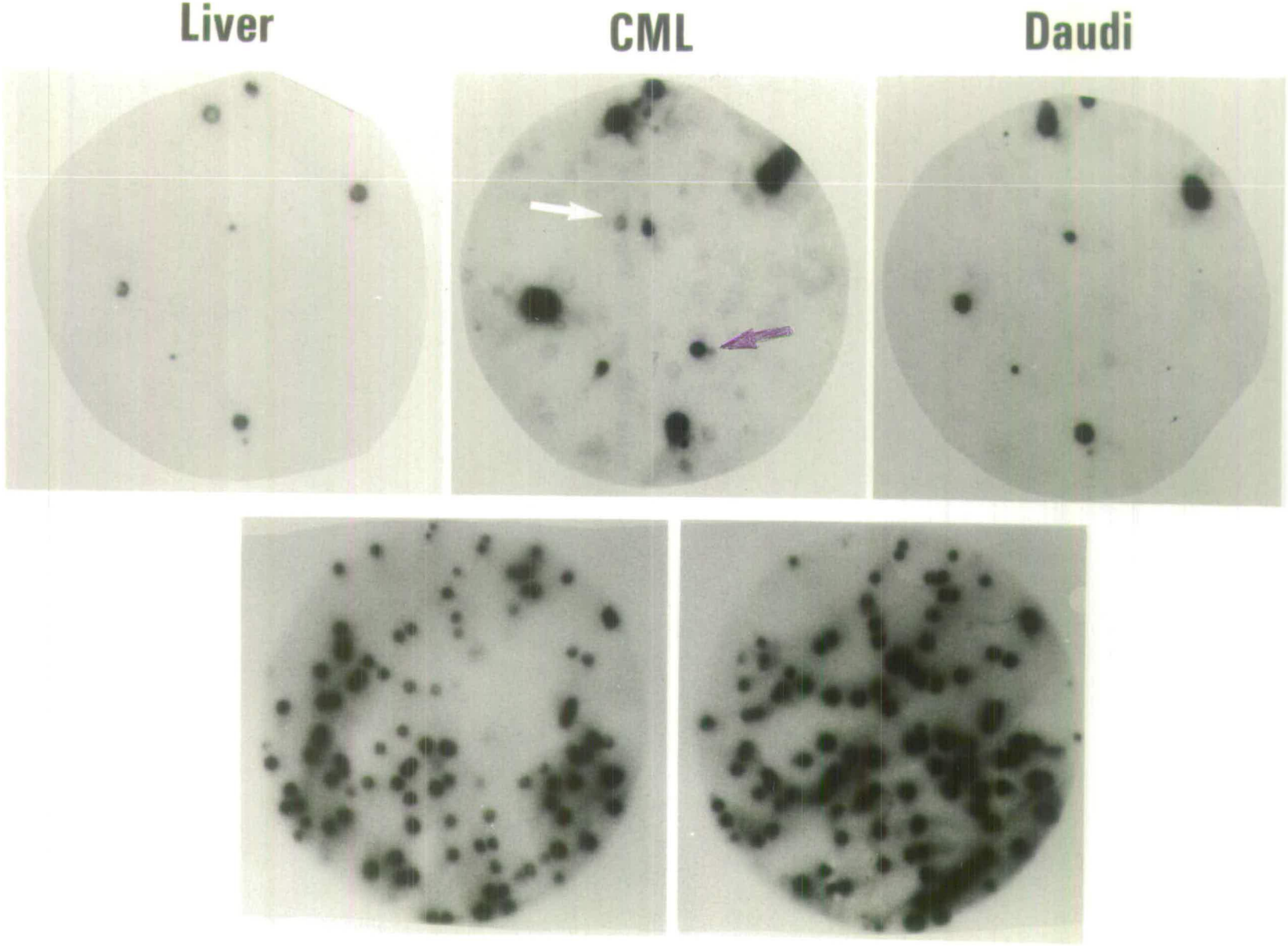

Liver
CML
Daudi

FIGURE 15

The top row shows an overnight autoradiograph of 3 nitrocellulose filters (top line) taken from CML cDNA library probed with CML, Daudi or liver cDNA. The arrows indicate clones only positive with CML cDNA. Filters are arranged in the same orientation.

The autoradiograph in the bottom row shows filters probed with cDNA from Daudi (left hand filter) and CML (right hand filter), to show the positive clone (marked with the blue arrow in the CML filter in figure 15 (top row)) was no longer CML specific when isolated and rescreened.

4. Screening with S1 nuclease treated human genomic DNA.

The human genome is interspersed with repeated sequences including the most abundant mid-repetitive sequences i.e the Alu family (Houck et al 1979), and ~20% of human midtrimester fetal tissue poly(A)$^+$ RNA have been reported to contain repetitive sequences (Kurnit et al 1982). This being the case, lead us to consider whether human repeats in transcripts were masking tissue specific clones. Any clone containing a repeat element in the untranslated region, irrelevant of its true abundance would hybridise strongly to cDNA probe from any tissue because 20% of the probe would have labelled cDNA species containing repeats. However not all these repeats will be of the same family or abundance.

Human DNA was enriched for repetitive sequences by isolating double-stranded DNA after annealing total human DNA to a C_0t of 1 and treating with S1 nuclease (gift of R.Meehan). If the pattern of hybridisation was the same with S1 nuclease treated DNA as with cDNA from a tissue, then it could be concluded that it is the abundant mid-repetitive sequences in the human genome that are hybridising to plaques irrespective of the unique sequence of the clone or abundance of the mRNA.

RESULTS

The pattern of hybridisation of the S1 nuclease treated genomic human DNA was compared with the pattern obtained with CML cDNA and liver cDNA. Figure 16 shows that of a plate of 5000 plaques, 300 clones hybridise to the CML cDNA, and 100 clones hybridise to the S1 nuclease treated human DNA. Of these S1 nuclease positive clones only 10 correlate with the CML cDNA positive plaques. The liver cDNA hybridised to all the same plaques to which the CML cDNA hybridised. This experiment was repeated several times and gave the same result indicating that highly abundant repeats were not responsible for the lack of success in isolating tissue specific clones, unless all the tissue specific clones contained repeats and these are the

10% of S1 treated DNA positive clones which cross-hybridise with CML cDNA. All the other CML cDNA positive clones which cross-hybridise with the liver cDNA must therefore be abundant non-tissue specific mRNA homologues. This seems an unlikely explanation of the results. Another person in our lab has found the same result using a λgt11 cDNA human liver library. Screening this library with CML cDNA, liver cDNA and Daudi cDNA, no difference was found in the pattern of hybridisation. Screening with S1 nuclease treated, repeat enriched DNA it was found that the hybridisation pattern was very different when compared to the liver cDNA hybridisation. Only 10% of the S1 nuclease treated DNA positive plaques were also positive with the liver cDNA plaques. If all the tissue specific clones contained highly repeated sequences then albumin (one of the liver's major transcripts) must contain such a repeat and both CML and liver cDNA would hybridise to the albumin clones in a liver cDNA library. This was shown not to be the case.

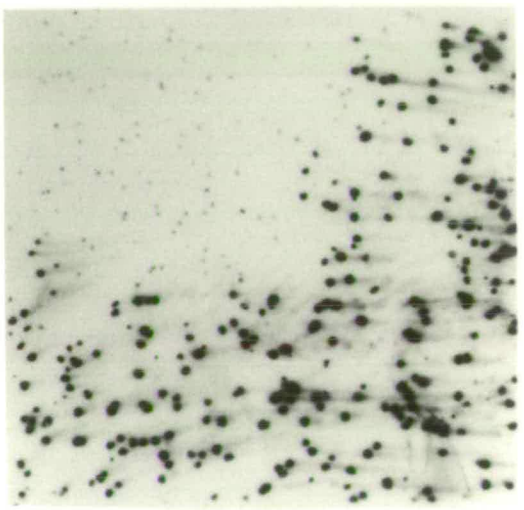
Those clones not hybridising to the S1 treated DNA represent clones containing single copy DNA, or clones containing repeat sequences below the limit of detection

In order to examine whether the method by which this technique was being carried out, or whether the general theory was at fault a positive control experiment was designed.

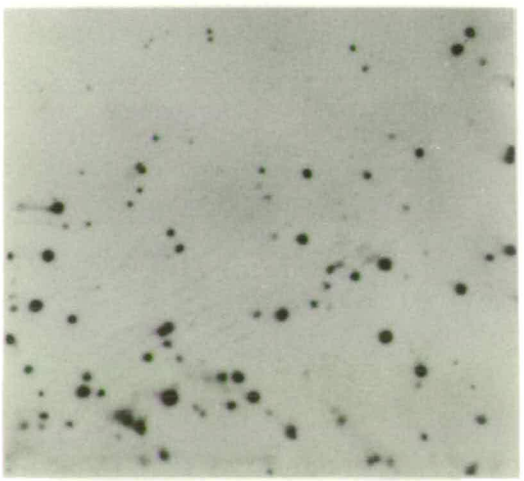
FIGURE 16

Autoradiographs of replica filters of CML cDNA library probed with liver cDNA, CML cDNA and nick translated S1 nuclease treated human DNA. The S1 nuclease treated human DNA gives a different hybridisation pattern to the library probed with the cDNAs. Filters were aligned in the same orientation.

Liver cDNA



CML cDNA



S1 nuclease treated DNA

5. Identification of albumin cDNA clones by differential screening

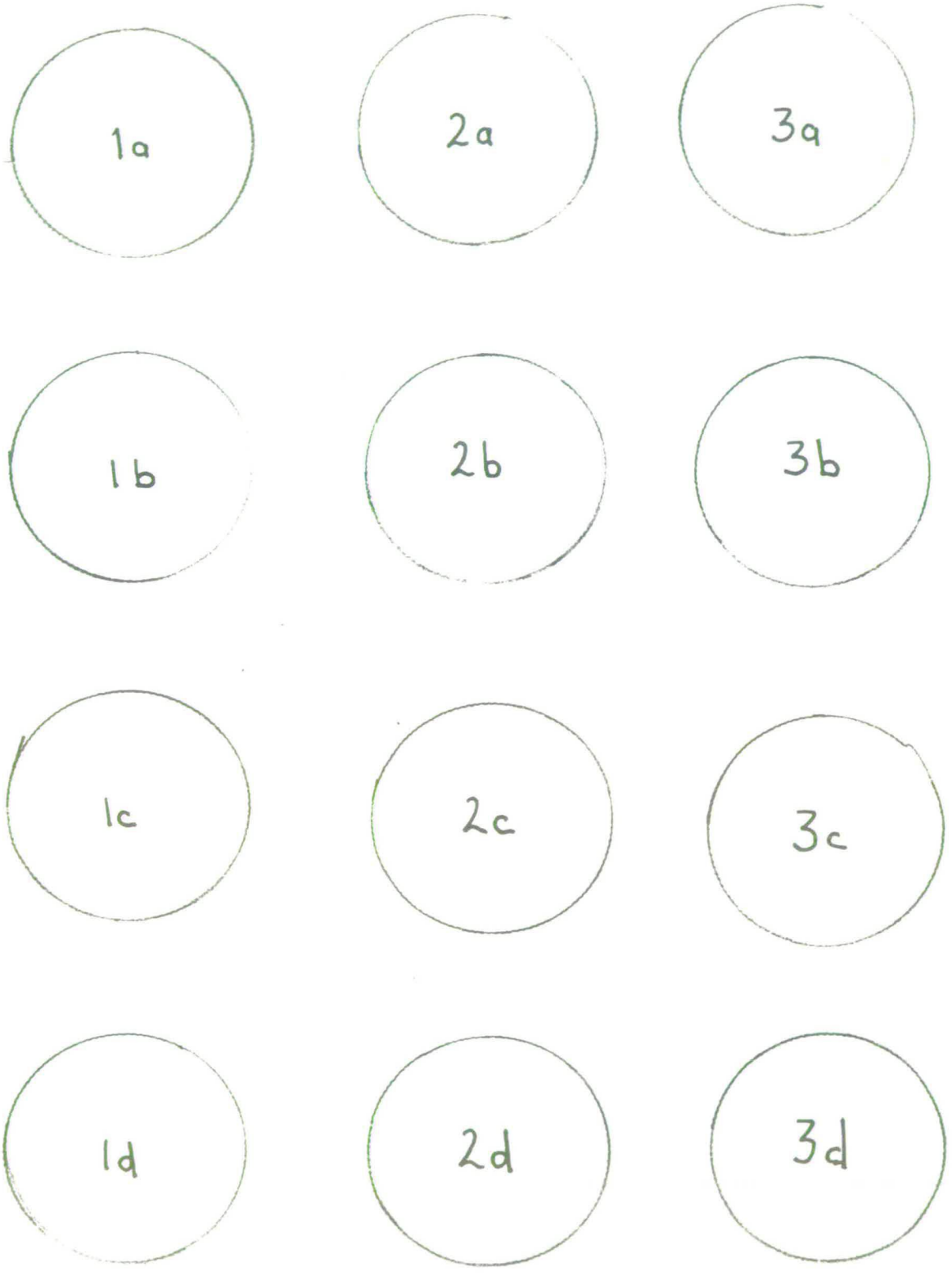
A λ gt11 cDNA clone was isolated with the human albumin cDNA probe. This clone was plaque purified and diluted into the λgt10 CML cDNA library at a ratio of 1:30 albumin:CML. The CML library and albumin contaminated CML library were plated out at a density of ~150 plaques per plate. A plate of 50 albumin only clones was also prepared. Four replica filters were then taken from each plate and hybridised with CML, Daudi, liver cDNA or the nick translated albumin cDNA probe. Figure 17 shows the experimental scheme and results.

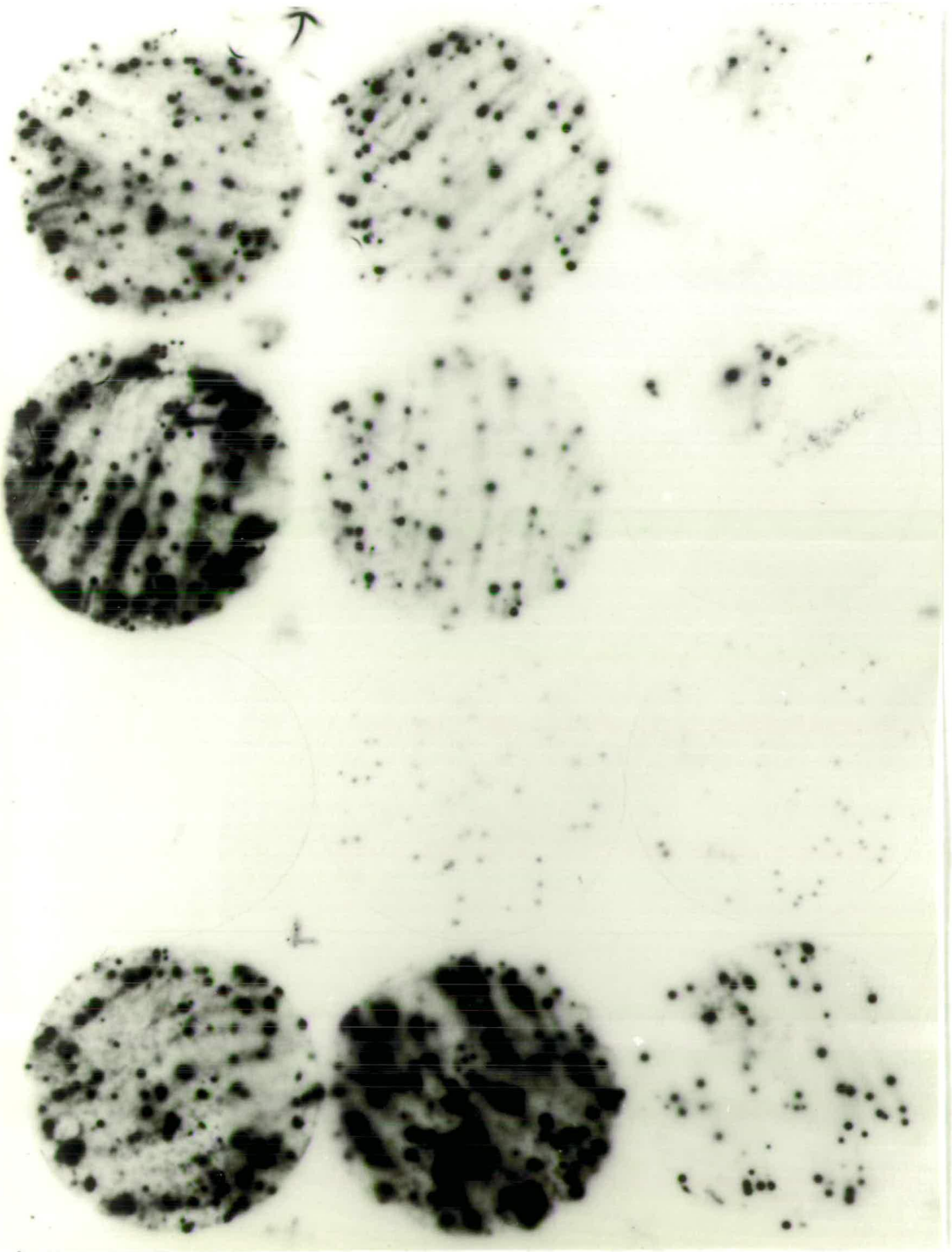
RESULTS

The albumin clones seeded in the CML library were identified with the mouse albumin probe. None of these clones were positive with the CML cDNA or the Daudi cDNA. The liver cDNA hybridised strongly to all the albumin clones either when plated out separately or when seeded in the CML library. Filter 1a (in figure 17) shows the liver cDNA identifying albumin clones, and 2a shows liver cDNA hybridising to albumin clones and CML clones. The CML cDNA however, does not hybridise to filter 1c (albumin clones) or the albumin clones seeded into the library (filter 2c). Apart from the albumin clones, the pattern of hbridisation of the three cDNAs to the CML library are identical. Thus the liver specific albumin clones, when seeded artificially in the CML library, can be identified by differential screening.

FIGURE 17

Schematic representation of experiment to show that albumin can be isolated by differential screening. The autoradiograph shows the filters corresponding to the filters in the diagram. Filters are aligned in the same orientation. Filters 1a-1d are albumin cDNA containing phage, only. Filters 2a-2d are CML cDNA clones, seeded with albumin cDNA clones. Filters 3a-3d are CML cDNA clones. Filters with the same numbers are taken from the same agar plate. Filters 1a, 2a and 3a were hybridised with human liver cDNA. Filters 1b, 2b and 3b were hybridised with nick translated mouse albumin cDNA clone. Filters 1c, 2c, and 3c were hybridised with CML cDNA. Filters 1d, 2d, and 3d were hybridised with Daudi cDNA. Filters 1c and 1d are completely negative, the radioactive marks on them which appear to resemble positive plaques are due to orientation holes in the filters, and can therefore be discounted as artefacts. The albumin plaques, are positive on filters 1a, 1b, 2a, and 2b. They are the only positive clones on filters 1a, 1b, 2b, and are therefore easily visible. They are smaller and fainter clones than the cDNA CML clones because the CML clones are in the phage λgt10, while the albumin clones are in λgt11. λgt10 gives larger plaques than γt11 and therefore the albumin positive plaques are difficult to visualize in the photograph of 2a, as the radioactive signal from the λgt 10 positive plaques tends to obscure that from the smaller λgt11 plaques.


1a
2a
3a
1b
2b
3b
1c
2c
3c
1d
2d
3d



6. Randomly Primed cDNA

In all the previous experiments the cDNA was synthesised from the mRNA, using oligo $(dT)_{12-18}$ as a primer. It has been demonstrated that highly repeated sequences such as those which would be detected with S1 nuclease treated genomic DNA are not responsible for the lack of identification of tissue specific clones. However if a repeat element present at an abundance below the resolution of S1 nuclease treated DNA is responsible, and that element is in the 3' region of the mRNA, then it will generally be represented if the oligo (dT) method of cDNA synthesis is used. The primer anneals to the poly$(A)^+$ tail of the mRNA and so cDNA is synthesised from this region and in theory all cDNA molecules should have the 3' region and extend variable distances 5' into the mRNA sequence. If the cDNA is of small size, then the 3' untranslated regions of the mRNA will be highly represented in the cDNA mix. To circumvent this problem and to get a better representation of the whole length of the mRNA molecules, it was decided to randomly prime the cDNA instead of using oligo (dT). Random priming uses 6 nucleotide stretches of single stranded DNA which will anneal to the mRNA at regions along the length of the molecules. cDNA primed in this way will thus represent the whole length of the mRNA as a mixed population.

A liver λgt11 cDNA library (Clontech) was screened with oligo (dT) primed cDNA and randomly primed cDNA synthesised from liver poly $(A)^+$ RNA. Initially the library was used to compare the hybridisation patterns of oligo (dT) primed cDNA from CML, liver and Daudi poly $(A)^+$. Identical patterns of hybridisation were obtained in each case.

500 pfu of the liver cDNA library were then plated out, and replica filters were probed with oligo (dT) primed liver cDNA or S1 nuclease treated DNA or randomly primed liver cDNA .

RESULTS

25 (10%) of the plaques hybridised to the liver oligo (dT) primed liver cDNA. Of these, 10 plaques hybridised to S1 nuclease treated genomic DNA and a further 17 plaques hybridised exclusively to the S1 nuclease DNA. The randomly primed cDNA gave a hybridisation pattern similar to the oligo (dT) primed pattern in that each plaque that was positive with the oligo (dT) primed cDNA was positive with the randomly primed cDNA but there were 12 additional plaques which hybridised only to the randomly primed cDNA. These 12 plaques did not hybridise to the S1 nuclease treated human DNA.

These results indicate that the randomly primed cDNA hybridises to liver cDNA sequences which are not picked out by the oligo (dT) primed cDNA. We have not found any λgt11 liver cDNA plaques which hybridise to liver oligo (dT) primed cDNA and do not hybridise to oligo (dT) primed CML or Daudi cDNA. It is hoped that the additional plaques which hybridise over and above the oligo (dT) primed cDNA are liver specific.

The experiment was repeated with the CML cDNA library being probed with randomly primed and oligo (dT) primed CML cDNA and S1 nuclease treated DNA. The result was the same as with the liver library in that the randomly primed cDNA gave ~6 more plaques in addition to the 50 plaques which hybridised in common with the oligo (dT) primed cDNA (see figure 18). The plaques unique to the randomly primed cDNA hybridisation pattern did not cross react with the S1 nuclease treated DNA.

5.4 Discussion

It has been shown here that randomly primed liver cDNA and CML cDNA, when hybridised to replica filters from the liver cDNA library, gave a common background pattern and some liver specific plaques, hybridising only to the liver cDNA. treated cDNA. The problems initially encountered, may have lain partially in the properties of the reverse transcriptase used. A colleague found subsequently that using new reverse transcriptase (nbl enzymes) which gave a very high incorporation of dTTP into the first

strand (30% compared to 1-3% with the old reverse transcriptase (Boehringer Mannheim) seemed to improve the differential screening results. Oligo (dT) primed cDNA from liver and CML when hybridised to the liver cDNA library gave distinctly different patterns of hybridisation when the new reverse transcriptase was used.

This work is incomplete as an oligonucleotide probe deduced from the amino acid sequence for CFAg became available. This meant that CFAg cDNA clones could be screened for directly and so differential screening for CML specific clones became redundant. The fact that randomly primed cDNA hybridises to more plaques than conventionally primed cDNA needs to be further investigated. The plaques which hybridise exclusively to the randomly primed cDNA need to be isolated and rechecked for their tissue specificity. There is no evidence that repeats are the reason for this result and the experiments with S1 nuclease treated human DNA indicate that it is not mid-repetitive sequences which are influencing the hybridisation pattern. The possibility exists that it could be repetitive elements of an abundance below the resolution of the S1 nuclease-treated DNA. This might possibly explain the different hybridisation patterns seen. If the CML cDNA hybridises to the CML library by virtue of homology with tissue specific sequence in that DNA sequence, then the liver cDNA should not hybridise unless the insert DNA also contains sequence which is not tissue specific and will hybridise to a population of the liver cDNA with homology to this piece of the sequence. This non tissue specific piece of sequence would necessarily be in the 3' untranslated region of the mRNA and would perhaps explain why more efficient reverse transcriptase extending further into the mRNA molecules would obviate the problem. This does not explain why the experiment where albumin was seeded into the CML library worked so successfully. As albumin was detected immunologically to be in the λgt11 liver library at ~5% of the clones, when J.Maule screened with liver versus CML cDNA at least 1 in 20 clones should have been liver specific!

