COMPARISON OF EXPRESSED SEQUENCES IN CELLS FROM NORMAL PERSONS AND CYSTIC FIBROSIS PATIENTS.

BY

BERWYN EWART CLARKE

a thesis submitted for the degree of

Doctor of Philosophy

in the

University of London

September, 1981

Department of Biochemistry,
St. Mary's Hospital Medical School,
London, W2 1PG

ABSTRACT

Cystic fibrosis is the most common autosomal disorder in most Caucasian populations. The basic biochemical defect has not been identified but the disease is manifested as a disorder of all the exocrine glands in the body. In order to detect the basic defect at the genetic level, expressed mRNA sequences from normal and cystic fibrosis patients in two types of cell were compared. This entailed the construction of bacterial libraries representing all the poly A^+ RNA molecules from these cell types using recombinant DNA technology. Poly A^+ RNA was isolated from normal and CF human lymphocytes, and normal and CF human fibroblasts. DNA copies were then made from these molecules and double-stranded for insertion into the plasmid vector pAT153. Several methods are available for inserting these sequences into the vector, including blunt end ligation, tailing and the use of synthetic restriction enzyme recognition sites or linkers. These methods were studied and in this laboratory the method used was tailing. cDNA from each cell type was inserted into the PstI site of pAT153 by G-C tailing and after transformation into E.coli HB101, representative recombinant libraries from normal lymphocyte and normal and cystic fibrosis fibroblasts were obtained. These libraries were then used in normal and competition hybridisation experiments to identify clones which are

differently expressed in cystic fibrosis cells as compared to normal cells.

Twenty-eight clones were picked which gave different intensities of signal when screened by the method of Grunstein and Hogness with normal and cystic fibrosis mRNA. These clones were analysed by similar experiments with a view to detecting a reproducible defect in cells from cystic fibrosis individuals.

To my parents and Eluned.

To Saccharomyces carlsbergensis, firstly, and Escherichia coli, secondly, without whom none of this would have been possible.

ACKNOWLEDGEMENTS

I would like to thank everyone in the department for their willing help and advice over the last three years but particularly Professor Bob Williamson for providing the opportunity to work in his laboratory.

I am indebted to Derek Woods for the groundwork of this project, for his supervision over the first two years, his constant encouragement and for putting up with a lot of abuse at times.

I am equally indebted to Julian Crampton for his supervision for the last year, for his considerable technical advice and assistance, for time spent deciphering my writing and for the hybridisation data in Chapters 3 and 5.

Thanks to Carol for the early typing and to everyone in Wales, Kent and London who helped to keep me sane.

Last, but not least, thanks to my family for their support and particularly to Jill for her encouragement, understanding and for being a source of strength and inspiration when times got rough. Also thanks for the bulk of the typing and the diagrams.

Finally, apologies to the other occupants of the office for continually lowering the tone, as I was often reminded.

ABBREVIATIONS

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
mRNA	messenger RNA
rRNA	ribosomal RNA
hnRNA	heterogeneous nuclear RNA
tRNA	transfer RNA
cDNA	complementary DNA
A,C,G,T,U	Adenine,Cytosine,Guanine,Thymine, Uridine.(Prefix d denotes deoxy)
E.coli	Escherichia coli
BSA	bovine serum albumin
EDTA	ethylenediaminetetraacetic acid
SDS	sodium dodecyl sulphate
PCA	perchloric acid
TCA	trichloroacetic acid
TRIS	2-amino-2(hydroxymethyl)-propane-1,3-diol
HEPES	N-2-hydroxyethyl piperazine-N-2-ethane sulphonic acid
DEAE	diethylaminoethyl
RNase	ribonuclease
DNase	deoxyribonuclease
EtOH	ethanol

ABBREVIATIONS continued

kb	kilobase pairs
dpm	disintegrations per minute
cpm	counts per minute
PPO	2,5-Diphenyloxazole
POPOP	1,4-Di-2-(5-phenyloxazolyl) benzene
M	molar
w/v	weight per volume
v/v	volume per volume
g	gravity
rpm	revolutions per minute
UV	ultraviolet
ABM	amino benzyl oxymethyl
DBM	diazobenzyloxymethyl

CONTENTS

Abstract

Dedication

Acknowledgements

Abbreviations

Chapter 1 Introduction

A. CYSTIC FIBROSIS

1.1 Genetics 1

1.2 Clinical features 8

1.2.1 Gastrointestinal. 8

1.2.2 Pulmonary. 10

1.2.3 Sweat electrolyte elevation. 11

1.2.4 Prognosis 12

1.3 Research on CF 14

1.3.1 Abnormalities in CF cells in culture. 15

1.3.2 The Cystic Fibrosis factors. 17

1.3.3 Membrane studies in CF. 19

1.3.4 Lysosomal enzymes and CF. 22

1.3.5 The Immune and Complement Systems. 23

1.3.6 Development of antenatal and neonatal screening for CF. 25

B.CONTROL OF GENE EXPRESSION

1.4 Introduction. 31

1.5 Sequence organisation of human nuclear DNA 31

1.6 The role of interspersed repetitive sequences in control of gene expression. 34

1.7 Gene rearrangement and repetitive sequences. 38

1.8 Sequence requirements for gene expression. 41

1.9 Other levels of transcriptional control. 44

1.10 Post-transcriptional regulation-hnRNA and mRNA. 46

1.10.1 Introns and splicing. 47

1.10.2 Polyadenylation. 50

1.11 Translational and post-translational control. 51

1.12 The concept of abundance classes. 52

1.13 Aims and objectives. 54

Chapter 2 Materials and methods.

2.1 Materials. 56

2.2 General Methods. 60

2.2.1 Restriction enzyme digestion. 60

2.2.2 Ethanol precipitation of nucleic acids. 60

2.2.3 Agarose gel electrophoresis. 61

2.2.4 Visualisation of DNA+RNA. 61

2.2.5 Gel filtration. 62

2.2.6 Deionization of formamide. 62

2.2.7 Autoradiography. 63

2.2.8 Determination of radioactivity. 63

2.3 Bacteriological Methods. 64

2.3.1 Bacterial strains and plasmid vectors. 64

2.3.2 Preparation of plasmid DNA. 64

2.3.3 Transformation of HB101. 66

2.3.4 Transformation of MC 1061. 68

2.3.5 Preparation of single colony lysates. 68

2.3.6 Replica plating. 70

2.3.7 Sealing nitrocellulose filters. 70

2.3.8 Storage of bacteria. 71

2.3.9 Containment and safety. 72

2.4 Specific Methods. 73

2.4.1 Isolation of lymphocyte RNA. 73

2.4.2 Oligo (dT) cellulose method of poly A^+ RNA isolation. 74

2.4.3 Synthesis of complementary DNA (cDNA). 77

2.4.4 Synthesis of double stranded cDNA. 77

2.4.5 Homopolymer tailing. 78

2.4.6 Annealing of tailed DNA. 79

2.4.7 Nick translation of DNA. 79

2.4.8 Labelling of RNA by T4 polynucleotide kinase. 80

2.4.9 Hybridisation of ^{32}P poly A^+ RNA to colony DNA on filters. 80

2.4.10 Hybridisation of ^{32}P cDNA to colony DNA on filters. 81

2,4.11 Hydroxyapatite chromatography. 81

2.4.12 RNA:cDNA hybridisations. 82

2.4.13 Southern transfers. 84

2.4.14 ^{3}H-Poly (U) assay of Poly A^{+} RNA. 85

2.5 Tissue Culture. 86

Chapter 3 Construction of cDNA libraries from normal lymphocytes, normal fibroblasts and Cystic Fibrosis fibroblasts. 88

3.1 Introduction. 88

3.2 Preparation of poly (A)$^{+}$RNA. 89

3.2.1 Use of ^{3}H poly (U) to assay poly A^{+} RNA. 93

3.3 Synthesis of complementary DNA (cDNA). 97

3.4 Synthesis of double stranded cDNA. 98

3.5 Cloning in the host/vector system E.coli HB101/pAT153. 104

3.6 Cloning cDNA libraries by blunt end ligation. 109

3.6.1 Preparation of the plasmid. 110

3.6.2 Ligation of linearized pAT153 and ds cDNA for cloning. 113

3.7 Cloning cDNA libraries using synthetic restriction enzyme recognition sites. 117

3.7.1 Methylation and blunt ending of cDNA. 118

3.7.2 Phosphorylation of linkers and ligation to DNA. 120

3.7.3 Preparation of the vector, ligation and transformation. 122

3.8 Cloning by homopolymer tailing. 127

3.8.1 Tailing the vector. 128

3.8.2 Tailing and cloning lymphocyte cDNA. 132

3.8.3 Tailing and cloning normal and CF fibroblast ds cDNA. 135

3.9 Storage and analysis of the libraries. 137

3.10 Stability of the libraries. 141

3.11 Discussion. 146

3.11.1 Synthesis of double stranded cDNA. 146

3.11.2 Relative merits of cloning methods. 150

3.11.3 Cloning and characterisation of the libraries. 156

3.11.4 Sequence representation in the libraries. 159

Chapter 4 Comparison of poly A^+ RNA sequences in CF and normal cells. 162

4.1 Introduction. 162

4.2 ^{32}P labelling of RNA with polynucleotide kinase. 162

4.3 Colony hybridisation to lymphocyte cDNA library. 163

4.4 Comparison of expressed sequences in lymphocyte RNA from different normal individuals. 164

4.5 Preselection of highly expressed lymphocyte clones. 170

4.6 Hybridisation of ordered arrays with CF and normal lymphocyte RNA. 173

4.7 Quantitation of abundance of CDL-3 and CDL-6 in CF and normal lymphocyte poly A^+ RNA. 178

4.8 Use of fibroblast libraries in competition hybridisation. 179

4.9 The use of ^{32}P cDNA in Grunstein-Hogness hybridisation. 182

4.10 Discussion.

4.10.1 Use of comparative hybridisation. 186

4.10.2 Use of competitive hybridisation. 189

4.10.3 Increasing sensitivity of screening. 191

Chapter 5.Summary and perspectives. 193

References. 196

LIST OF FIGURES.

Figure No.		Page
1.2.1	Pathogenesis of cystic fibrosis.	9
1.4	Biosynthetic pathway of typical protein.	32
3.2.1	2% agarose/SDS gel of total RNA.	91
3.2.1.1	Titration of mouse globin RNA with ^{3}H poly U.	94
3.2.1.2	Fractionation of poly A^{+} RNA on 15-30% sucrose gradients.	96
3.3.1	1% agarose gels of ^{32}P single stranded cDNA.	99
3.4.1	Analysis of double stranded lymphocyte cDNA on 1% agarose gels.	102
3.4.2	Sucrose gradient profiles of normal and CF fibroblast ds cDNA.	103
3.4.3	Autoradiograph of sucrose gradient fractions of lymphocyte double stranded cDNA analysed on 1% agarose gel.	105
3.4.4	Analysis of double stranded lymphocyte cDNA on 1% agarose gel.	106
3.4.5	Analysis of fibroblast and CF fibroblast double stranded cDNA on 1% agarose gel.	107
3.6.1.1	1% agarose gel of pAT153 DNA.	111
3.6.2	Size analysis of blunt end ligation clones on 1% agarose gel.	116
3.7.1.1	Use of EcoR1 digestion of lambda DNA to monitor methylation of cDNA.	119

3.7.1.2 Protection of EcoR1 sites in SmaI digested lambda DNA by methylation. 121

3.7.2.1 Ligation of EcoR1 digested pAT153. 123

3.7.2.2 Digestion of lambda DNA with EcoR1 in the presence of EcoR1 linkers. 124

3.7.3.1 1% agarose gel showing ligation of EcoR1 pAT153 before and after alkaline phosphatase treatment. 126

3.8.1.1 1% agarose gel of ligated and redigested PstI pAT153. 129

3.8.1.2 Varying parameters of tailing reaction. 131

3.8.1.3 a) Autoradiograph of 1% agarose gel of ^{32}P dGTP tailed pAT153. 133

b) and c) TaqI digestion of ^{32}P tailed PstI pAT153. 133

3.9.1 Purification of plasmid DNA on caesium chloride gradient. 139

3.9.2 1% agarose gel electrophoresis of DNA from single colony bacterial lysates. 140

3.9.3 Southern transfer of lymphocyte cDNA library plasmid. 142

3.9.4 Southern transfer of CF fibroblast library plasmid. 143

3.10.1 1% agarose gel of ampicillin resistant recombinants. 145

3.10.2 Growth curves of individual clones. 147

4.3.1 Grunstein-Hogness hybridisation of lymphocyte cDNA library. 165

4.3.2 Use of Grunstein-Hogness hybridisation in the presence of 10% dextran sulphate. 166

4.4.1 Hybridisation of lymphocyte cDNA clones with ^{32}P RNA from normal individuals. 169

4.4,2 Southern transfer of male and female DNA with a putative 'Y' specific clone. 171

4.5.1 Grunstein-Hogness hybridisation of lymphocyte cDNA library. 174

4.6.1 Grunstein-Hogness hybridisation of preselected lymphocyte clones with CF and normal lymphocyte RNA. 176

4.6.2 Rescreen of selected clones with ^{32}P normal and CF lymphocyte RNA. 177

4.8.1 Use of competition hybridisation. 181

4.8.2 Grunstein-Hogness hybridisation after pre-hybridisation of probe and competitor. 183

4.9.1 Grunstein-Hogness hybridisation using cDNA. 185

1. INTRODUCTION

A. CYSTIC FIBROSIS

Cystic fibrosis (CF) is an inherited disease, the biochemical defect of which is unknown. Clinically it is characterised by chronic pulmonary disease, pancreatic insufficiency and elevated sweat electrolytes.

1.1. Genetics

Cystic fibrosis is the most common lethal genetic disease of most Caucasian populations (di Sant Agnese, 1976), incidences at birth varying from 1 in 620 in South West Africa (Dutch descendants) (Super,1975) to a more general figure of 1 in 2000 (di Sant Agnese,1967) (see Table 1). Although fewer incidence figures are available for non-Caucasians the relative infrequency of reported cases (Table 1) suggests that the disease is rare enough in these populations to be due to sporadic mutation. The disease affects males and females with equal frequency and its incidence in live births corresponds to heterozygote frequency of 1:20 if a single gene defect is involved (Conover 1976). Genetic studies have been hampered for two main reasons:-

A) Only CF females are fertile, males being aspermatic (Denning 1968) and only recently has life expectancy become great enough for females to reach reproductive age.

B) CF has only been recognised as a discreet entity for the past

Table 1 Incidence of CF in various populations

Incidence	Population	Reference
1:90,000	United States (Hawaiian Asians)	Wright (1968)
1:17,000	Washington D.C. (Negroes)	Kulozycki (1974)
1: 3,700	United States	Wright (1968)
1: 3,300	Germany	Vivell (1963)
1: 3,200	France	Bernheim (1961)
1: 3,000	England	Carter (1967)
1: 2,450	Australia	Danks (1965)
1: 2,400	England	Hall (1968)
1: 2,000	Italy	Romeo (1981)
1: 1,000	United States	Goodman (1952)
1: 620	South West Africa	Super (1975)

40 years, being classified previously under at least 13 different names (Anderson 1959).

In spite of this, numerous large studies have been carried out into the mode of inheritance of this disease. Most of these studies (Steinberg 1960, Merritt 1962) have concluded that CF is an autosomal recessive disease but several studies indicated that the number of sibs of CF patients who also had the disease exceeded the theoretical figure of 25 per cent which would be expected for an autosomal recessive disease. (Brunecky, 1972, Bernheim 1968). For this reason an alternative model was proposed (Schaap + Cohen, 1976) suggesting that CF may be determined not by a pair of recessive alleles, but by dominant alleles at two polymorphic autosomal loci with interaction between them. However a recent study in Italy (Romeo 1981) has produced the most conclusive evidence so far that the defect behaves as an autosomal recessive. In a recessive defect the parents of the patient must be heterozygotes, by definition. Obviously if a heterozygote were to marry a member of his own family eg. a cousin , the frequency of the disease would be greater than in the general population. Conversely an increased frequency of a particular disease associated with consanguineous marriages is indicative of an autosomal recessive disease. In Italy, consanguineous marriages require special dispensations from the Catholic Church. By careful survey of the records and relationships with the

occurrence of cystic fibrosis, an average threefold increase in frequency of CF children was observed among the offsprings of consanguineous marriages, with respect to the general population, which supports the autosomal recessive mode emphatically.

The very high incidence of CF in many Caucasian populations poses a major genetic problem. A population frequency in this range is more characteristic of random polymorphisms and is about 10 times too high to be explained in terms of recurrent mutation resulting in deleterious genes. Any explanation of the high frequency must take into consideration that the mechanism does not operate in non-Caucasians. Several theories have been put forward, one being that a mutation at any one of multiple sites within a gene might produce CF (Conneally, 1973). Most genetic disorders are, in fact, found to be heterogeneous, arising from multiple mutant alleles at a single locus, i.e., from different alterations of a single gene product, or mutation at several different loci, each capable of producing the same or similar clinical appearance. There is some biochemical and clinical evidence for heterogeneity. Studies of skin fibroblast cultures from CF patients have shown that the patients can be divided into two classes on the basis of the presence or absence of metachromasia (a staining property of cationic dyes with negatively charged macromolecules) in the cultured cells. (Danes and Bearn, 1968). Within the metachromasia four morphologically

distinguishable classes are present and heterozygous parents have the same metachromatic class as their CF children, suggesting that multiple loci do exist. Clinically, CF is manifested by a wide variation of expression.