Had I continued on this line of work, the next step that I should have taken

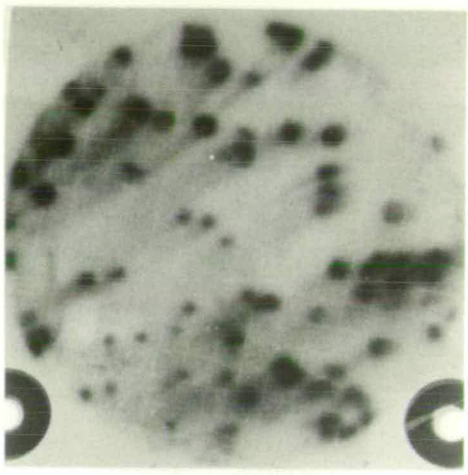
would have been to attempt subtractive hybridisation. This is a technique which has been used very successfully by Littman et al (1985) where the gene encoding the T_8 cell surface glycoprotein was isolated. Basically the strategy was to produce mouse L cell cotransformants which expressed the human T_8 cell surface marker. cDNA was synthesised to the mRNA from the cells and hybridised to a vast excess of mRNA isolated from nontransformed mouse L cells. The resultant probe, highly enriched for T_8 sequences, was then used to screen a cDNA library of human peripheral T cells and the relevant cDNA was isolated. However in this instance, Littman et al (1985) have a highly expressed gene and the non transformed cells provide the perfect source of cDNA to remove the non specific sequences.

I feel that the " repeat problem " could be removed from our experiments by absorbing out the molecules containing repeat elements from the labelled first strand cDNA. This could be accomplished by hybridising with an excess of "cold" total human DNA (Sealy et al 1985) before screening the library.

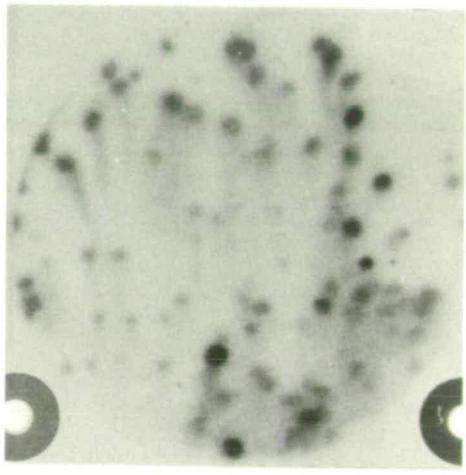
Another possibility which could account for the anomalous results obtained is the presence of mitochondrial mRNA in poly $(A)^+$ preparations from most tissues. The abundance of the mitochondrial mRNA in the library and in the various tissue cDNAs determines how tissue specific the initial overall pattern will appear. This does not explain the difference between random and oligo (dT) primed cDNA hybridisation patterns. The presence of mitochondrial cDNA sequences could be quantified by probing the CML cDNA library with mitochondrial specific probes.

FIGURE 18

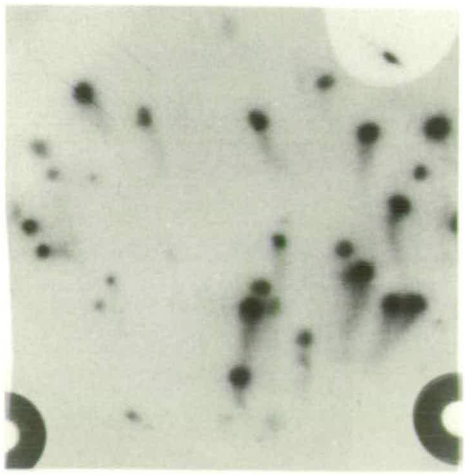
Photograph of autoradiographs of replica filters from λgt10 CML cDNA library probed with a) oligo (dT) primed CML cDNA b) randomly primed CML cDNA c) S1 nuclease treated human DNA nick translated.



CML cDNA



(random primed cDNA)



S1 nuclease treated DNA

CHAPTER 6

ISOLATION OF THE cDNA FOR CFAg

6.1.a Design of oligonucleotide probes to screen the λgt10 CML cDNA library.

As described in the introduction, several IgG monoclonal antibodies (McAbs) were raised against CFAg (Hayward et al 1986) from CML lysate. The McAbs fell into two classes which recognised different epitopes on the antigen. C557, the lowest affinity McAb, was used in the immunopurification of the antigen (Novak et al 1987). The antibody was coupled to CNBr-activated Sepharose and used to purify CFAg from the plasma of CF patients, lysate of CML cells, and lysate from DMSO stimulated HL-60 cells. The CFAg purified from all these sources was compared by N-terminal amino acid sequence determination, SDS PAGE immunoblotting, and Ouchterlony double diffusion using the anti-CML antiserum. By all these criteria, the antigen appeared to be identical from whichever source it was purified. The greatest length of amino acid sequence was obtained from antigen purified from normal granulocytes. In this instance 53 amino acids were sequenced. The partial sequencing of the first 30 residues of the antigen from CML cells, enabled the corresponding coding sequence to be searched for a suitable oligonucleotide sequence. Owing to the redundancy of the genetic code, most amino acids are encoded by four different codon triplets and three of the commonest amino acids each have a total of 6 different codons. Thus a happy medium between length of oligomer (the longer the more specific) and the number of alternative probes required to cover all codon possibilities must be reached. Oligonucleotide mixes of 4 (Stetler et al 1982); 8 (Sood et al 1981); 12 (Breslow et al 1982);

64 (Orkin et al 1983); and 384 (Whitehead et al 1983) have been successfully used. In this case 2 probes were designed as shown below: one of 15 nucleotides with 144 variations and one of 17 nucleotides with 64 variations.

```
Amino acid    12  13  14  15  16
              I   I   D   V   Y

        5'  AUU AUU GAU GUU UAU  3'
              C   C   C   C   C
              A   A       A
                          G
```

Anti-sense sequence used

SY-284

15mer Mix 144

```
         5' ATA AAC ATC AAT AAT  3'
            G   G   G   G   G
                T       T   T
                C
```

Predicted mRNA sequence from protein sequence

```
Amino acid    23  24  25  26  27  28
              K   G   N   F   K   A

         5'  AAA GGU AAU UUU AAA GC   3'
               G   C   C   C   G
                   A
                   G
```

Anti-sense sequence used

SY-312

17mer Mix 64

```
        5'  GC TTT AAA ATT ACC TTT  3'
              C   G   G   G   C
                          T
                          C
```

6.1.b Screening the λgt10 CML cDNA library with oligonucleotide probes.

The oligonucleotide mixes were end-labelled (section 2.2) with 200 μCi (γ-^{32}P ATP) per reaction and 5pmoles of 5' ends. For a 17 nucleotide probe (assuming the average molecular weight of a base to be 300), the molecular weight of the probe is 300 x 17 = 5,100. Thus 1 mole, or 6.2 x 10^{23} molecules weighs 5,100g. 1μg therefore = 1.2 x 10^{14} molecules. 1.2 x 10^{14} molecules = (1/6.2 x 10^{23}) x 1.2 x 10^{14} = 1.9 x 10^{-10} moles (196 pmoles). Therefore, 5 pmoles of 5' ends = 25.5ng of 17mer DNA. 100,000 plaques over 10 plates were screened in duplicate with each oligo mix (section 2.23). The optimum temperature at which nitrocellulose filters are hybridised with an oligonucleotide probe, can be calculated as follows. The ideal temperature for hybridisation to an oligonucleotide probe is where perfect hybrids are stable, but mismatched hybrids are not. This is usually 5°C below the temperature (Td) at which a perfectly-matched hybrid will be half-dissociated (Suggs et al 1981). Td = 4°C per GC base pair + 2°C per AT base pair. For an oligonucleotide mix, the lowest calculable Td is used i.e hybridisation is at 5°C below the lowest Td. For SY 284, this temperature was 31°C and for SY 312 it was 37°C. A final wash was given at the Td temperature.

Seven plaques were positive on duplicate filters for the 15mer SY-284 but none of these were positive with SY-312. There were no plaques positive on duplicate filters for SY-312. This was somewhat disturbing at the time, because the position of SY-312 was only 18bp from SY-284 and was more 3' in the mRNA. It seemed unlikely that cDNA synthesised from the 3' end of the mRNA should contain sequence homologous to the oligonucleotide SY-284 and not have sequence homologous to SY-312. As SY-284 had such a varied sequence and was only 15 bp long there was some anxiety that just by chance there would be a sequence in the library that would match one of the oligomer variations. The fact that the other oligomer did not hybridise added emphasis to this worry. However the 7 positive clones were picked, replated and reiso-

lated until the plaques were all positive. It was subsequently discovered by M. Novak that the amino acid sequence to which the oligomer SY-312 had been made was incorrect and the residue at position 27 was histidine and not lysine.

6.1.c Insert size of the lambda clones

DNA was made from the 7 positive lambda clones (Section 2.24) using a 20x20 cm clear plate lysate. The inserts were released from the lambda by restricting the DNA with *Eco R1*, and visualised by running on 0.8% agarose gel. Clones 4 and 8, had inserts of ~350bp. Clones 3-16, 3-14, A, B and C, had insert sizes of ~250bp. The *Eco R1* cut lambda gt10 cDNA clones 4, 8, 3-14, 3-16 A, B, C were then shotgun cloned (Section 2.30) into pUC-9 (Messing et al 1982). The lambda arms are too large (32kb and 10kb) to be ligated into the plasmid and pUC-9 contains the β-lactamase gene and Col E1 origin of DNA replication of pBR322 ligated to a portion of the lacZ gene of *E.coli*. Introduced into lac minus *E.coli*, the plasmid will give rise to blue colonies when grown on the appropriate indicator plates. Cloning into the unique restriction sites within the lacZ gene fragment, inactivates the gene and gives rise to white colonies.

Cloning into a plasmid enables large amounts of cloned DNA to be grown up and used as a probe, either labelled as the whole plasmid, or as insert DNA. pUC-9 is only 2769bp whereas lambda gt10 is 43.34kb in size. A 300bp insert is thus ~10% of the plasmid DNA, but only 0.75% of the lambda gt10 genome.

Clone 4 and 8 were successfully subcloned into pUC-9. Ligations and transformations were carried out as described in the materials and methods (section 2.30, 2.31). Transformed *E.coli* JM 83 were identified as ampicillin resistant bacteria. Bacteria carrying plasmid with insert DNA were white (non-blue) colonies when grown on X-gal. 18 white colonies were picked from each experiment and small scale DNA preparations were done on 1ml cultures (section 2.32). Cutting these small scale preparations of plasmid DNA with *Eco R1*, revealed that all of the samples contained insert DNA

of the correct size. λgt10 cDNA clones 8 and 4 subclones in pUC-9, (8-9 and 4-7) were grown up in large scale plasmid preparations (section 2.33) and insert DNA was isolated from the plasmids (section 2.34) by cutting with *Eco R1* and then excising the relevant band from a low melting point agarose gel and then carrying out phenol extractions (section 2.34).

The insert fragments were used in Northern blots, to look for the abundance and tissue specificity of the mRNA homologous to CFAg cDNA. In order to verify the fact that the isolated cDNA clones were identical, the insert DNA of clone 8-9 was used to probe *Eco R1* lambda gt10 digests of the other 6 clones (3-1, 3-2, 4-7, A, B, C). Clone 8-9 hybridised with all the other clones' insert DNA, but not with the lambda vector's arms. Therefore all the clones isolated, were probably from the same cDNA. As clone 8-9 was the largest clone, this was used in the further characterisation.

6.1.d Verification of clone 8-9 being cDNA for CFAg.

Initial evidence of the identity of clone 8-9 was present in Northern blot data. When various tissues were probed with insert DNA from clone 8-9, hybridisation signal was only seen with mRNA samples from CML cells and induced but not uninduced HL-60. Further, there was hybridisation with RNA extracted from a somatic cell hybrid expressing CFAg but not with another hybrid which did not express CFAg. This evidence is an indication of the identity of clone 8-9 but is somewhat circumstantial (see later for Northern blot data in detail). The chromosomal location of the clone, also provided circumstantial evidence as to the identity of clone 8-9 (see later for chromosomal mapping data). Sequencing the DNA of clone 8-9 however, provided uneqivocal evidence that the cDNA for CFAg had been cloned.

6.2 Sequencing data

In order to sequence clone 8-9, the insert fragment was cloned into the *Eco R1* site of M13 mp9. The M13 lifecycle can be exploited for the preparation of single

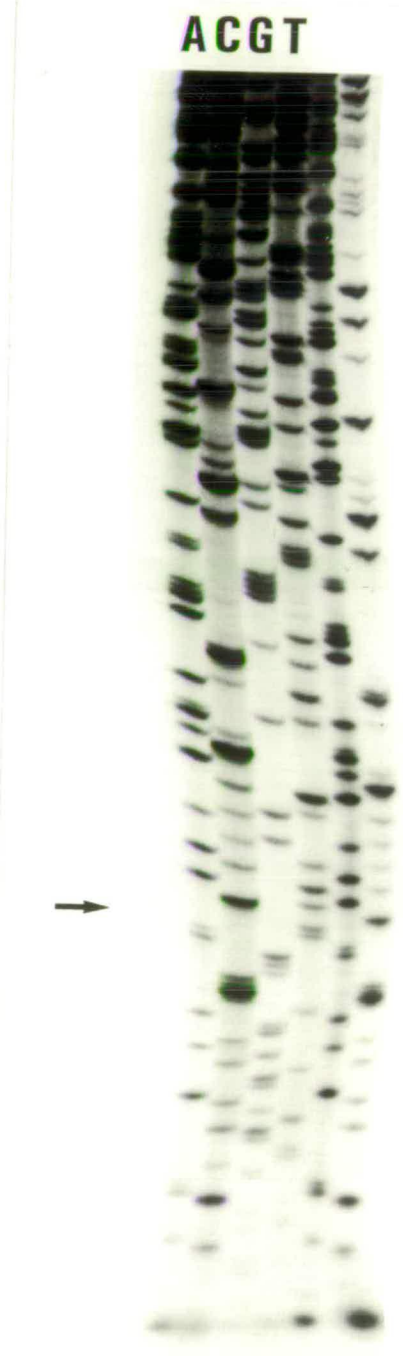
stranded recombinant DNA which then lends itself to sequencing using the Sanger dideoxy sequencing method. The general protocol is as follows (for experimental details see section 4.8). In its replicative form, M13 exists as a double stranded species. This enables foreign DNA to be cloned into a suitable restriction site. The M13 can then be introduced into a competent bacterial host by a transfomation step. The phage will then produce both double stranded replicative forms and single stranded mature virus forms. Large fragments of DNA can be cloned in M13 because it is a filamentous phage and has no constraints on the size of molecule due to packaging restrictions. However 300-900bp inserts appear to be most successful in practice. M13 does not lyse the host cell but it does retard the growth so that infected cells appear as turbid plaques on a lawn of uninfected cells. In order to identify recombinants from nonrecombinants the Messing and Vieria (1982) M13 vectors have been designed to incorporate the *E.coli* lacZ gene under the control of the lac operon. The unique cloning sites are located within the lacZ gene and in the presence of the gratuitous inducer IPTG only nonrecombinant phage will give a blue colour when plated on the colourimetric substrate X-gal. This is because the lacZ gene product will only be functional, and able to hydrolyse X-gal when it is not interrupted by the insertion of foreign DNA. The recombinant phage having been identified, are amplified in small cultures and the protein coat removed by phenol treatment. The single stranded template is now ready for sequencing. This is carried out by annealing a 17mer oligonucleotide homologous to a region of the M13 genome adjacent to the insertion site. The primer is then extended in the 5‘ to 3‘ direction by the Klenow fragment of *E.coli* DNA polymerase I. This lacks the 5‘ to 3‘ exonuclease activity of the whole enzyme but still retains the polymerase activity and will incorporate a dideoxynucleoside instead of a deoxynucleotide although at reduced efficiency. In order to continue chain synthesis there must be a 3‘ hydroxyl group on the last nucleotide to use in the phosphodiester bond to the next nucleotide. Dideoxynucleosides lack a 3‘ hydroxyl group

so if one is incorporated into the growing DNA strand, termination will occur at that point. This is exploited in Sanger's sequencing method. Four reactions are carried out, each with one of the dideoxynucleosides (ddATP, ddGTP, ddCTP, or ddTTP), as well as dNTPs. In each reaction the enzyme will occasionally use a ddNTP and terminate the DNA synthesis of some growing strands but in most cases a dNTP will be incorporated and the DNA will be extended further. The result is a series of DNA fragments terminated at different ddNTPs. This can be visualised by the incorporation of a radiolabelled dNTP into all the reactions and the denatured DNA then resolved on a polyacrylamide gel and autoradiographed.

M13 clones were isolated with the 8-9 insert fragment in both orientations, which allowed the DNA to be sequenced in both directions. A 19mer oligonucleotide was synthesised to a region of the insert DNA from 142 to 160 in both the sense and antisense so that clones of both orientations could be sequenced to verify the data. Figure 19 shows an example of DNA sequencing by the Sanger dideoxy technique (Sanger et al 1977). The arrow indicates the *Eco R1* site of the M13, and the insert sequence is read from that point up the gel. The resulting 338bp sequence (see figure 20) was translated to protein sequence. The predicted amino acid sequence was found to agree almost exactly with the protein sequencing data obtained by M.Novak. The residue at position 27 was found to be histidine and not lysine which explained why the oligomer SY-312 had not hybridised. At amino acid 42 the protein sequencing revealed a threonine where the DNA sequence predicted a cysteine and at position 50 the protein analysis identified a threonine where the DNA sequence indicated a glycine. These differences were resolved on subsequent protein sequencing and the deduced sequence was correct.

FIGURE 19

Autoradiograph to show an example of Sanger dideoxy sequencing. The method was carried out as described in section 2.35. The arrow indicates the *Eco R1* site in the *Eco R1* linker. The letters at the top of the tracks, indicate which dideoxy NTP was included in the reaction mix. The track to the right of the T track is from another reaction.



The sequence of clone 8-9 (figure 20) includes an open reading frame of 249bp which encodes 83 amino acids stretching from the first amino acid used for construction of the oligonucleotide probe to the carboxyterminal residue before the chain termination codon. The sequence encoding the first 11 N-terminal amino acids known from the protein sequencing data and the 5' untranslated region is missing from the clone. The molecular weight of the protein calculated from this deduced sequence is 10,938 kDa and this is in approximate agreement with the observed molecular weight from the immunoprecipitation of the antigen. There do not appear to be any N-glycosylation sites (N X S, N X T) in CFAg from the predicted sequence and there is no evidence that CFAg is a glycoprotein. The predicted protein is rich in lysine residues (14%) and leucine (10.5%).

Sense strand:

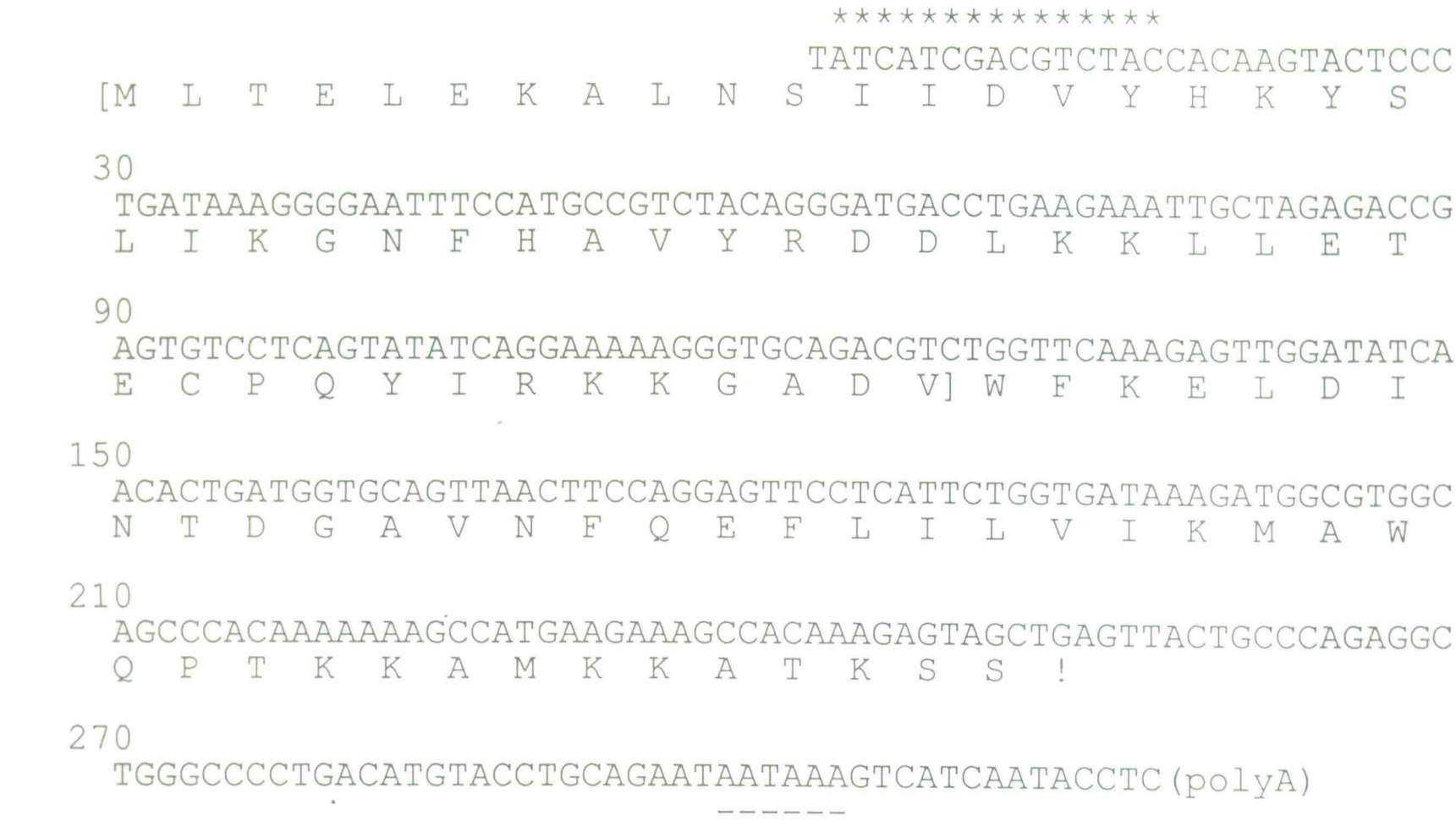


FIGURE 20

Nucleotide sequence of the cDNA clone 8-9 with the predicted amino acid sequence. The amino acid sequence determined by protein sequencing (Novak et al 1987) is enclosed in brackets. The asterisks mark the position of the oligonucleotide used to isolate the clone. The putative polyadenylation site is underlined.

M. Novak used a modification of the method of Hamlyn et al (1978) to sequence a region of the mRNA of CFAg 5' to that present in clone 8-9. Initially oligomer SY-284 was used to prime cDNA synthesis from DMSO induced HL-60 cells as a source of CFAg mRNA. This enabled us to design a unique 24bp oligomer to a region 8bp 5' to the start of the sequence present in clone 8-9. This was then used as a primer and enabled M.Novak to complete the sequencing of the 5' region of CFAg cDNA. The new oligonucleotide was also used to screen the lambda clones 4-7, 3-1-14, 3-1-16, A,B,and C. Unfortunately none of these clones hybridised to the oligonucleotide and so it was assumed that they contained only insert sequence already represented by clone 8-9. Combining the two sources of sequence, we could compile the complete sequence of CFAg cDNA (see figure 21).

Sense strand:

```
          -51
          CTCTTGTCAGCTGTCTTTCAGAAGACCTGGTGGGGNAAGTCCGTGGGCATC

1                                        G***************          60
ATGTTGACCGAGCTGGAGAAAGCCTTGAACTCTATCATCGACGTCTACCACAAGTACTCC
[M  L  T  E  L  E  K  A  L  N  S| I  I  D  V  Y  H  K  Y  S
                                |
                                CFA 8-9
                                                                120
CTGATAAAGGGGAATTTCCATGCCGTCTACAGGGATGACCTGAAGAAATTGCTAGAGACC
 L  I  K  G  N  F  H  A  V  Y  R  D  D  L  K  K  L  L  E  T

                                                                180
GAGTGTCCTCAGTATATCAGGAAAAAGGGTGCAGACGTCTGGTTCAAAGAGTTGGATATC
 E  C  P  Q  Y  I  R  K  K  G  A  D  V] W  F  K  E  L  D  I

                                                                240
AACACTGATGGTGCAGTTAACTTCCAGGAGTTCCTCATTCTGGTGATAAAGATGGCGTGG
 N  T  D  G  A  V  N  F  Q  E  F  L  I  L  V  I  K  M  A  W

                                                                300
CAGCCCACAAAAAAAGCCATGAAGAAAGCCACAAAGAGTAGCTGAGTTACTGCCCAGAGG
 Q  P  T  K  K  A  M  K  K  A  T  K  S  S  !

CTGGGCCCCTGACATGTACCTGCAGAATAATAAAGTCATCAATACCTC(polyA)
                          ------
```

FIGURE 21

Nucleotide and corresponding amino acid sequence of the cDNA clone CFA 8-9. The probable polyadenylation signal of the mRNA is underlined. The asterisks mark the position of the oligonucleotide used to isolate the clone. Brackets enclose the N-terminal amino acids determined by protein sequencing (Novak et al 1987). The nucleotide sequence from -51 to +33 was determined by reverse transcriptase primer extension using induced HL-60 cell mRNA as a template.

Figure 22.

The restriction map of CFAg clone 8-9, showing relative positions of common restriction enzyme sites. 1cm represents 20 base pairs.

6.3 Tissue distribution, transcript size and abundance of CFAg mRNA

The CML cDNA library was rescreened with the insert DNA from clone 8-9 and it was found that approximately 0.5% of the recombinants were positive. This reflects the abundance of the mRNA for CFAg in CML cells. Replica filters were screened with the 24mer N-terminal oligomer and 30% of the clones positive with 8-9 were also positive with the N terminal oligomer.

6.3.a Tissue distribution

The insert 8-9 DNA was used in Northern blot analysis to examine the tissue distribution of the mRNA, (figure 23a and 23b). RNA was prepared from a variety of tissues and cell lines (section 2.14). Hep G2 is a hepatoma cell line which expresses the majority of liver functions (Aden et al 1979). EMBK is an embryonic kidney cell line (ATCC No. CRL 1573). MNNG-HOS is an osteosarcoma cell line (ATCC No. CRL 1547). Daudi is a Burkitt lymphoma derived B-lymphoblastoid line (Klein et al 1968). CML #1 and #2 are RNA samples from CML cells from two different leukophoresis samples. The mouse myeloid stem cell WEHI-TG (Geurts van Kessel et al 1983) was used to fuse with CML cells from different patients and the resulting hybrids were analysed to see which chromosomes segregated with CFAg (van Heyningen et al 1985). WEH 3127.1.9 contains the human chromosomes 1,9,14, and 21. It expresses immunologically detectable amounts of CFAg. WEP 1342.2.4 contains human chromosomes 7,10,13, and 20 and does not express immunologically detectable levels of CFAg.

The blot in figure 23a demonstrates that clone 8-9 hybridises to RNA from CML cells, induced but not uninduced HL-60, and the CFAg expressing hybrid (WEH 3127.1.9). There is weak hybridisation with poly(A)$^+$ RNA isolated from normal liver. This is attributed to blood contamination of the tissue since it has been demonstrated that there is turnover of CFAg in normal peripheral granulocytes and immunohistochemistry of liver sections with anti-CFAg monoclonal antibodies reveals antigen only

in granulocytes and not in hepatocytes. Further, there is no signal with Hep G2 cells which express the majority of liver functions, and there is no hybridisation with hepatoma sample RNA isolated from a sample which was relatively free from blood contamination.

6.3.b Transcript size

Figure 23b shows that in CML and HL-60 mRNA probed with CFAg cDNA clone 8-9, there is a major transcript of ~550bp and a minor transcript of 1kb. The Southern blotting data indicates that there is probably only one gene for CFAg, and therefore the two bands may be due to post transcriptional processing, although there is no evidence for this. The size of the transcripts was determined using size markers provided by the cutting the phage phi X174 with the restriction enzyme *Hae III*, denaturing, and running the digest on the gel. The Northern blot was probed with a mixture of radiolabelled phi X174 and clone 8-9.

6.3.c Abundance of CFAg mRNA in myeloid tissues.

The Northern blot in figure 23b, compares the abundance of CFAg mRNA from HL-60 cells treated with DMSO for 5 days, with HL-60 cells induced for 2 days with RA, and with CML cell RNA. The blot was reprobed with a mouse actin probe (Minty et al 1981) to ensure that equal amounts of mRNA were loaded in each case and to check that the RNA was not degraded. The actin hybridisation intensity (figure 23b) indicates that the RNA amounts were roughly equal and this is also the case in figure 23c although the actin dat for this blot is not shown. The signal obtained with clone 8-9 (figure 23c) hybridised to HL-60 induced with DMSO indicates that the abundance of the mRNA for CFAg is high (equal to the abundance in CML) whereas in HL-60 induced for 2 days with RA it is relatively low (figure 23b). There was no detectable signal observed when clone 8-9 was hybridised to uninduced HL-60 (figure 23c). These data agree with the protein data in Chapter 3 i.e CFAg is abundantly expressed in

CML cells, expressed at a lower level in HL-60 cells induced for 2 days with RA, and is undetectable in uninduced HL-60 cells grown in serum free medium.

It is not clear why HL-60 induced with DMSO gives such an increased expression of antigen compared to RA-induced HL-60. It may be related to the degree of maturation that the cells attain with the two different chemicals. After 5 days of RA treatment 30% of the cells can be identified as mature granulocytes whereas among DMSO treated cells, only 10% are recognisable as mature cells. The remaining cells in the population are at varying stages from promyelocytes to metamyelocytes (Breitman et al 1980; Ferrero et al 1983). Thus at day 2 of RA treatment the HL-60 cells may be at the optimum stage of differentiation for antigen production and this stage is attained at day 5 of DMSO stimulation. The difference in message abundance may be ascribed to more cells of the optimal maturation stage for CFAg expression being present in the DMSO-treated cells at 5 days than in the RA-treated cells at 2 days.

FIGURES 23a and 23b

RNA (Northern) blot analysis of various human cell lines and tissues. *Lane a,* Hep G2(hepatoma) (38); *b,* EMBK (embryonic kidney, ATCC No. CRL 1573); *c,* MNNG-HOS (osteosarcoma, ATCC No. CRL 1547); *d,* Daudi (Burkitt lymphoma derived B-lymphoblastoid line)(39); *e,* hepatoma; *f,* CML #1; *g,* CML #2; *h,* uninduced HL60; *i,* HL60 induced (2 days RA); *j,* normal liver; *k,* WEHI TG (mouse myeloid stem cell) (13); *l,* WEP 1342.2.4 (mouse x human myeloid cell hybrid containing human chromosomes 7,10,13,20) (2); *m,* WEH 3127.1.9 (myeloid cell hybrid containing human chromosomes 1,9,14,21 and expressing CFAg) (2).

Lane 1, CML #1; *2,* HL60 induced (2 days RA); *3,* HL60 induced (5 days DMSO); 4, CML #2. Both blots were reprobed with mouse actin to ensure comparable sample loading (shown only in lanes 1-4).

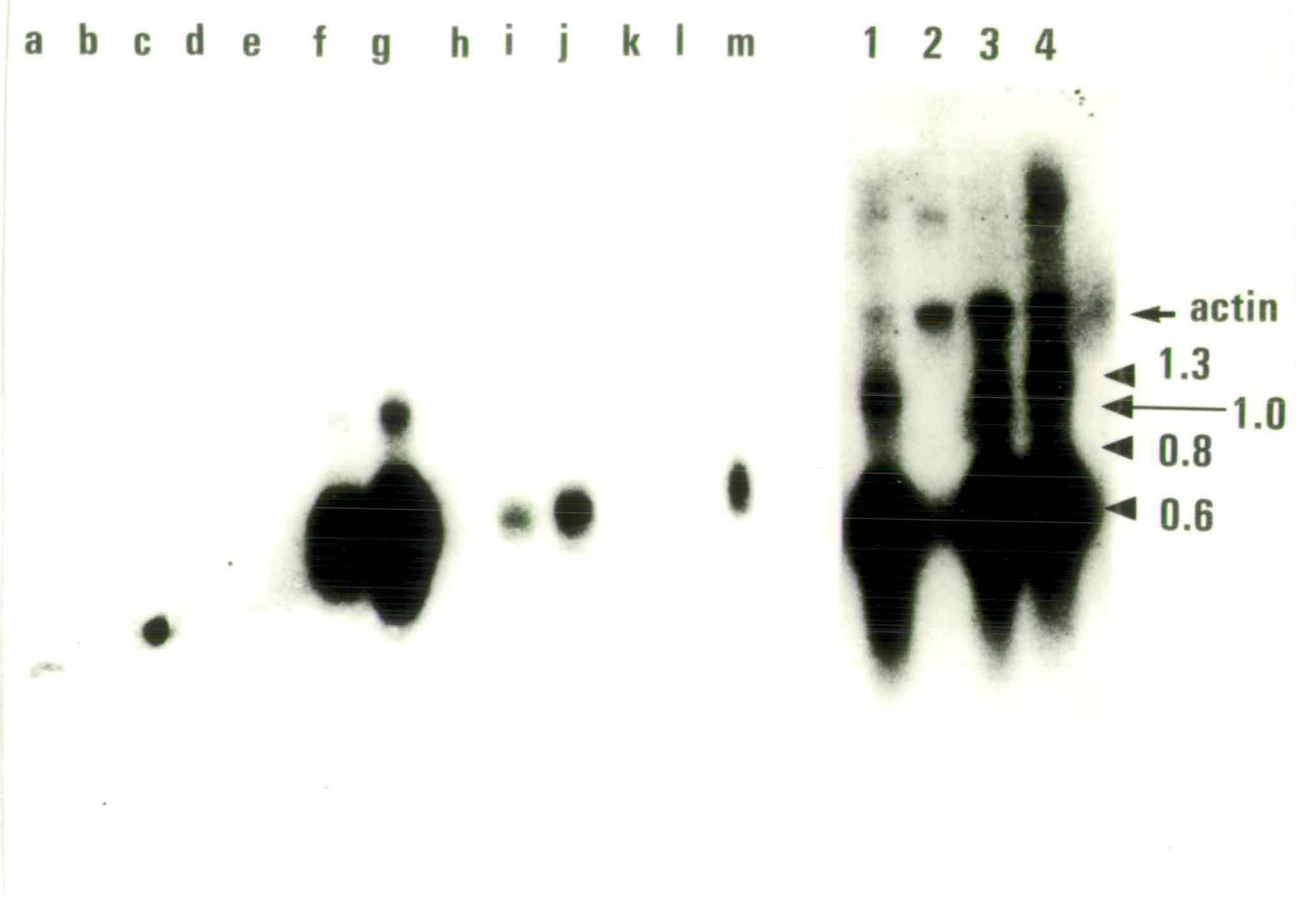
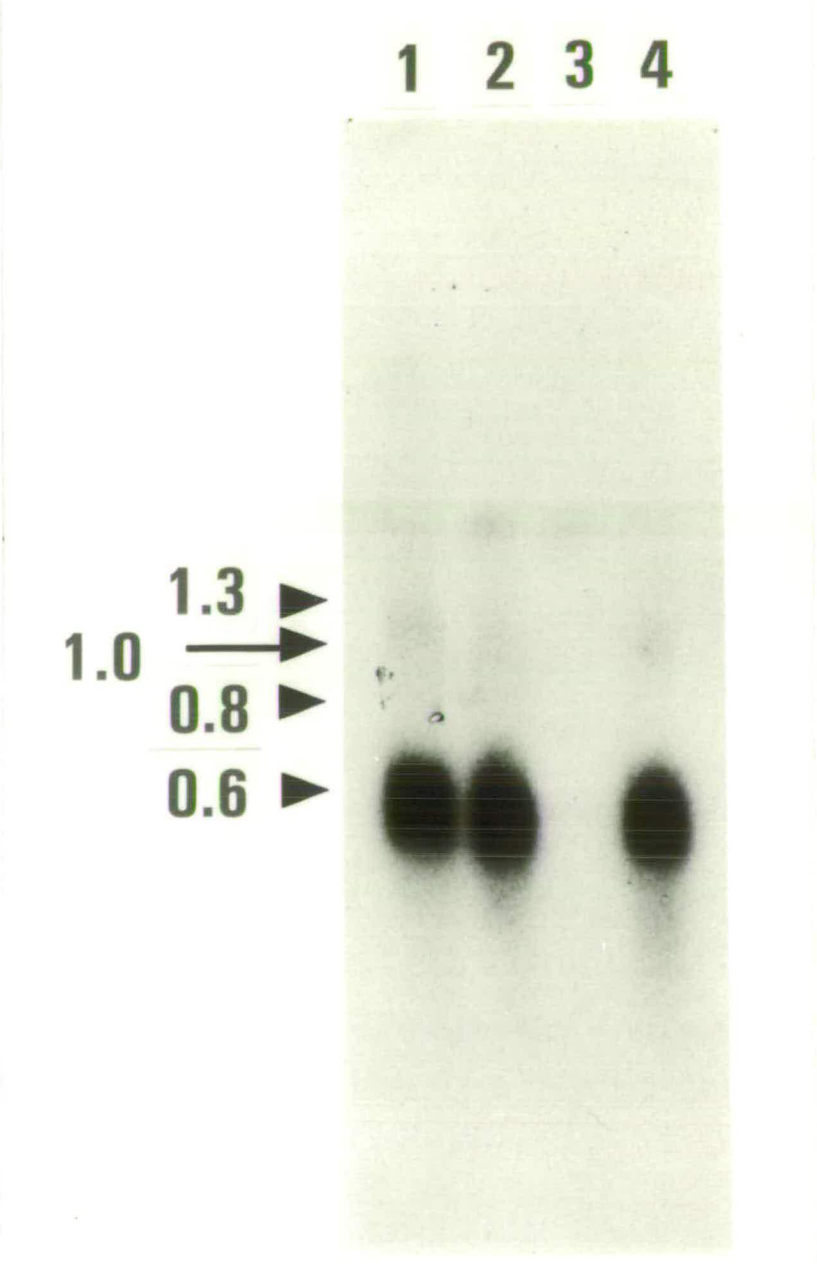


FIGURE 23c

RNA blot analysis to show the relative abundance of CFAg mRNA in equal amounts of RNA from different myeloid tissues. Lanes 1 to 4 contain 1μg of poly $(A)^+$ RNA from various tissues. *lane 1*, CML #2; *lane 2*, HL-60 induced 5 days with DMSO; *lane 3;* HL-60 uninduced; *lane 4*, CML #1.



6.4 Southern blot analysis of genomic DNA with CFAg cDNA clone 8-9

Isolation and digestion of high molecular weight DNA, and Southern analysis were carried out as described in sections 2.36, 2.28 and 2.37. Genomic DNA from two CML patients (including the patient from whom the cDNA library had been made) and a normal male and female were digested with the following enzymes- *Eco R1, Pst I, Bam H1, Hind III, and Bgl II.* The first four enzymes gave single bands of the following sizes

Eco R1	12.5Kb
Pst I	4.5 Kb
Bam HI	8.5 Kb
Hind III	18 Kb

Bgl II gives three bands of sizes 3.9, 4.4, 5.2 Kb.

This is consistent with a single gene for CFAg. It would be necessary for the gene to contain two Bgl II sites within introns for the model to be correct.

Further evidence that the gene is single copy is that, when the cDNA clone 8-9 was used to screen a Bam H1 EMBL 3 genomic library, only one positive plaque was obtained in screening 600,000 plaques. This genomic clone has been isolated and its identity verified by its reactivity with the oligomers to the CFAg DNA sequence which are available. It has not been used for mapping intron/exon organisation as yet.

6.5 Conservation of CFAg between species

In order to examine how conserved the cDNA sequence was, clone 8-9 was used to probe genomic DNA from various species. In Southern blot analysis under very low stringency conditions (5xSSC and 55^{o}C hybridisation conditions; and 2xSSC and 55^{o}C washes) a hybridisation signal was only observed with mouse and human and not with equivalent amounts of DNA from chick, trout, xenopus or yeast. Even at this low stringency, only one hybridising band was seen with the human and mouse genomic

DNAs. Blots hybridised with human CFAg cDNA clone 8-9 could also be washed at 68°C in 2xSSC and rodent bands were as strongly hybridising as human bands.

6.6 Chromosomal localisation of the gene for CFAg

The gene for CFAg has already been mapped to human chromosome 1 by somatic cell genetics (van Heyningen et al 1985). Owing to the availability of rodent x human somatic cell hybrids it was possible to map the CFAg gene to a small region on chromosome 1. This was not previously possible because the detection of the gene relied on the expression of CFAg and its immunological detection. The somatic cell hybrids used by van Heyningen et al consisted of a mouse myeloid parent cell fused to human CML cells. The fact that the mouse parental cell was of myeloid origin was important because it allowed the the continued expression of the granulocyte specific protein. The hybrid cells which were used to map the gene using the cDNA clone did not require expression of the gene and so the fact that the parental rodent cells were fibroblast cells was unimportant.

The first important step was to confirm that CFAg gene was located on chromosome 1. This was done by using a mouse x human hybrid WEH 3127.1.9 which contained human chromosomes 1, 9 and 14. This hybrid was then FACS sorted for the expression and non-expression of a chromosome 1 encoded cell surface marker (Andrews et al 1985). The hybrid was provided by V van Heyningen and the FACS sorting and chromosome analysis was kindly carried out by J. Fletcher. After the FACS sorting was completed, chromosome analysis showed that the clone WEH 3127.1.9 T7 contained the human chromosomes 1,9,14; and the clone WEH 3127.1.9 T8 contained the human chromosomes 9 and 14. It could be shown in Southern analysis that the cDNA clone 8-9 hybridised to genomic DNA from both clones but a human specific band was only seen in the clone WEH 3127.1.9 T7 (Figure 24a).

A series of somatic cell hybrids were then used which have different fragments of

chromosome 1 contained in them. In the case of NYX 5.6 and NYX 5.6.11 only the chromosome 1 fragments have been analysed and data for the remaining human chromosomes are not available. The hybrid NYX 5.6.11 is a subclone of NYX 5.6 but the chromosome 1 has broken in the process of subcloning. Table 1 summarises the chromosome data on the various hybrids and shows that the human specific bands hybridising to clone 8-9 are only present when the region of chromosome 1 from q12 to q21 is also present. Figure 24a and 24b show the Southern blot analysis used to determine the presence or absence of the CFAg gene in various hybrids. Hybrids, Rag J8, NYX 5.6, NYX 5.6.11 and WEH 3127.1.9 T7,8, are mouse/human hybrids. SK 81 and 82 are chinese hamster/human hybrids. The hybrid Rag J8, contains the short arm of the human chromosome 1 and a region of the long arm from the centromere to q21. This hybrid contains DNA which gives the human band size with CFAg. SK 82 however, contains the short arm of the human chromosome 1, but the region of the long arm from the centromere to q12. This hybrid does not appear to contain human DNA which hybridises to CFAg. These results imply that CFAg lies in the region q12 to q22 on human chromosome 1.

It should be noticed that no reaction has ever been observed with the guinea pig antiserum raised against CFAg and any mouse tissue. The human CFAg cDNA clone does not hybridise even under very low stringency washes to mRNA from a mouse myeloid cell (WEHI-TG) however a mouse specific hybridising band is observed in genomic DNA blots, probed with clone 8-9, and the mouse myeloid cell line (WEHI-TG) does allow expression of human CFAg in somatic cell hybrids. In order to verify the chromosomal assignment further, in situ hybridisation was attempted by J.Gosden using the CFAg cDNA clone but no specific region to which the clone hybridised was identified. This was probably due to the small size of the clone, and will be repeated with the genomic clone.

FIGURE 24a

Southern blot analysis to illustrate mapping of CFA 8-9 in rodent/human somatic cell hybrids. Chromosome analysis is shown in Table 1. 10µg of genomic DNA digested with *Bgl II* was loaded in each case. Lambda DNA cut with *Hind III* is also shown with the size of the bands indicated in kilobases. *Lane 1,* NYX5.6.11; *2,* F9.6 ; *3,* NYX5.6 ; *4,* CML#3; *5,* WEH 3127.1.9 T7 ; *6,* WEH 3127.1.9 T8; *7,* WEH 3127.1.9 T7; *8,* WEH 3127.1.9 T8; *9,* WEHI-TG.

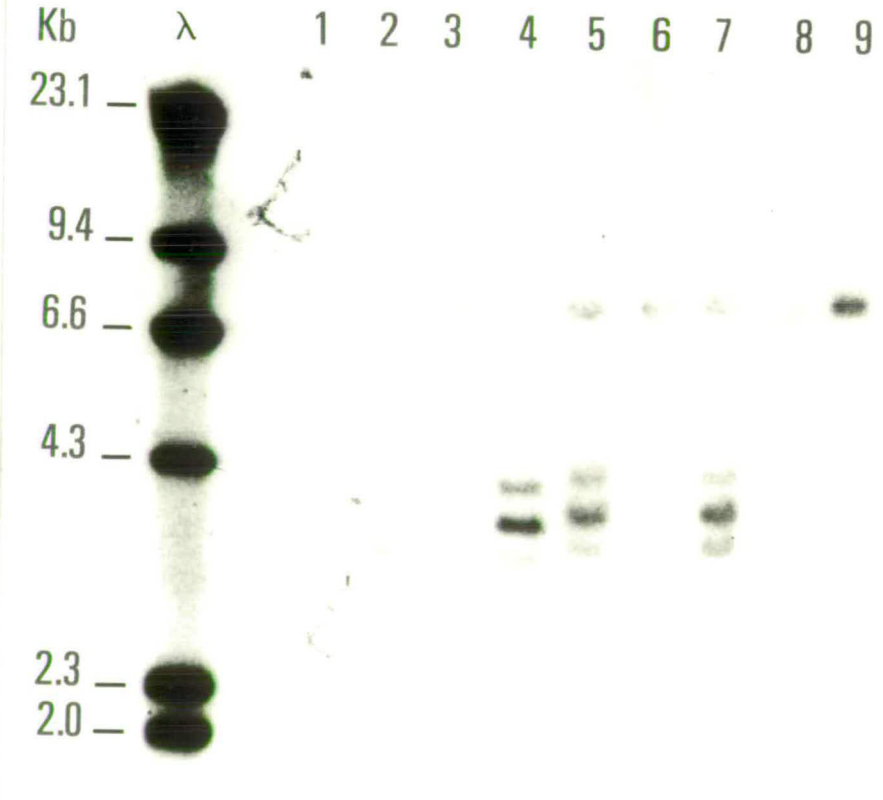
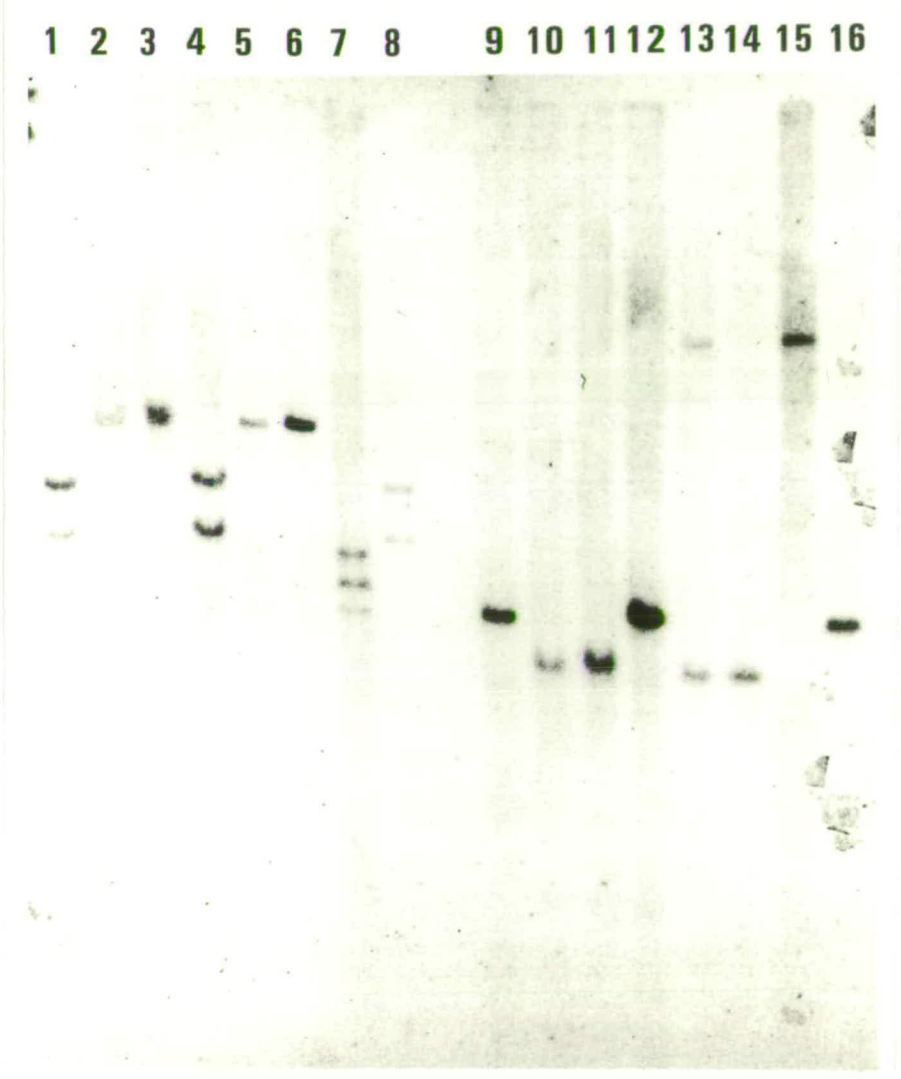


FIGURE 24b

Southern blot analysis to illustrate mapping of CFAg in rodent/human somatic cell hybrids. Chromosome analysis shown in Table 1. 10μg of DNA was loaded in each case. DNA in lanes 1 to 8 are cut with *Bgl II*, while DNA in lanes 9 to 16 are cut with *Eco R1*.Lanes 1-8 Lanes 1 and 9 contain SK 81; lanes 2 and 10 contain WEHI-TG, lanes 3 and 11 contain CML #1; lanes 4 and 12 contain SK 82; lanes 5 and 13 contain RAG J8; lanes 6 and 14 contain WEHI-TG; lanes 7 and 15 contain CML #2; and lanes 8 and 16 contain chinese hamster DNA.