One of the major diagnostic features of the disease, however is the presence of elevated sweat electrolytes. Recently a study (Stern et al.,1978) has described seven patients where even this parameter does not hold, the patients having intermediate electrolyte levels and apparently normal exocrine pancreatic function. These patients are postulated to be either homozygous for a portion of a "closely linked multigenic CF locus" or that they may have some kind of modifier genes that allow the expression of the pancreatic and sweat chloride defect. Another study (Davis,1980) has described a patient with an electrolyte level which is actually significantly lower than normal and yet has all other symptoms of CF. These studies therefore strongly suggest that there may be genetic variants of CF. On theoretical grounds, if CF could result from mutation at any one of several different loci, the observed incidence would represent the sum of the incidence of the specific disorders.

The hypothesis favoured by most workers on CF is that of heterozygote advantage and numerous theories have been proposed to explain this. One study (Knudson et al.,1967) has suggested that

females heterozygous for the CF gene have a greater reproductive fitness than normal individuals while another (Brackenridge, 1978) has noted that children born to CF heterozygote parents are conceived with seasonal distribution, which may have relevance if the pattern shows increased levels of reproductive performance as compared with the general population.

The most favoured theory of heterozygote advantage, however, is of increased resistance to disease just as malaria resistance has contributed to the high heterozygote frequency of sickle cell anaemia (Allison, 1954). Resistance of heterozygotes to tuberculosis has been suggested (Crawfurd, 1972) as this used to be endemic in Europe and accounted for many deaths among children and young adults at a time when half of all newborns failed to reach the age of 20 years. Family studies (Crawfurd, 1975) have shown that there is a lower incidence of TB among parents of CF patients than among parents of controls, so this is a very good candidate if disease resistance is the cause of heterozygote advantage.

Clearly identification of the chromosomal location of the CF gene(s) would be useful in several ways: it could settle the question of the number of gene loci involved and it might allow for prenatal diagnosis in some cases, if a close linkage to a polymorphic marker could be found. A few attempts have been made to establish linkage relations of CF (Goodchild et al, 1976,

Herrod, 1979) but no strong evidence for linkage, with any of the markers used, was observed, although linkage with the ABO, Rh or HLA loci was excluded. The most promising work on localisation of the CF gene has come from somatic cell genetic studies (Mayo et al 1980). A protein fraction displaying biological activity in mucociliary assays has been associated with CF genotypes; it has been detected in plasma (Conover, 1973), saliva (Bowman et al, 1969) urine (Mc. Neely et al 1980), and medium of cultured diploid fibroblasts (Bowman et al, 1973) derived from CF heterozygotes, with its concentration being about one half the value in heterozygotes that it is in homozygotes. Previous studies (Mayo 1976) using somatic cell hybrids indicated that this cystic fibrosis mucociliary inhibitor (CF MI) phenotype was associated with the presence of one or more human autosomes. It is possible to construct somatic cell hybrids in tissue culture between human and mouse cells which retain single human chromosomes. By correlation of the CF MI phenotype with the presence of specific human chromosomes a possible linkage was observed between CF MI and human chromosomes 2, 4, 6, 10 and 18 with the strongest chance of linkage existing for chromosome 4. The implications of this finding will be discussed later (See Discussion).

1.2. Clinical Features

Heterozygotes for CF are asymptomatic and the symptoms are extremely variable and complex. This makes it extremely difficult to define a symptom as a primary result of the defect or as a secondary symptom such as hepatic cirrhosis. However, it is well established that the primary diagnostic characteristic involves abnormal secretion from the exocrine glands (di Sant Agnese, 1976) These abnormal secretions give rise directly or indirectly to all the symptoms shown in Fig. 1.2.1. Although the severity of the disease is so variable, several features are now accepted to be pathognomonic of the disease ie. meconium ileus (blockage of the intestine at birth), elevated sweat electrolyte level and more recently the presentation, in males, of chronic pulmonary disease and infertility (Colten et al, 1981). As shown in Fig. 1.2.1. there are three main areas of diagnostic characterization and these will be briefly described in turn.

1.2.1. Gastrointestinal.(GI)

These complications are mainly due to pancreatic blockage caused by mucus precipitating in the pancreatic ducts.

a) Meconium ileus - this is the earliest problem associated with CF and is usually recognized between 24 and 48 hours after birth. It is an intestinal obstruction characterized by dilatation of the ileum and thickening of the intestinal walls. It is not

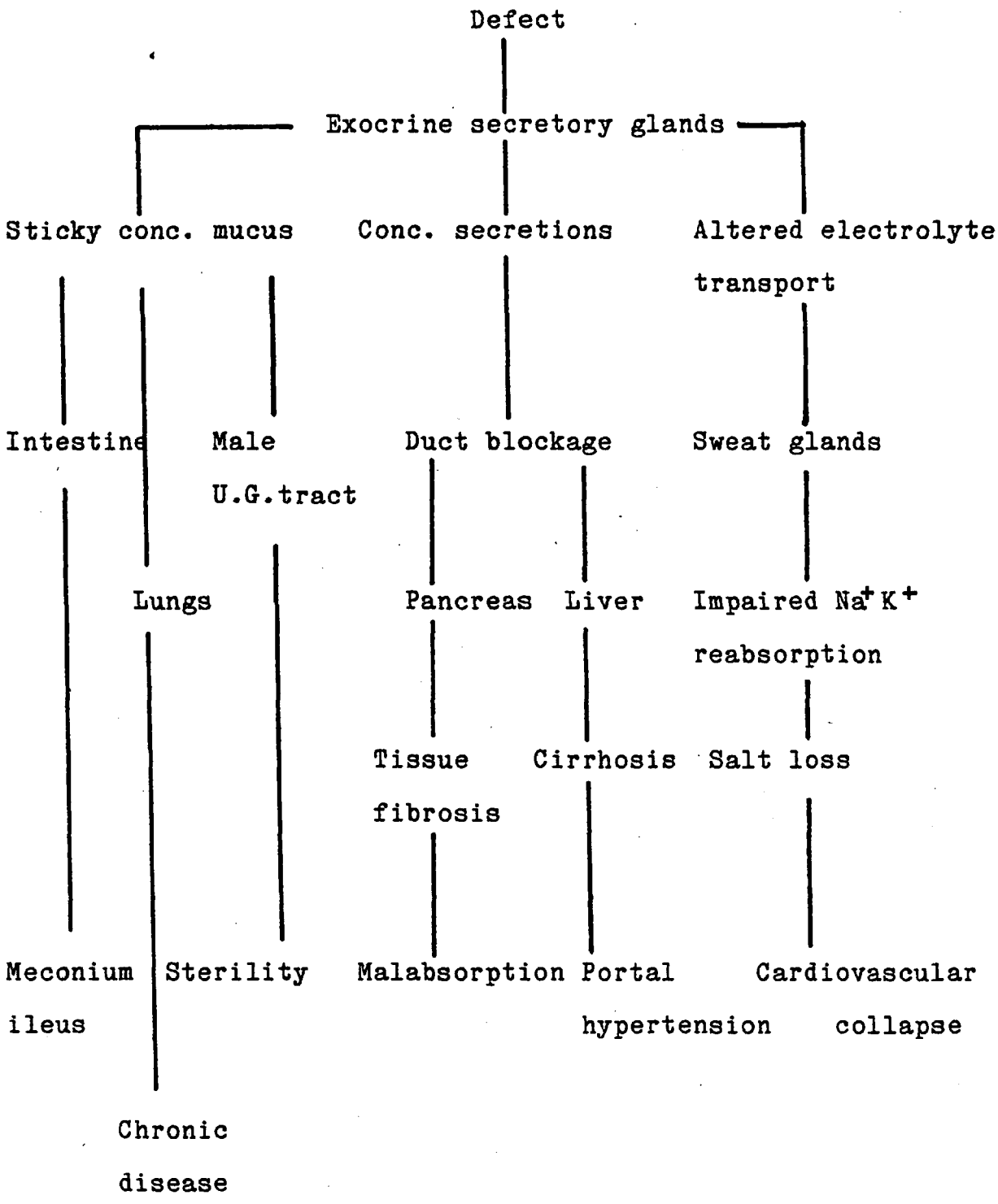


Fig.1.2.1. Pathogenesis of cystic fibrosis

from Anderson and Goodchild,1976.

always present in CF but is extremely rare in any other syndrome. Surgical treatment is usually necessary but the blockage does not usually recur.

b) Pancreatic insufficiency - early in life CF children show signs of pancreatic insufficiency characterized by malabsorption and severe steatorrea (fat excretion). This is a result of the inability of pancreatic products to reach the digestive tract, because of blockage, and emulsify fats etc, for absorption. As a result patients have ravenous appetites but poor weight gain and this may also lead to symptoms of vitamin A deficiency such as xerophthalmia and corneal ulceration, (Denning, 1968). Blockage of the bile duct and formation of gall stones is typical and contributes to localized tissue damage in the pancreas and liver. Sonographic techniques have been developed (Willi et al, 1980) which can visualize the extent of the bile duct obstruction, cholelithiasis and liver fibrosis and so specific treatment is now possible. Otherwise, treatment is restricted to strict dietary control according to the degree of malabsorption (Weihofen + Pringle, 1969) and the use of oral pancreatic enzyme therapy (Nassif et al, 1981) and anabolic steroids to enhance development and physical strength.

1.2.2. Pulmonary

At birth the CF patient has morphologically normal lungs but

respiratory difficulty usually arises during the 1st year. The predominant symptoms involve coughing, purulent sputum and wheezing, and unless the disease process can be arrested at this stage by effective antibiotic treatment a vicious cycle of obstruction, chronic infection and tissue damage results with progressive loss of pulmonary function and eventual death. Pulmonary disease accounts for 95% of deaths in CF patients and, so, much effort has been directed into alleviating blockage of airways and reduction of infection. The most successful measures have been the use of water aerosal therapy (Matthews et al, 1967) which has also been used prophylactically (Doershuk et al, 1965). Particularly effective as antibiotics for treatment of acute bacterial and viral infections have been chloramphenicol, the semi synthetic penicillins and aminophylline (Larson et al, 1980). The most persistent infections have been those of Pseud, aeruginosa and it has been suggested that this bacterium may be resistant as a result of the production of an inhibitor of the patients immune response or else a specific lymphocyte sub population may be impaired in some cystic fibrosis patients (Sorensen, 1979). Antibiotics directed against this species were unsuccessful, until recently when some improved results have been reported. (Marino, 1981, Hodson, M.E. personal communication).

1.2.3. Sweat electrolyte elevation

The most constant and clearly defined abnormality that has

been detected in cystic fibrosis is increased sodium and chloride concentrations in eccrine sweat. (Lobeck + Huebner 1962). This characteristic finding is present at birth and provides the basis for the only diagnostic test for the disease in children (Gibson + Cooke, 1959). Increases in sodium and chloride concentrations in CF sweat probably arise from decreased sodium reabsorption in the ducts, since saliva and sweat from CF patients can block sodium reabsorption by a normal sweat duct (Mangos, 1967).

Clinical manifestations of the sweat gland abnormality are related to the high rate of salt loss. Excessive loss of salt and fluid may lead to portal hypertension, hyponatremia and hyperkalemic alkalosis (Gottleib, 1971) and cardiovascular collapse. (Marino, 1980). The use of salt replacement therapy has now alleviated most problems in this area.

1.2.4. Prognosis

The prognosis for a CF patient has improved markedly over the last 10-15 years. This is mainly due to a better understanding of the disease and the development of effective means of treating it.

Most figures regarding prognosis etc are based on the results of the U.S. CF patients register, a body which was set up in 1966. Before that time there was very little co-ordination regarding treatment of CF and consequently the best methods of treatment were not universally applied. This register accumulated data from

every doctor treating CF patients and gradually an accurate picture of the state of the CF population in the U.S. was formed. Over the last ten years it has shown a marked increase in the age distribution and survival rate for CF patients, as CF moves from an exclusively paediatric disease into the field of general medicine.

1.3 Research on CF

In this section I shall review some of the work which has been carried out in the field over the last 30 years. This is important because the volume of literature is so vast and the areas of study have been so diverse that it really emphasizes the need for studying the defective gene directly. The research which has been carried out could be loosely classified into three groups:-

A) Improving methods of treatment

B) Study of biochemical abnormalities

C) Development of methods of diagnosis both postnatally and antenatally.

It seems unlikely that further substantial improvement in survival can be obtained through refinements in antibiotic or pulmonary therapy (Littlefield, 1981) so it is the latter two classes which I shall consider here.

To say that the current state of CF research is chaotic is an understatement. A recent review article (Benke, 1977) described reports that:- (cell membrane surfaces are abnormal, and normal; beta glucuronidase is abnormal, and normal; fatty acids are abnormal, and normal; serum calcium binding is abnormal and normal; polyamines in blood are high, and normal; RNA methylation is low, and normal; calcium or magnesium ATPase and sodium transport is affected, and not affected by CF materials; complement levels are

elevated, and normal; the Tamm-Horsfall glycoprotein is aggregated, and normal; glycosyl transferases are low, high, and normal; the cellcycle is abnormal and normal; arginine is important, and unimportant; Kallikrein is abnormal and normal; a CF "factor" can be demonstrated, and not demonstrated.) and so on for a considerable number of other parameters (all references in Benke, 1977). To review such wide ranging reports is beyond the scope of this thesis so I shall limit myself to those areas which I consider to have the most consistent data and which may prove to have greatest relevance to the elucidation of the basic defect .

1.3.1 Abnormalities in CF cells in culture

In studies of many diseases, research has been considerably aided by the availability of animal models reflecting the human disease (eg. diabetes). The search for a similar model in CF has been singularly unsuccessful. It was reported (Green, 1972) that a new recessive neurological mutation of mice called cribiform degeneration had similarities to CF with respect to electrolyte imbalance and more recently a further line of mice has been described which . . . also have some hereditary alterations resembling CF (Pivetta, 1980; Pivetta, 1981). This line does seem to have altered pulmonary response to bacterial challenge and altered pancreatic juice content (both characteristic of CF) but it remains to be established as a true model of CF.

The lack of an animal model led to the study of CF cells in tissue culture (reviewed by Ward, 1976). CF fibroblasts in culture do not differ in shape, number or refractility from normal fibroblasts. However, after staining with toludine blue, fibroblasts of many CF patients and heterozygotes show intracellular metachromasia (Danes+Bearn, 1968). Though this has also been shown in cultured leukocytes from homozygotes and heterozygotes it does not appear in fibroblasts from all CF patients and so could not be used for diagnosis of the disease. It did seem to suggest, however, that the mutant gene could express itself in cells other than those of exocrine glands, and so justified the study of CF in cells in culture. Metachromasia is associated with cellular glycosaminoglycan (GAG) metabolism and initial experiments showed a 2-10 fold increase in total GAG content (Matalon+Dorfman, 1968) but subsequent studies (Wiesmann, 1970, Welch 1975) could show no consistent defect in CF fibroblasts. The protein content of CF fibroblasts was shown to be significantly lower than in normal fibroblasts during exponential growth (Farrell et al, 1975). Hyperactivity of glycosyl transferases in CF tissues when compared with normals was reported (Luisot et al, 1973) but as so often in CF research normal levels were subsequently reported from another laboratory (Butterworth, 1974). Methionine deprivation in laboratory animals results in lung and pancreatic abnormalities similar to thoses found in CF and, similarly, CF children cannot

tolerate soybean milk which contains no methionine. Ribonucleic acid was therfore studied (Rennert, 1973) and found to be undermethylated in both CF and heterozygous cell lines but again this undermethylation could not be reproduced by other workers (Klagsbrun, 1974; Hodes, 1974). It seems unlikely that such an important process as this would be defective and only affect exocrine cells. In general, with such contradictory and widespread findings tissue culture studies have not been as enlightening as expected.In fact, recently, an extremely thorough ultrastructural and cytochemical comparison has showed no differences whatsoever between normal and CF fibroblasts (Spicer, 1980). However, some abnormalities have been observed in CF cells in culture which have been reproducibly observed and these are the so called CF factors.

1.3.2. The Cystic Fibrosis factors

In the blood and biological fluids of CF patients are found a number of humoral abnormalities known as 'CF factors'. In 1967 a factor was reported to be present in serum from patients with cystic fibrosis which inhibited the inherent rhythmic ciliary motion of rabbit tracheal explants. (Spock et al, 1967). This ciliary dyskinesis factor (CDF) has subsequently been assayed in other systems such as oyster gill cilia (Bowman, 1969), fresh water mussel cilia (Besley, 1969) and by an assay involving agglutination of Proteus Vulgaris (Cohen, 1974). This factor was however also

found in serum of some patients with autoimmune and allergic disorders and so could not be used diagnostically. This factor has been shown by several workers to be associated with the lg G fraction in serum (Bowman et al, 1969) and only discharged by CF cells in culture which exhibit metachromasia (Danes+Bearn, 1972). Although this factor has been characterized and shown to be a protein of M. WT. 4,000 many of the details of its charateristics have been contradictory. It is also present in the fluid of CF heterozygotes and it may be the protein which Brock and Wilson are studying for antenatal diagnosis (See Section 1.3.6).

An activity which copurifies with the CDF is that of degranulation. This activity causes release of lysosomal enzymes from cells and is more active in CF homozygotes than heterozygotes. It has been suggested that the basic defect in CF is in an enzyme that normally inactivates a family of particular proteins which includes these factors. These factors are thought normally to facilitate extrusion of material out of the cell and if not inactivated, mucus is oversecreted, becomes thick and sticky so blocking the gland or duct (Conover, 1974). This hypothesis was given some foundation (Rao and Nadler, 1975) when it was shown that one of the two components of the enzyme arginine esterase was reduced to 30% of its normal activity in plasma from CF cells. This work has subsequently been developed into the best current method of antenatal diagnosis for CF and will be discussed in detail

at a later stage. A second hypothesis for the CF factors is that they are aberrant gene products which interfere with membrane metabolisim and function. There is a great deal of evidence that CF may be related to a generalized membrane defect (eg. Galey. 1980) and this will now be discussed.