CHROMOSOMAL LOCALIZATION OF THE GENE FOR CFAg.

Human Chromosomes present / Hybrid	Fragment of 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	CFAg
Rag J8	pter-q21								■				■												+
SK82	pter-q12				■												q					■			-
SK81	pter-p31					■		■							■			■		■	■	■			-
NYX 5.6	complete	not fully characterised																							+
NYX 5.6.11	pter-q23	not fully characterised																							+
F9.6	complete	■	■					■		■		■		■		■	■	■			■		■		+
WEH 3127.1.9 T_7	complete								■					■											+
WEH 3127.1.9 T_8	absent								■					■											-

TABLE 1

Segregation of the CFA 8-9 hybridising, human specific restriction fragment with chromosome 1 or fragments of it.

WEH 3127.1.9 T7 and T8 (van Heyningen et al 1985) are subclones of the same hybrid FACS sorted for expression and non-expression respectively of a chromosome 1 encoded cell surface marker (Andrews et al 1985). Black squares denote the presence of chromosomes in >50% of cells scored. The restriction fragments corresponding to the human gene for CFA 8-9 are seen whenever the entire chromosome 1 is present and also in hybrid RAG J8 with human chromosome 1 from the tip of the short arm to band q21 on the long arm (pter-q21) present. No human bands are present in hybrid SK 82 containing only region pter-q12 of chromosome 1. The gene must therefore lie between bands q12 and q22 on chromosome 1. Other hybrid cell data support this result.

6.7 Conclusion

The full cDNA sequence of CFAg has been determined by combining the sequence data from the cDNA clone isolated from the CML cDNA library and cDNA synthesised from DMSO induced HL-60 mRNA template. The total length of the cDNA is 404 bp which agrees with the size of the mRNA visualised on agarose gels (~550bp) assuming a poly $(A)^+$ tail of 150-200bp. The open reading frame is 282bp long stretching from the ATG start codon to the base pair at position 252 (see figure 20) preceeding the TGA stop codon. The size of the protein from the deduced amino acid sequence is 10,938 kDa, which is in approximate agreement with the size observed on SDS PAGE (Novak et al 1987). Deduced protein sequence data for clone 8-9 and the protein sequence of the immunopurified CFAg (Novak et al 1987) agree completely and verify the identity of the clone. The results of Northern blot analysis with CFAg clone 8-9, establishes the expression of CFAg in myeloid tissue which agrees with the immunohistochemical data available for CFAg (Wilkinson et al 1986). Furthermore, using Southern analysis of genomic DNA from somatic cell hybrids it was possible to assign the location of CFAg clone 8-9 to chromosome 1 which agrees with the chromosomal assignment of CFAg using immunological detection of expression in a panel of somatic cell hybrids by van Heyningen et al (1985). The availability of hybrids containing fragments of chromosome 1 enabled CFAg to be mapped to the region q12-q22.

CHAPTER 7

Comparison of CFAg Deduced Sequence to other Protein Sequences.

7.1 Introduction

The protein sequence deduced from the nucleotide sequence of clone 8-9 (including the 11 amino acids missing from the sequence but identified by protein sequencing) was compared with all the sequences in the National Biomedical Research Foundation library (NBRF) data base (version 8, 809,386 residues) by a method which allows high stringency comparison with all the sequences present (Coulson et al 1986). Coulson, Lyall and Collins at Edinburgh University, Department of Molecular Biology kindly analysed our deduced protein sequence using an algorithm which exhaustively searched the data base for similarities. Normally the assessment of significance of detected similarities relies upon the statistical behaviour of randomly compiled implausible sequences, because the best exhaustive algorithms (Needleman-Wunsch-Sellers; (Collins and Coulson 1986)) cannot be efficiently used due to the large number of known protein sequences. The exhaustive searches necessary to discriminate genuinely similar proteins, run on conventional serial mainframe computers are impracticably slow and, consequently, most searches are carried out initially at a superficial level. For strong similarities approximate methods eg FASTP (Lipmann and Pearson 1985) should reveal the same alignments as an exhaustive search. However, more distantly related proteins require a more thorough search. In our case, the best local similarity alignment of Smith and Waterman (1981) was used in conjunction with the I.C.L. 4096 distributed array processor (DAP). The full potential of the multi processor

machine is exploited and the program takes 1-2 seconds per residue of test sequence to search the entire protein database. The score for each amino acid pairing was assigned and log odds tables for a range of evolutionary distances were used. Indels (insertions or deletions) were scored as a single fixed value penalty. An inbuilt measure of significance is the expected number of alignments with a certain score when compared to the vast majority of unrelated proteins in the database. The log of the frequency of a particular score declines linearly with the score itself, i.e the greater the score the less is the expected frequency of alignments. Alignments with scores expected less than once require examination, and closely related sequences will produce alignments with scores expected to occur many orders of magnitude less often.

7.2 Proteins with sequence similarity to CFAg

The database search with the clone 8-9 protein sequence was carried out using each reading frame separately, and reading through stop codons in case the sequencing had been incorrect. Only one reading frame produced significant alignments and this corresponded to the CFAg sequence from protein sequencing. The four alignments were :-

Bovine Intestinal calcium binding protein (ICaBP).....Expected frequency 1.1×10^{-5}.

Porcine Intestinal calcium binding protein.....Expected frequency 2.8×10^{-5}.

Bovine S-100 α protein alpha chain..... Expected frequency 5.8×10^{-4}.

Bovine S-100 β protein beta chain..... Expected frequency 1.4×10^{-3}.

The search was repeated using scoring tables for 5 evolutionary distances. Depending on the evolutionary distance one sequence is from another, the scores assigned to particular residue changes must also change. Dayhoff, (Dayhoff 1978) compiled score tables for amino acid changes at different PAMs (accepted number of point mutations per 100 residues), where the PAM is a measure of evolutionary dis-

tance between two protein sequences. The scores and expected frequencies are shown below for the best four local similarity alignments using the 80 PAM similarity table. Indels were scored -10.

Bovine S-100a protein alpha chain.....score 189.....Exp Freq 1.7×10^{10}

Bovine ICaBP.....Score 167..... Exp Freq 1.2×10^{-8}

Bovine S-100b protein beta chain.....score 163.....Exp Freq 2.6×10^{-8}

Porcine ICaBP.....Score 153.....Exp Freq 1.8×10^{-7}

The best alignment of CFAg and S-100α protein is shown below. * indicates residue identity and . indicates pairs of residues with a positive score in the similarity table. There is 43% homology over a stretch of 72 amino acids. CFAg sequence is given below the S-100α protein sequence.

```
SELETAMETLINVFHAHSGKEGDKYK_LSKKELKELLQTELSGFLDAQKDADAVDKVMKELDEDGDG
.*** *. ..*.*.*  *   *. .    . .** **          * ** *    **** . **
TELEKALNSIIDVYHKYSLIKGN_FHAVYRDDLKKLLETECPQYI_RKKGAD_V_W_FKELDINTDG


EVDFQEYVVLV
 *.***. .**
AVNFQEFLILV
```

The database search revealed that CFAg had significant homology to S-100 α and β chains and ICaBPs from the rat, cow and pig. These proteins are all members of the troponin C super family characterized by a relatively low molecular mass and the ability to bind calcium . Twenty four of the top twenty five alignments in the search involved calcium-binding proteins such as calmodulin and troponin C.

7.3 Calcium-binding proteins

The proteins with the highest significant similarity to CFAg are calcium-binding proteins. No physiological function has been assignd to any of these proteins, but the intestinal CaBP are known to be vitamin D inducible (Thomasset et al 1979), and may be involved in the absorption of calcium across the gut. The bovine S-100 α and β proteins are predominantly located in the brain (Eldik et al 1982), as are the human forms (Donato 1986). Bovine ICaBP has not been cloned, but the protein has been isolated and sequenced (Hofmann et al 1979, Fullmer and Wasserman 1981. cDNA clones encoding the rat and chicken ICaBP have been isolated and sequence data published (Desplan et al 1983, Desplan et al 1983a, Wilson et al 1985,). The three dimensional structure of the minor A form (i.e the form lacking the first two N-terminal amino acids of the native protein) of bovine ICaBP has been determined by X-ray crystallography (Szebenyi et al 1981). ^{1}H NMR studies have also been performed on the minor A form of bovine ICaBP (Dalgamo et al 1983) as well as the native porcine protein (Shelling et al 1983). Recently a synthetic gene encoding the known amino acid sequence of bovine ICaBP has been cloned into an expression vector in *E.coli* in an attempt to examine and define the domains important in the structure and function of the protein (Brodin et al 1986). The ICaBPs belong to a larger family of calcium-binding proteins (see discussion) including parvalbumin, troponin c, and calmodulin. They all contain the same helix-loop-helix structure of their calcium-binding sites, the so called EF-hand structure (Kretsinger 1980). Parvalbumin (Kretsinger 1973), bovine

ICaBP (Szebenyi et al 1981), troponin C (Herzberg and James 1985; Sundaralingham et al 1985) and calmodulin (Babu et al 1985) have all been characterized by X-ray crystallography. It was from the amino acid sequence and crystallographic structure of parvalbumin that Kretsinger (1980) stated the essential features of the EF-hand calcium-binding region. Parvalbumin consists of a loop-helix structure and it is helix 5 and 6 or E and F which are involved in the calcium binding together with the interhelix loop which contains the oxygen atoms to ligand the calcium ion. This model can be demonstrated with an outstretched forefinger, thumb and clenched middle finger leading to the term "EF-hand".

For a region to satisfy the constraints of an EF-hand certain amino acids must have oxygen atoms in their side chains available to ligand the calcium ion.

Residue 10, 12, 14, 18, 21 supply side chain oxygens and residue 16 uses a peptide oxygen. The putative EF-hand domain for CFAg determined by amino acid sequence comparison to the other calcium-binding proteins is shown below. Numbering the residues with the standard domain numbering scheme devised for parvalbumin, it can be seen that the residues 10, 12, 14, 18, 21 all have side chain oxygen molecules. According to Kretsinger's model, residues 2, 5, 6, 9, 22, 25, 26 and 29 must have hydrophobic side chains as these positions are on the inside of the molecule. Asp, Asn, Glu, Gln, Ser or Thr must occupy positions 10, 12, 14, 18, and 21. CFAg obeys these rules also. Finally the conformation and packaging of the loops imposes two constraints. Firstly the bend at residue 15 is so tight that only Gly can assume the accquired angle and residue 17 must be able to anchor the loop to the hydrophobic core of the molecule.

G A D V W F K E L D I N T D G A V N F Q E F L I L V I K M A W
1_____5_____10_____15_____20_____25_____30

Calmodulin, troponin C and parvalbumin contain only EF-hand type of calcium binding domain whereas S-100 and ICaBPs have one EF-hand region and another distinct calcium binding structure at a more N-terminal position in the protein. The putative EF-hand region of CFAg shows significant similarity to the EF hand region of various calmodulins, but the other, non EF-hand, calcium-binding region of CFAg is not significantly similar to any region of calmodulin. This supports the theory that, like S-100 and ICaBP, CFAg contains one EF-hand type calcium binding region and another distinct calcium binding structure. The constraints of this region have not been defined, and the calcium ion uses oxygen atoms from the peptide backbone as ligands (van Eldik 1982). The mammalian ICaBPs bind calcium with different affinity to their two types of site (Bryant and Andrews 1984).

Very recently, the human cDNA sequence for a growth factor inducible gene has been determined (Calabretta et al 1986). It is 434 nucleotides in length and has an open reading frame of 270 nucleotides coding for a putative polypeptide of 90 amino acids. The closest homology at the nucleotide level of this protein 2A9, is to the S-100 β protein with 55% homology in the coding region. At the amino acid level the homology is most noticeable in the calcium-binding regions of S-100. In the first (non EF-hand site) the homology is 88% and at the second (EF-hand site) the homology is 61% to the S-100 α subunit.

The EF-hand calcium binding motif is the most highly conserved region between CFAg, S-100, ICaBP,and 2A9. Over the whole length of the 2A9 predicted protein the homology to CFAg is 28.8%, but over the 30 amino acids involved in the EF-hand, the homology is 40%. For S-100α and bovine ICaBP the homology in this region is 50%.

The function of 2A9 has not yet been elucidated and the protein has not yet been isolated. However Calabretta et al (1986) believe it may be involved in cell proliferation as the mRNA corresponding to the cDNA clone is induced in quiescent fibroblasts by various growth factors.

S-100 has been reported to be closely related to p11 (Gerke and Weber 1985) p11 is an 11 kDa protein, whose function appears to be as a regulatory subunit of a larger molecule p36. This 36 kDa protein is the major cytoplasmic target of avian sarcoma virus, tyrosine-specific protein kinases (Greenberg et al 1984), and has moderate calcium binding capacity. The cellular function of p36 is unknown but it may be involved in microfilament association and binding to lipids and membranes. In some cell types, such as intestine, p36 exists as a heterodimer in association with the small p11 subunit (Gerke and Weber 1985). In human epidermal cells, the analagous protein p35 exists mainly as a monomer and *in vitro* phosphorylation seems to occur preferentially on this form. This has led to the proposition that p11 is a regulatory element (Gerke and Weber 1985).

p11 has a lower observed similarity to ICaBP than to S-100 and no calcium binding has been shown for p11 although attempts have been made to show this (Gerke and Weber 1985). The introduction of a gap of 3 consecutive residues in the p11 95 amino acid residues allows for 43 identically placed residues when comparing to the S-100 α subunit, plus 15 highly conserved amino acid changes. The 3 deleted residues however lie within the non EF-hand calcium binding site. The EF-hand region of p11 also differs from true calcium binding site as defined by Kretsinger (1980). The possibility exists that S-100 may have a regulatory function similar to that proposed for p11.

Comparing CFAg to p11 there are 23 conserved amino acids over a stretch of 94 amino acids, the most striking homology being in the proposed EF-hand region of calcium binding (36%). Calabretta et al (1986) postulate that as calcium and tyrosine

kinases have an important role in cellular proliferation, perhaps 2A9 mediates this regulation in fibroblasts by calcium-binding and interaction with a cellular substrate for tyrosine kinase in a similar way to p11 regulation of p36.

S-100 and p11 both exist as dimers but S-100 has been shown to bind 2 moles of calcium whereas no calcium binding has been detected for p11. Further, the analagous sites for calcium binding in p11 appear to be altered. It is possible that S-100 has a protein ligand related to p36. CFAg on native gel Western blotting appears to exist predominantly as a dimer.

Figure 25 shows the alignment of ICaBPs from pig, rat and cow and bovine S-100α and β subunits, human 2A9 and porcine p11 with CFAg protein sequence. Conserved amino acids are boxed and the calcium binding ligands for bovine ICaBP are marked.

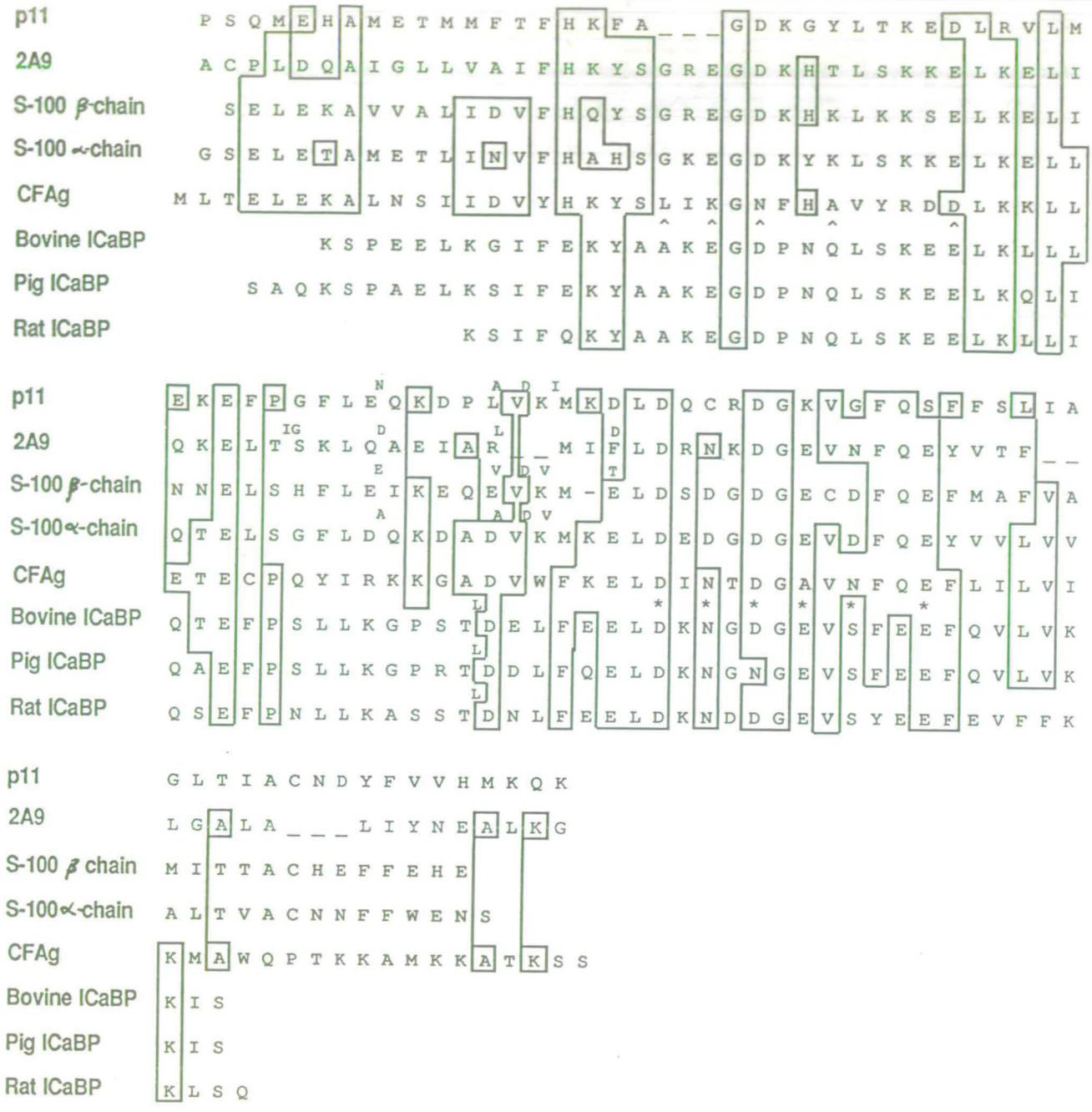

p11
2A9
S-100 β-chain
S-100 ∝-chain
CFAg
Bovine ICaBP
Pig ICaBP
Rat ICaBP
P S Q M E H A M E T M M F T F H K F A _ _ _ G D K G Y L T K E D L R V L M
A C P L D Q A I G L L V A I F H K Y S G R E G D K H T L S K K E L K E L I
S E L E K A V V A L I D V F H Q Y S G R E G D K H K L K K S E L K E L I
G S E L E T A M E T L I N V F H A H S G K E G D K Y K L S K K E L K E L L
M L T E L E K A L N S I I D V Y H K Y S L I K G N F H A V Y R D D L K K L L
K S P E E L K G I F E K Y A A K E G D P N Q L S K E E L K L L L L
S A Q K S P A E L K S I F E K Y A A K E G D P N Q L S K E E L K Q L I
K S I F Q K Y A A K E G D P N Q L S K E E L K L L I
p11
2A9
S-100 β-chain
S-100∝-chain
CFAg
Bovine ICaBP
Pig ICaBP
Rat ICaBP
E K E F P G F L E Q K D P L V K M K D L D Q C R D G K V G F Q S F F S L I A
Q K E L T S K L Q A E I A R _ M I F L D R N K D G E V N F Q E Y V T F _ _
N N E L S H F L E I K E Q E V K M - E L D S D G D G E C D F Q E F M A F V A
Q T E L S G F L D Q K D A D V K M K E L D E D G D G E V D F Q E Y V V L V V
E T E C P Q Y I R K K G A D V W F K E L D I N T D G A V N F Q E F L I L V I
Q T E F P S L L K G P S T D E L F E E L D K N G D G E V S F E E F Q V L V K
Q A E F P S L L K G P R T D D L F Q E L D K N G N G E V S F E E F Q V L V K
Q S E F P N L L K A S S T D N L F E E L D K N D D G E V S Y E E F E V F F K
p11
2A9
S-100 β chain
S-100∝-chain
CFAg
Bovine ICaBP
Pig ICaBP
Rat ICaBP
G L T I A C N D Y F V V H M K Q K
L G A L A _ _ _ L I Y N E A L K G
M I T T A C H E F F E H E
A L T V A C N N F F W E N S
K M A W Q P T K K A M K K A T K S S
K I S
K I S
K L S Q

FIGURE 25

Alignment of amino acid sequences of ICaBP from rat, pig and cow, and bovine S-100α and β subunits, human 2A9 and porcine p11 with CFAg protein sequence. Amino acids which occur in CFAg and one other sequence are boxed. The standard one letter code is used to abbreviate the amino acids. Calcium binding ligands as determined by X-ray crystallography for the EF-hand region of bovine ICaBP (Kretsinger 1980), are marked with *. The calcium binding ligands for the second distinct region are marked ^.

CHAPTER 8

DISCUSSION

This thesis describes the isolation of cDNA clones for CFAg from a CML library. The identity of the clone has been established by comparing the predicted amino acid sequence deduced from the DNA sequence of the clone with the 53 amino acid sequence obtained for the immunopurified protein (Novak et al 1986). The cDNA clone has been used to demonstrate that the major mRNA transcript in CML cells is an abundant species of ~550bp in size. The CFAg protein can be induced in HL-60 cells with RA and to a greater extent by DMSO. Both these chemicals induce the cells to differentiate towards myeloid cells.

Using Southern blot analysis with the cDNA clone for CFAg, there appears to be only one gene. Using very low stringency conditions the gene will only hybridise to mouse and human DNA and not to bird, fish, amphibian or yeast DNA.

The location of the gene for CFAg has been confirmed to be on chromosome 1 and sublocalized to the region q12-q22.

The predicted amino acid sequence for CFAg was compared to other known protein sequences in the NBRF database and was found to have significant homology to calcium binding proteins from mammalian intestine and the S-100α and S-100β subunits which are most abundant in brain. Similarity to the predicted amino acid sequence of the human cDNA clone, 2A9 (Calabretta et al 1986) and to the porcine p11 protein has also been established. These latter protein sequences were not present in the database due to their recent identification. In order to postulate a function for CFAg it is

useful to examine the function of these homologous proteins.

CFAg has significant homology to a family of calcium-binding proteins whose functional importance is not completely understood. Baba et al (1984) established a phylogenetic reconstruction where all 50 proteins of the calmodulin family protein members could be divided into 7 lineages established during more than 1 billion years of evolution. The lineages originate from a one domain polypeptide of 36-40 amino acid residues with a central 12 amino acid calcium binding site, and a series of tandem duplications result in the various family members.

Troponin C, calmodulin, parvalbumin, myosin light chains, S-100 and intestinal calcium binding proteins are all members of different classes of calcium modulated proteins (van Eldick et al 1982). Troponin C and calmodulin both bind 4 calcium ions per molecule while parvalbumin binds only 2. The type of calcium binding region in these proteins has been termed EF-hand structure and was first characterised by Kretsinger (1975) for parvalbumin (1975a). Calmodulin has four such regions as does troponin C. S-100 and ICaBPs however (although they bind two calcium ions each) have only one region which satisfies the constraints of the EF hand model. Bovine ICaBP, which has the second highest homology to the CFAg predicted sequence, has been examined by X-ray crystallography (Szebenyi et al 1981) to reveal the different binding regions. The porcine ICaBP has been studied using NMR (Shelling et al 1985) and the two binding sites appear to have random binding with equivalent association constants.

8.1 Calmodulin

Calmodulin is the best characterised calcium modulated protein, and has homology to CFAg in the EF hand region of calcium binding. Calmodulin is a highly conserved, ubiquitous protein contained mainly but not exclusively in the soluble fraction of cells. It is not necessarily advantageous to look at specific functions of calmodulin

when attempting to assign a putative function to CFAg, but a general idea of the types of function calmodulin fulfills may be of some help. No definite function has been assigned to the calcium binding-proteins to which CFAg is most significantly homologous.

Calmodulin itself is not active but when complexed to calcium it becomes an active species (Cheung 1980). Although it has no demonstrable enzyme activity it has phosphodiesterase activator activity including the calcium-dependent stimulation of cyclic nucleotide phosphodiesterase, adenylate cyclase, myosin light chain kinase, skeletal muscle phosphorylase kinase and other protein kinases (Kennedy et al 1981). However not all of these functions are necessarily calmodulin specific as many preparations are impure and calmodulin can, for example, substitute for troponin C and activate actinomyosin ATPase *in vitro* (Dedman et al 1977). However Hathaway et al (1981) have shown a physiological role for calmodulin in the regulation of the myosin light chain kinase and Pichard et al (1981) have shown muscle phosphorylase kinase regulation by using purified protein. Calmodulin is highly conserved throughout evolution and the recently cloned yeast homologue has 60% homology to the bovine counterpart (Davis et al 1986).

8.2 Intestinal calcium-binding proteins

The biosynthesis of these proteins which have been found in the small intestine of cow, pig, rat and chicken is dependent on vitamin D. They are soluble species and appear to be of two major types: type I proteins with MW of ~28,000 and type II proteins with a molecular weight of ~10,000. The type II proteins to which CFAg has been found to be homologous have only been detected so far in mammalian species. Figure 25 shows the calcium-binding sites and Syebenyi et al (1981) determined that there were 4 alpha helices. Two of the helices III and IV and the calcium binding loop connecting them form a calcium-binding structure similar to the EF hand structure of parvalbumin. The explanation of these binding structures is in Chapter 7. The III

and IV site calcium (EF hand) appears to more readily removed than the I and II site calcium. Syzebenyi suggests that the I, II site may be permanently saturated with calcium whereas the III, IV site is a candidate for a regulatory site which could easily bind and release calcium in the same way that has been demonstrated for calmodulin. Antibodies have been raised against mammalian type I and type II ICaBP and chick type I proteins. The two types are immunologically distinct and although chick ICaBP will cross react with mammalian type I ICaBP from kidney and brain, mammalian type II proteins appear to be species specific (Wasserman et al 1977). When vitamin D is given to rachitic rats an immediate response is observed in the increase in intestinal absorption of calcium followed by a delay of several hours and the production of ICaBP (Thomasset et al 1979).

8.3 S-100

Another class of calcium-binding proteins, separate from those regulated by vitamin D, is a group of closely related small acidic proteins. S-100 are members of this group and are so called because of their solubility in 100% ammonium sulphate (Moore 1968). The S-100 proteins are dimers composed of the α chain and the β chain (S-100a protein) or two α chain (S-100a0 protein) or two β chain (S-100b) subunits. S-100 is the overall term used for these proteins which are abundant in the brain and contained in well defined cell types of both neuronal and non-neuroectodermal origin. In the mammalian brain S-100a and S-100b are confined to glial cells while S-100a0 is neuronal in localization (Donato 1986). Norepinephrine and dibutrylcyclic AMP increase the level of S-100 in C6 rat glioma cells both by stimulation of synthesis and inhibition of degradation of S-100 (Labourette et al 1980). Taubuchi et al (1982) showed that forskolin and prostaglandin E plus theofillin known to activate adenylate cyclase, also induces an increase in the S-100 content of rat glioma cells and therefore S-100 may be regulated by cAMP (Higashida et al 1985). S-100 a and b proteins possess two types of calcium binding site. An EF hand conventional site in the carboxy

terminus and an unconventional calcium binding site in the N terminal part of the polypeptide chain. This site has a lower calcium binding affinity due to the high content of basic amino acids (Baudier et al 1983). The rat, human, (Kuwano et al 1984) and bovine (Isobe et al 1978) protein sequences are homologous except for 4 residues. Although S-100 appears to be mainly present in the soluble fraction of glial cells *in vitro* it appears to interact, with high affinity, with several natural membranes isolated from brain, including synaptic vesicles and nuclear membranes (Donato et al 1986). Natural membranes possess high and low (Kd=7nM and Kd =0.4μM) affinity binding sites for S-100, the high affinity binding site requiring the presence of calcium (Donato 1986). S-100 has been detected immunologically in human glial cells of the central nervous system, some neurons, Schwann cells of the peripheral nervous system, melanocytes, Langerhans cells of the epidermis, sweat glands and ducts of the skin, myoepithelial cells and occasional ducts of salivary glands and breast, serous glands, and hyaline cartlidge of bronchi and the larynx (Kahn et al 1983). The presence of S-100 can also be used as a marker for malignant melanoma and neoplasms derived from Schwann cells (Donato 1986). Thus S-100 proteins are clearly important and more widely distributed than previously thought, although not ubiquitous like calmodulin. The function of S-100 remains far from clear, the majority of evidence being related to the acquisition of behavior patterns (Hyden et al 1970). It potentiates the microtubule disassembling affect of high calcium concentrations (Donato 1986) and also interacts *in vitro* with some of the proteins known to be a target of calmodulin (Burgess et al 1984). It has been shown to be phosphorylated by brain nuclear kinases and inhibits the phosphorylation of a 73 kDa cytoplasmic polypeptide of brain in a calcium dependent way (Patel et al 1982). S-100 has also been reported to bind zinc very tightly and in a calcium independent manner (Baudier et al 1983).

8.4 p11

As mentioned in chapter 7, S-100a and b subunits have homology to p11, a regulatory subunit of p36, the substrate for tyrosine specific protein kinases . p11 exists as a dimer as does S-100. Calcium binding was searched for in unsucessfully in p11 and although the homology of p11 to S-100 α chain is significant, the region of p11 homologous to the first calcium-binding region of S-100 α lacks 3 residues, and the EF-hand region has two amino acids which are not conserved and may affect calcium binding. The possibility of zinc binding for p11 is being examined, as is the possibility of a larger subunit to complex with S-100 (Gerke and Weber 1985).

8.5 2A9

We have also noted that CFAg has sequence similarity to the predicted protein sequence of the cDNA clone 2A9 The 2A9 cDNA clone was isolated by screening transformed human fibroblasts with a hamster homologue, which itself was isolated by differential screening for cell cycle variable messenger RNAs. The human clone was overexpressed in human acute myeloid leukaemia cells. The protein equivalent to the 2A9 cDNA has not been isolated. The homology of 2A9 to S-100 is particularly remarkable in the putative calcium binding regions. The mRNA transcript size of 2A9 is 0.6kb and the transcript becomes apparent in rat, mouse, and human diploid fibroblasts only when these are stimulated with various growth factors. It is not detectable in G*o* cells. 2A9 is detectable in human bone marrow cells and cells from acute and myeloid leukaemias (Calabretta et al 1985; 1986). Its inducibility is not affected by inhibition of protein synthesis and Calabretta et al (1986) postulate that 2A9 may have a role in cell cycle regulation. Calcium is required for growth of cells in culture, especially fibroblasts, and levels of intracellular free calcium are markedly increased when quiescent fibroblasts are stimulated to proliferate by platelet derived growth factor (PDGF) and epidermal growth factor (EGF) but not when stimulation is by insulin. 2A9 is induced by PDGF and EGF but not by insulin. Further, the 2A9 predicted

amino acid sequence has significant homology to p11 which is part of a complex which is a substrate for tyrosine kinases. Tyrosine kinases are also involved in the cellular response to growth factors (Cooper et al 1982). Wilkinson et al (personal communication) have shown that the only normal cells in which CFAg can be detected immunologically are non keratinised squamous epithelial cells of the tongue, oesophagus and cervix. However squamous cell carcinoma of the skin and lung are strongly positive. This may indicate some involvement of CFAg in cell cycle control and when these cells are rapidly proliferating CFAg levels increase. However it may reflect an expansion of the particular cell type in which CFAg is present.

8.6 CFAg as calcium-binding protein

The predicted sequence of CFAg from the cDNA data indicates that it is a calcium binding protein. Results of experiments using radiolabelled calcium by C.Hayward and M.Novak indicate that CFAg does bind calcium. ICaBPs are not directly involved in the uptake of calcium but are associated with this process (Corradino R et al 1976). The control of intracellular calcium levels is extremely important and although no clear function has been assigned to these calcium modulating proteins they may be involved in controlling the level and hence in transducing calcium signals (van Eldvik 1982). A rise in calcium levels may result in calcium binding to these proteins and in turn activate enzymes in a way similar to that for calmodulin. However the tissue distribution of this family of proteins suggests a more restricted role than that of the ubiquitous calmodulin.

8.7 Signal transduction

a) cAMP acts as a second messenger in mammalian cells. Regulated by adenylate cyclase levels, an increase in cAMP (generated from ATP) will result in the activation of protein kinases. The activation of protein kinases can be calcium dependent or independent. The stimulation of the adenylate cyclase occurs by the binding of an

agonist to the cell membrane receptor which then is coupled to a G protein which binds GTP when activated.

The multiple steps in the β-adrenergic system permit other signals to activate the cascade at many different points. Some protein kinases e.g glycogen phosphorylase kinase, are activated both by phosphorylation and calcium. The phosphorylation is mediated by the cAMP-dependent protein kinase and phosphorylation increases the affinity of the calcium-binding subunit (calmodulin) for calcium. In fact, independently of cAMP signalling, a sufficient increase in calcium will result in an unphosphorylated and yet active enzyme (Lodish et al 1986).

Calcium-binding proteins may inter-relate with the cAMP second messenger pathway at the level of protein kinases or at other points in the cascade. Calmodulin for example is involved in regulation of the cAMP response. The calcium-calmodulin complex binds to and activates phosphodiesterase which converts cAMP to AMP and thus terminates the effect of the agonist.

b) The phosphoinositol pathway is responsible for the release of calcium (trigger calcium) into the cytosol from stores in the endoplasmic reticulum (ER) (Berridge et al 1984). External agonists eg vasopressin, glucagon, neurotransmitters bind to receptor and stimulate two second messengers (Higoshida et al 1986). Phosphatidyl(4,5)bisphosphate, a precursor lipid, is located within the inner leaflet of the plasma membrane and is cleaved in response to agonists which stimulate phosphodiesterase (phospholipase c). This yields the second messengers diacylglycerol and inositol(1,4,5)triphosphate. Diacylglycerol activates protein kinase c, while phosphatidylinisitol is converted to phosphatidylinisotol (4) phosphate and then to phosphatidyl (4,5) bisphosphate which is cleaved to inositol (1,4,5) triphosphate and thus mobilizes calcium. Protein kinase c phosphorylates one set of proteins while inositol trisphosphate elevates calcium to activate calmodulin-dependent protein phosphorylation and other calmodulin-dependent reactions. It may be that S-100,and ICaBPs are activated by this system and regulate

enzymes in a similar way to calmodulin but in a more defined manner suggested by their tissue specificity.

8.8 Relationship of CFAg to the gene defect in CF

The relationship of CFAg to the basic gene defect in CF is as yet unknown. The recent data, which indicate the biochemical fault lies in control of chloride ion permeability in various exocrine tissues may indicate the metabolic pathway in which the two gene products (on chromosome 1 and 7) must inter-relate. The β-adrenergic response in CF exocrine tissue appears to be somehow defective (Welsh and Liedtke 1986; Frizzel et al 1986; Sato and Sato 1984; MacPherson et al 1986) and β-agonists stimulate chloride channel activity in normal, but not CF epithelial cells (Welsh and Liedtke 1986; Frizzel et al 1986). This does not however preclude inositol trisphosphate or diacylglycerol also being involved.

8.9 Defective chloride channel activity in CF

The most recent results obtained independently by Welsh and Liedtke (1986) and Frizzel et al (1986) using patch clamp techniques, on trachea epithelial cells indicate that

1. The activity of isolated chloride channels is equivalent in normal and CF samples.

2. Whole cells stimulated with β-adrenergic agonists demonstrate chloride channel activity only in samples from normals and not in CF individuals.

3. Whole cells stimulated with 8-bromo-cAMP (cAMP analogue) or forskolin (activator of adenylate cyclase) only demonstrate chloride channel activity in normal and not CF samples.

4. cAMP accumulation in response to β-adrenergic agonists is normal in CF nasal and tracheal epithelial cells.

This places the defect in CF chloride ion reabsorption at a point distal to cAMP accu-

mulation but not intrinsic in the channel itself. However this does not rule out the possibility that it could be a defect in the phosphorylation of an important site on the cytoplasmic side of the channel for example. Sato and Sato (1983) also found normal cAMP regulation in sweat glands from CF individuals although sweat secretion was only stimulated by isoproterenol in normals. McPherson et al (1985, 1984) reported a decreased stimulation of CF submandibular glands in CF patients by β-adrenergic agonists, measured by a lack of mucin or amylase secretion from CF exocrine acinar cells. However noradrenaline (a mixed α and β-agonist) did stimulate CF cells to secrete mucin and amylase. They observed however that cAMP was increased to the same amount in normal and CF β-agonist stimulated tissues.

Davis et al (1983), reported that intact leukocytes from patients with CF produce significantly less cAMP than normal lymphocytes and granulocytes in response to isoproterenol. Prostaglandin E however produces a normal response in CF white cells and the basal level of cAMP is normal. In opposition to their results Galan et al (1981) had shown that whole granulocytes from patients with CF had reduced numbers of beta-adrenergic receptors, and heterozygotes had normal numbers of receptors although they also displayed a reduced response to isoproterenol

If the defect in CF is in the control of anion channel function, this also explains the reduced HCO_3^- secretion from CF pancreatic ducts. Previous theories involving the chloride channel being intrinsically faulty in CF individuals did not explain this observation.

However as has been recently demonstrated by Wakelam et al (1986), in the case of glucagon it is not always clear whether only one pathway is involved in eliciting a hormonal response. A wealth of data accumulated to show that glucagon increased the cAMP levels via adenylate cyclase stimulation and caused glycogen breakdown , probably mediated by increased Ca^{2+} levels. Wakelam et al used a glucagon analogue (TH-glucagon), which does not activate adenylate cyclase or increase cAMP, to show

that there was still gluconeogenesis, glycogenolysis, and urea synthesis stimulation, and the production of inositol phosphates occured with this analogue. This has been shown for other signal hormones including vasopressin (which acts through V1 and V2 receptors (Creba et al 1983)), and has been reported to be the case for secretin in pancreatic acinar cells (Peterson and Bear 1986). Consequently it may be too rash to attribute chloride secretion stimulation completely to a cAMP mediated pathway. Further, these two systems are not autonomous, and it appears that in the case of glucagon (Wakelam et al 1986) adenylate cyclase activation can inhibit inositol phosphate production, and similarly, activation of protein kinase C desensitises the glucagon-stimulated cAMP pathway.

8.10 Evidence for the involvment of a calcium-binding protein in CF

McPherson et al (1985;1984;1986b) discovered that the phosphodiesterase inhibitor 3,isobutyl-1-methyl-xanthine partially restored the defective β-adrenergic response in the CF submandibular glands. McPherson believes that in the CF exocrine acinar cells the mucin and amylase secretion has switched from β to α adrenergic function. In a scientific correspondance to Nature (1986a) McPherson et al postulate that as cAMP phosphodiesterase represents a point of interaction between cAMP and the calcium regulatory pathway it may be a Ca^{2+} dependent regulatory protein such as calmodulin which is affected in CF. An overactivity of phosphodiesterase (which is responsible for the hydrolysis of cAMP to AMP) could be the result of an alteration in a calcium-dependent regulatory protein. Mangos (1981) also observed increased β-adrenergic secretory responses in CF parotid acinar cells in the presence of a phosphodiesterase inhibitor. Gnegy et al (1981) have found that CF fibroblasts have an increased ability to stimulate phosphodiesterase activity and by inference have increased calmodulin levels. Vladutiu (1984) screened CF patients, their parents and normals by measuring the calmodulin concentration in dried blood spots in a radioimmunoassay. He found a significant elevation of calmodulin levels in both CF and

heterozygotes compared to controls. However the study was small (7 CF and 7 heterozygotes) and the overlaps between groups did not allow diagnosis based on the results. A more recent study carried out by McPherson et al (1986b) showed that calmodulin from CF and normal submandibular glands appeared identical when analysed on reverse phase high pressure liquid chromatography. Therefore the possible calcium -binding regulator does not appear to be calmodulin, although calmodulin activation of cAMP phosphodiesterase by crude extracts from CF cells is increased relative to control extracts. MacPherson (1986) postulates the alteration of a specific modulator of calmodulin function in CF cells

Banschbach et al 1978 discovered that a two fold increase in $^{45}Ca^{2+}$ labelling was observed when normal human leukocytes were incubated in the presence of serum from CF patients. The degree of calcium labelling of patients' leukocytes was significantly greater than that observed for normal leukocytes. This experiment showed that high levels of intracellular calcium may be responsible for the reduced cAMP stimulation seen by Davis et al (1983). Rasmussen et al (1975) and Shapiro et al(1978) had already reported that CF fibroblasts take up significantly more $^{45}Ca^{2+}$ than do normal fibroblasts.

8.11 CFAg

How does CFAg fit into these data? Clearly intracellular calcium levels are important in mediating second messenger pathways, both in cAMP and inositol phosphate stimulation as already discussed. If CFAg is shown physiologically to be a calcium-binding protein (and there is some unpublished data to support this (C.Hayward pers comm)) then it could act to regulate either one of these systems by directly controlling Ca^{2+} levels or binding calcium and specifically activating a protein. The major tissue for CFAg is granulocytes, and this cell type is the most likely source of CFAg when it is present in serum.

The aberrant gene for CF having been located on chromosome 7, means that CFAg cannot be the mutant gene product. However the gene product of the chromosome 7 gene might interact directly with CFAg. The mutation in the CF gene may result in lack of function of its protein product, and as a consequence CFAg levels alone or together with levels of other proteins may then become elevated in serum. The level of CFAg in serum may be raised as a consequence of the CF gene mutation by :-

a) Affecting the localisation of CFAg.

In I-cell disease, fibroblast acid hydrolases become elevated in the plasma due to inappropriate localisation, the enzymes become extracellular rather than intralysosomal. In the same way, CFAg may be mislocalised due to the deficiency of the CF gene product in CF individuals.

The CF gene may affect the granulocyte membrane in some way and increase the level of CFAg in the serum of CF patients due either to leakage of the protein out of the cells or perhaps increased cell fragility.

b) Alternatively the chromosome 7 gene product may have a control function and the aberration of this function in CF, results in misregulation of CFAg.

It is possible that mislocalisation or misregulation of CFAg may result in the disruption of a calcium mediated signal involved in anion channel control. Alternatively, the abnormal accumulation of CFAg in CF patients may be a secondary symptom of the perturbation in ion channel control, or may be an indirectly produced cause of abnormal channel regulation.

8.12 Use of CFAg in heterozygote detection

Hayward et al (1987), have assessed the serum concentrations of CFAg in cystic fibrosis patients and normals, using monoclonal antibodies against CFAg in a sensitive, two-site immunoassay. Unfortunately, although the median values of the serum

concentration of CFAg in homozygotes was significantly higher than heterozygotes, and heterozygotes higher than normals, there was an unacceptable degree of overlap between the different ranges of values. Because CFAg can also be found at high concentration in the serum of patients with myeloproliferative disorders such as CML, the possibility that CFAg presence in serum is due to granulocyte proliferation was investigated. Inividuals with CF have chronic respiratory infections and so it is conceivable that they might have increased granulocyte turnover. This is not true however of CF heterozygotes. Hayward et al (1987) tried to answer this point and improve the heterozygote test by measuring serum levels of another granulocyte protein (lactoferrin) to control for increased granulocyte number or damage, and C-reactive protein to obtain a measure of inflammation. The results could not improve the separation of the three CF genotypes and raised the question whether CFAg has any direct relevance to the CF gene or can be attributed to increased granulocyte proliferation and active tissue damage. However, CFAg is significantly raised in the serum of CF heterozygotes compared to normals but lactoferrin and C-reactive protein are not. Disease controls have raised serum CFAg levels but also have significantly raised lactoferrin and C-reactive protein levels. Hayward et al conclude that the elevated levels of CFAg in the serum of CF heterozygotes is due to the presence of a single copy of the CF gene, while in CF homozygotes there are two elements contributing to the raised level of CFAg. One element due specifically to the CF genotype of the individual and the other resulting from increased granulocyte turnover found in active inflammation. However, as the test stands at the moment, it regretfully cannot be used in heterozygote detection.

8.13 Granulocytes as an affected tissue

Granulocytes have not been closely examined as an affected tissue. Davis et al (1983) has shown that there is reduced β-adrenergic response in granulocytes from CF and heterozygotes. The activity of chloride channels in CF granulocytes has not been examined and, as the tissue is easily available, this seems to be an obvious step to

take. The availability of peripheral blood means that if granulocytes were identified as an affected tissue, physiological studies could be carried out on virtually unlimited supplies of tissue. Samples would be readily available from CF patients, heterozygotes and normals, and would allow clearer analysis of disease associated changes. Furthermore if granulocytes expressed the CF gene defect then the possibility of designing a heterozygote test using these cells would be feasible.

von Tscharner et al (1986) have shown that receptor-agonist stimulated inositol triphosphate causes transient openings of calcium gated cation channels in neutrophils. Inositol trisphosphate triggers the release of calcium from neutrophil intracellular stores (Spat et al 1986) and when the cytosolic free calcium rises above 10^{-7}M the ion channels examined by von Tscharner et al (1986) are opened. However, the channels are inactive when the free calcium concentration falls. It is thus possible to hypothesise that an increase in the level of a protein which binds available free calcium could damp down the calcium signal.

It would be interesting to see if other calcium-binding proteins (possibly as yet undiscovered) have raised levels in other CF tissues, especially in the supposedly affected tissues, i.e pancreas, sweat gland, respiratory tract epithelia. S-100, for example, is present in sweat glands and serous glands of normals (Kahn et al 1983).

The function of CFAg requires further investigation. The HL-60 cell system provides an excellent cell culture system to examine the effect of manipulation of the gene, (overexpression, or absence of a functional protein). S-100 has been reported to associate with a variety of proteins *in vitro* and particularly those with which calmodulin is known to bind. Discovery of specific proteins which bind CFAg may lead to a clear indication of the types of processes in which it is involved and by analogy elucidate the function of the other classes of calcium-binding proteins with sequence similarity to CFAg.

S100, is reported to be highly conserved between mammals and it would be interesting to isolate the rodent homologue of CFAg and determine the degree of species divergence for this gene.

At present, we have not ascertained which tissue in the mouse expresses CFAg. One would predict that mouse myeloid tissue would express the gene because the mouse myeloid cell line WEHI-TG, allowed expression of the human gene in somatic cell hybrids (van Heyningen et al 1985). A variety of mouse tissue RNAs are being analysed by Northern blot at this time.

We have information (Calabretta et al 1987; unpublished) that the gene for 2A9 is located on human chromosome 1, in the region q21-q25. This region overlaps the region in which CFAg lies. It seems very probable that CFAg, and 2A9 are members of the same gene family and could have arisen from an ancestral gene by tandem duplication at some point in evolution. This theory is supported by the fact that the two genes appear to be, physically, very close in the genome. It would be interesting to see if S-100, and the ICaBPs are also located within the centromere-proximal region of the chromosome 1 long arm. This would identify a cluster of genes coding for a group of ancestrally related calcium-binding proteins, and possibly facilitate isolation of other genes within the same family.

ABBREVIATIONS

The following is a list of abbreviations used frequently or not explained fully within the text.

AMP.....ampicillin
bp.....base pair
^{o}C.....degrees centigrade
Ca^{+}.....calcium
cAMP.....adenosine 2':3' cyclic monophosphate
cDNA.....complementary DNA
Ci.....curie
CF.....cystic fibrosis
CFAg.....cystic fibrosis antigen
Cl^{-}.....chloride
CML.....chronic myeloid leukaemia
cpm.....counts per minute
dATP.....deoxyadenosine 5' triphosphate
dCTP.....deoxycytidine 5' triphosphate
dGTP.....deoxyguanosine 5' triphosphate
dNTP.....deoxynucleoside 5' triphosphate
DNA.....deoxyribonucleic acid
dpm.....disintegrations per minute
dTTP.....deoxythymidine 5, triphosphate
E.coli.....*Escherichia coli*
EDTA.....ethidium diaminetetra-acetic acid
g.....grammes
ITS.....insulin, transferrin, sodium selenite
kb.....kilobase
kDa.....kilodalton
µl.....microlitres
McAb.....monoclonal antibody
mRNA.....messenger ribonucleic acid
β-NAD.....β-nicotinamide adenine dinucleotide
Na^{+}.....sodium
poly(A)$^{+}$.....polyadenylated ribonucleic acid
RA.....retinoic acid
RNA.....ribonucleic acid
SDS.....sodium dodecyl sulphate

Tris.....tris (hydroxymethyl) aminomethane
rpm.....revolutions per minute

REFERENCES

Aden, D., Fogel, S., Damjanov, I., Plotkin, B. and Knowles, B. (1979). Controlled synthesis of hepatitis B antigen in a differentiated liver carcinoma-derived cell-line. *Nature*, 282, 615.

Adshead, P., Martinez, J., Kilburn, K. and Hess, R. (1975). Ciliary inhibition and axonemal microtubule alterations in freshwater mussels. *Ann. NY. Acad. Sci.* 253, 192.

Ansah, T. and Katz, S. (1980). Evidence for a calcium transport deficiency in patients with cystic fibrosis. *Cell Calcium* 1,

Arrand, J., (1986). Preparation of nucleic acid probes. In *Nucleic acid hybridisation.* (Eds. Haimes, B and Higgins, S.) pp20-21. Publisher, IRL press.

Athens, J. (1970). In *Regulation of Haematopoiesis,* vol 2. (Ed Gordon, A.), pp1143-1166. Publisher, Appleton, century, crofts. New York.

Aviv, H. and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. U.S.A.* 69, 1408.

Axel, R. and Lemke, G. (1985). Isolation and sequence of a cDNA encoding the major structural protein of a peripheral myelin. *Cell*, 40. 501-508.