1.3.3 Membrane studies in CF

All evidence suggests that CF is primarily a disease of exocrine glands. It appears that in the most commonly involved gland, the eccrine sweat gland, the major cause of increased sodium and chloride concentration is reduction in sodium reabsorption in the duct of the gland. In 1967, a substance was identified in the sweat and saliva of CF patients (Mangos, 1967,1968) which inhibited sodium reabsorption in the duct of the rat parotid gland during perfusion studies. More recently (Mangos, 1973) in vitro microfusion studies have shown that the duct of the sweat gland in CF behaves normally when perfused with normal sweat, whereas both normal and CF sweat gland ducts behave abnormally when perfused with cystic fibrosis sweat. Although this tends to suggest a disorder in the sweat rather than the membrane, several groups have studied transport in CF membranes with interesting results. Breslow and Epstein (1981) showed that cultured fibroblasts from patients with CF accumulated less Na in the presence of ouabain than did normal cells. They also showed that CF fibroblasts,

when compared with normal fibroblasts, had enhanced survival after exposure to several different classes of cytrotoxic drugs including ouabain, steroid hormones and cyclic AMP (Epstein, 1977; Breslow, 1978; Breslow, 1980). Despite vigorous disagreement from several laboratories (Kurz, 1979; Liedtke, 1981) they used their dexcamethasone cytotoxicity test and their ion-transport assay in a double blind study on CF. Although their first 8 attempts gave 100% success subsequent work has shown this assay to be not reproducible and their findings have been retracted (J. Breslow, personal communication).

Many of the processes that transport ions in and out of cells are closely interrelated. Thus the increase in sodium in the sweat of CF patients and the abnormalities in the uptake of sodium by CF fibroblasts might be secondary effects related to defects in the transport of other ions. Much attention has been focused on the role of Ca^{++} in CF for two reasons:-

1) Increased calcium ion concentration in some CF secretions (Botelho 1973, Blomfield 1976)

2) There is an enlarged intracellular Ca^{++} pool in CF homozygote and heterozygote fibroblasts (Feigal, 1979 c)

Using isolated organelles it was shown (Feigal + Shapiro, 1979 a) that the increased intracellular Ca^{++} was due entirely to increased uptake of calcium by the mitochondria. Mitochondrial

calcium uptake and release is vital to calcium homeostasis in many cells. In such cells calcium dependent processes are fundamentally under the influence of the mitochondrial calcium pool. Release of calcium by mitochondria gives a rapidly available calcium source and sequestration by mitochondria provides a major calcium sink. It is possible to suggest a link between the altered mitochondrial calcium in CF fibroblasts and the major signs associated with the disease. Much of the pathology of CF is caused by secretion of highly viscous mucus which can be attributed to high levels of Ca in these secretions causing decreased solubility of proteins. The enlarged intracellular Ca pool may therefore result in increased levels of calcium in secretions. Mitochondrial Ca uptake is driven by the electron transport chain and Feigal + Shapiro (1979 a) showed that both CF and heterozygote fibroblasts consumed oxygen at a significantly higher rate than control fibroblasts. This work was then extended (Feigal, 1979 b) using electron transport chain inhibitors to show that the relative inhibition of O_2 consumption was CF homozygotes > Heterozygotes > Controls and that the site of the abnormality was at the position of NADH Dehydrogenase in the transport chain. Although the NADH dehydrogenase system is extremely complex, studies on the Km, V max and pH optima show absolutely consistent differences between CF and normal cells. Although this system requires further

substantiation (particularly by independent workers!), it seems extremely promising and provides a very reasonable explanation of CF symptoms. A defect in NADH dehydrogenase would mean a low yield of ATP in respiration (two ATP instead of three). Thus CF patients would use more oxygen to produce the same amount of ATP as controls and so their election transport chain must overwork. Since the electron transport chain drives the mitochondrial calcium pump there is increased uptake of calcium leading to increased intracellular calcium and high calcium in the secretions. This would result in the observed increased viscosity. Subsequent studies (Katz 1980, Foder 1980) have also reported defects in this pump mechanism by directly studying (Mg^{++} Ca^{++}) ATPase in membranes from CF and normal patients. Other reports of a generalized membrane defect have suggested altered numbers of insulin receptors (Lippe, 1980) and altered response to β-adrenergic stimulation and cyclic AMP in CF membranes (Roscher, 1980; Davis, 1980 b). Although all these studies may provide a pointer towards the basic defect they cannot be envisaged as providing reliable diagnostic tests.

1.3.4 Lysosomal enzymes and CF

Because of the presence of metachromatic granules in CF cells (Danes + Bearn 1968) several investigations into the activity of lysosomal enzymes in CF cells have been made. Kraus et al.(1971) carried out a study of four lysosomal enzymes using fibroblasts

cultures. No difference was found in the activities of beta galactosidase, beta glucuronidase, arylsulphatase and acid phosphatase in CF and normal control cells. A more recent study of lysosomal enzymes in cultured CF fibroblasts was carried out by Hosli and Vogt (1977) using microtechniques. They measured enzyme activities in the cells and in the medium and suggested that a leakage of lysosomal enzymes was taking place into the extracellular fluid due to abnormal recognition sites on the lysosomal enzymes and alkaline phosphatase. The co-ordinate regulation of the synthesis of lysosomal enzymes and alkaline phosphatase would counteract the leakage and thus result in the observed high intracellular activity of alkaline phosphatase in CF homozygotes. The same group also found that the lysosomal enzymes and phosphatases could be induced in CF cells with Tamm-Horsfall glycoprotein (Hosli, 1976) and that this 7-fold induction could discriminate between CF homozygotes and heterozygotes. In a double blind trial (Hosli, 1978) 24 specimens of CF homozygotes, heterozygotes and normals were all diagnosed correctly. This method of antenatal diagnosis has been found to be difficult to reproduce because differences in culture conditions, preparation of homogenates and methods of enzyme assay make comparison of the results from different groups difficult.

1.3.5 The Immune and Complement Systems

Recurrent infection is an important feature of progressive

pulmonary involvement in CF. Therefore the possibility of a defect in the immune system of CF patients has been closely examined. Many abnormalities of the immunological system have been identified: Ig G, IgA and IgM levels have been shown to be raised (Schwartz, 1966) but this is probably a secondary cause of chronic infection. Indeed, it is particularly difficult to determine whether immune defects are primary or secondary effects. Independent reports (Bigger et al, 1971; Boxerbaum et al, 1973) have observed that CF serum has an inhibitory effect on rabbit alveolar macrophage ingestion and destruction of Pseudomonas, but whether this is due to a deficiency in serum or to the presence of an inhibitor has not been resolved.

Since the complement system plays a major role in nonspecific host defence mechanisms, it has come under close scrutiny in CF. Increased levels of C3 component in the serum were reported by Conover et al (1973 b) and this was not related to age, therapy or degree of pulmonary involvement in the CF patient (Hanns et al, 1974). Conover et al (1974) suggested that the primary defect could be a deficiency in, or a defect of, a carboxypeptidase like enzyme which is involved in the activation of the C3a component and that the CF factor may be C3a. This theory was backed by the arginine esterase deficiency, as previously described (Rao + Nadler 1974), since activation of the C3a component involves cleavage of the terminal arginine.

1.3.6 Development of antenatal and neonatal screening for CF

The possibility of diagnosing any disease antenatally has only arisen during recent years. As a result attention has been directed for the past 20 years at developing methods of unequivocally diagnosing CF neonatally. There is some dispute as to the benefit of early diagnosis in CF since there is no effective prophylactic treatment available (Gibson, 1980). In spite of this and particularly because of the singular lack of success of developing an effective antenatal test, much research is being carried out into developing absolutely reliable, cheap and sensitive neonatal assays.

First developments followed the observation of high salt concentrations in CF fluids (di Sant' Agnese 1953) and a basic sweat test (Schwachman, 1955) and the chloride agar plate test (Schwachman, 1956) arose. The sweat test was modified and simplified by a method known as quantitative pilocarpine iontophoresis (Gibson, 1959) which is still the most useful clinical test in CF. However, this method has recently been shown to be not totally reliable (Denning, 1980) and is not effective for patients less than three months old. Until recently it was particularly unreliable for adults but in combination with suppression by fluorohydrocortisone it is now more effective (Hodson, 1980). To facilitate diagnosis before the age of 3 months

two further tests have been developed. Serum immunoreactive trypsin concentrations in CF newborns was found to be extremely high (Crossley, 1979) and to fall to subnormal levels in older patients (Dandona, 1979, Dandona, 1981 a). This assay was developed into a diagnostic test (King, 1979) but is extremely expensive. The second test, which is more simple and inexpensive though less reliable involves measurement of raised levels of protein, particularly albumin, and increased lactase activity in the meconium of infants (Berry, 1980). By prescreening children with the meconium test and then screening those which are positives by the trypsin method a satisfactorily reliable result may be obtained (Dandona, 1981 b)

The development of a reliable test for antenatal diagnosis is a far more difficult task but recently some optimistic results have been reported. As previously described (Section 1.3.2) one of the two components of the enzyme arginine esterase was reduced to 30% of its normal activity in plasma from CF cells. The activity was assayed using methylumbelliferyl guanidinobenzoate (MUGB), an artificial compound which is an active site titrant of many serine proteases. Because of these observations and the potential value of amniotic fluid for the prenatal diagnosis of CF, these studies were extended to amniotic fluid and the activity of the MUGB-reactive protease was found to be markedly reduced in CF fluid, similar to the relative deficiency

observed in plasma and cultivated skin fibroblasts from patients with CF (Walsh - Platt et al 1978). In addition, after isoelectric focusing on polyacrylamide gels subsequently reacted with MUGB a single band (p I 5.1-5.4) was consistently absent in the CF amniotic fluid. This observation was subsequently confirmed by two other groups (Brock + Hayward, 1979, Dann + Blau, 1979). The same group also showed that normal amniotic fluids contain two major populations of MUGB - reactive protease sequences by gel filtration. One had a molecular weight of 200,000, the other of 100,000 but CF fluid lacked the species of 200,000 molecular weight and the 100,000 species was unstable. (Nadler, 1980 a). By a combination of these three methods ie:-

1) Active site assay
2) Isoelectric focusing
3) Gel filtration

a test study was performed on 13 pregnancies from women who already had CF children. Three fetuses were diagnosed as having CF, the rest normal and all diagnoses proved correct (Nadler 1980 b).

Subsequent studies have shown that this test is not perfect but is now available as a predictive test for those families who have a CF child and request antenatal diagnosis before starting another pregnancy (See Discussion).

The first isoelectric focusing (IEF) study in CF was carried out by Schmoyer et al (1972) working on the cystic fibrosis factor and they observed that the ciliary dyskinesis activity migrated as a single band with an isoelectric point of 8.54. Wilson, et al (1973) carried out IEF of whole serum and observed a band at pI 5.48 which was absent in CF and a band at pI 8.41 which was present in CF but not in normals. Wilson and Fudenberg (1976) studied the protein with pI 8.41 (the so called CF protein) and in a study of 65 CF homozygotes, 66 heterozygotes and 110 normal controls, 96% CF homozygotes, 64% obligate heterozygotes and 8% controls were CF positive. They also showed that the protein (CFP) was associated with Ig G and on gel filtration was associated with two molecular weight fractions 90,000-180,000 and 1,000-13,700, which may indicate that the factor occurs in association with Ig G and also in a free form in serum.

The CFP has several characteristics in common with the ciliary dyskinesis factor and Wilson et al (1977) showed that the protein did indeed cause asynchrony in the rabbit tracheal system. The existence of this protein has been confirmed by several workers, however reproducible high quality IEF is not easy and even the best gels cannot distinguish homozygotes from heterozygotes. For this reason Manson + Brock (1980) removed the band from an IEF gel

injected it into a guinea pig and collected the antiserum. When tested on CF serum by immunprecipitation a positive reaction with sera from 16 of 17 homozygotes, 8 of 9 obligate heterozygotes and only 1 of 15 normal subjects was observed. Precipitin peaks were larger with sera from homozygotes than from heterozygotes which suggests very strongly that this could be related closely to the basic defect. If the antibody reacts only with this factor and not with any other factor in serum, and if it could be made in sufficient qualities (as a monoclonal antibody) then it may be possible to detect CF protein in aminocentesis samples by radioimmunoassay.

Recently a serum lectin has been identified in heterozygote and homozygote serum (Lieberman, 1979) and its detection is extremely simple, quick and inexpensive. This lectin has been characterised (Nordstrom, 1980) and appears to have extremely similar properties to the CF protein. If these factors are the same entity then this may develop into a very promising diagnostic tool. It seems therefore that there are good prospects for a reasonably reliable antenatal diagnostic procedure for 'at risk' pregnancies in the near future. However the parameters being studied may not necessarily be the primary effect ot the basic defect and this can only be elucidated when the defective gene itself can be analyzed and the molecular basis of the disease

characterised. In recent years recombinant DNA technology has been developed such that there are now ways to study genes directly at the molecular level and the work described in this thesis is the initial approach to the problem.

B. CONTROL OF GENE EXPRESSION

1.4 Introduction

Genetically determined diseases in humans present two major questions to molecular biologists. Firstly, such diseases by definition represent defects in either a gene, its expression, control of gene expression, or in the subsequent processing of genetic information. Secondly, having elucidated the nature of the defect is it possible to treat or alleviate the disorder? Initially then, it is essential to elucidate the nature of the defect. In Fig. 1.4 a schematic outline is presented of the biosynthetic pathway of a typical protein. Dysfunction may occur at any of the five main levels shown, and a brief review of the current understanding of gene expression is now presented, so that the potential areas in which the CF defect may occur, can be understood.

1.5 Sequence organisation of human nuclear DNA

One of the most prominent features of eukaryotic DNA is the presence of repeated sequences (Britten & Kohne,1968) which are distributed in non-random fashion within the genome. When analysed by denaturation-renaturation techniques at least three kinetic components are distinguishable : a very rapidly renaturing one, a fast renaturing one and a slowly renaturing component. The very rapidly renaturing component is repeated in the order of 10^6

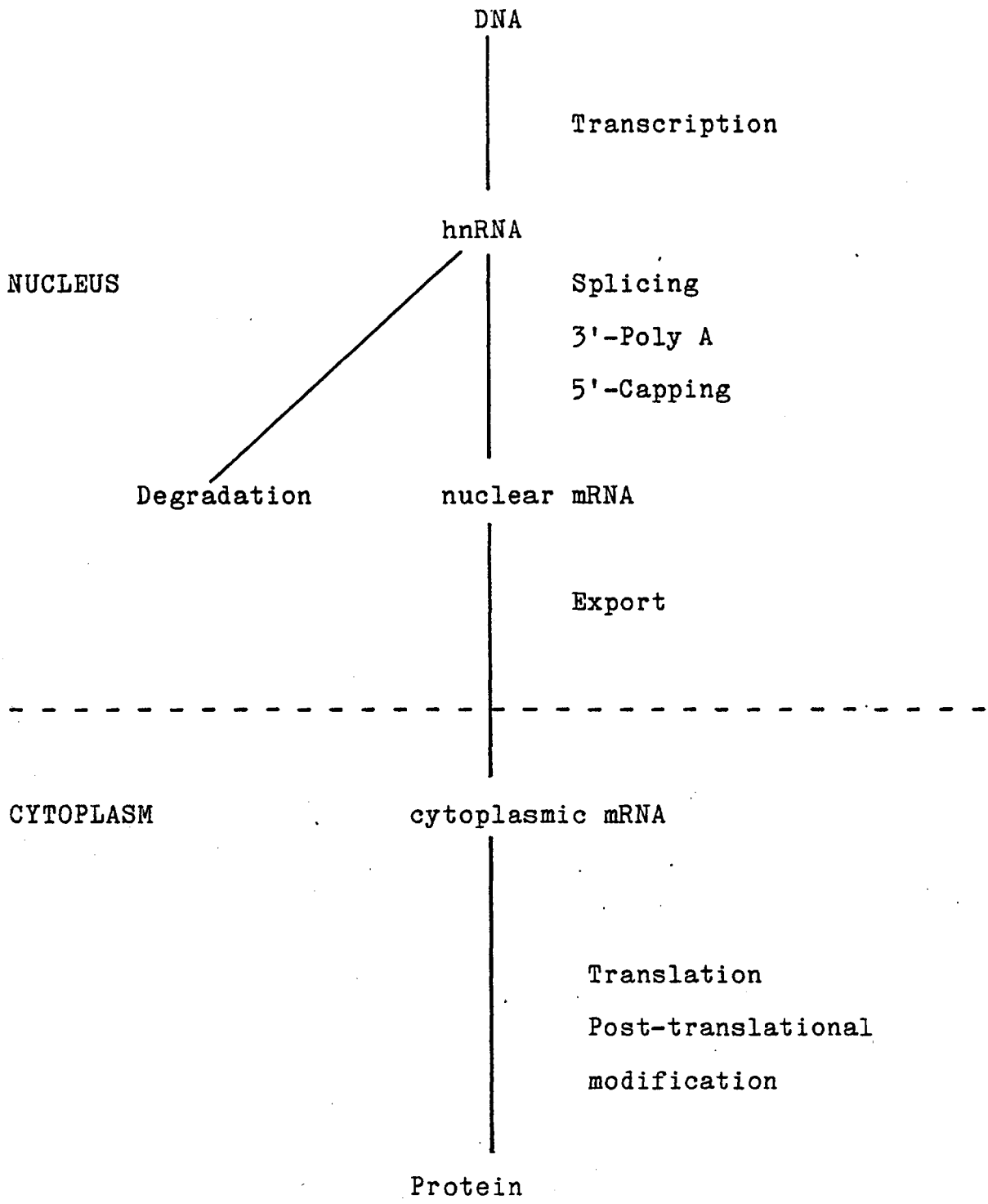


Fig.1.4 Biosynthetic pathway of typical protein

times per genome and is often found in clustered arrangements on chromosomes (Cooke,1979). In addition this fraction may contain foldback sequences ie. sequences capable of forming intrastrand duplexes which represent 6% of the human genomic DNA (Schmid et al.,1975) and are often found in the 5' region of eukaryotic and prokaryotic genes (see Nishioka & Leder,1979).