Baba, M., Goodman, M., Berger-Cohn, J., Demaille, J. and Matsuda, G. (1984). The early adaptive evolution of calmodulin. *Mol. Biol. Evol.* 16, 442-455.

Babu, Y., Sack, J., Greenhough, T., Bugg, C., Means, A., and Cook, W. (1985). Three dimensional structure of calmodulin. *Nature*: 315, 37-40.

Baggioloni, M., Bretz, U. and Dewald, B. (1978). Subcellular localization of granulocyte enzymes. In: *Neutral proteases of human polymorphonuclear leukocytes.* (Eds Havemann, K. Janoff, A.). Publisher: Urban and Schwarzenberg, Baltimore/Munich. pp3-17.

Bansbach, M., Karam., Love, P. and Hilman, B. (1978). Cystic Fibrosis serum promotes (^{45}Ca) uptake by normal human leukocytes. *Biochem. Biophys. Res. Commun.* 84, 4. 922-926.

Baudier, J. and Gerard D. (1983). Ions binding to S-100 proteins: structural changes by calcium and zinc on S-100a and b proteins. *Biochemistry 22,* 3360-3369.

Bedrossian, C., Greenberg, S., Singer, D., Hansen, J. and Rosenberg, H. (1976). The lung in cystic fibrosis. *Hum Pathol* 7, 195.

Benton, W. and Davis, R. (1977). Screening λgt recombinant clones by hybridisation to single plaques in situ. *Science*, 196, 393.

Berridge, M. and Irvine, R. (1984). Inositol triphosphate, a novel second messenger in cellular signal transduction. *Nature* 312, 315-321.

Bogart, B., Conod, E., and Conover, J. (1977). The biologic activities of cystic fibrosis serum. I. The effects of cystic fibrosis sera and calcium ionophore A 23187 on rabbit tracheal explants. *Pediatr. Res*, 11, 131.

Bogart, B., Conod, E., Gaerlan, P., Denning, C. and Conover, J. (1979). Biological activities of cystic fibrosis serum. III. CF serum induced uptake of $^{45}Ca^{2+}$ By rabbit tracheal explants. *Biochem, Biophys. Res Commun.* 88, 1398.

Bowman, B., Lockhart, L. and McCombs, M. (1969). Oyster ciliary inhibition by cystic fibrosis factor. *Science*, 165, 325.

Bowman, B., Barnett, D., Matalon, R., Danes, B. and Bearn, A. (1973) Cystic fibrosis. Fractionation of fibroblast media, demonstrating ciliary inhibition. *Proc. Natl. Acad. Sci U.S.A.*, &0, 548.

Bowman, B., Lankford, B., McNeely, M., Carson, S., Barnett, D. and Berg, K. (1977). Cystic Fibrosis. Studies with the oyster ciliary assay. *Clin Genet* 12, 333.

Breitman, T., Collins, S. and Keene, B. (1981). Terminal differentiation of human promyelocytic leukaemia cells in primary culture in response to retinoic acid. *Blood*, 57. 1000.

Breitman, T., Selonic, S. and Collins, S., (1980). Induction of differentiation of the human promyelocytic line HL-60 by retinoic acid. *Proc. Natl. Acad. Sci. U.S.A.* 77. 2936.

Breslow, J., Epstein, J., Fontaine, J. and Forbes, G. (1978). Enhanced dexamethasone resistance in cystic fibrosis cells: potential use for heterozygote detection and prenatal diagnosis. *Science*, 201, 180-182.

Breslow, J. McPherson, J. (1981). Sodium transport in cystic fibrosis- fibroblasts not different from normal. *N. Engl. J. Med.* 305, 98.

Breslow, S., Ross, D., McPherson, J., Williams, H., Kurmit, D., Nussbaum, A., Karathanasis, S. and Zannis, V. (1982). Isolation and characterization of cDNA clones for human apolipoprotein A-1. *Proc. Natl. Acad. Sci. U.S.A.*, 79, 6861-6865.

Bretz, U. (1978). Subcellular localization of plasminogen activator in human neutrophils. *Experientia.* 34, 934.

Brodin, P., Grundstrom, T., Hofman, T., Drakenberg, T., Thulin, E., and Forsen, S. (1986). Expression of bovine intestinal calciuym binding protein from the synthetic gene in *Escheria coli* and characterization of the product. *Biochemistry.* 25, 5371-5377.

Brock, D., Hayward, C. and Super, M. (1982). Controlled trial of serum isoelectric focussing in the detection of the cystic fibrosis gene. *Human Genetics*, 60, 30-31.

Brock, D., Bedgood, D., Hayward, C., (1984). Prenatal diagnosis of cystic fibrosis by assay of amniotic fluid microvilar enzymes. *Human Genetics* 65, 248-251.

Bryant, D and Andrews, P. (1984). Investigation of the binding of Ca^{2+}, Mg^{2+}, Mn^{2+} and K^{+} to the vitamin-D dependent calcium-binding protein from pig duodenum. *Biochemistry. J.* 219, 287-292.

Bullock, S., Hayward, C., Manson, J., Brock, D. and Raeburn, A. (1982). Quantitative immunoassays for diagnosis and carrier detection in cystic fibrosis. *Clinical Genetics*, 21, 336-341.

Burgess, W., Watterson, D. and van Eldik, L. (1984). Identification of calmodulin-binding proteins in chicken embryo fibroblasts. *J.Cell.Biol.* 99, 550-557.

Calabretta, B., Kaczmarek, L., Mars, W., Ochoa, D., Gibson, C., Hirschhorn, R. (1985). Cell-cycle -specific genes differentially expressed in human leukaemias. *Proc. Natl. Acad. Sci. U.S.A.* 82, 4463-4467.

Calabretta, B., Battini, R., Kaczmarek, L., Riel, J. and Baserga, R. (1986). Molecular cloning of the cDNA for a growth factor inducible gene with strong homology to S-100, a calcium-binding protein. *J. Biol. Chem.* 261, 12628-12632.

Carbarns, N., Gosden, C. and Brock, D., (1983). Microvillar peptidase activity in amniotic fluid: possible use in prenatal detection of cystic fibrosis. *Lancet* i 329-331.

Case, M. (1984). Calcium, stimulus-secretion coupling and cystic fibrosis. In *Cystic Fibrosis: Horizons.* (Ed Lawson, D.) pp53-67, J. Wiley and sons.

Case, M. (1986). Chloride ions and cystic fibrosis. *Nature*, 322, 407.

Ceder, O and Roomans, G. (1983). Effects of culture medium on cystic fibrosis and normal fibroblasts, studied by X-ray microanalysis. *Ultrastructural Pathology*, 4, 305-309.

Ceder, O., Hosli, P., Vogt, E., and Kollberg. (1983a). Diagnosis of cystic fibrosis homozygotes and heterozygotes from plasma and fibroblast cultures. A three generation family study. *Clinical Genetics.* 23, 298-303.

Champlin, R. and Golde, D. (1985). Chronic Myelogenous Leukemia: Recent advances. *Blood.* 65, 1039.

Chaplinski, T. and Niedel, J. (1982). Cyclic nucleotide-induced maturation of human promyelocytic leukaemia cells. *J. Clin. Invest.* 70, 953.

Chaio, J., Freitag, W., Steinmetz, J. and Andreeff, M. (1981). Changes of cellular markers during differentiation of HL-60 promyelocytes to macrophages as induced by T lymphocyte conditioned medium. *Leukaemia Res* 5, 477-489.

Chase, H., Long, M. and Lavin, M. (1979). Cystic fibrosis and malnutrition. *J. Pediatr* 95, 337.

Cheung, W. (1980). Calmodulin plays a pivotal role in cellular regulation. *Science*, 207, 19-27.

Chikkappa, G., Borner, G., Burlington, H., Chanana, A., Cronkite, E., Pzavelec, S. and Robertson, J. (1976). Periodic oscillation of blood leukocytes, platelets and reticulocytes in a patient with chromic myelocytic leukemia. *Blood*, 47, 1023.

Chirgwin, J., Pryzbyla, R., MacDonald, R. and Rutter, W. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*, 18, 5294.

Chung, S., Hayward, C., Brock, D. and van Heyningen, V. (1986). A monoclonal antibody-based immunoassay for human lactoferrin. *J. Immunol. Meth.* 48, 45-50.

Collins, J. and Coulson, A. (1986). In *Nuclic Acid and Protein Sequence Analysis: A Practical Approach.* Eds Bishop, M. and Rawlings, C. Publisher, I.R.L. Press Oxford.

Collins, S., Gallo, R. and Gallagher, R. (1977). Continuous growth and differentiation of human myeloid leukaemia cells in suspension culture. *Nature.* 270, 347.

Collins, S., Bodner, A., Ting, A. and Gallo, R. (1980). Induction of morphological and functional differentiation of human promyelocytic leukaemia cells (HL-60) by compounds that induce differentiation of murine leukaemia cells. *Int. J. Cancer.* 25, 213.

Coleman, D., Tuet, I. Widdicombe, J. (1984). Electrical properties of dog tracheal epithelial cells grown in monolayer culture. *Am. J. Physiol*, 246, 355-359.

Conner, B., Reyes, A., Morin, C., Itakura, K., Teplitz, R. and Wallace, R. (1983). Detection of sickle cell β^S-globin allele by hybridisation with synthetic oligonucleotides. *Proc. Natl. Acad. Sci.* 80, 278.

Conod, E., Conover, J. and Gaerlan, P. (1977). Separation of serum ciliary dyskenesis substances from cystic fibrosis subjects. *Pediatr Res*, 11, 45.

Cooper, J., Bowen-Pope, D., Raines, E., Ross, R. and Hunter, T. (1982). Similar effects of platelet-derived growth factor and epidermal growth factor on the phosphorylation of tyrosine in cellular proteins. *Cell*, 31, 263-273.

Corradino, R., Fullmer, C. and Wasserman, R. (1976). Embryonic chick intestine in organ culture: stimulation of calcium transport by exogenous vitamin-D induced calcium binding protein. *Arch. Biochem. Biophys.* 174, 738-743.

Coulson, A., Lyall, A. and Collins, J. (1987). Identifying significant similarities between protein sequences in large databases. In Preparation.

Creba, J., Downes, C., Hawkins, P., Brewster, G., Michell, R. and Kirk, C. (1983). Rapid breakdown of phosphatidylinositol-4-phosphate and phosphatidylinositol 4, 5-bisphosphate in rat hepatocytes stimulated by vasopressin and other calcium mobilizing hormones. *Biochem. J.* 212, 733.

Dale, I., Fagerhol, M. and Naesgaard, I. (1983). Purification and partial characterization of a highly immunogenic human leukocyte protein, the L1 antigen. *Eur. J. Biochem.*, 134, 1-6.

Davis, T., Urdea, M., Masiarz, F. and Thomer, J. (1986). Isolation of the yeast calmodulin gene: Calmodulin is an essential protein. *Cel* 47, 423.

Davis, P., Dieckman, L., Boat, T., Stern, R. and Doershuk, C. (1983). Beta Adrenergic Receptors in Lymphocytes and granulocytes from patients with cystic fibrosis. *J. Clin. Invest.* 71, 1787-1795.

Davis, P and Nadel, G. (1983). In *Textbook of cystic fibrosis.* (Eds, J. Lloyd-Still.) Publisher, San Fransisco Press Inc. San Fransisco.

Dayhoff, M., Schwartz, R. and Orcutt, B. (1978). In *Atlas of protein sequence and structure.* vol 5, supplement 3. Ed Dayhoff, M. Publisher, N.B.R.F. Washington, pp 345-352.

Dedman, J., Potter, J. and Means, A. (1977). Biological crossreactivity of rat testis phosphodiesterase activator protein and rabbit skeletal muscle troponin-C. *J. Biol. Chem.*

Desplan, C., Heidmann, O., Lille, J., Auffray, C. and Thomasset, M. (1983). Sequence of rat intestinal vitamin D-dependant calcium-binding protein derived from a cDNA clone. *J. Biol. Chem.*, 258, 13502-13505.

Desplan, C., Thomasset, M. and Moukhtar, M. (1983a). Synthesis, Molecular cloning, and restriction analysis of DNA complementary to vitamin D-dependant calcium binding protein mRNA from rat duodenum. *J. Biol. Chem.* 258, 2762-2765.

Di Sant'Agnese, P. and David, P. (1976). Research in cystic fibrosis. *New Engl. J. Med.* 295, 481.

Dominick, H., Bassewitz, D. and Avens, P. (1976). Ultrastructure of the liver in cystic fibrosis. In *Proceedings of the Seventh International Cystic Fibrosis Congress, Paris.* pp474-477.

Donato, R. (1986). S-100 Proteins. *Cell Calcium.* 7, 123-145.

Donnison, A., Shwachman, Y. and Gross, R. (1966). A review of 164 children with meconium ileus seen at the Children's Hospital Medical Centre, Boston. *Pediatrics*, 37, 833.

Dworkin, M. and David, I. (1980). Use of a cloned library for the study ofabundant poly(A)$^+$ RNA during *Xenopus laevis* development.

Efstratiadis, A., Kafatos, F., Maxam, A. and Maniatis, T. (1976). Enzymatic *in vitro* synthesis of globin genes. *Cell*, 7, 279.

Feinberg. A, and Vogelstein. B., (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem.* 132. 6-13.

Ferrero, D., Pessano, S., Pagliardi, G. and Rovera, G. (1983). Induction of differentiation of human myeloid leukaemias; surface changes probed with monoclonal antibodies. *Blood*, 61, 171.

Fleming, N. and Sturgess, J. (1981). Stimulation of glycoprotein secretion in dispersed rat submandibular gland acini by cystic fibrosis serum. *Experimentia.* 37, 139.

Fondacaro, J., Heubi, J., Kellogg, F. and Ingberg, R. (1980). Bile acid malabsorption in cystic fibrosis. In *Perspectives in cystic fibrosis.* Eds Strurgess, J. Toronto Canadian Cystic Fibrosis Foundation, pp 14a.

Frazier, M., Mars, W., Florine, D., Montagna, R. and Saunders, G. (1983). Efficient extraction of RNA from mammalian tissue. *Molecular and Cellular Biochemistry.* 56, 113-122.

Friedman, D. (1976). Role of cyclic nucleotides in cell growth and differentiation. *Physiol Rev.* 56. 652.

Friedman, E., and Rosbach, M. (1977). The synthesis of high yields of full-length reverse transcripts of globin mRNA. *Nucleic, Acids, Res.* 4, 3455.

Friend, S., Bernards, R., Rogelj, S., Weinberg, R., Rapaport, J., Albert, D. and Dryja, T. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323, 643-646.

Frizzell, R., Rechkemmer, G. and Shoemaker, R. (1986). Altered regulation of airway epithelial cell chloride channels in cystic fibrosis. *Science*, 233, 558-601.

Fullmer, C. and Wasserman, R. (1981). The amino acid sequence of bovine intestinal calcium binding protein.
J. Biol. Chem., 256, 5669-5674.

Galant, S., Norton, L., Herbst, J. and Wood, C. (1981). Impaired beta adrenergic receptor binding and function in cystic fibrosis neutrophil. *J. Clin. Invest.* 68. 253-258.

Galau, G., Klein, W., Britten, R. and Davidson, E. (1976). Significance of rare mRNA sequences in liver. *Arch. Bioch. Biophy.* 179, 584-599.

Gallagher, R., Collins, S., Trujillo, J., McCredie, K., Gilhearn, M., Tsai, T., Metzgar, R., Hulakh, G., Ting, R., Ruscetti, F. and Gallo, R. (1979). Characterization of the continuously differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukaemia. *Blood*, 54, 713.

Gerke, V. and Weber, K. (1985). The regulatory chain in the p36-kd substrate complex of viral tyrosine-specific protein kinases is related in sequence to the S-100 protein of glial cells. *Embo J.* 11, 2917-2920.

Geurts van Kessel, A., Tetteroo, P., von dem Borne, A., Hagemeijer, A. and Bootsma, D. (1983). Expression of human myeloid associated surface antigens in human-mouse myeloid cell hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 80, 3748-3752.

Gibson, L. and Cooke, R. (1959). A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis. *Pediatrics*, 23, 545.

Glenney, J. and Tack, B. (1983). Amino-terminal sequence of p36 and associated p10: identification of the site of tyrosine phosphorylation and homology with S-100. *Proc. Natl. Acad. Sci. U.S.A.*, 82, 7884-7888.

Gnegy, M., Erickson, R. and Markovac, J. (1981) Increased calmodulin in cultured skin fibroblasts from patients with CF. *Biochemical. medecine.* 26. 294-298.

Greenberg, M., Brackenberg, R. and Edelman, G. (1984). Changes in the distribution of the 34 Kd tyrosine kinase substrate during differentiation and maturation of chicken tissues *J. Cell. Biol.* 98, 473-486.

Gubler, U. and Hoffman, B. (1983). A simple and very efficient method for generating cDNA libraries. *Gene.* 25, 263.

Hamlyn, P., Brownlee, G., Cheng, G., Gait, M. and Milstein, C. (1978). Complete sequence of constant and 3' noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing. *Cell*, 15, 1067-1075.

Handler, J., Steele, R., Sahib, M., Wade, J., Preston, A., Lawson, N. and Johnson, J. (1979). Toad urinary bladder epithelial cells in culture: maintenance of epithelial structure, sodium transport, and response to hormones. *Proc. Natl. Acad. Sci.*, 76, 4151-4155.

Harris, A., Benson, P. and Fensom, A. (1979). Detection of cystic fibrosis with plasma hydrolases. *Lancet*, 2, 1371.

Hastie. N. and Bishop, J. (1976). The expression of three abundance classes of messenger RNA in mouse tisues. *Cell*, 9, 761-774.

Hathaway, D., Eaton, C. and Adelstein, R. (1981). Regulation of human platelet derived myosin light chain kinase by the catalytic subunit of cyclic AMP-dependent protein kinase. *Nature*. 291, 252-254.

Hayward, C., Chung, S., Brock, D. and van Heyningen, V. (1986). Monoclonal antibodies to cystic fibrosis antigen. *Jou. Immunol. Meth.*, 91, 117-122.

Hayward, C., Glass, S., van Heyningen, V., Brock, D. (1987). Serum concentrations of a granulocyte derived calcium-binding protein in cystic fibrosis patients and heterozygotes. *Clin Chim Acta* submitted.

Heinz-Erian, P., Dey, R., Flux, M. and Said, S. (1985). Deficient vasoactive intestinal peptide innervation in the sweat glands of cystic fibrosis patients. *Science*, 229,

Helfman, D., Feramisco, J., Fiddes, J., Thomas, G. and Hughes, S. (1983). Identification of clones that encods chicken tropomyosin by direct immunological screening of a cDNA expression library. *Proc. Natl. Acad. Sci. U.S.A.* 80, 31-35.

Helm, C., Graham, M., Dutchick, J. and Olson, M. (1984). A new method of purifying Lambda DNA from phage lysates. *DNA*. 4. 39-49.

Higashida, H., Sano, M. and Kato, K. (1985). Forskolin induction of S-100 protein in glioma and hybrid cells. *J. Cell. Physiol.* 122, 39-44.

Higashida, H and Brown, D. (1986). Two polyphosphatidylinositide metabolites control two K^{+d} currents in a neuronal cell. *Nature*, 323. 333.

Hofmann, T., Kawakami, M., Hitchman, A., Harrison, J. and Dorrington, K. (1979). The amino acid sequence of porcine intestinal calcium-binding protein.
Can. J. Biochem., 57, 737-748.

Hosli, P and Vogt, E. (1979). Detection of cystic fibrosis homozygotes and heterozygotes with plasma. *Lancet*, 2, 543-545.

Houck,C., Rinehart, F. and Schmid, C. (1979). A ubiquitous family of DNA repeated sequences in the human genome. *J. Molec. Biol.* 132, 289-306.

Hoyt, M., Knight, D., Das, A., Miller, H. and Echols, H. (1982). Control of phage λ development by stability and synthesis of CII protein. Role of the viral CIII and host hfl A, him A, and him D genes. *Cell* 31, 565.

Huberman, E. and Callahan, M. (1979). Induction of terminal differentiation of human promyelocytic leukaemia cells by tumor-promoting agents. *Science*. 76, 293.

Hultberg, B., Ceder, O. and Kollberg, H. (1981). Acid hydrolases in sera and plasma from patients with cystic fibrosis. *Clinica. Chimica. Acta.* 112, 167-175.

Huynh, T., Young, R., and Davis, R. (1985). Constructing and screening cDNA libraries inλgt10 and λgt11. In *DNA cloning volume I* (ed Glover, D.) pp49-79. Publisher: IRL Press Limited, Oxford, Washington DC.

Hyden, H. and Lange, P. (1970). S-100 brain protein: correlation with behaviour. *Proc. Natl. Acad. Sci. U.S.A.* 67, 1959-1966.

Isobe, T. and Okuyama, T. (1981). The amino acid sequence of the α-subunit in bovine brain S-100 protein. *Eur. J. Biochem.* 89, 379-388.

Jamieson, A., Aitken, D., Cooke, A., Yaquoob, M., Mckinley, E. and Ferguson-Smith. M, (1984). Investigation of serum proteins in cystic fibrosis using isoelectric focussing and immunoaffinity. In *Cystic Fibrosis: Horizons* (Eds Lawson. D.) pp192, Wiley, J. and sons publisher.

Johansen, P., Anderson, C., Hadorn, B. (1968) Cystic fibrosis of the pancreas: A generalized disturbance of water and electrolyte movement in exocrine tissue. *Lancet*, 1, 455.

Kahn, H., Marks, A., Thom, H. and Baumal, R. (1983). Role of antibody to S-100 protein in diagnostic pathology. *Am. J. Clin. Pathol.* 79, 341-347.

Katznelson, D., Blau, H. and Sack, J., (1981). Detection of cystic fibrosis genotypes. Lancet, 2, 622.

Kennedy, M. and Greengard, P. (1981). Two calcium/calmodulin dependent protein kinases which are highly concentrated in brain, phosphorylate protein I at distinct sites. *Proc. Natl. Acad. Sci. U.S.A.* 78, 1293-1297.

Klein, E., Klein, G., Nadkarni, J., Nadkarni, J., Wigzell, H. and Clifford, P. (1968). Surface IgM-Kappa specificity on a Burkitt lymphoma cell *in Vivo* and in dervived culture lines. *Cancer. Res.* 28, 1300-1310.

Knowles, M., Gatzy, J. and Boucher, R. (1981). Increased bioelectric potential difference across respiratory epithelia in cystic fibrosis. *New Eng. J. Med.*, 305, 1488-1492.

Knowles, M., Gatzy, J. and Boucher, R. (1983a). Relative ion permeability of normal and cystic fibrosis nasal epithelium. *J. Clin. Invest.*, 71, 1410-1417.

Knowles, M., Stutts. M., Spock, A., Fisher, N., Gatzy, J. and Boucher, R. (1983b). Abnormal ion permeation through cystic fibrosis respiratory epithelium. *Science*, 221, 1067-1070.

Knowlton, R., Cohen-Haguenauer, O., Van Cong, N., Frezal, J., Brown, V., Barker, D., Braman, J., Schumm, J., Tsui, L., Buchwald, M., Donnis-Keller, H., (1985). A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* 318, 380-382.

Koeffler, P., Bar-Eli, M. and Territo, M (1981). Phorbol ester effect on differentiation of human myeloid leukaemia cell lines blocked at different stages of maturation. *Cancer. Res* 41, 919.

Koeffler, P. (1983). Induction of differentiation of human acute myelogenous leukaemia cells. Therapeutic implications. *Blood*, 62, 709.

Kretsinger, R. and Nockolds, C. (1973). Carp muscle calcium-binding protein II. Structure determination and general description. *J.Biol. Chem.* 248, 3313-3326.

Kretsinger, R and Barry, C. (1975). The predicted structure of the calcium-binding component of troponin. *Biochim. Biophys. Acta.* 405, 40.

Kretsinger, R. (1980). Structure and evolution of calcium-modulated proteins. *CRC, Critical Rev. Biochem.* 8, 119-174.

Kretsinger, R. (1980a). Crystallographic studies of calmodulin and homologs. *Ann. N.Y. Acad.Sci.* 356, 14-16.

Kuwano, R., Usui, H., Maeda, T., Fukui, T., Yamanari, N., Ohtsuka, E., Ikehara, M. and Takahashi, Y. (1984). Molecular cloning and the complete nucleotide sequence of cDNA to mRNA for S-100 protein of rat brain. *Nucleic. Acids. Res.* 12, 7455-7465.

Kunkel, L. et al (1986). Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. *Nature* 322, 73-77.

Kurnit. D., Wentworth, B., Delong, L. and Villa-Komaroff, L. (1982). Construction of cloned libraries from RNA of human fetal tissues. *Cytogenet. Cell. Genet.*. 34. 193-203.

Labourette, G. and Mandel, P. (1980). Effect of norepinephrine and dibutryl cyclic AMP on S-100 protein level in C6 glioma cells. *Biochem. Biophys. Res. Commun.* 96, 1702-1709.

Lieberman, J., Kaneshiro, W. and Costea, N. (1981). Presence of a serum hamaggluytinin (lectin like factor) in cystic fibrosis homozygotes and heterozygotes. *J. Lab. Clin. Med.* 97, 646-653.