The DNA renaturing with intermediate rate contains sequences repeated in the order of 10 to 10^5 times and may be clearly subdivided into distinct kinetic subclasses (Schmid & Deininger, 1975). The slowly renaturing genome component contains sequences that are usually present only once per haploid genome and this component contains most of the structural genes. However, various factors suggest that the number of structural genes is in the order of 10^4 (Ohno,1971; Kiper,1979) which means that in the human genome less than 5% of the single copy DNA consists of "classic" genes. The existence of a class of low-repetitive sequences renaturing with single-copy properties has been described (Deininger & Schmid,1979) but nevertheless a large proportion of the genome is present in single copy and at present has no defined function.

The development of recombinant DNA techniques which permit clonal replication of eukaryotic DNA segments in bacteria has brought a rapidly increasing insight into the actual organization of these repetitive and non-repetitive sequences with respect to

each other. Interspersion of short (0.2 -0.4 Kb.)repetitive sequences with short (2 Kb.) non-repetitive sequences has been found to be a prominent feature of most eukaryotic genomes including man (Schmid & Deininger, 1975). In addition, interspersion of longer nonrepetitive sequences with repetitive sequences occurs in smaller parts of most of these genomes (Schmid & Deininger , 1975; Pearson et al, 1978). In view of these patterns of sequence organization, a role has been suggested for these interspersed repeated sequences in gene expression (Britten & Davidson, 1969, 1971; Davidson et al ,1971; Davidson & Britten, 1979) and the next section summarizes the current hypothesis.

1.6 The role of interspersed repetitive sequences in control of gene expression

A fundamental difference in prokaryotic and eukaryotic genome organization is the physical location of functionally related structural genes. In bacteria genes for proteins from a common synthetic pathway are usually physically linked and the primary transcript is read off as a giant polycistronic messenger RNA. In the human genome however related genes are usually dispersed throughout the genome, the genes for the first six main steps in glucose metabolism being on six different chromosomes (Donald & Hamerton, 1978).It is apparent that a model

that is suitable for transcriptional control of gene expression must show how genes at different genomic sites are regulated together.

Relying on the concept of Jacob & Monod (1961) according to which recognition elements map adjacent to structural genes they control, Britten & Davidson (1969) suggested that genes which are functionally related share homologous repetitive sequences. These sequences would be sensitive to stimulation by RNA molecules from "integrator" genes, themselves regulated by hormonal or other external stimuli. This model has now been revised (Davidson & Britten, 1979) to fit new data. The vast majority of structural genes appear to be transcribed continuously at the same rate in any one cell type, with post transcriptional control determining the nature of the cytoplasmic mRNA. It also appears that the basic transcription unit contains structural genes linked to short interspersed repetitive sequences (see Davidson et al.,1975; Kuroiwa & Natori, 1979; Jagodzinski et al ., 1979). Other regions of the genome containing repetitive elements but no structural genes are transcribed in a tissue specific manner (Scheller et al., 1978 ;Costantini et al., 1978) the concentration of repeat sequence transcripts varying over two orders of magnitude. The new model therefore postulates that specific RNA-RNA duplexes form between homologous repetitive sequences on the structural genes and the tissue specific RNA

so protecting those transcripts from degradation or, more likely to promote messenger RNA processing (Goldberg et al.,1973; Klein et al. ,1974; Campo & Bishop, 1974 ; Rabbits & Milstein, 1975). Thus co-ordinately controlled expression may mean the presence of common repeat sequences on specific structural genes (Kindle et al., 1979). As recombinant DNA technology reveals more specific genes with repeated sequences in their flanking regions (eg. Fritsch, 1980; Shen, 1980; Kaufman, 1980) a more precise model may be developed. Recently, a ubiquitous family of repetitive sequences has been identified (Houck et al., 1979) This family was shown to be repeated at least fifty thousand times in the human genome and to correspond to 7% of the genome. It was characterised and found that half its members contained the recognition site for the restriction enzyme Alu I (hence the Alu family.)

Approximately one-third of interspersed repeated sequences in humans exist as inverted repeats (ir-DNA) (Schmid & Deininger, 1975; Deininger & Schmid, 1976,) Likewise heterogeneous nuclear RNA (hn RNA), the precursor of cytoplasmic RNA, contains some interspersed repetitive sequences of the same length which can also exist as inverted repeats (Federoff, 1977). It therefore appeared that ir DNA might be a transcription template for this part of hnRNA (Jelinek, 1977).Base sequence comparisons of the Alu family members and hnRNA suggests that this is true (Jelinek,

1980). A specific sequence of low molecular weight RNA was observed in Chinese hamster cells (Jelinek & Leinwand, 1978) which also had complementarity to a part of the Alu family sequence and which appeared to be hydrogen bonded to nuclear and cytoplasmic poly A^+ RNA molecules. Tashima et al. (1981) also reported the presence of Alu sequences in hnRNA but showed them to be greatly reduced in cytoplasmic mRNA, suggesting a role for the Alu family in RNA processing as speculated by Robertson(1977) and Jelinek(1972). A 7S RNA moiety has also been identified(Weiner,1980) which is complementary to the Alu sequence and whose function in RNA processing has also been suggested (Lerner,1980; Rogers,1980). The observations of the frequency and distribution of the Alu family imply that most genomic clones from mammalian cells will contain such sequences. Recent reports (Duncan,1979;Elder,1981) uphold this and show that the Alu sequences function as templates for transcription by RNA polymerase III in vitro. Perhaps either transcription of these sequences by RNA polymerase III or the interaction of these sequences with the low molecular weight RNA, or both, play a role in the production of structural gene transcripts. There is currently no avaliable data to decisively test these possibilities Several other reports (Duncan, 1981; Bell, 1980) show some Alu family members to be flanked by short direct repeats in a similar arrangement to the δ sequence of yeast. The δ sequence is a

mobile element so the possibility that Alu family members are mobile transcriptional promoters, is intriguing. Finally, it should be noted that the Alu family sequences have similarities to DNA sequences of BK virus, SV 40 and polyoma virus at or near the origins of replication of these viral genomes. Members of the Alu family may therefore be the normal cellular DNA replication initiation sites. In mammalian cells DNA replication proceeds bidirectionally, simultaneously from many sites (Huberman & Riggs, 1968) and thus the initiation sites might be expected to be interspersed repetitive sequences with two fold rotational symmetry as shown by the inverted repeats of the Alu family.

1.7 Gene rearrangement and repetitive sequences

The Alu family of repeated sequences is not the only family of repeated sequences found in mammalian genomes but in most cases their function has not been decisively eluicidated. For example tandem repeats of RNA genes in mouse have been shown to contain three different repetitive families (Kominami, 1981) while five different classes have been shown surrounding the chicken lysozyme gene by an adaption of the Southern transfer technique (Baldacci, 1981). Similarly families of repetitive sequences have been identified on the X chromosome (Schmeckpeper,1979) some of which are found on other autosomes and some specific to the X chromosome. The functions of some repetitive sequences are

however now becoming clear and these are involved in an area of expression which had hitherto not been realized, namely gene rearrangement. This arose as a result of the study of immunoglobulin genes and as early as 1965 it was proposed that an antibody polypeptide was encoded by two genes that were joined together during differentiation of antibody producing cells (Dreyer & Bennett, 1965), one gene constituting the so-called variable or V region and the other the constant or C region. This idea was proved correct by restriction analysis of DNA (Hozumi & Tonegawa ,1976) and by cloning and sequencing experiments (Bernard et al ,1978). Most of the early work was carried out in mice and it was shown that light chains were encoded by three distinct DNA elements: variable (V), joining (J) and constant (C). (Max et al. 1979; Sakano et al. 1979a). It has been demonstrated that there are at least 2000 V light chain germ line genes (Zeelon, 1981) and that during differentiation of antibody producing cells, a V and J region are joined together to generate a specific V gene (Seidman et al.,1979) thus generating the potential for a vast amount of antibody diversity. Recently the genomic regions surrounding the light chain V and J genes have been sequenced (Steinmetz et al.,1980) and it is apparent that homologus repetitive DNA sequences are found near the V and C genes suggesting a role for the repeats in the recombinational event leading to the gene rearrangement. The situation regarding mouse heavy chains is more complicated

because a fourth region (the D region) is also involved in the recombinational event (Early et al.,1980; Sakano et al., 1979b) thus generating further potential for antibody diversity. Secondly it appears that there are at least eight different classes of constant region gene and that a particular V gene generated by the first recombinational event may be then switched sequentially to different classes of constant region (Davis et al. 1980a; Kataoka et al.,1980; Maki et al.,1980). Once again sequence data of the germline DNA has identified homologous repetitive sequences of a 250 bp repeating unit 5' to the various constant region genes (Tyler et al.1980; Nikaido et al.1981; Kataoka et al.1981) and several feasible models for the mode of recombination involving these so-called 'S' regions have been suggested (for review see Kataoka,1981; Davis,1980b). Two other eukaryotic systems, yeast mating types (Hicks et al.1979) and trypanosome coat proteins (Hoeijmakers et al.1980) also use DNA rearrangements as a fundamental mechanism of gene expression. Apart from the immunoglobulin system a recent report suggests that rearrangement may play a similar role in the expression of different isozymes of the single pyrurate kinase gene from liver and red cells (Marie,1981). However,control of immunoglobulin expression is not purely through DNA rearrangement and post-transcriptional RNA processing is clearly also involved (Early, 1980b; Moore 1981).

1.8 Sequence requirements for gene expression

The ability to purify specific genes from genomic DNA and analyse their sequences has led to greater understanding of sequences involved in initiation and termination of transcription. Systems of in vitro transcription have been developed for genes transcribed by RNA polymerase II (pol II) (structural genes) and RNA polymerase III (pol III) (5S RNA , tRNA etc.). The pol III systems have been developed during the last three years and have the ability to initiate and terminate correctly an exogenous gene normally transcribed by pol III (a pol III gene) which is added (Birkenmeir et al. 1981, Ng et al. 1979, Wu 1978, Wormington et al.,1981). The pol II systems developed can also initiate correctly, but no efficient termination system is avaliable (Manley et al., 1980; Weil et al., 1979). Various cellular methods have also been developed such as microinjection (Mertz & Gurdon, 1977; French Anderson et al.,1980) transfection of cloned segments into cells by SV 40 (Goff & Berg, 1979) and trans- formation of tissue culture cells (Pellicer et al.,1978; Mantei et al.,1979; Wigler et al.,1979). By using techniques such as in vitro mutagenesis it is possible to make deletions or substitutions of specific bases in the putative transcriptional control regions (see Shortle & Nathans,1978; Hutchinson et al. 1978) or by genetic engineering to make larger deletions, rearrangements or insertions. The effect of these alterations on

the transcription of the gene is tested by analysing the products of transcription in the in vitro system. Experiments such as these have led to considerable insight into which regions of the specific genes are involved in transcription.

Much of the work on RNA pol III transcription has been carried out on the Xenopus oocyte 5S RNA gene. Using elegant deletion manipulations it was possible to delete the entire 5' flanking regions of the gene and the first 50 5' nucleotides of the gene itself with no effect on transcription (Sakonju et al., 1980). Similarly, deletions from the 3' end of the gene could be made up to nucleotides 80-83 before transcription was abolished (Bogenhagen et al.,1980) suggesting an internal control region in the gene between nucleotides +50 and +80. More recently a 5S RNA transcription factor has been isolated by several groups (Pelham & Brown, 1980; Pelham et al.,1981; Honda & Roeder, 1980) which has been shown to interact specifically with this region of the DNA (Engelke et al.,1980; Sakonju et al.,1981), supporting this observation. Peterson et al. (1980) observed a sequence AAAAGT at -12 to -17 from the putative initiation codon which was conserved in 5 different Xenopus 5S genes but which could be deleted without affecting transcription. Defranco et al. (1980) switched 5' flanking regions between an actively transcribed gene and a slow gene for a eukaryotic tRNA and found that the slow gene was actively transcribed and vice versa. Kressman et al.(1979)

supported this and suggested the generalized view that RNA polymerase III genes have two regions of control, one which is internal controls transcription ultimately, the other in the 5' flanking region modulates the rate. The control of termination by pol III is not so well characterized but Bogenhagen & Brown (1981) has shown that it occurs at a cluster of T residues, however the adjacent nucleotides also appear to affect the efficiency.

For genes transcribed by RNA polymerase II, a characteristic sequence TATAAA very much resembling the promoter site for prokaryotic polymerase (Pribnow,1975) has been found in all animal genes investigated so far. This sequence, known as the "TATA box" is about 23 nucleotides upstream from the cap site (see Section 1.8). It has been shown to play a role in the initiation of transcription for several genes in vitro (Wickens & Lasky,1981). A further region of homology has been observed in many pol II genes, the sequence CCAAT occuring around 70 base pairs to the 5' side of the gene. Dierks et al.(1980) have shown that in vitro transcription gives 60% correct initiation of a rabbit gene if both sequences are present, 25% if the -70 region is deleted and none at all if the "TATA box" is deleted. Similarly several other studies have shown that the only DNA which is indispensable is the "TATA box" (Corden et al.,1980; Tsai et al., 1981; Grosveld et al.,1981) suggesting that the -70 region is not important for correct initiation but may modulate efficiency of

transcription (Dierks et al.,1981). Base substitutions in the "TATA box" for the conalbumin gene drastically reduces specific in vitro transcription (Corden et al.,1980, Wasylyk et al.,1981) and Grosschedl and Birnsteil (1980) have found that it may direct the RNA polymerase to the transcription initiation site. It is possible that there may be other control signals, for example internal control signals for pol II as there are for pol III, and indeed a small pol II gene has recently been reported which is transcribed constitutively and has no "TATA box" at all (Roop et al.,1981). Similarly, developmentally regulated globin genes in goats have been shown to have only 8 variable bases in the 275 bases preceding the gene suggesting that developmental control of expression may involve another sequence (Haynes et al.,1981). As previously mentioned, pol II expression systems do not terminate accurately, however Benoist et al.(1980) has shown a consensus sequence of TTTTCACTGC to be near many (but not all) putative pol II termination sites.

1.9 Other levels of transcriptional control

Apart from the processes which have already been discussed there are clearly many other factors whose function is not yet fully understood. At the level of the chromosome it is clear that specific active genes are in a special chromatin conformation, that makes them more susceptible to several enzymatic probes

and especially to DNase I (Weintraub & Groudine, 1976; Wu, 1981; Stalder, 1980a). In tissues that do not express these functions the genes are not in this conformation (Felber, 1981). This sensitivity to DNase I is a general phenomenon that seems to be true for all active genes regardless of there extent of expression (Stalder, 1980b; Weisbrod et al., 1980). In addition to chromosomal alterations there is strong evidence that chemical modification at the level of DNA may be correlated with gene expression. By using methylation sensitive restriction enzymes to probe specific sequences it has been shown in several systems that the methylation pattern of a particular gene is tissue specific, with a general correlation between high levels of gene expression, and under-methylation (Naveh- Many & Ceder, 1981; Van der Ploeg et al. 1980; Weintraub et al.,1981). These differences at the chromosomal level may be important in CF,particularly since metachromasia has been observed (see section 1.3).

The influence of external factors on expression has been extensively investigated, particularly the effect of hormones. Using co-transfection techniques it has been shown (Kurtz et al., 1981) that DNA regions required for a specific hormonal response are located near the structural gene involved and sequence homologies have been observed 5' to three genes of the ovalbumin cluster which are co-ordinately regulated, (Royal et al.,1979). In the case of the steroid hormones, steroid-membrane receptor

complexes interact with chromatin to control transcription (Palmiter et al.,1976) but in the case of polypeptide hormones the interaction with the receptor triggers off modulation of the concentration of intracellular messengers (eg. cAMP. Ca^{++}) which are then thought to stimulate transcription (Evans et al., 1981; Chung et al.,1981). Although the rate limiting step in hormonal response is at the transcriptional level (Chan et al., 1973) both steroid and polypeptide hormones also effect both post-transcriptional and translational control (Le Meur et al.,1981; Cox,1977; Robins & Schimke,1978).

1.10 Post-transcriptional regulation-hnRNA and mRNA

Many, but not all, classes of RNA are transcribed as precursors that are larger than the mature mRNAs themselves (see Perry,1976). Initial experiments concerned with the relationship between heterogeneous nuclear RNA (hnRNA) and mRNA dealt with fractionated populations which were complex in structure and heterogeneous in size (Bishop et al.,1974; Galau et al.,1974). Only after the introduction of techniques allowing the isolation of specific eukaryotic mRNAs did the study of individual pre-mRNA molecules become possible. Initially it was assumed that post transcriptional processing of these larger pre-molecules involved trimming from the 5' end of the molecule. These views were radically altered by the discovery of intervening sequences

within genes.

1.10.1 Introns and splicing

Introns, insertions or intervening sequences are regions of genes which are transcribed but which are subsequently removed to give the functional mature RNA. First observed in the 28S rRNA gene of Drosophila (White & Hogness,1977; Glover & Hogness, 1977), they were thought to be inactivated genes but their presence in an active eukaryotic gene was first reported for the rabbit β globin gene (Jefferys & Flavell,1977). They have subsequently been reported in many genes such as yeast tRNA (Johnson et al.,1980), human growth hormone (De Noto et al.,1981) chick ovalbumin (Dugaiczyk et al.,1979; Garapin et al.,1978), Xenopus vitellogenin (Ryffel et al.,1980; Wahli et al.,1980), mouse alpha-fetoprotein (Gorin & Tilghman,1980), rat prolactin and growth hormone (Chien & Thompson,1980) and various human globin genes (Flavell et al.,1978; Mears et al.,1978; Little et al. 1979; Tuan et al.,1979). Clearly there is no species preference and no limit to the number of introns in a gene, ranging from 1 in the rat insulin gene (Lomedico et al.,1979) to 50 in the chicken proα-2 collagen (Wozney et al.,1981). Similarly there is no limit to the size of the insertions, the dihydrofolate reductase gene in mouse occupying 42 Kb. of DNA of which only 1 Kb. is coding sequence (Nunberg et al.,1980). It seems that

almost all protein coding genes in eukaryotes have introns (Crick,1979; Maxwell et al.,1980) except histones (Schaffner et al.,1978) and the interferons (Nagata et al.,1980; Lawn et al., 1981; Houghton et al.,1981). Although intervening sequences differ considerably in length and sequence even for the same gene in different species, there is remarkable conservation of the regions flanking the coding sequences (reveiwed by Sharp,1981). At the splice junction the prototype sequence

GT - Intron - AG

is characteristic of all mRNA intron/exon junction investigated to date. This, together with the ability of cells to carry out correct splicing of pre-mRNA molecules from other species (Hamer & Leder,1979a; Breathnach et al.,1980) suggests that the mechanism of splicing is universal. In vitro systems for pre-tRNA splicing have been developed (Knapp et al.,1979) and the enzymes involved well characterised (review by Altman,1981). Subcellular fractionation of HeLa cells has been reported to yield four fractions with mRNA processing activity (Ferrari et al.,1980) but no in vitro system for pre-mRNA processing is currently available. In spite of this some insight into the mechanism of splicing has been achieved using other techniques particularly R-looping (Woolford & Rosbash,1979) and S_1 nuclease mapping (Berk & Sharp,1977). Roop et al. (1978) showed the presence of 7 precursor RNAs for ovalbumin by gel electrophoresis and"Northern"

blotting suggesting that there was an ordered, rather than random manner of removing the introns. This has been subsequently verified for ovomucoid (Tsai et al.,1980), duck globin (Reymond et al.,1980) and cytochrome b mRNA (Halbreich et al.,1980). Avvedimento et al. (1980) have proposed that introns at the 3' end of the RNA are excised before those at the 5' end, while Grosveld et al. (1981b) have shown for rabbit β-globin that the smaller of the two introns is removed initially, and that each intron can be removed in smaller pieces. (See also Kinniburgh & Ross,1979).

Several theories as to the biological significance of discontinuous genes and RNA splicing have been advanced (Gilbert,1978; Tiemeier et al.,1978; Darnell,1978; Crick,1979). Recent evidence for the physiological role of splicing has arisen from several laboratories. Hamer and Leder (1979a) constructed hybrid SV 40 viruses containing chromosomal mouse β-globin genes and infected monkey kidney cells. Fragments were transcribed in the monkey cells regardless of the orientation of the globin fragment but the introns were only excised if the mouse fragment was in the correct orientation. Subsequently it was shown that the presence of at least one intron is a necessary prerequisite for correct processing and stabilization of a eukaryotic mRNA (Hamer et al.,1979; Hamer & Leder,1979b). One plausible model explaining this observation is that splicing is linked to RNA

transport from the nucleus. If this is the case, genes which have no introns may be transported by independent mechanisms. A further interesting observation is that by using alternative processing pathways, ie. joining different sections of the mature mRNA together, discrete mRNA sequences can be derived from the same gene. This has been reported for mouse α-amylase (Young et al. 1981), a sea urchin maternal mRNA (Lee et al.,1980) γ-immunoglobulins (Rogers et al.,1980b) and human growth hormone (De Noto et al.,1981). This raises the possibility that the actual diversity of proteins coded by messages could be significantly higher than has been calculated on the assumption that each distinct RNA sequence codes for a single protein. Implicit in this observation is that specific post-transcriptional processing controls cytoplasmic expression of some sequences.

1.10.2 Polyadenylation

Apart from removal of intervening sequences, most eukaryotic pre-mRNAs are modified in the nucleus by the addition of tracts of poly(A) residues to the 3' end of the transcript. The 3' end of these RNAs consists of an untranslated region and in all eukaryotic mRNAs studied to date the pentanucleotide AAUAA is found approximately 10-20 nucleotides before the poly(A) starting site, (Valenzuela,1981; Proudfoot & Brownlee,1976). It seems that a signal for transcription termination must be situated

downstream to the poly(A) attachment site, giving a large non polyadenylated pre-mRNA molecule that must be trimmed at its 3' end before poly(A) can be added. Such a mechanism has been observed for adenovirus (Fraser et al.,1979) and more recently for mouse β globin (Hofer & Darnell,1981). The role of the poly(A) tail in mammalian RNA has been and remains one of the unsolved mysteries of molecular biology (reviewed by Brawerman 1981). Recently Huez et al.(1981) have injected native and deadenylated globin mRNA into HeLa cells and the poly(A) trail clearly stabilizes the molecule. It is also clear that the addition of poly(A) is essential for maturation and the cytoplasmic appearance of most eukaryotic mRNAs. However it cannot be the only mechanism because some RNAs are not polyadenylated and yet are still found in the cytoplasm (eg. histone mRNA). Similarly poly(A) tails have been observed on HeLa cell mitrochondrial mRNA and some viral RNAs. This suggests a non nuclear role for poly(A) since these RNAs are exclusively cytoplasmic.

1.11 Translational and post-translational control

The mechanisms of gene expression already discussed have been concerned with presentation of mature mRNA molecules to ribosomes for protein synthesis. Although there are two further levels of control, namely translational and post-translational

a defect in CF occuring at one of these levels would not be observed by the techniques used in this study. Excellent reviews of this area of control of gene expression are available (Kozak, 1980; Clemens,1979).

1.12 The concept of abundance classes

It should be clear from the preceding discussion that control of gene expression exists at transcriptional and post-transcriptional levels. Much work has been carried out into which level is rate limiting in particular systems. In muscle it has been shown that increased synthesis of cell-specific proteins during differentiation is predominately transcriptionally controlled (Paterson & Bishop,1977). Similarly in liver most moderately abundant tissue-specific sequences are regulated at the level of transcription (Derman et al.,1981). On the other hand Jacobs and Birnie (1980) have shown by hybridization kinetics that some sequences at radically different concentrations on the polysomes of liver and brain are at very similar concentrations in their nuclei, implying that the differences arise post-transcriptionally. The importance of post-transcriptional regulation of gene expression has similarly been described for sea urchin (Wold et al.,1978; Davidson & Britten,1979) mouse L cells (Meyuhas & Perry,1978) and even tobacco (Kamalay & Goldberg, 1980).

Wherever the rate limiting step in the control of expression maybe, in the steady state situation any particular cell type will have a certain number of mature mRNAs in the cytoplasm at specific concentrations characteristic for that cell type. Analysis of the kinetics of hybridisation of complementary DNA (cDNA) copies of messenger RNA to a vast excess of its template RNA permits a determination of not only the number of different sequences present as mRNA, but the relative abundance of these sequences within the mRNA population as well. Initially employed by Bishop et al. (1974) to determine the abundance classes and genetic complexity of cytoplasmic poly(A)-containing RNA from HeLa cells it has been subsequently used to analyze many eukaryotic cell systems (Hastie & Bishop, 1976; Monahan et al.,1976; Birnie et al.,1974; Williams et al.,1977). Much of this work indicated that poly(A) containing RNA from a variety of species can be generally divided into two or three classes with the most genetically complex class containing the fewest copies of each gene sequence, and the lowest complexity class containing the most copies. Reports on various mammalian cells vary with respect to the number of abundance classes and genetic complexities which range between 8,000 and 15,000 sequences in most mammalian cells (Bishop et al.,1974; Hastie & Bishop,1976; Young et al.,1976; Ostrow et al.,1979).

1.13 Aims and objectives

Since CF is an inherited disease it is highly likely that the defect is represented in the messenger RNA content of CF cells, whether it be a change in abundance of a single sequence or set of sequences, or merely a structural alteration. In view of this and the lack of any knowledge of chromosomal location or a specific abherrant gene product it was decided to compare the expressed sequences from cells derived from CF and normal individuals, to try to detect any differences between them. The cell type we initially decided to study were fibroblasts for three main reasons:-

A) CF is assumed to be manifest in all cells (see section 1.3.1)

B) Primary fibroblasts grow reasonably well in culture without transformation and so their expressed sequences should, as near as possible, reflect the in vivo situation.

C) Availability of CF material

The third criterion is particularly important because there are ethical problems concerning withdrawal of large quantities of blood from CF children, in order to isolate RNA from their lymphocytes, for example. However it takes time to accumulate enough fibroblast material to carry out the experiments in mind. On the other hand a comparatively good source of normal human blood was available when this study commenced, and so normal human lymphocytes were used as a model system to develop the

technology.

The aim of the study was to compare the expressed sequences from normal and CF fibroblasts. This involved using recombinant DNA techniques to construct bacterial "libraries" of expressed sequences of normal lymphocytes, initially, and normal and CF fibroblasts subsequently. Once the libraries were constructed it would then be possible, by various methods, to compare the expressed sequences they represented. Within the limits of sensitivity of these methods it would be possible to detect any sequences which were differentially expressed in CF cells. These sequences could then be used to study CF homozygotes and heterozygotes to establish their relationship to the basic defect, and possibly to develop diagnostic tests for CF.

2. MATERIALS and METHODS

2.1. MATERIALS

2.1.1) Chemicals

All chemicals were Analar grade and supplied by B.D.H. Poole, England unless otherwise stated.

2.1.2) Buffers

Trizma base (TRIS), HEPES and sodium cacodylate were obtained from Sigma Chemical Co, St Louis, Missouri, U.S.A.

2.1.3) Isotopes

All radioisotopes were obtained from the Radiochemical Centre Amersham, England.

2.1.4) Chromatography, electrophoresis, etc

Acrylamide and bis acrylamide for preparative gels were from Bio-Rad laboratories, Richmond, California, U.S.A. Ficoll and Sephadex were supplied by Pharmacia, Upsalla, Sweden and poly vinyl pyrrolidone and bovine serum albumin (BSA) were from Sigma Chemical Co. Agarose was from Miles laboratories Inc, Kankakee, Illinois, U.S.A. Oligo (dt) cellulose (grade T3) was obtained from Collaborative Research Inc, Waltham, Massachusetts, U.S.A. Ethidium bromide and Orange G were obtained from Sigma Chemical Co. Hypaque Sodium was obtained from Winthrop laboratories Surbiton-upon-Thames, Surrey.

2.1.5) Photography and Autoradiography

FP4 film, Phenisol and PQ Universal developers, and Hypam fixer were from Ilford, Basildon, England. RX Medical X-ray film and Mach 2 intensifying screens were from the Fuji Photo Co, Ltd, Tokyo, Japan. Ultraviolet light source was a short wave transilluminator from UV Products Inc, San Gabriel, California U.S.A.

2.1.6) General Enzymes

DNA Polymerase I (Klenow fragment), Deoxyribonuclease I, Proteinase K and Calf Intestinal Phosphatase were from Boehringer Mannheim, West Germany. Terminal deoxynucleotidyl transferase and polynucleotide kinase (T4) were obtained from PL Biochemicals, Milwaukee, Wisconsin, U.S.A. T4 DNA Ligase was obtained from New England Biolabs, Beverly, Massachusetts, U.S.A. S_1 nuclease and lysozyme were obtained from Sigma Chemical Co. Bacterial alkaline phosphatase & Sepharose linked bacterial alkaline phosphatase (BAP-MATE) were obtained from Bethesda Research Laboratories Inc (BRL), Rockville, Maryland, U.S.A. Eco R1 methylase was a gift from Lesley Woods, Imperial College, London. RNA dependent DNA polymerase (Reverse transcriptase) was a gift from Dr. J. Beard, through the Biological Carcinogenesis Branch, Viral Oncology Program, National Cancer Institute.

2.1.7) Restriction Enzymes

Pst I was obtained from Boehringer. All other restriction

enzymes were obtained from New England Biolabs Inc, unless otherwise stated.

2.1.8) Nucleic Acids

Phage λ DNA was obtained from Boehringer. Herring sperm DNA, polyadenylic acid and deoxyribonucleotides were obtained from Sigma Chemical Co. Oligo dT and synthetic Eco RI and Bam HI recognition sites were obtained from Collaborative Research Inc.

2.1.9) Scintillation Counting

PPO and POPOP were obtained from Fisons, Loughborough, England. Triton X-100 was obtained from Koch-Light Laboratories Ltd, Colinbrook, England.

2.1.10) Filters

GF/A, GF/C and No.71 were obtained from Whatman Ltd, Maidstone, Kent. Nitrocellulose filters and ABM paper were obtained from Schleicher and Schuell, Dassell, West Germany.

2.1.11) Bacterial Culture

Yeast extract and tryptone were obtained from Oxoid Ltd, Basingstoke, England. Agar was obtained from Difco Laboratories, Detroit, Michigan, U.S.A. Streptomycin, ampicillin, chloramphenicol, tetracycline and cycloserine were obtained from Sigma Chemical Co. All antibiotics were filter sterilized using individual filter units obtained from Millipore S.A. Molsheim,

France. 90 mm petri dishes and micro titre plates were obtained from Sterilin Ltd, Teddington, England.

2.1.12) Miscellaneous

Rat liver ribonuclease inhibitor was a gift from G D Searle, High Wycombe, England. Nuclease free BSA was obtained from B.R.L. Inc, and guanidine hydrochloride from Sigma Chemical Co Ltd.

2.2. General Methods

2.2.1 Restriction enzyme digestion

DNA was dissolved in sterile distilled water at a concentration of 200 μg/ml or less. One tenth volume of 100 mM TRIS-HCl pH 7.5, 60 mM $MgCl_2$, 60 mM β - mercaptoethanol, 1mg/ml BSA was added to the DNA solution. This reaction mix was augmented with an appropriate amount of NaCl or KCl as recommended by the enzyme suppliers. Enzymes were added to give an estimated 2-3 fold over digestion within the allocated digestion time (eg.- one μg of DNA plus three units of enzyme for one hour).

Digests were performed in eppendorf tubes at 37°C. After incubation, digests were placed on ice and small aliquots were electrophoresed on 1 or 2% agarose gels to analyse the extent of the reaction. Preparative samples were subsequently phenol extracted to remove protein and ethanol precipitated.

2.2.2 Ethanol precipitation of nucleic acids

Nucleic acids were routinely precipitated by the addition of one tenth volume of 4M NaCl and two volumes of ethanol. Samples were chilled overnight at -20°C or for 10-15 minutes on a dry ice/ethanol bath. The sample was then centrifuged in a Sorvall HB4 head at 16,000g for 10 minutes or for 10 minutes in an Eppendorf microfuge (Eppendorf 5412). When very small quantities of material were being precipitated, the centrifugation was performed in a

Sorvall OTD-50 Ultra centrifuge at 40,000 r.p.m. in and SW40 Ti rotor. In each case the supernatant was removed and the pellet washed in 70% ethanol. After respinning, the supernatant was removed and the pellet dried under vacuum. The DNA was redissolved in an appropriate volume of water or the required buffer.

2.2.3 Agarose gel electrophoresis

DNA and RNA samples were analysed by agarose slab gel electrophoresis. The appropriate percentage of agarose was dissolved with heating in "E" buffer (40 mM TRIS, 20 mM Sodium acetate, 2 mM EDTA pH 7.5) for general use or in T buffer (40 mM TRIS, 250 mM sodium acetate, 1 mM EDTA, 0.25 μg/ml ethidium bromide, pH 7.7 with glacial acetic acid) for gels for transfer hybridisations. For RNA samples 0.1% w/v SDS was included in both the gel and electrophoresis buffer. Gels were poured at 60°C allowed to set, then the sample well formers removed and the tank chambers filled with the appropriate buffer. Samples were loaded onto gels in E buffer plus 2% (w/v) ficoll with a small amount of orange G as a marker. Electrophoresis was at 20 V until the sample could be seen to be in the matrix of the gel and subsequently at 100-150 V until the orange G marker had moved to the bottom of the gel.

2.2.4 Visualisation of DNA+RNA

Following electrophoresis gels were stained in E buffer

containing 0.5 μg/ml ethidium bromide for 10-15 minutes. In the case of RNA, gels were soaked for 30-60 minutes in E buffer containing no SDS to remove the SDS which inhibits the binding of ethidium to the RNA. The nucleic acid was then visualized with short u.v light and photographed on Ilford FP4 film using a Polaroid MP-4 camera. Films were developed in PQ Universal developer and fixed in Hypam fixer.