Lipmann, D. and Pearson, W. (1985). Rapid and sensitive protein similarity searches. *Science*, 227, 1435-1441.

Lippe, B., Kaplan, S.,, Neufeld, N., Smith, A. and Scott, M. (1980). Insulin Receptors in cystic fibrosis. Increased receptor number and altered affinity. *Pediatrics*, 65, 821.

Littman, D., Yolene, T., Maddon, P., Chess, L. and Axel, R. (1985). The isolation and sequence of the gene encoding T_8: A molecule defining functional classes of T lymphocytes. *Cell*, 40, 237-246.

Lodish, H., Darnell, J. and Baltimore, D. (1986). In *Molecular Cell Biology*. pp689. Publisher, Scientific American Books.Inc.

Lotem, H. and Sachs, L. (1979). Regulation of normal differentiation in mouse and human myeloid leukaemia cells by phorbol esters and the mechanism of tumor promotion. *Proc. Natl. Acad. Sci. U.S.A.* 76, 5158.

Mangos, J. (1973a). Transductal fluxes of Na^+, K^+, and water in the human eccrine sweat gland. *Am. J. Physiol*, 224, 1235-1240.

Mangos, J. (1973b). Microperfusion study of the sweat gland abnormality in cystic fibrosis. *Tex Rep Biol Med* 31, 651.

Mangos, J. and Donnelly, W. (1981). Isolated parotid acinar cells from patients with cystic fibrosis: functional characterization. *J. Dent Res.* 60, 797-804.

Mangos, J. (1983). Cystosolic Ca^{2+} measurements in buccal epithelial cells from CF patients and controls. Abstract, in *Proceedings of the 12th Annual Meeting of the European Working Group for Cystic Fibrosis*, Athens, Greece, pp21.

Mangos, J. and Boyd, R. (1984). Characteristics of the CF cell. In *Cystic Fibrosis: Horizons.* (Eds Lawson, D.) pp29-50. Johm Wiley and sons.

Maniatis, T., Fritsch, E. and Sambrook, J. (1982). In *Molecular Cloning, A laboratory Manual.* Publisher, Cold Spring Harbour Laboratory, New York.

Manson, J. and Brock, D. (1980). Development of a quantitative immunoassay for the cystic fibrosis gene. *Lancet*, 1, 330-331.

Markovics, B., Gyurkovitz, K., Laszlo, A. and Szabo. L. (1982). Altered peroxide metabolism in erythrocytes from children with cystic fibrosis. *Clin. Chimica. Acta.*, 125, 199-306.

Martinez, J. (1982), Alteration in salivary gland structure and function in cystic fibrosis. In *Fluid and Electrolyte Abnormalities in Exocrine Glands in Cystic Fibrosis.* (Eds Quinton, Martinez and Hopfer), pp125-142. San Fransisco Press, Inc, San Fransisco.

Mayo, B., Klebe, R., Barnett, D., Lankford, B and Bowman B. (1980). Somatic cell genetic studies of the cystic fibrosis mucociliary inhibitor. *Clinical Genetics*, 18, 379-386.

McPherson, M., Dormer. R., Dodge, J. and Goodchild, M.(1984). Autonomic secretory responses of control and CF submandibular acinar cells. In *Cystic Fibrosis: Horizons.* (Ed Lawson, D) pp51, (J.Wiley and sons).

McPherson, M., Dormar, R., Dodge, J. and Goodchild, M. (1985). Adrenergic secretory responses of submandibular tissues from control subjects and CF patients. *Biochem Soc Trans* 13, 967-968.

McPherson, M., Dormar, R., Dodge, J., and Goodchild, M. (1986a). Biochemical basis of cystic fibrosis. *Nature* 323. 400.

McPherson, M., Dormar, R., Dodge, J. and Goodchild, M. (1986b). Defective β-adrenergic secretory responses in submandibular acinar cells from CF patients. *Lancet* 2, 1007.

Messing, J. and Vieria, J, (1982). A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. *Gene*, 19, 269.

Minty, A., Caravatti, M., Benoit, R., Cohen, A., Daubas, P., Weydert, A., Gros, F. and Buckingham, M. (1981). Mouse actin messenger RNAs. *J.Biol. Chem.* 256, 1008-1014.

Monaco, A., Neve, R., Colletti-Feener, C.., Bertelson, C., Kurnit, D. and Kunkel, L. (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* 323, 646-650.

Moore, B., Perez, V. and Gehring, M. (1968). Assay and regional distribution of a soluble protein characteristic of the nervous system. *J. Neurochem.* 15, 265-272.

Murray, N., Brammar, W. and Murrat, K. (1977). Lamboid phages that simplify the recovery of *in vitro* recombinants. *Mol. Gen. Genet. 150, 53.*

Nevin, G., Nevin, N., Redmond, A., Young, I. and Tully, G (1981). Detection of CF homozygotes and heterozygotes by serum isoelectric focusing. *Hum Genet*, 56, 387.

Neufeld, E. and McKusik, V. (1983). Disorders of lysosomal enzyme synthesis and localization. In *Metabolic Basis Of Inherited Disease.* (eds Stanbury, J., Wyn gaarden, J., Fredrickson, D., Goldstein, J. and Brown, M.) pp778-787.

Novak, M., Dorin, J., Hayward, C., Northorp, F., Brock, D., Walker, J., van Heyningen, V. and Secher, D. (1987). Immunopurification and characterization of a cystic fibrosis-associated protein. In preparation.

Okayama, H. and Berg, P. (1982). High Efficiency Cloning of full-length cDNA. *Mol Cell Biol.* 2, 161-170.

Olsson, I., Breitman, T.and Gallo, R. (1982). Priming of human leukaemia cell lines HL-60 and U-937 with retinoic acid for differentiation effects of cyclic adenosine 3':5' monophosphate inducing agents and a T-lymphocyte derived differentiation factor. *Cancer Res* 42. 3928.

Olofsson, T., Odeberg, H. and Olsson, I., (1976). Granulocytic function in chronic granulocytic leukemia. *Blood.* 48, 581.

Oppenheimer, E. and Easterly, J. (1970). Observations on cystic fibrosis of the pancreas. IV. The uterine cervix. *j. Pediatr*, 77, 991.

Oppenheimer, E. (1981). Similarity of the tracheobronchial mucous glands and epithelium in infants with and without CF. *Hum Pathol* 12, 36.

Oppenheimer, E. and Esterly, J. (1973). Cystic fibrosis of the pancreas. Morphologic findings in infants with and without diagnostic lesions. *Arch Pathol*, 96, 149.

Orkin, S., Daddona, P., Shewach, D., Markham, A., Bruns, G. Goff, S., and Kelley, W. (1983). Molecular cloning of human adenosine deaminase gene sequences. *J.Biol. Chem*, 258. 12753-12756.

Ouchterlony, O. (1967). In *Handbook of experimental immunology*, pp 655. Publisher, Elsevier. Amsterdam.

Patel, J. and Marangos, P (1982). Modulation of brain protein phosphorylation by S-100 protein. *Biochem. Biophys. Res. Commun.* 109, 1089-1093.

Patrick, A. and Ellis, R. (1979). Plasma hydrolases in cystic fibrosis. *Lancet*, 2, 1015-1016.

Paul, T., Brandt, I., Elsas, L., Jackson, C., Mamunes, P. and Nance, C., and Nance, W. (1978). Phenylketonuria heterozygote detection in families with affected children. *Am. J. Hum. Genet.* 30. 293.

Pheiffer, B. and Zimmerman, S. (1983). Polymer stimulated ligation: enhanced blunt-ended or cohesive-end ligation of DNA or deoxyribo-oligonucleotides by T_4 DNA ligase in polymer solutions. *Nucleic Acids Res.* 11. 7853-7871.

Peterson, O. and Bear, C. (1986). Two glucagon transducing systems. *Nature* 323. 18.

Pichard, A., Daegelen-Proux, D., Alexander, Y. and Dreyfus, J. (1981). Calmodulin ligands. The interaction of muscle phorylase kinase with phosphodiesterase. Comparison of calmodulin ligands in muscle extracts from normal and phosphorylase kinase-deficient mice. *Biochim. Biophys. Acta.* 657, 84-93.

Pivetta, O., Labal, M. and Sordelli, D. (1979). Serum ciliotoxic activity in mutant mice with some hereditary alterations resembling cystic fibrosis. *Pediatr Res*, !3, 1160.

Quinton, P. and Bijman, J. (1983a). Higher bioelectric potentials due to decreased chloride absorption in the sweat glands of patients with cystic fibrosis. *New Engl. J. Med.* 308, 1185-1189.

Quinton, P. (1983b). Chloride impermeability in cystic fibrosis. *Nature*, 301, 421-422.

Rado, T., Bollekens, J., St Laurent, G., Parker, L. and Benz, E. (1984). Lactoferrin biosynthesis during granuloctopoiesis. *Blood*, 64, 1103-1109.

Rebagliati, M., Weeks, L., Harvey, R. and Melton, D. (1985). Identification and cloning of localised maternal RNAs from Xenopus eggs. *Cell.* 42, 769-777.

Retzel, E., Collet, M. and Faras, A. (1980). Enzymatic synthesis of deoxyribonucleic acid by the avian retrovirus reverse transcriptase in vitro: Optimum conditions required for transcription of large ribonucleic acid templates. *Biochemistry*, 19, 513.

Richardson, C. (1971). Polynucleotide kinase from *Escherichia coli* infected with bacteriophage T_4. *Proc. Natl. Acad. Sci. U.S.A.* 2, 815.

Rigby, P., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labelling deoxyribonucleic acid to a high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113, 237.

Rogers, A. (1983). In *Cells and tissues: an introduction to histology and cell biology.* London Academic Press. pp93-95.

Romeo, G., (1984). Cystic fibrosis; A single locus disease. In "*Cystic Fibrosis Horizons*" (eds Lawson, D.,), pp 155-165. John Wiley and sons.

Rosenthal, S., Canellos, G., DeVita, V. and Gralnick, H., (1977). Blast crisis of chronic granulocytic leukaemia morphologic variants and therapeutic implications. *Am. J. Med.* 63, 542.

Rovera, G., O'Breine, T. and Diamond, L. (1979). Induction of differentiation in human promyelocytic leukaemia cells by tumor promotors. *Science*. 204, 868.

Rovera, G., Santoli, D. and Damsky, C. (1979a). Human promyelocytic leukaemia cells in culture differentiate into macrophage like cells when treated with a phorbol ester. *Proc. Natl. Acad. Sci. U.S.A.* 76, 2779-2783.1

Rowley, J. (1975). Nonrandom chromosomal abnormalities in haematologic disorders of man. *Proc. Natl. Acad. Sci. U.S.A.* 72, 152.

Sack, J., Katznelson, D., Frass, M. and Stur, O. (1980). Heat inactivation of plasma hydrolases in cystic fibrosis. *Lancet*, i, 268.

Sanger, F., Nicklen, S. and Coulson, A. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A*, 74, 5463.

Sanguinetti-Briceno, N. and Brock, D. (1982). NADH dehydrogenase in cystic fibrosis. *Clinical Genetics*, 22, 308-311.

Sato, K. and Sato, F. (1983). Cholinergic potentiation of isoproproterenol-stimulated cyclic AMP accumulation in the sweat gland. *Am. J. Physiol.* 245, (Cell physiol. 14): C189-C195.

Sato, K. and Sato, F. (1984). Defective beta-adrenergic response of cystic fibrosis sweat glands *in vivo* and *in vitro*. *J. Clin. Invest.* 73, 1763-1771.

Scambler, P., Law, H., Williamson, R. and Cooper, C. (1986). Chromosome mediated gene transfer of six DNA markers linked to the cystic fibrosis locus on human chromosome seven. *Nucleic Acids Res* 14, 7159-7174.

Scholey, J., Applegarth, D., Davidson, A. and Wong. L. (1978). Letter to the editor. Detection of cystic fibrosis protein by isoelectric focussing. *Pediatr. Res. 12; 800.*

Schulz, I., (1969). Micro puncture studies of the sweat formation in cystic fibrosis patients. *J. Clin Invest*, 48, 1470.

Schwachman, H., Kulczycki, L., Mueller, H. and Flake, C. (1962). Nasal polyposis in patients with cystic fibrosis. *Pediatrics*, 30, 389.

Sealy, P., Whittaker, P. and Southern, E. (1985). Removal of repeated sequences from hybridisation probes. *Nucleic. Acids. Res.* 13, 6.

Shapiro, B. and Lam, L. (1982). Calcium and age in fibroblasts from control subjects and patients with cystic fibrosis. *Science*, 216, 417-419

Shapiro, B., Lam, L. and Feigal, R. (1982a). Mitochondrial NADH dehydrogenase in cystic fibrosis: Enzyme kinetics in cultured fibroblasts. *American Journal Human Genetics.*, 34, 846-852.

Shelling, J. and Sykes, B.(1985). ^{1}H nuclear magnetic resonance study of the two calcium binding sites of porcine intestinal calcium-binding protein. *J.Biol. Chem.* 260. 8342-8347.

Smith, T. and Waterman, M. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195-197.

Sood, A., Pereira, D. and Weissman, S. (1981). Isolation and partial nucleotide sequence of a cDNA clone for human histocompatabimity antigen HLA-DR by use of an oligonucleotide primer. *Proc. Natl. Acad. Sci. U.S.A.*, 78, 616-620.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503.

Spat, A., Bradford, P., McKinney, J., Rubin, R. and Putney, J. (1986). A saturable receptor for ^{32}P-inositol-1,4,5-trisphosphate in hepatocytes and neutrophils. *Nature*. 319, 514.

Spock, A., Heick, H., Cress, H. and Logan, W. (1967). Abnormal serum factor in patients with cystic fibrosis of the pancreas. *Pediatr Res*, 1, 173.

Spicer, S., Briggman, J. and Baron, D. (1982). Morphological and cytochemical correlates of transport in sweat glands of normal and cystic fibrosis subjects. In *Fluid and Electrolyte Abnormalities in Exocrine Glands in Cystic Fibrosis.* (Eds Quinton, Martinez and Hopfer), pp11-34, San Fransisco Press, Inc, San Fransisco.

Stetler, D., Das, H., Nunberg, J., Saiki, R., Sheng-Dong, R., Mullis, K., Weissman, S. and Erlich, H. (1982). Isolation of the cDNA clone for the human HLA-DR antigen α chain by using a synthetic oligonucleotide as a hybridisation probe. *Proc. Natl. Acad. Sci. U.S.A.*, 79, 5966-5970.

Stutts, J., Cotton, C., Yankaskas, J., Cheng, E., Knowles, M., Gatzy, J. and Boucher, R. (1985). *Proc. Natl. Acad. Sci.*, 82, 6677-6681.

Suggs, S., Hirose, T., Miyake, T., Kawashima, E., Johnson, M., Itakura, K. and Wallace, R. (1981). In *Developmental Biology using purified genes.*Eds Brown, D. and Fox, C. p683.

Sundaralingham, M., Drendel, W. and Greaser, M. (1985). Stabalization of the long central helix of troponin-C by intrahelical salt bridges between charged amino acid side chains. *Proc. Natl. Acad. Sci. U.S.A.* 82. 7944.

Super, M., Swindlehurst, C., and Super. M, (1984). Isoelectric focusing of serum in genetic counselling of families with cystic fibrosis. In *Cystic Fibrosis: Horizons.* (Ed Lawson, D.) pp 193, Wiley, J. and sons, Publisher.

Szebenyi, D., Odendorf, S. and Moffat, K. (1981). Structure of vitamin D-dependant calcium-binding protein fron bovine intestine. *Nature*, 294, 327-332.

Talamo, R., Rosenstein, B. and Berninger, R. (1983). Cystic Fibrosis. In *The Metabolic Basis Of Inherited Disease.* (Eds Stanbury, J., Wyngaarden, J., Fredrickson, D., Goldstein, J. and Brown, M.) pp1889-1917. McGraw-Hill, New York.

Tabuchi, K., Imada, M. and Nishimoto, A. (1982). Effect of cyclic AMP on S-100 protein level in C6 glioma cells. *J. Neurol. Sci* 56, 57-63.

Taussig, L., Lobeck, C., Di Sant'Agnese, P., Ackerman, D. and Kattwinkel. (1972). Fertility in males with cystic fibrosis. *N. Engl. J. Med*, 287, 586.

Thomassen, M., Demko, C., Wood, R., Tandler, B., Dearborn, D., Boxerbaum, B. and Kuchenbrod, P (1980). Ultrastructure and function of alveolar macrophages from cystic fibrosis patients. *Pediatr Res*, 14, 715.

Thomasset, M., Cuisinier-Gleizes, P. and Mathieu, H. (1979). 1,2, Dihydroxycholecalciferol: dynamics of the stimulation of duodenal calcium-binding protein, calcium transport and bone calcium mobilization in vitamin D and calcium deficient rats. *FEBS Lett*, 107, 91-94.

Tourain, A. and Sidbury. (1983). Phenylketonuria and hyperphenylalanemia. In *The metabolic basis of inherited disease..* (Eds Stanbury, J., Wyngaarden, J., Fredrickson, D., Goldstein, J. and Brown, M.) pp270.

Tully, G., Nevin, G., Young, I. and Nevin, N. (1979). Letter to the editor. Detection of cystic fibrosis protein by isoelectric focusing of serum. *Pediatr Res* 13, 1078.

van Eldik, L., Zendegui, J., Marshak, D. and Watterson, D. (1982). Calcium-binding proteins and the molecular basis of calcium action. *Int Rev Cytol.* 77, pp 1-61.

van Elsen, A., Leroy, J., Vanneuville, F. and Vercruyssen, A.(1976). Isoenzymes of serum N-acetyl-beta-D-glucosaminidase in the I-cell heterozygote. *Hum Genet* 31, 75.

van Heyningen, V., Brock, D., and van Heyningen, S. (1983). A simple method for ranking the affinities of monoclonal antibodies. *Journal of Immunological Methods.* 62, 147-153.

van Heyningen, V., Hayward, C., Fletcher, J. and McAuley, C. (1985). Tissue localization and chromosomal assignment of a serum protein that tracks the cystic fibrosis gene. *Nature*, 315, 513-515.

van Heyningen, V., Boyd, P., Seawright, A., Fletcher, J., Fantes, J., Buckton, K., Spowart, G., Porteous, D., Hill, R., Newton, M. and Hastie, N. (1985a). Molecular analysis of chromosome 11 deletions in aniridia-Wilms tumor-syndrome. *Proc. Natl. Acad. Sci U.S.A.*, 82, pp8592-8596.

Vidgoff, J. and Buist, N., (1977). Serum hexosaminidase activity in I-cell disease carriers. *Hum Genet* 36, 307.

Vieria, J. and Messing, J. (1982). The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene.* 19, 269.

Vladutiu, G. (1984). Calmodulin quantitation in cystic fibrosis. *IRCS Med. Sci.* 12, 515-516.

von Tscharner, V., Prod'hom, B., Baggiolini, M. and Reuter, H. (1986). Ion channels in human neutrophils activated by a rise in cytosolic free calcium concentration. *Nature* 324. 369.

Wainwright, B., Scambler, P., Schmidtke, J., Watson, E., Law, H., Farrall, M., Cooke, H., Eiberg, H. and Williamson, R., (1985). Localisation of cystic fibrosis locus to human chromosome 7cen-q22.

Nature 318, 384-385.

Wakelam, M., Murphy, G., Hruby, V. and Houslay, M. (1986). Activation of two signal-transduction systems in hepatocytes by glucagon. *Nature*. 323. 68.

Wasserman, R. and Feher, J. (1977). In *Calcium-binding proteins and calcium function.* (Eds Wasserman, R. Corradino, R., Carafoli, R., Kretsinger, R., MacLennan, D. and Siegal, F.) pp293-302. Elsevier Amsterdam.

Welsh, M. and Liedtke, C. (1986a). Chloride and potassium channels in cystic fibrosis airway epithelium. *Nature*, 322, 467- 470.

Welsh, M. (1986b). An apical membrane chloride channel in human tracheal epithelium. *Science* 232, 1648.

Welsh, M. and McCann, J. (1985). Intracellular calcium regulates basolateral potassium channels in a chloride secreting epithelium.*Proc. Natl. Acad. Sci. U.S.A.* 82, 8823-8826.

Whitehead, A., Goldberger, G., Woods, D., Markham, A. and Colten, H. (1983). Use of a cDNA clone for the fourth component of human complement (C4) for analysis of a genetic deficiency of C4 in guinea pigs. *Proc. Natl. Acad. Sci. U.S.A.*, 80, 5387-5391.

Wieslander, L. (1979). A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. *Anal. Biochem* 98, 305.

White, R., Woodward, S., Leppert, M., O'Connell, P., Hoff, M., Herbst, J., Lalouel, J., Dean, M. and Vande Woude, G., (1985). A closely linked genetic marker for cystic fibrosis. *Nature* 318, 382-384.

Widdicombe, J., Welsh, M. and Finkbeiner, W. (1985). Cystic fibrosis decreases the apical membrane chloride permeability of monolayers cultured *Proc. Natl. Acad. Sci. U.S.A.*, 82, 6167-6171.

Wiesmann, U., Boat, T. and Di Sant'Agnese, P. (1972). Flow-rates and electrolytes in minor salivary gland saliva in normal subjects and patients with cystic fibrosis. *Lancet* 2, 510.

Williams, J. and Lloyd, M. (1979). Changes in the abundance of polyadenylated RNA during slime mould development measured using cloned molecular hybridisation probes. *J. Mol. Biol* 129, 19-35.

Wilson, A., Harding, M. and Lawson, D. (1985). Putative amino acid sequence of chick calcium-binding protein deduced from a complementary DNA sequence. *Nucleic Acids Res.* 13, 8867-8881.

Wilson, G., Fundenberg, H., and Jahn, T. (1975), Studies on cystic fibrosis using isoelectric focussing: an assay for detection of cystic fibrosis homozygotes and heterozygote carriers in serum. *Paediatric Research*, 9, 635-640.

Wilson, G and Fundenberg, H. (1977). Ciliary dyskenesis factor in cystic fibrosis and asthma. *Nature*, 266, 463.

Wilson, G., and Fundenberg, H. (1977b). Studies on cystic fibrosis using isoelectric focusing. IV. Distinction between ciliary dyskenesia activity in cystic fibrosis and asthmatic sera and association of cystic fibrosis protein with the activity in cystic fibrosis serum. *Pediatr. Res.*, 11, 317.

Wilson, G., Arnaud, P. and Fundenberg, H. (1978). Improved method for the detection of cystic fibrosis protein in the serum, using the LKB multiphor electrofocussing apparatus. *Paediatr Research*, 12, 801-804.

Wilson, G. and Bahm, V. (1980). Synthesis and secretion of cystic fibrosis ciliary dyskenesia substances by purified sub-populations of leukocytes. *J. Clin. Invest.* 66, 1010.

Wood, R, and diSant'Agnese, P. (1973). Bioassays of cystic fibrosis factor. *Lancet*, 2, 1452-1453.

Yankaskas, J., Knowles, M., Gatzy, J. and Boucher, R. (1985). Persistance of abnormal chloride ion permeability in cystic fibrosis nasal epithelial cells in heterologous culture. *Lancet*, 1, 954-956.

Young, R. and Davis, R. (1983). Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad. Sci. U.S.A.* 80, 1194-1198.

Zebronski, A., Billing, R., Mikulski, S, Gale, R. and Terasaki, P. (1977). Human B-lymphocyte antigens expressed by lymphocyticand myelocytic leukaemia cells: lymphocyte dependent antibody studies with rabbit antiserum. *Leuk Res* 1, 13-18.

Zimarkoff, J., Hoiby, N., Rosendal, K. and Guilbert, J., (1983). Antibiotic treatment of *Pseudomonas aeruginosa* of CF. *Jou Hosp Inf*, 4, 31-40.

A clue to the basic defect in cystic fibrosis from cloning the CF antigen gene

Julia R. Dorin*, Michal Novak†§, Robert E. Hill*, David J. H. Brock‡, David S. Secher† & Veronica van Heyningen*

* MRC Clinical and Population Cytogenetics Unit, and ‡ Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK
† MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The metabolic basis of the autosomal recessive disease cystic fibrosis (CF) remains unidentified. Elevated levels of a serum protein in CF homozygotes and obligate heterozygotes have been described[1]. As heterozygotes are clinically unaffected, any consistently observed abnormality in these individuals is a likely pointer to the aetiology of the disease. The gene for this serum protein, called cystic fibrosis (CF) antigen, has been mapped to chromosome 1 (ref. 2). It is not the gene that is mutant in CF because this has since been assigned to chromosome 7 by cosegregation of the disease with closely linked DNA markers in CF families[3–5]. CF antigen is a product of normal and leukaemic granulocytes[2] and is inducible in the promyelocytic cell line HL60 (M.N., J.D., C. Hayward, F. Northrop, D.J.H.B., J. Walker, V. van H. and D.S.S., manuscript in preparation). We have isolated cDNA clones for this protein from a library constructed with messenger RNA from chronic myeloid leukaemia (CML) cells. The complete nucleotide sequence was obtained from the cDNA clone and by primer extension of mRNA. We have confirmed that the gene encoding CF antigen is on chromosome 1 and have localized it to a particular region. RNA blot analysis shows a 550-bp major transcript in CML cells and in induced HL60. The amino-acid sequence predicted from the nucleotide sequence shows significant homology with intestinal and brain calcium-binding proteins. Abnormal accumulation of such a protein in CF is a clue which must be pursued now that evidence is gathering that the basic defect in CF is in pathways controlling chloride channel activity[6,7].