2.2.5 Gel filtration

Gel filtration was used routinely to remove unincorporated nucleotides from cDNA syntheses, nick translations etc. Columns used were either pasteur pipettes plugged with glass wool or hollow glass rods when larger volumes were to be filtered. The matrix of the columns was either Sephadex G-50 or G-100 as indicated in the text. The Sephadex was pre swollen as advised by the distributors and then stored in 0.2% sodium azide. Before use each column was pre-equibilibrated with the required buffer and the void volume calculated by the elution of Dextran Blue from the column. Samples were collected in 1.5 ml eppendorf tubes, monitored, and the excluded peak of radioactivity pooled.

2.2.6 Deionization of formamide

Formamide was deionized using 2g (100 mls) of mixed bed resin and swirling gently for approximately 1 hour. The formamide was filtered through a Whatman No.1 filter and the conductivity was

less than 10 mho. Formamide (a potent teratogen) was stored in the dark at -20°C for up to two months.

2.2.7 Autoradiography

Agarose gels used to size radioactively labelled molecules were dried under vacuum and mounted on cardboard sheets. Nitrocellulose filters from hybridisation reactions were dried in air and similarly mounted. All samples were wrapped in clingfilm and autoradiographed at -70°C using intensifying screens. Films were sensitised by preflashing if required.

2.2.8 Determination of radioactivity

^{32}P radioactivity was usually measured by Cerenkov counting and the efficiency of counting was approximately 30%. TCA precicpitated ^{32}P labelled was immobilised on Whatman GF/C or Millipore filters (for cDNA), dried and counted in toluene based scintillant (TBS: 0.5% w/v PPO, 0.03% POPOP in toluene). ^{3}H and aqueous samples were counted in TXBS (2 vols. TBS and 1 vol. Triton -X) and the counting efficiency was also 30%.

2.3. Bacteriological Methods

2.3.1. Bacterial strains and plasmid vectors

Bacterial hosts used in all recombinant work in this thesis were E.coli strains HB101 and MC 1061. HB101 was a gift from Dr. H. Boyer and its genotype is str^R, rB$^-$, rec A$^-$, pro$^-$, gal$^-$,(Boyer & Roulland - Dussoix, 1969). MC 1061 was a gift from Dr. R. Dalgleish. The plasmid vector pAT153, to which both strains are hosts, is a derivative of pBR322 and was a gift from Prof. D. Sherratt (Twigg and Sherrat, 1980).

2.3.2. Preparation of plasmid DNA

Plasmid DNA was prepared by two methods. The first method involved detergent lysis of bacteria (Clewell & Helinski, 1969) followed by centrifugation in ethidium bromide/caesium chloride density gradients (Radloff *et al.*,1968). The second method was devised by D. Ish-Horowitz and eliminates the cleared lysate procedure and all phenol extractions.

2.3.2(i) Liquid cultures were grown in L broth (5g/l yeast extract 10g/ltryptone, 5g/l NaCl, 1g/l D-glucose pH 7.2. Autoclaved broth was augmented with antibiotics as appropriate (see text).

50 ml cultures were grown overnight at 37^{o}C, and diluted 1 in 100 into the required volume of L broth (typically 1 litre). Cultures were grown to an O.D.$_{650}$ of 0.8 and chloramphenicol was added to a final concentration of 0.17 mg/ml and the plasmids

allowed to amplify overnight.

Bacteria were harvested by centrifugation in a sorvall GSA rotor for 10 minutes at 8000g at 4°. They were then washed in 25% (w/v) sucrose, 50mM TRIS, pH 8, repelleted and resuspended in a small volume of the sucrose/tris solution. Cells were lysed by the addition firstly of 450µg/ml freshly dissolved lysozyme for 5 minutes on ice. The suspension was made 100 mM TRIS pH 8, 70 mM EDTA and left for 5 minutes on ice. Finally the cells were lysed by the addition of Triton X-100 to 0.2%. After 20 minutes on ice lysis was usually complete, as indicated by the suspension becoming very viscous. If lysis was not apparent, 5 minutes at 37° was usually sufficient to complete it.

Cell debris and the chromosomal DNA were pelleted in a Beckman SW 40 rotor at 150,000g at 4° for 30 minutes and the resulting supernatant contained the plasmid DNA. This was phenol extracted several times before digesting with 50 ng/ml RNase at room temperature for 1 hour. The plasmid was again phenol extracted and finally chloroform extracted. The supernatant was made 8 mM with K_2HPO_4, pH 7.5, 1 mg/ml ethidium bromide and solid caesium chloride to 1.56 g/ml. The solution was cleared by centrifugation at 30,000g for 30 minutes in the SW 40 rotor and centrifuged to equilibrium at 140,000g in the SW 40 for 40 hours or the Sorvall vertical rotor TV-865B for 16 hours. The plasmid band was visualised under UV illumination and collected by

piercing the tube with a syringe needle. Ethidium bromide was removed by extraction with caesium chloride saturated isoamylalcohol and the DNA dialyzed extensively against TRIS/EDTA and ethanol precipitated. Plasmids were subsequently analysed by gel electrophoresis.

2.3.2.(ii) The second method of plasmid preparation was suggested by D. Ish-Horowitz. Bacteria were grown, amplified and harvested exactly as described in the previous section. Each pellet was then suspended in a small volume of 50 mM glucose, 25mM TRIS-HCl pH 8 10 mM EDTA and 5 mg/ml fresh lysozyme. This solution was kept at room temperature for 10 minutes when twice the volume of 0.2M NaOH 1% SDS were added. This suspension was shaken gently on ice for 5-10 minutes before the solution was made to 1M potassium acetate pH 4.8. This suspension was vortexed thoroughly and kept on ice for 30 minutes. Cell debris and chromosomal DNA were then removed by centrifugation in a Sorvall GSA rotor for 10 minutes at 8000g. The resultant supernatant contained almost exclusively supercoiled plasmid DNA which was precipitated at room temperature for 5 minutes by the addition of 60% volume of propan-2-ol. This was subsequently centrifuged at 10,000 r.p.m. in an HB-4 rotor dissolved in 10 mM TRIS/1mM EDTA and spun to equilibrium in a caesium chloride density gradient as in the previous method.

2.3.3 Transformation of HB101

E.coli strain HB101 was made competent for transformation by

an adaption of the method of Wensink et al (1974).

An overnight culture of HB101 was diluted 1:100 into L broth plus streptomycin and grown, with shaking, at 37°C to an OD 650 of 0.5. The cells were harvested by centrifugation at 6000g in a Sorvall HB-4 rotor and resuspended in a volume of freshly made 80 mM $CaCl_2$ equal to the original culture volume. After 15 minutes on ice the cells were concentrated 10 fold by pelleting again and resuspending in one tenth the original volume of 80 mM $CaCl_2$, 15% glycerol. The cells were dispensed into 200 µl aliquots. The standard transformation method used 2 volumes of competent cells and 1 volume of the transforming DNA (1 ng/ml) in TEN buffer.

Competent bacteria and DNA were mixed, left on ice for 10 minutes and then given a - 55°C (isopropanol:dry ice) cold shock for 100 seconds. The mixture was allowed to thaw at room temperature then placed on ice for a further 5 minutes, followed by incubation at 42°C for 1 minute, and then by 2 minutes at 0°C. 700µl of L broth was added and the final transformation mixture was incubated at 37°C for 30 minutes. This allows the cells to recover from the $CaCl_2$ treatment (Lennox, 1955).

Following transformation the cells were filtered directly onto nitrocellulose filters to give an average colony density of 5000 transformants per 9 cm^2 diameter filter. The filters were incubated at 37°C for 1 hour on L-agar plates containing no antibiotics and then transferred to L-agar containing antibiotics

in order to select for transformants. The presence of the plasmid pAT 153 confers resistance to tetracycline and ampicillin except when foreign DNA fragments are inserted in the antibiotic resistance genes.

2.3.4. Transformation of MC 1061

A streak of MC 1061 growing on an L-agar plate plus streptomycin were inoculated into a small volume (5 mls) of L broth + streptomycin. This was grown at 37^{O} with vigorous shaking until turbulent and then diluted 1:100 into a larger volume (100-250 mls) of L broth and streptomycin. The bacteria were grown to an OD 650 of 0.2 and chilled on ice water for 15-20 minutes. They were harvested by centrifugation at 0^{O}C, 5000 r.p.m. in a Sorvall HB-4 rotor for 5 minutes. Each pellet derived from 50 mls of original culture was resuspended in 10 mls of sterile 60 mM $CaCl_2$ 10 mM pipes 15% glycerol and chilled on ice water for 20 minutes before being pelleted as before. Bacteria were finally resuspended in 2 mls of the $CaCl_2$ solution per 50 mls of culture and kept on ice water for 15-20 minutes before use. Cells which were not used immediately could be snap frozen in a dry ice/methanol bath and stored for up to 1 week at -70^{O}C without significant loss of efficiency.

2.3.5. Preparation of single colony lysates

For rapid identification and sizing of plasmids from single colonies two very similar procedures have been described (Telford

et al.,1977; Barnes, 1977).

In both methods amplification of the plasmid within the bacteria is necessary for visualization. The amplification procedure depends on whether the colonies are growing on filters or on agar plates.

Filters

When bacteria were about 1 mm in diameter the filters were transferred to L agar plates containing chloramphenicol (170 µg/ml) and incubated overnight at 37°C. Single colonies were then picked into 20µl of 50 mM TRIS, 5mM EDTA pH 7.6 and suspended by vortexing vigorously. 1 µl of 15% SDS and 50µl of Proteinase K (1 mg/ml in 50 mM TRIS, 5 mM EDTA pH 8) were added and the mix incubated at 37°C for 2 hours. The sample was next split in half and to half was added 50 ng of supercoiled pAT 153 DNA as a marker. Sucrose was then added to each sample to 1% and bromophenol blue marker. 20µl of molten 0.8% (w/v) agarose was added to each sample prior to loading onto a 0.8% (w/v) agarose gel. Electrophoresis was at 10 V for approximately 6 hours.

Plates

A colony was inoculated into 0.5 mls. of warm L broth and incubated with shaking for 2 hours at 37°C. Chloramphenicol was added to a final concentration of 170 µg/ml and the sample was incubated overnight at 37°C. The sample was then centrifuged in a microfuge for 5 minutes and the pellet resuspended in 20µl of 50

mM TRIS, 5 mM EDTA pH 8.0. The lysis was then continued as prevously described.

2.3.6. Replica plating

Preparation of nitrocellulose filters containing replicas of bacterial colonies was essentially as described by Hanahan and Meselson (1980). Bacteria were grown on filters until colonies were 0.5 mm in diameter. Two sterile Whatman filters were laid on an EtOH washed glass plate and the master filter laid face up on top of these. 3 or 4 orientation marks were made within the master filter by pen in order to identify any bacteria required for further study.

The replica filter was prewetted on a new agar plate inverted and placed dry side down on top of the master filter. Two further Whatmen filters were then laid on top of this sandwich and pressed down firmly with a glass plate for even distribution of pressure. The replica filter was then marked to align with the master. The two filters were carefully peeled apart and the replica grown, colonies up , as usual. A second replica was usually taken from the master which was finally regrown for 1-1½ hours before more replicas were made.

2.3.7. Sealing nitrocellulose filters

Sealing nitrocellulose filters were basically as described by Grunstein and Hogness (1975) except that the filters were not

treated with proteinase K unless otherwise stated.

Colonies were lyzed by placing each nitrocellulose filter onto large Whatman filter sheets saturated in 0.5 M NaOH for 15 minutes. The alkali was neutralized by transferring the nitrocellulose disc to a similar Whatman sheet saturated with 1M TRIS pH 7.6 for 2 minutes. The DNA was then immobilized on the filters by 2 minutes on 1.5 M NaCl/0.5M TRIS pH 7.6 followed by 2 minutes on 0.3M NaCl and finally by baking for 2 hours at 80°C.

2.3.8. Storage of bacteria

2.3.8.(i) Liquid culture

Bacterial cultures were stored in the long term by growing to confluence in L broth and the addition of an equal volume of sterile glycerol. This suspension was thoroughly mixed and could be stored at -20°C indefinitely. Bacterial colonies growing on nitrocellulose discs could be similarly stored purely by shaking the bacteria in a volume of sterile L broth and addition of an equal volume of glycerol.

2.3.8.(ii) Filters

Colonies on filters could be maintained indefinitely by quick freezing at -55°C. Filters were prepared by filtering as described in Section 2.3.3. but colonies were grown on F plates instead of L agar plates.

F plates

3.6 mM $K_2 HPO_4$

1.3 mM $K H_2 PO_4$

2 mM Na Citrate

1 mM $Mg SO_4$

4.4% glycerol

1.5% Agar

Alternatively pre existing template filters were incubated for a few hours on F plates. Replica filters were wetted on fresh F plates prior to use and replication was carried out as described in 2.3.6 but the two filters were left together, sandwiched between several dry Whatman filters, plus one wet Whatman filter to maintain humidity. This stack was then placed inside a freezer bag sealed, laid on a bed of dry ice/isopropanol and when frozen it was then stored at -55°C or below. When needed, the bag was removed and brought to room temperature, the filters were peeled apart, laid on fresh plates and incubated until small distinct colonies appeared.

2.3.9. Containment and Safety

All genetic manipulations were carried out under C II containment conditions in accordance with G.M.A.G advice for recombinant DNA research as based on the Williams report 1976.

Williams, R.E.O 1976 Report ot the working party on the Practice of Genetic Manipulation, Cmnd 6600, H.M.S.O, LONDON.

2.4. Specific Methods

2.4.1 Isolation of lymphocyte RNA

a) Sterilization of glassware and solutions

i) Glassware

Glassware was washed thoroughly, rinsed with distilled water containing 0.1% Diethlpyrocarbonate (DEPC), covered with tin foil and autoclaved.

ii) Solutions

Solutions were made up using high purity chemicals, made 0.1% with DEPC and autoclaved.

N.B. DEPC a) breaks down to CO_2 and ethanol on boiling.

b) reacts with RNA with resultant loss of biological activity.

0.1-1% SDS is an extremely good inhibitor of ribonuclease and was used in all solutions whenever possible. The presence of SDS allowed work to be carried out at room temperature. Great care was taken to remove all traces of SDS from the RNA preparations before use in transcription experiments. 2-3 washes of ethanol precipitated RNA with 70% ethanol was sufficient to remove all traces of SDS when precipitated with the RNA.

b) Fractionation of lymphocytes from blood

Human blood was obtained as buffy coat residues from the National Blood Transfusion Service or as heparinized blood samples

from normal and cystic fibrotic patients.

Lymphocytes were separated by centrifugation through gradients of Ficoll-Hypaque.

Gradients were prepared as follows:-

a) 9% Ficoll-400 was prepared by dissolving 45g in 500mls of distilled water.

b) 34% Hypaque was prepared by dissolving 68g in 200mls of distilled water.

480mls of Ficoll-400 and 200mls of Hypaque were mixed and 80 mls. was aliquoted into sterile repelcoted glass MSE bottles. Blood samples were diluted using normal room temperature saline and 150 mls. of blood was carefully layered on top of each 80 ml. gradient using an autoclaved rubber tube. The tubes were balanced and spun in an MSE centrifuge for 30 minutes at 1300 r.p.m. at 18-20^{o}C with steady acceleration and deceleration. The plasma layer was removed with a suction pump and lymphocytes were removed from the top of the gradient by pipette. Lymphocytes were then washed thoroughly with normal saline.

c) <u>Extraction of RNA from lymphocytes</u>

The washed lymphocytes were placed in 10-20 volumes of a solution made up of 19 volumes of 6M guanidine hydrochloride (Gu HCl), 1 volume 2M potassium acetate (K Ac) pH 5.0. The mixture was then sonicated until non-viscous to shear the DNA. (3x15 second bursts was usually sufficient). The mixture was

centrifuged at 10K for 5 minutes at 4°C to remove any insoluble debris and ½ volume of -20°C 95% ethanol was added to selectively precipitate the RNA. After being left overnight at -20°C this was centrifuged at 10K, -10°C for 15 minutes and resuspended in one tenth the original volume of 9.5 vol. 6M Gu HCl, 0.5 vol. 2M KAc. pH 5, 0.5 vol. 500mM EDTA. Once in solution (usually requiring gentle homogenization) the RNA was again selectively precipitated by addition of ½ vol. of 95% ethanol at -20°C overnight. After spinning at 10K, -10°C for 15 minutes the pellet was resuspended in 2-5mls of sterile ribonuclease free 20mM EDTA with gentle homogenization. The solution was then extracted with an equal volume of chloroform-octanol (4:1) and shaken at room temperature for several minutes. After spinning at 10K 5 minutes at room temperature the phases were separated and the process repeated until no protein interface was visible. The pooled aqueous phases were made to 0.4M NaCl, two volumes of EtOH were added and left to precipitate at -20°C overnight. The suspension was then spun at 10K, 10 minutes, 0°C and the pellet of total RNA washed in 70% EtOH and dissolved in the required buffer.

2.4.2 Oligo (dT) cellulose method of poly A^+ RNA isolation

Poly A^+ RNA was prepared from total RNA by oligo (dT) cellulose affinity chromatography essentially as described by Aviv and Leder (1972).