Quantification of serum levels of CF antigen by immunoassay shows that the three genotypes (CF homozygotes, heterozygotes and normals) fall into three overlapping classes[1]. Monoclonal antibodies, recognizing two epitopes on this protein, have been produced[8]. These antibodies have now been used for immunopurification and sequence-comparison of the antigen from a variety of sources (M.N. *et al.*, in preparation).

Partial amino-acid sequence analysis of the purified protein allowed us to synthesize a highly redundant oligonucleotide probe, corresponding to the region indicated in Fig. 1 (M.N. *et al.*, in preparation), which was used to isolate the cDNA clone CFA 8-9. The library screened was constructed in λgt10, using mRNA prepared from CML cells. Antigen turnover had previously been demonstrated in such cells by the specific immunoprecipitation of labelled protein following ^{35}S-methionine incorporation. The identity of CFA 8-9 was confirmed by matching the predicted and known amino-acid sequence outside the oligonucleotide parent region.

The nucleotide sequence of CFA 8-9 is shown in Fig. 1. The insert of 338 nucleotides includes an open reading frame of 250 base pairs (bp) which encodes 83 amino acids stretching from the first amino acid used for the construction of the oligonucleotide probe, to the C-terminal residue before the chain-termination codon. The sequence encoding the first 11 N-terminal amino acids, known from the protein sequence data, is also shown in Fig. 1. The nucleotide sequence for this and the 5′-untranslated region (Fig. 1) was determined by oligonucleotide primer-extension using mRNA from DMSO-induced HL-60 as template. The relative molecular mass (M_r) of 10,938 deduced for the complete sequence of 94 amino acids, is in agreement with the approximate subunit M_r observed for immunoprecipitated antigen (M.N. *et al.*, in preparation). There are no *N*-glycosylation sites predicted from the amino-acid sequence and there is no evidence that CF antigen is a glycoprotein. The 3′ untranslated region of the mRNA is unusually short. On re-screening the CML cDNA library with CFA 8-9 as probe, approximately 0.5% of the recombinants were positive, suggesting that the gene product is relatively abundant in these cells.

RNA blot analysis (Fig. 2) on a variety of tissues and cell lines was carried out to examine message size and abundance. In all cases where mRNA hybridizing to the CFA 8-9 insert was observed, two bands were seen. The major band of approximately 550 bp was of at least 20-fold higher intensity than the second 1 kilobase transcript. Total mRNA levels loaded were monitored by reprobing with an actin cDNA probe[9]. Poly(A)$^+$ mRNA isolated from two CML samples, one of which had been used for the construction of the cDNA library, contained very high levels of CFA 8-9 transcripts. Equivalent amounts of mRNA isolated from uninduced HL60 cells gave no detectable signal, whereas HL60 cells induced to differentiate along the myeloid pathway[10] possess both size transcripts. However, retinoic acid (RA) and dimethyl sulphoxide (DMSO) induce

```
Sense strand:
        -51
        CTCTTGTCAGCTGTCTTTCAGAAGACCTGGTGGGGNAAGTCCGTGGGCATC

1                                G**************              60
ATGTTGACCGAGCTGGAGAAAGCCTTGAACTCTATCATCGACGTCTACCACAAGTACTCC
(M  L  T  E  L  E  K  A  L  N  S  I  I  D  V  Y  H  K  Y  S

                                    CFA 8-9
                                                              120
CTGATAAAGGGGAATTTCCATGCCGTCTACAGGGATGACCTGAAGAAATTGCTAGAGACC
 L  I  K  G  N  F  H  A  V  Y  R  D  D  L  K  K  L  L  E  T

                                                              180
GAGTGTCCTCAGTATATCAGGAAAAAGGGTGCAGACGTCTGGTTCAAAGAGTTGGATATC
 E  C  P  Q  Y  I  R  K  K  G  A  D  V] W  F  K  E  L  D  I

                                                              240
AACACTGATGGTGCAGTTAACTTCCAGGAGTTCCTCATTCTGGTGATAAAGATGGCGTGG
 N  T  D  G  A  V  N  F  Q  E  F  L  I  L  V  I  K  M  A  W

                                                              300
CAGCCCACAAAAAAAGCCATGAAGAAAGCCACAAAGAGTAGCTGAGTTACTGCCCAGAGG
 Q  P  T  K  K  A  M  K  K  A  T  K  S  S  !

CTGGGCCCCTGACATGTACCTGCAGAATAATAAAGTCATCAATACCTC(polyA)
          ------
```

Fig. 1 Nucleotide and corresponding amino-acid sequence of the cDNA clone CFA 8-9. The probable polyadenylation signal of the mRNA is underlined. Asterisks, the oligonucleotide used to isolate the clone; !, termination codon; vertical bar, start of probe; brackets, N-terminal amino acids determined by protein sequencing (M.N. *et al.*, in preparation). The nucleotide sequence from −51 to +33 was determined by reverse transcriptase primer extension using induced HL-60 cell mRNA as a template.
Methods. RNA was prepared[30] from 10^{11} CML cells collected by leukophoresis. Poly(A)$^+$ RNA was isolated[31] and used as a template for cDNA synthesis[32]. The resulting double-stranded cDNA was cloned in λgt10 by standard procedures[33]. Titration on *Escherichia coli* Hfl strain revealed a library of 70,000 independent recombinants. The amplified library was screened by the method of Benton and Davis[34]. The 144-fold redundant 15-mer oligonucleotide probe was labelled with polynucleotide kinase using 10 pmol of 5′ ends and 200 μCi of [γ-^{32}P]ATP per reaction[35]. Filters were washed in 2×SSC containing 0.1% SDS for 90 min at 37 °C. Positive plaques were recloned repeatedly and then subcloned into M13mp9 for sequencing. Both strands of the clone with the largest insert (CFA 8-9) were sequenced by the dideoxy chain termination method[36]. Reverse transcriptase mediated primer extension and dideoxy sequencing of first strand cDNA was carried out following Novak *et al.* (manuscript in preparation).

§ Permanent address: Institute of Virology 81703, Bratislava, Czechoslovakia.

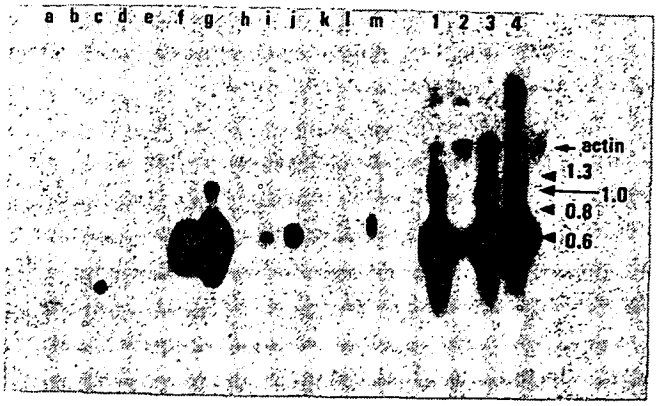


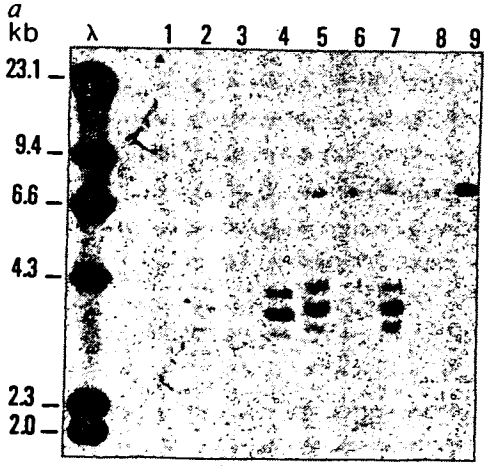
Fig. 2 RNA (Northern) blot analysis of various human cell lines and tissues: a, Hep G2(hepatoma)[37]; b, EMBK (embryonic kidney, ATCC No. CRL 1573); c, MNNG-HOS (oesteosarcoma, ATCC No. CRL 1547); d, Daudi (Burkitt lymphoma derived B-lymphoblastoid line)[38]; e, hepatoma; f, CML #1; g, CML #2; h, uninduced HL60; i, HL60 induced (2 days RA); j, normal liver; k, WEHI TG (mouse myeloid stem cell)[12]; l, WEP 1342.2.4 (mouse × human myeloid cell hybrid containing human chromosomes 7, 10, 13, 20)[2]; m, WEH 3127.1.9 (myeloid cell hybrid containing human chromosomes 1, 9, 14, 21 and expressing CFAg)[2]. Lane 1, CML #1; 2, HL60 induced (2 days RA); 3, HL60 induced (5 days DMSO); 4, CML #2. Both blots were probed with CF antigen gene cDNA, and reprobed with mouse actin to ensure comparable sample loading (reprobe shown for lanes 1–4 only).

Methods. Cells were grown in RPMI 1640 medium supplemented with 5% fetal calf serum, except for uninduced HL60 which was grown in serum-free medium supplemented with insulin, transferrin and selenite (ITS, Sigma). HL60 induction was by continuous culture with 1.2% (v/v) DMSO or 10^{-6} M RA. RNA was isolated as in Fig. 1. Then poly(A)$^+$ RNA (2 μg in lanes a, b, c, k, l, m), total RNA (20 μg in lanes d–j) or poly(A)$^+$ RNA (1 μg in lanes 1–4) was separated on a 1.2% agarose/formaldehyde gel and transferred directly to nitrocellulose[35]. Hybridization was at 68 °C with the insert from CFA 8-9 randomly primed with a mixed sequence hexadeoxynucleotide (Pharmacia), filled in using Klenow enzyme and radiolabelled with [α-^{32}P]dTTP. Filters were washed in 2×SSC/0.1% SDS at 68 °C for 90 min.

low and high relative message abundance respectively. A possible explanation for this may be that cells treated with RA and DMSO do not reach the same degree of maturation along the myeloid pathway. After equivalent treatment regimes only 10% of DMSO-treated cells are recognizable as mature neutrophils[10,11], compared with 30% for RA treatment. Messenger RNA isolated from a variety of other malignant human cell lines (embryonic kidney, osteosarcoma, hepatoma and Burkitt's lymphoma) gave no detectable signal. Human liver mRNA gave a weak signal which we have assumed to be due to blood contamination of the original tissue, because we have demonstrated antigen turnover in normal granulocytes, and immunohistochemical analysis of liver sections reveals no antigen in hepatocytes (V. van H. *et al.*, unpublished). No detectable signal was seen with mRNA from the mouse myeloid stem cell line WEHI-TG[12], or in somatic cell hybrids between WEHI-TG and human CML cells unless human chromosome 1 was present in such hybrids[2]. It appears therefore that there is no synthesis of a homologous cross-hybridizing mRNA in this mouse myeloid line. No signal distinguishable from the human 550-bp band was found in the 'switched on' chromosome 1 containing hybrids. Thorough immunohistochemical analysis of normal tissues revealed CF antigen only in a few non-keratinizing squamous epithelia (tongue and oesophagus), in addition to granulocytes (manuscript in preparation).

Human and mouse DNA digested with restriction enzymes *Pst*I, *Bam*HI, *Eco*RI and *Hin*dIII give a single hybridizing band with CFA 8-9 insert even with low stringency washes. Digestion with *Bgl*II gives three bands but of small size, consistent with restriction sites within the gene. This suggests the existence of only one gene for this protein in both these species, and the absence of any closely related cross-hybridizing genes. Mouse–human somatic cell hybrids, regardless of whether they express CF antigen[2], exhibit a band of human mobility only when chromosome 1 is present at high frequency (Fig. 3*a*). The confirmation that this gene is on chromosome 1 is additional evidence that the cloned insert corresponds to the immunologically recognized protein[1,2]. The availability of cell hybrids containing defined fragments of chromosome 1 has allowed us to sublocalize the gene locus to the region between bands q12 and q22 on the long arm (Fig. 3*b*).

a



b

Hybrid \ Human chromosomes present	Fragment of 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	CFAg
	Chromosomal localization of the gene for CF antigen																								
Rag J8	pter-q21								■				■												+
SK82	pter-q12				■												q					■			-
SK81	pter-p31					■		■								■		■		■	■	■			-
NYX 5.6	complete	not fully characterized																							+
NYX 5.6.11	pter-q23	not fully characterized																							+
F9.6	complete	■	■					■		■		■		■	■	■		■			■		■		+
WEH 3127.1.9 T_7	complete								■					■											+
WEH 3127.1.9 T_8	absent								■					■											-

Fig. 3 *a*, Southern blot analysis[39] to illustrate mapping of CFA 8-9 in rodent/human somatic cell hybrids. Lane 1, NYX5.6.11[40]; 2, F9.6 (unpublished data); 3, NYX5.6[40]; 4, CML#3; 5, WEH 3127.1.9 T7[2]; 6, WEH 3127.1.9 T8; 7, WEH 3127.1.9 T7; 8, WEH 3127.1.9 T8; 9, WEHI-TG. Size markers (λ fragments) in kb. *b*, Summary of chromosome analysis of the location of the gene for CF antigen (CFAg). Segregation of the CFA 8-9 hybridizing, human-specific restriction fragment with chromosome 1 or fragments of it. WEH 3127.1.9 T7 and T8 are subclones of the same hybrid[2], FACS sorted for expression and non-expression respectively of a chromosome-1-encoded cell surface marker[41]. Black squares denote the presence of chromosomes in >50% of cells scored; q, long arm only present. The restriction fragments corresponding to the human gene for CFA 8-9 (+ in CFAg column) are seen whenever the entire chromosome 1 is present and also in hybrid RAG J8 with human chromosome 1 from the tip of the short arm to band q21 on the long arm (pter-q21) present. No human CFA8-9 bands are present (− in CF Ag column) in hybrid SK 82 which contains only region pter-q12 of chromosome 1. The gene must therefore lie between bands q12 and q22 on chromosome 1. Other hybrid cell data support this result.

Methods. In *a*, ~10 μg of genomic DNA was digested with *Bgl*II, fractionated on 0.8% agarose gels and transferred to nitrocellulose filters. After hybridization with labelled cDNA probe CFA 8-9 insert, blots were washed in 2×SSC/0.1% SDS at 68 °C.

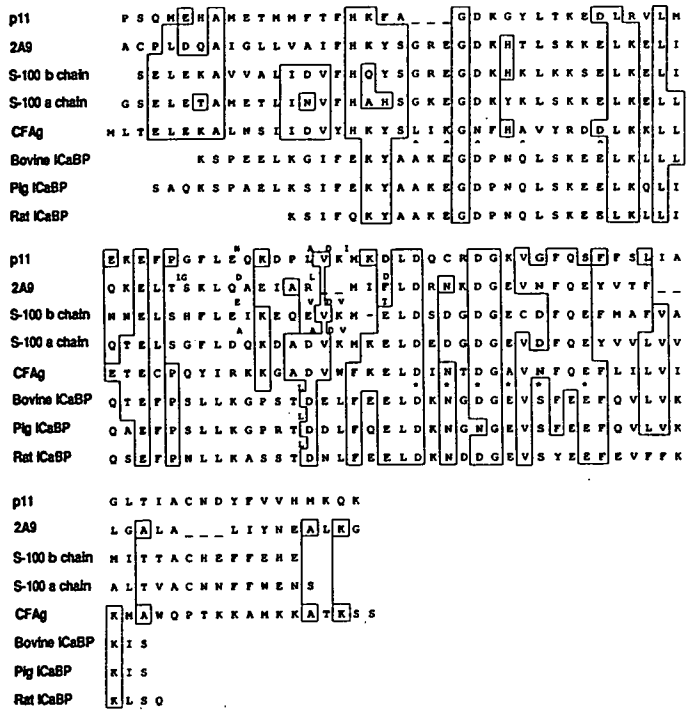


Fig. 4 Alignment of the predicted amino-acid sequence of CF antigen with amino-acid sequences of ICaBP from cow, pig and rat; bovine S100a and b subunits; porcine p11 and human 2A9. The calcium binding ligands, as determined by X-ray crystallography[20] for the EF-hand region of bovine ICaBP, are marked with asterisks. Calcium-binding ligands marked with triangles are from the second distinct type of calcium-binding region in ICaBPs and S100. Amino-acid residues which are identical in CF antigen and at least one of the other sequences are boxed.

The deduced amino-acid sequence of the cloned CF antigen gene was compared with all sequences in the NBRF database by a method which permits high stringency comparison with all the sequences present[13]. Highly significant homology (optimized score: 189, 43% identity over a stretch of 72 amino acids) was obtained with bovine S100a subunit, a brain-associated calcium-binding protein[14], and the similarity extends to the homologous S100b subunit. Strong homology was also seen with bovine intestinal calcium-binding protein (BICaBP)[15] and its pig and rat homologues. These proteins all belong to a class of calcium-binding proteins which are postulated to be the major transducers of biological calcium signals[16]. Like S100 proteins, but unlike the intestinal calcium-binding proteins which are monomeric, native cystic fibrosis antigen is predominantly in dimeric form under non-reducing conditions (M.N. *et al.*, manuscript in preparation). Amino-acid comparisons are shown in Fig. 4. Two additional sequences, not yet in the NBRF database but recently shown to be similar to S100 proteins, are included. These are: the sequence for porcine p11 (or p10), a regulatory subunit of the protein complex which is a major cellular target for tyrosine kinase[17,18]; and the deduced sequence for a cell-cycle specific human cDNA molecule, 2A9[19], the protein product of which has not yet been studied.

In the well studied calcium-binding proteins there are two types of calcium-binding sites[20]. Amino acids which act as ion-binding ligands in the conventional EF hand-like site[20] are marked with asterisks in Fig. 4. These are well conserved in cystic fibrosis antigen, suggesting that calcium binding remains a normal function of this protein. The criteria for the conservation of calcium-binding capacity at the second site (marked with triangles) are less clearly defined[20]. Equilibrium binding studies, using $^{45}Ca^{2+}$ in competition with unlabelled CA^{2+}, have shown that CF antigen binds two atoms of calcium per molecule of protein and with an affinity similar to that shown by S100 protein (D.J.N.B. and C. Hayward, unpublished data).

The most widely accepted abnormality demonstrable at the cellular level in CF is defective transport of chloride ions across epithelia[21]. However, recent patch-clamp analyses of chloride channel activity in epithelial cells of the airway, a tissue affected in CF, suggest that the conductance characteristics of isolated, cell-free chloride channels from CF tissue are indistinguishable from those observed for normal cells[6,7]. When similar studies are carried out with whole cells, the β-adrenergic stimulation of chloride channel activity observed in normal cells is absent in CF cells. As cAMP generation is unaltered in such CF cells[6,22], deviation from normal function might be sought further on in the cAMP pathway or in other signal transducing systems. Activation of the phosphatidyl inositol pathway[23] leads to mobilization of intracellular calcium. Alteration of any component of this system is a candidate for the basic defect in CF tissue. Recent investigations have suggested the existence of more than one receptor, capable of responding through different pathways to signals by certain ligands[24,25]. Control of chloride channel activity may fall into this category.

The most likely tissue source for the elevated serum levels of cystic fibrosis antigen is granulocytes. This would imply that myeloid cells are an affected tissue in CF. There have been suggestions that this is the case[26,27], although no clinically relevant granulocyte abnormalities have been reported[28]. Investigation of ion transport with particular reference to calcium metabolism in granulocytes might be fruitful. Calcium-modulated activation of ion channels has been demonstrated in normal human neutrophils[29]. Confirmation that the basic defect is expressed in granulocytes would be an important step forward for CF research because these cells are relatively easily accessible from affected and heterozygous individuals.

Demonstration of granulocytes as a tissue affected in CF would suggest that elevated serum levels of cystic fibrosis antigen can be attributed to the absence (in homozygotes) or reduction (in heterozygotes) in these cells of the normal product of the CF gene on chromosome 7 which may be necessary for correct assembly of a multi-molecular complex.

We thank Dr Andrew Coulson for running the database comparison, Dr John Goldman for supplying leukophoresis samples, Richard Buckland for supplying some hybrid-cell DNA, Judy Fletcher for the chromosome analyses, Richard Meehan and Nick Hastie for invaluable advice and discussion and Professor H. J. Evans for support. J.R.D. is the recipient of an MRC Training Award. M.N. is a Royal Society Exchange Fellow from the Slovak Academy of Sciences, Bratislava. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00278.

Received 25 November 1986; accepted 11 February 1987.

1. Bullock, S., Hayward, C., Manson, J. C., Brock, D. J. H. & Raeburn, J. A. *Clin. Genet.* **21**, 336-341 (1982).
2. Van Heyningen, V., Hayward, C., Fletcher, J. & McAuley, C. *Nature* **315**, 513-515 (1985).
3. Knowlton, R. G. *et al. Nature* **318**, 380-382 (1985).
4. White, R. *et al. Nature* **318**, 382-384 (1985).
5. Wainwright, B. J. *et al. Nature* **318**, 384-385 (1985).
6. Welsh, M. J. & Liedtke, C. M. *Nature* **322**, 467-470 (1986).
7. Frizzel, R. A., Rechkemmer, G. & Shoemaker, R. L. *Science* **233**, 558-560 (1986).
8. Hayward, C., Chung, S., Brock, D. J. H. & Van Heyningen, V. *J. immun. Meth.* **91**, 117-122 (1986).
9. Minty, A. J. *et al. J. biol. Chem.* **256**, 1008-1014 (1981).
10. Breitman, T. R., Scolnick, S. E. & Collins, S. J. *Proc. natn. Acad. Sci. U.S.A.* **277**, 2936-2940 (1980).
11. Ferrero, D., Pessando, S., Pagliardi, G. & Rovera, G. *Blood* **61**, 171-179 (1983).
12. Geurts van Kessel, A. H. M., Tetteroo, P. A. T., von dem Borne, A. E. G., Hagemeijer, A. & Bootsma, D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3748-3752 (1983).
13. Lyall, A., Hill, C., Collins, J. F. & Coulson, A. F. W. in *Parallel Computing '85* (eds Feilomeier, M., Joubert, G. & Scheudel, U.) 235-240 (North-Holland/Elsevier, Berlin, 1986).
14. Isobe, T. & Okuyama, T. *Eur. J. Biochem.* **116**, 79-86 (1981).
15. Fullmer, C. S. & Wasserman, R. H. *J. biol. Chem.* **256**, 5669-5674 (1981).
16. Van Eldrick, L. J., Zendegui, J. G., Marshak, D. R. & Watterson, D. M. *Int. Rev. Cytol.* **77**, 1-61 (1982).

17. Gerke, V. & Weber, K. *EMBO J.* **4**, 2917–2920 (1985).
18. Glenney, J. R. & Tack, B. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7884–7888 (1985).
19. Calabretta, B., Battini, R., Kaczmarek, L., de Riel, J. K. & Baserga, R. *J. biol. Chem.* **261**, 12628–12632 (1986).
20. Kretsinger, R. H. *Ann. N.Y. Acad. Sci.* **356**, 14–19 (1980).
21. Case, M. *Nature* **322**, 407 (1986).
22. Sato, K. & Sato, F. *J. clin. Invest.* **73**, 1763–1771 (1984).
23. Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315–321 (1984).
24. Michell, R. H. & Houslay, M. *Trends biochem. Sci.* **11**, 239–241 (1986).
25. Wakelam, M. J. O., Murphy, G. J., Hruby, V. J. & Houslay, M. *Nature* **323**, 68–71 (1986).
26. Galant, S. P., Norton, L., Herbst, J. & Wood, C. *J. clin. Invest.* **68**, 253–258 (1981).
27. Davis, P. B., Dieckman, L., Boat, T. F., Stern, R. C. & Doershuk, C. F. *J. clin. Invest.* **71**, 1787–1795 (1983).
28. Talamo, R. C., Rosenstein, B. J. & Berninger, R. W. in *The Metabolic Basis of Inherited Disease* (eds Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S.) 1889–1917 (McGraw-Hill, New York, 1983).
29. Von Tscharner, V., Prod'hom, B., Baggiolini, M. & Reuter, H. *Nature* **324**, 369–372 (1986).
39. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
31. Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412 (1972).
32. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263–269 (1983).
33. Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning: A Practical Approach* Vol I (ed. Glover, D.) 49–78 (IRL Press, Oxford, 1985).
34. Benton, W. & Davis, R. *Science* **196**, 180–182 (1977).
35. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
36. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
37. Aden, D. P., Fogel, A., Damjanov, I., Plotkin, B. & Knowles, B. *Nature* **282**, 615 (1979).
38. Klein, E. *et al. Cancer Res.* **28**, 1300–1310 (1968).
39. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
40. Van Heyningen V. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8592–8596 (1985).
41. Andrews, P. W. *et al. Ann. hum. Genet.* **49**, 31–39 (1985).