Buffers

TS = 10mM TRIS pH 7.4, 0.1% SDS

NTS = 0.5 M NaCl, 10mM TRIS pH 7.4, 0.1% SDS

a) Preparation of column

Oligo dT cellulose was suspended in TS buffer and 1-2mls were added to a small syringe plugged with DEPC treated glass wool. The column was washed with several bed volumes of TS buffer followed by 0.1 M NaOH. This was allowed to stand at room temperature for 30 minutes and then washed out with several bed volumes of NTS buffer. The column was then ready for use.

b) Fractionation of poly A^+ RNA

The RNA was suspended in 5 mls of NTS buffer and heated to 65°C for 5 minutes to overcome aggregation. This was then applied to the column, the eluate collected and reapplied twice more. The column was then washed with NTS buffer (approx. 20 mls) until no RNA (ie. A_{260}) could be detected in the eluate. The bound RNA was then eluted using 5 mls of TS buffer and the $O.D._{260}$ measured to estimate yield.

The sample was then made up to 0.5 M NaCl (ie. NTS) and reapplied to the column and treated exactly as before. After two cycles of chromatography the poly A^+ RNA was recovered by ethanol precipitation and taken up in sterile ribonuclease free buffer as required

2.4.3 Synthesis of complementary DNA (cDNA)

cDNA was synthesised in a reaction mixture containing the following: 50mM TRIS/HCl pH8.2; 50mM KCl; 6mM $MgCl_2$; 5mM dithiothreitol; 500μM each of TTP,dCTP,dATP and dGTP; 25 units/ml rat liver ribonuclease inhibitor (Searle); 1 unit/ml. oligo(dT), 250 units/ml reverse transcriptase and 10-50μg/ml polyA$^+$ RNA. The reaction mixture was incubated at 42^{o}C for 1 hour and stopped by addition of one tenth volume of 100mM EDTA. RNA was degraded by addition of one fifteenth volume of 5M NaOH followed by incubation at 60^{o}C for 20-30 minutes. The solution was neutralised with 1N HCl and the cDNA desalted on a G-50 Sephadex column in 5mM TRIS pH 7.5, 5mM NaCl. ^{3}H cDNA (10^7 c.p.m./μg) was synthesised using 1 mCi/ml of ^{3}H dCTP (22 Ci/mMole) and the 500μM dCTP was reduced or excluded from the incubation. The ^{32}P labelled cDNA used in the cloning experiments was synthesised by adding 100 Ci/ml. ^{32}P dCTP (400 Ci/mMole) and reducing or eliminating the unlabelled dCTP concentration.

2.4.4 Synthesis of double stranded cDNA

cDNA was used as the template for the double stranded cDNA (ds cDNA) synthesis using the Klenow fragment of E.coli DNA polymerase I (Klenow et al.,1971). Reaction conditions were as follows: 30 mM TRIS/HCl pH 7.5; 4 mM $MgCl_2$; 0.5 mMβ - mercapto-ethanol; 70 mM KCl; 500μM each of dCTP, dGTP, TTP, dATP; 1-15 μg/ ml of ^{32}P labelled cDNA and 75 units of DNA polymerase I

(Klenow fragment); incubation was at 18°C for two hours. After incubation the mixture was made 10mM with EDTA and passed over a G-50 Sephadex column equilibrated with S_1 buffer (70 mM Na acetate, 2.8 mM Zn_2SO_4, 140 mM NaCl pH 4.5). To the excluded volume 40 units of S_1 nuclease were added and the solution incubated at 37°C for 1 hour. The incubate was extracted with phenol/chloroform (1:1) and the aqueous phase fractionated on a 15-30% linear sucrose gradient. Aliquots of the gradient fractions were analysed on a 1% agarose gel and those fractions containing the ds cDNA (9S and greater) were pooled and precipitated with ethanol.

2.4.5 Homopolymer tailing

DNA tailing reactions were carried out in incubations containing 100 mM TRIS/HCl pH 7.3, 140 mM sodium cacodylate, 1 mM β-mercaptoethanol, 1 mM cobalt chloride, BSA nuclease free 100μg/ml, DNA termini 10-50 nM and dCTP 100μM or dGTP 1 mM. Analytical tailing reactions were carried out to determine suitable nucleotide concentrations and times for each batch of DNA used. The poly dG tailing reaction of the Pst 1 digested plasmid, was incubated typically for 1 hour at 37°C at a dGTP concentration of 1 mM and DNA terminus concentration of 80 nM. Double stranded cDNA tailing reactions were typically at 37°C for 5 minutes at a dCTP conc. of 100μM and DNA concentration of 5μg/ml. After tailing, all

DNA preparations were phenol extracted, precipitated with ethanol and dissolved in TEN buffer (100 mM NaCl; 10 mM TRIS/HCl pH 7.5; 1m EDTA).

2.4.6 Annealing of tailed DNA

Homopolymer tailed plasmid and ds cDNA in TEN buffer were annealed by mixing at a 5:1 (wt/wt) plasmid: cDNA ratio at a plasmid concentration of 2.5 μg/ml. The mixture was then incubated at 65° for 2 hours and allowed to cool slowly overnight (average rate 2°/hour). Alternatively the mixture was heated at 65°C for 10 minutes at 42°C for 1 hour and then cooled to room temperature (4-5 hours). Samples were then used directly in transformations as described (Sections 2.3.3 and 2.3.4).

2.4.7 Nick translation of DNA

To label DNA to high specific activity for use as hybridisation probes, nucleotides labelled at the position were incorporated into DNA by nick-translation (Rigby et al, 1977). DNA at a concentration of 15 g/ml was incubated at 15° in a small volume containing 120 units/ml. E. coli DNA polymerase I, 170 pg/ml DNase I, 4 μM dATP, 4 μM dGTP 2.6 μM α^{32}-P dCTP, 2.6 μM α^{32}-P TTP (both of specific activity 400 Ci/m mol), 10 mM 2-mercapto-ethanol, 5 mM $MgCl_2$ and 50 mM TRIS pH 7.5. After 90 minutes the reaction was stopped by addition of one tenth volume 100 mM EDTA. After phenol extraction the unincorporated nucleotides were removed by gel filtration and the excluded peak pooled.

2.4.8 Labelling of RNA by T4 polynucleotide kinase

In order to increase incorporation of label, RNA was hydrolysed to an average length of 100-200 bases using 0.1 N NaOH at 0°C for 30 minutes before labelling with γ-^{32}P labelled ATP. (Van de Sande,1973). After alkali cleavage, RNA was neutralized with HCl and ethanol precipitated. After precipitation and centrifugation the RNA was resuspended in sterile water at 50 μg/ml. The conditions for the enzyme reaction were: 1 μg RNA, 10 mM $MgCl_2$, 0.1 M TRIS pH 7.6, 20 mM β -mercaptoethanol and 50 units/ml T4 polynucleotide kinase. The mixture was incubated for 1 hour at 37°C. The reaction was made to 10 mM EDTA and unincorporated nucleotides removed by gel filtration in 0.15 M NaCl, 0.015 M Na citrate, 0.1% SDS of Sephadex G-50. The eluted peak could then be used directly in hybridisation.

2.4.9 Hybridisation of ^{32}P poly A$^+$ RNA to colony DNA on filters

Filters to be hybridised were prewashed in 50% formamide, 0.1% SDS, 5xSSC, 50 μg/ml carrier DNA, 10 μg/ml poly (A) and 1 x Denhardt solution (Denhardt,1966) at 42°C. Hybridisation with ^{32}P RNA was carried out in the same prewash buffer (1 ml per 9 cm filter disc) in heat sealed plastic bags at 42°C for 16 hours. After hybridisation the filters were washed four times in 2 x SSC 0.1% SDS at 65°C and twice in 0.1 x SSC. 0.1% SDS at 65°C. The filters were then air dried and autoradiographed as described (Section 2.2.7).

2.4.10 Hybridisation of ^{32}P cDNA to colony DNA on filters

Filters to be hybridised were prewashed in 3 x SSC, 0.1% SDS , 50 μg/ml carrier DNA, 10 μg/ml poly (A) and 1 x Denhardt solution at 65°C. Hybridisation was carried out in heat sealed bags at 65°C for 16 hours (10^7 cpm/ml). After hybridisation the filters were washed at 65°C four times in 2 x SSC, 0.1% SDS and finally were again air dried and autoradiographed.

2.4.11 Hydroxyapatite chromatography

Hydroxyapatite chromatography of RNA:DNA hybrids was carried out as described by Old et al. (1976) and in detail by P. Tolstoshev (personal communication).

Preparation of hydroxyapatite (HAP)

25g of DNA grade HAP. was suspended in 100 mls of 1M sodium phosphate buffer pH 6.8 and allowed to settle. The buffer was decanted and the HAP was resuspended in 100mls of 0.16 M sodium phosphate buffer pH 6.8 and allowed to settle. The HAP was then resuspended in the same buffer and boiled for 15 minutes. After cooling the HAP was washed once with distilled water and then twice with 0.03 M sodium phosphate buffer pH 6.8 and stored in this buffer at 4°C. (At all stages of preparation the resuspension of the HAP was carried out by gently swirling the flask).

Preparation and use of the column

A purpose built perspex water bath was available which allowed

the use of 5ml. syringes as columns while maintaining a constant water temperature of 60°C around them. Columns were prepared by pouring 1-2 mls. of HAP slurry into the syringe which had been plugged with a GF/C filter. Hybridisation samples were usually diluted 20-fold with 0.15M NaCl in order to stabilise hybrids and dilute any formamide present. 1 ml. aliquots were then allowed to sit on the column for 2-3 minutes to equilibrate to 60°C before being eluted under air pressure. The column was subsequently washed four times with 0.03M sodium phosphate and single and double stranded molecules eluted by increasing the phosphate concentration. Actual concentrations necessary to elute these molecules was deduced by precalibration of the column. Samples were usually counted by radioactivity in PCS as previously described. (Section 2.2.8).

2.4.12 RNA:cDNA hybridisations

Hybridisations were carried out essentially as described by Getz et al.(1975). The required amounts of RNA (in distilled water) and 2.5µg of E.coli rRNA were placed into small siliconised glass tubes and freeze dried. The amount of cDNA necessary for a set of reactions was treated similarly. The cDNA was then dissolved in hybridisation buffer (see below) and aliquots of this mixture were used to dissolve the lyophilised RNA samples. The hybridisation mixtures (2-20µls.) were then sealed in

siliconized glass capillaries.

Hybridisation buffer (0.5M NaCl, 25 mM HEPES, 0.5 mM EDTA, pH 6.8, 50% (v/v) formamide) was prepared as follows:- 4 x concentrated buffer was prepared, passed over Dowex chelating resin, treated with diethyl pyrocarbonate, autoclaved, and mixed with two parts deionized formamide, and one part sterile distilled water.

Capillaries containing hybridisation mixtures were heated to 80°C for 2 minutes, and then incubated at 43°C. Reactions were stopped by immersing the capillaries into ethanol at -20°C.

SI Nuclease assay

The contents of each capillary were flushed out with 250 µls of nuclease assay buffer (70 mM sodium acetate, 2.8 mM Zn_2SO_4, 140 mM NaCl pH 4.5) containing 10 µg/ml of heat denatured calf thymus DNA. 100 µls of SI nuclease (A.oryzae)(2500 units/ml in nuclease assay buffer) was added and the mixture incubated at 37°C for 90 minutes. A 100 µl aliquot of the incubate was removed and the radioactivity determined (Total d.p.m.) A 200 µl aliquot of the incubate was removed and to it was added 50 µls of a 1 mg/ml BSA solution and 50 µls of 6N perchloric acid. The mixture was chilled at 0°C for 30 minutes and then centrifuged at 12,000 x g for 2 minutes. A 200 µl aliquot of the supernatant was removed and the radioactivity was determined (Soluble d.p.m.). From the soluble and total d.p.m. the percentage of the cDNA hybridised was

calculated using the formula $100 - \left(\frac{0.75 \times \text{Soluble d.p.m.}}{\text{Total d.p.m.}} \right)$

2.4.13 Southern transfers

DNA for Southern transfers was fractionated on 0.8-1% agarose gels with DNA size markers. After electrophoresis the gel was photographed and then denatured to render the DNA single-stranded. Denaturation was for 2 hours in 0.5M NaOH, 1.5 M NaCl. The gel was neutralized in 0.5M TRIS-HCl pH 5.5, 1.5 M NaCl. (The DNA strands do not renature during the neutralisation step). DNA was blotted overnight in 20 x SSC onto a nitrocellulose filter using the apparatus described by Southern (1975). The nitrocellulose filter, now with single stranded DNA bound to it, was washed for 10 minutes in 2 x SSC. DNA was immobilised on the filters by baking at 80°C for 2 hours. Non-specific binding sites on the filters were saturated by pre-treatment of the filters prior to hybridisation. All treatments were at 65°C. Filters were soaked for 30 minutes in 3 X SSC, then for 3 hours in 3 x SSC plus 0.2% w/v ficoll, poly vinyl pyrrolidone, and bovine serum albumin. This solution is 3 x SSC, 10 x Denhardts (Denhardt 1966). Filters were washed for a further 60 minutes in 3 x SSC, 10 x Denhardts plus 0.1% w/v SDS, 50 μg/ml sonicated herring sperm DNA and 10 μg/ml polyadenylic acid. The same solution was used for hybridisations. Nick translated probes were denatured by immersing in boiling water for 5 minutes and then added directly to the above solution. Hybridisation times were dependent on the nature of the probe

being used. After hybridisation, the filters were washed repeatedly (6 times) in 3 x SSC plus 0.1% SDS. If appropriate a low salt (high stringency) wash was added to the protocol. (2 x 30 minutes).

2.4.14 ^{3}H-Poly (U) assay of Poly A^{+} RNA

Hybridisation of ^{3}H-poly (U) to poly A^{+} RNA was carried out in 50 µl assays in 2 X SSC. To increasing amounts of RNA (0-500 ngms.) were added 10 ngms. of ^{3}H-poly (U) (Specific Activity=3621 d.p.m./ ngm.) on ice. After incubation on ice for 30 minutes at 45°C the mixture was transferred to ice water for 5 minutes. 40µl of pancreatic RNase was added and incubated on ice for 20 minutes. After addition of 10 µl of 1 mg./ml. BSA carrier a 75 µl was removed and acid precipitated.

2.5 Tissue Culture

Cystic fibrosis fibroblasts were obtained from skin biopsy samples and were gifts from Dr. J. Breslow (Boston) and Dr. A. Harris (Guy's, London). Normal fibroblasts were a gift from Dr. J. Crampton. Cells were routinely maintained as monolayers in 75 cm^2 plastic tissue culture flasks in BHK - GMEM supplemented with 10% foetal calf serum (see below). Cells were grown at 37°C in 5% CO_2 atmosphere until confluent. Medium was then removed from plates by suction, cells were washed with warm phosphate buffered saline and a small volume of warm trypsin solution was added. After incubation at 37° for a few minutes an excess of fresh medium was added and this mixture was then divided into three new flasks and cultured as before. When harvested for RNA extraction cells were trypsinized, neutralized and pelleted in a bench centrifuge, washed several times in phosphate buffered saline and RNA was then extracted as described in Section 2 for lymphocytes.

Bicarbonate - buffered BHK Eagles

450 mls 1 x Baby Hamster Kidney Medium.
22 mls Sodium bicarbonate (85 mgs/ml).
5 mls L- Glutamine (200 mM).
0.5 mls 50 mg/ml Kanamycin
5 mls Non essential amino acids.

Medium was gassed with CO_2 and 50 mls. of foetal calf/newborn calf serum was added.

3. CONSTRUCTION OF cDNA LIBRARIES FROM NORMAL LYMPHOCYTES, NORMAL FIBROBLASTS AND CYSTIC FIBROSIS FIBROBLASTS.

3.1 Introduction

The essence of molecular cloning or recombination in vitro is the joining together of two or more DNA fragments. One fragment, the vector, is capable of replication in a host organism and the other, the cloned fragment, can be passively replicated when joined to the vector. Hybrid molecules are taken up by a host organism by a process called transformation. The discovery of reverse transcriptase in 1970 (Baltimore 1970; Temin & Mizutani, 1970) to make complementary DNA (cDNA) copies of the mRNA and the development of plasmid cloning vehicles (see Bolivar & Backman, 1979) enabled DNA copies of mRNA sequences to be cloned in this way. A complete cDNA clone library from any particular cell therefore is a population of bacterial transformants each containing a plasmid with a single cDNA insert, and with a sufficiently large number of individual transformants that every mRNA molecule is represented at least once in the bacterial population. Any DNA cloning procedure has four essential parts: a method for generating DNA fragments; a method of joining the fragments to vector DNA molecules; a means of introducing the composite DNA molecule into a host cell in which it can replicate and a method of selecting clones of recipient cells

that have acquired recombinants. The initial step in cloning a cDNA library is the reverse transcription of the mRNA population followed by the double stranding of these cDNA molecules. Various methods are avaliable for joining the fragments to vector DNA molecules such as homopolymer tailing, blunt end ligation and the use of synthetic restriction enzyme recognition sites or linkers. In this chapter I shall describe the isolation of poly A^+ RNA from the various cell types, construction of double stranded cDNA copies and the construction of the recombinant libraries with an assessment of the relative merits of some of the cloning methods.

3.2 Preparation of poly $(A)^+$RNA

As discussed previously, initial cloning experiments were carried out using lymphocyte RNA as a model system. However, for the sake of clarity I shall consider the construction of each library in parallel although they were carried out at different periods of time.

The major problem in isolating cellular RNA is removal of protein in such a way that the nucleic acids are not degraded. Many commonly employed methods of nucleic acid deproteinization are based on the use of phenol as a protein denaturant (reviewed by Taylor,1979). High levels of endogenous RNase in human lymphocytes resulted in degradation of RNA isolated by these

standard extraction procedures even in the presence of RNase inhibitors. Alternative procedures have been developed which employ concentrated salt solutions to dissociate ribonucleoprotein complexes (Deeley et al.,1977), and the procedure of Strohman et al.(1977) was found to be effective for lymphocytes. Lymphocytes were initially fractionated from buffy coat residues as in 2.4.1, cells were solubilized in 6M Gu HCl and the total RNA separated from DNA and protein by selective ethanol precipitation. Normal fibroblasts and CF fibroblasts were harvested as in 2.5 and the RNA isolated similarly. Yields of RNA varied considerably but typical values are shown in Table 3.2.

Cell type	Yields of total RNA
Lymphocyte	2mg./10^9 cells
Normal fibroblast	2mg./2x10^8 cells
CF fibroblast	2mg./7x10^8 cells

TABLE 3.2

Under these conditions undegraded total RNA was isolated as judged by analysis on 2% agarose/SDS gels as shown in Fig. 3.2.1. Clearly there are large amounts of undegraded 28S and 18S rRNA indicating that little degradation has occurred during the extraction. The fibroblast samples show the presence of RNA

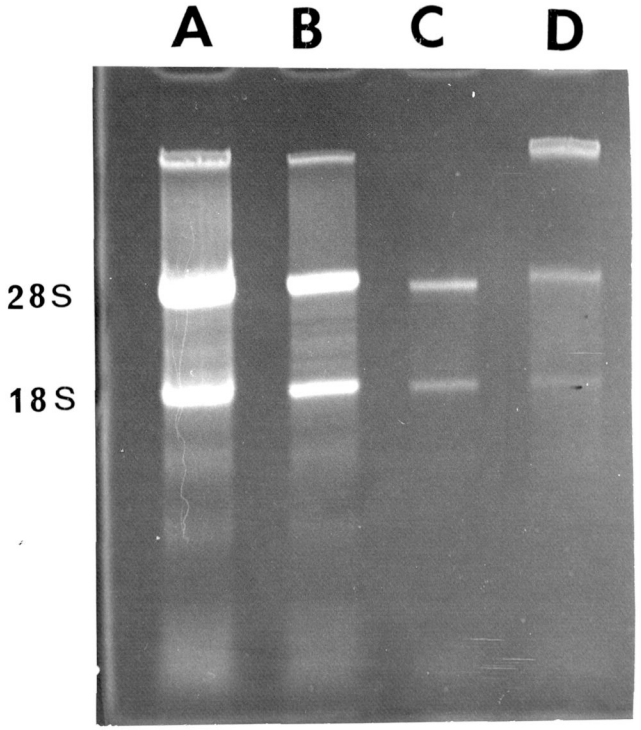


Figure 3.2.1 2% agarose/SDS gel of total RNA

Track a: 20 μg of normal fibroblast RNA.

Track b: 10 μg of CF fibroblast RNA.

Track c: 2 μg of ribosomal RNA.

Track d: 5 μg of lymphocyte RNA.

extending to the top of the gel which is almost certainly due to aggregation of molecules induced by the phenol extraction (Taylor,1979). There is clearly a heterogeneous size distribution for the RNAs but in the fibroblast samples there appear to be superimposed bands suggesting discrete size classes in the RNA. This has also been observed in lymphocyte preparations and will be discussed later.

In almost all cDNA cloning experiments the cDNA synthesis is primed by the use of oligo dT tracts which bind to poly A regions on the poly A^+ RNA. Since rRNA does not contain poly A^+ regions it is possible to copy total cellular RNA in the presence of oligo dT and obtain a cDNA preparation consisting almost entirely of poly A^+ RNA transcripts suitable for cloning (King 1979). However rRNA can be copied at a low efficiency (Williams & Lloyd,1979) and total RNA preparations can inhibit reverse transcriptase activity. For this reason poly A^+ RNA was isolated by oligo (dT) cellulose chromatography (Aviv & Leder,1972) as in 2.4.2. All samples were heated to 65°C before loading in order to reduce any aggregation (see fig 3.2.1) and were passed through two cycles of chromatography. 50% of the total RNA was retained after one cycle and between 2-5% after two cycles of chromatography. This RNA was then ethanol precipitated and dissolved in sterile ribonuclease free distilled water. This

poly A^+ RNA contains, as a result of the isolation procedure, both cytoplasmic and nuclear poly $(A)^+$ RNA.

3.2.1. Use of ^3H poly (U) to assay poly A^+ RNA.

Accurate size distribution of poly $(A)^+$ RNA was determined by hybridizing ^3H poly (U) to RNA after fractionation on a sucrose gradient as described by Rosbash & Ford (1974) and Bishop et al. (1974). The rationale of this procedure is that the ^3H poly (U) hybridizes to the poly A tracts on the poly A^+ RNA and since it is radioactively labelled to high specific activity, extremely small amounts of mRNA can be detected. The reaction was initially calibrated by hybridization of fixed amounts of ^3H poly (U) (10ng.) to increasing amounts of mouse globin mRNA (0-600ng) and assayed by the proportion of the label rendered resistant to pancreatic ribonuclease digestion (ie. in hybrids). Fig 3.2.1.1shows this calibration and shows that 10ng of input ^3H poly(U)is hybridized completely by 200ng of globin mRNA. This means that there is 10ng of poly (A) in 200ng of globin mRNA and if the RNA preparation is 100% pure the poly A represents 5% of the poly A^+ RNA. Similarly Fig 3.2.1.1 shows that the reaction over the range 0-50 ng poly A^+ RNA is linear and so the amount of poly A^+ RNA in any sucrose gradient fraction could be quantitated by the amount of ^3H poly (U) hybridized up to a limit of 50ng poly A^+ RNA. In view of this 3μg of poly A^+ RNA from lymphocyte, normal and CF fibroblasts

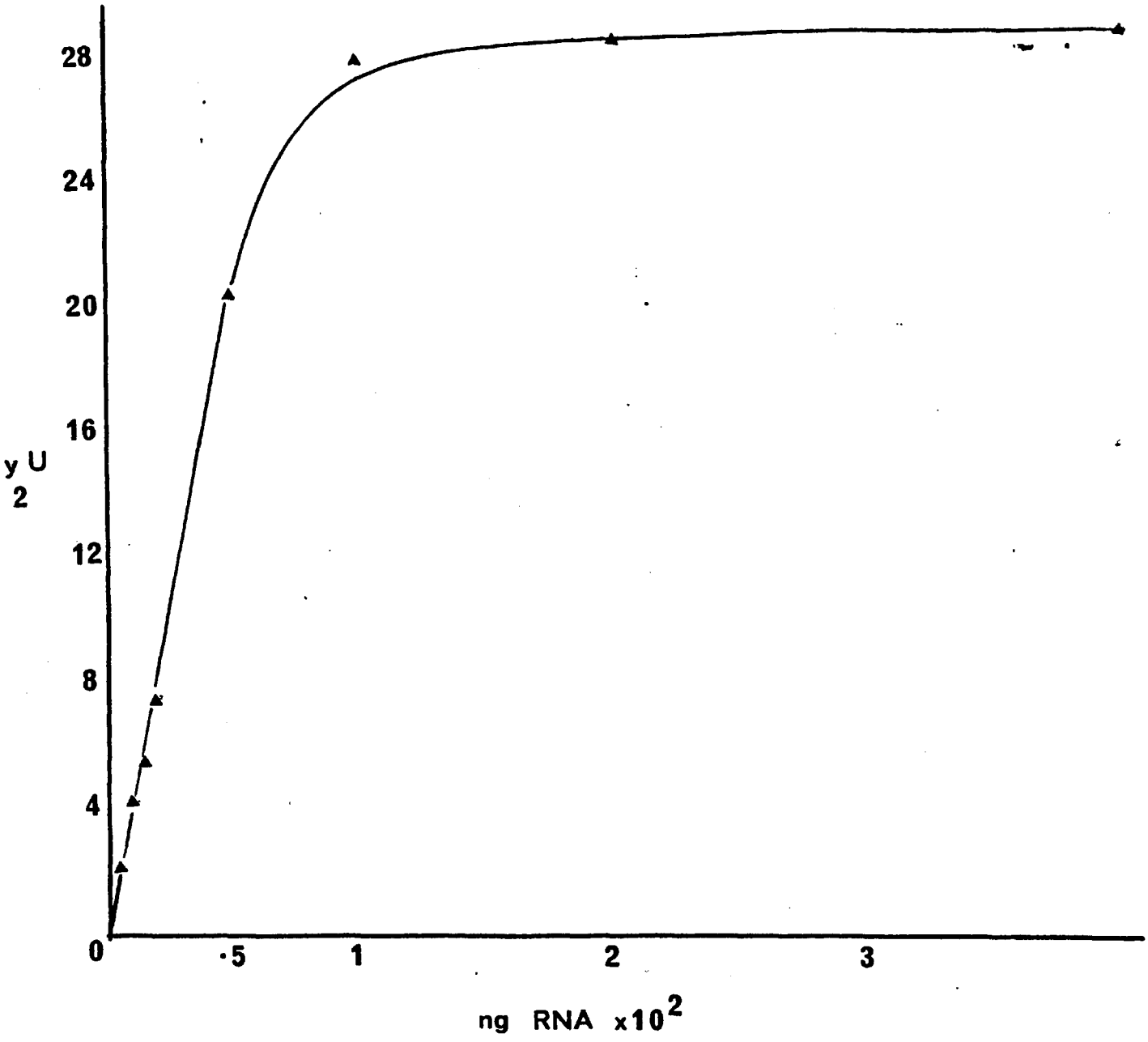


Figure 3.2.1.1 Titration of mouse globin RNA with ^{3}H poly U.

Titrations were carried out as described in the text.

were sized on 15-30% sucrose gradients. The gradients were fractionated in ½ml aliquots of which 40 µl was used for the ^{3}H poly (U) assay. Assuming that the poly $(A)^{+}$ RNA preparations were 100% poly A^{+} RNA this would mean that even if the RNA was present in one ½ml. fraction only 240 ng poly A^{+} RNA would be present. In view of the size heterogeneity on agarose gels and the fact that the preparations were unlikely to be 100% poly A^{+}RNA the ^{3}H poly (U) assay should be quantitative for any fraction. The profiles of the gradients are shown in Fig.3.2.1.2. Clearly there is a heterogeneous size distribution for all samples with an average around 18S for lymphocyte and approximately 20S for the fibroblast samples. This corresponds to average poly A^{+} RNA lengths of 2000 nucleotides for lymphocyte and 2500 nucleotides for the fibroblast samples.

Assuming the average poly (A) length to be 100 nucleotides (Brawerman,1981) the poly(A) moiety in these preparations comprises approximately 5% of the length of the mRNA, which correlates well with other tissues (Hastie & Bishop,1976). Using this figure it was possible to calculate the degree of poly $(A)^{-}$ and rRNA contamination in the RNA preparations. Increasing amounts of RNA from each cell type (0-50ng) were hybridized to 10ng of ^{3}H poly (U). From the point of saturation of ^{3}H poly (U) it was possible to calculate the amount of poly (A) in the preparation and assuming the poly (A) to be 5% of the poly A^{+}RNA the actual

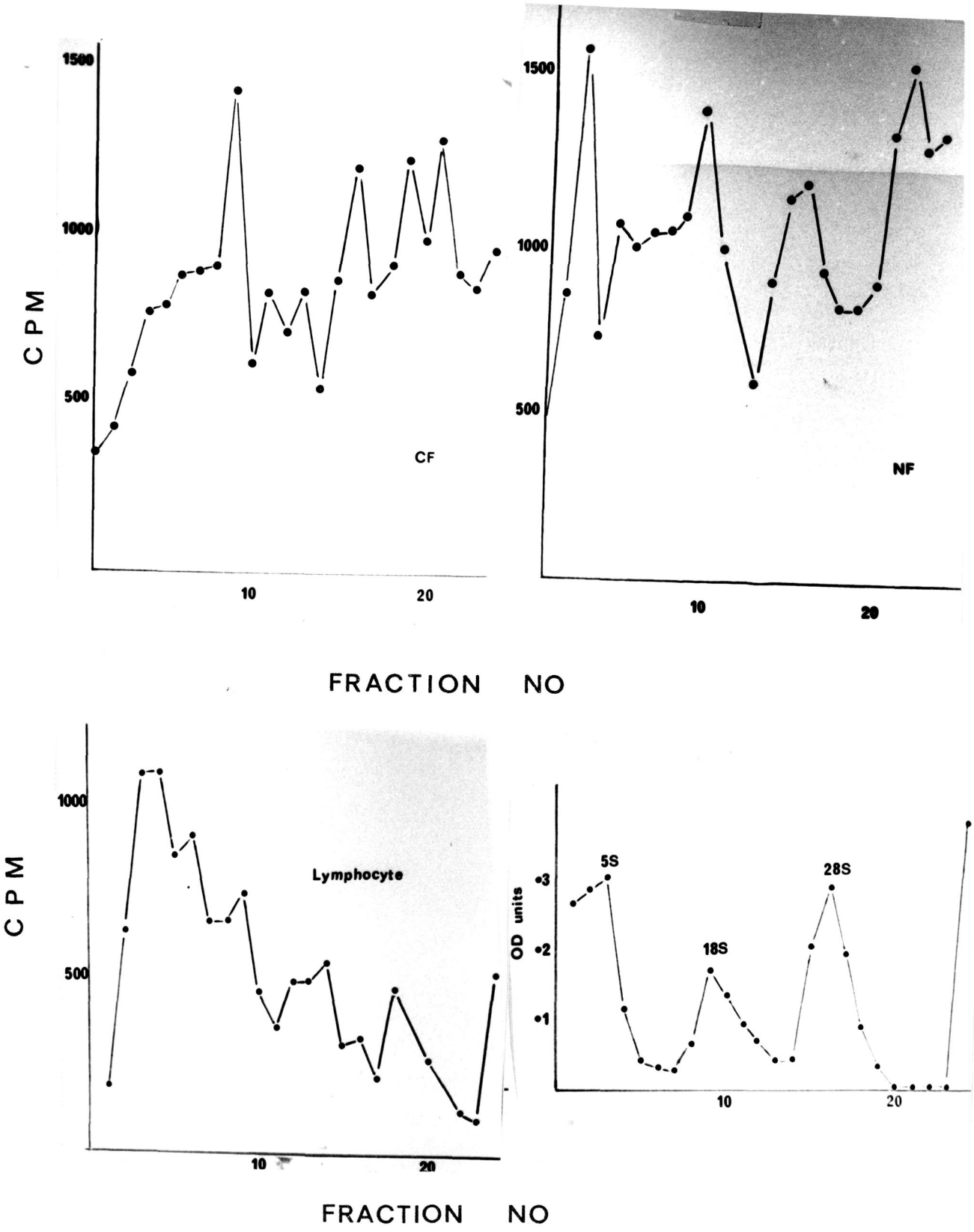


Figure 3.2.1.2. Fractionation of poly A^+ RNA on 15-30% sucrose gradients.

3 μg of poly A^+ RNA and 15 μg rat liver ribosomal RNA markers were centrifuged in 12 ml. gradients at 20°C for 6 hours in an SW40 rotor at 40,000 r.p.m. Aliquots were analysed as described in text.

amount of poly A^+ RNA was calculated. Throughout the course of this thesis poly A^+ RNA populations of varying "purities" were isolated, usually in the range of 25-40% being poly A^+ RNA but occasionally as high as 60-70%. This was probably due to the efficiency of different batches of oligo (dT) cellulose. These figures gave an estimate of the concentration of poly A^+ RNA which was important for subsequent kinetic hybridization studies (see later).

3.3 Synthesis of complementary DNA (cDNA)

Single stranded cDNA was synthesized exactly as described in section 2.4.3 for lymphocyte mRNA. In order to monitor sizes of products, yields etc. ^{32}P was usually used in the synthesis of cDNA for cloning. This, of course, meant that the whole cloning procedure should be performed as rapidly as possible since as the ^{32}P decays nicks will be introduced in the DNA. Synthesis of the normal and CF fibroblast libraries was essentially the same except that the alkali step was reduced to 10 minutes (L. Kunkel, personal communication). Clearly the mRNA:cDNA hybrid will be denatured by the alkali almost immediately, and most of the mRNA should be degraded. Prolonged exposure to alkali does cause damage to DNA so reducing the length of time in alkali should be beneficial. After alkaline hydrolysis all samples were neutralized and passed through Sephadex G-50 to remove unincorporated

nucleotides and the peak of radioactivity pooled. Yields of between 10 and 30% of the input RNA were routinely obtained correlating well with the observation of RNA purity. The size distribution of the ^{32}P labelled cDNA was analyzed on 1 or 2% agarose gels as indicated in Fig 3.3.1. The lymphocyte cDNA was shown to range from approximately 400 to 2000 nucleotides in length. Both fibroblast cDNAs appeared to be somewhat larger than the lymphocyte on average, with some molecules as large as 5 or 6 Kb. It is apparent that, as in the mRNAs, in both the CF fibroblast and lymphocyte cDNA discrete bands are visible. When desalted the cDNAs were eluted from the column with low salt buffer and lyophilized.

3.4. Synthesis of double stranded cDNA

The synthesis of double stranded cDNA was carried out using the Klenow fragment of E. coli DNA polymerase I essentially as described by Humphries *et al.* (1977). The lyophilized sample was resuspended in sterile distilled water and brought to correct reaction conditions by addition of 1/10 volume of 10 times concentrated reaction buffer. Initial double stranding experiments were unsuccessful since the reaction volume was too small to dilute the low salt buffer in which the single stranded cDNA had been lyophilized. Final concentrations of 100 mM NaCl greatly

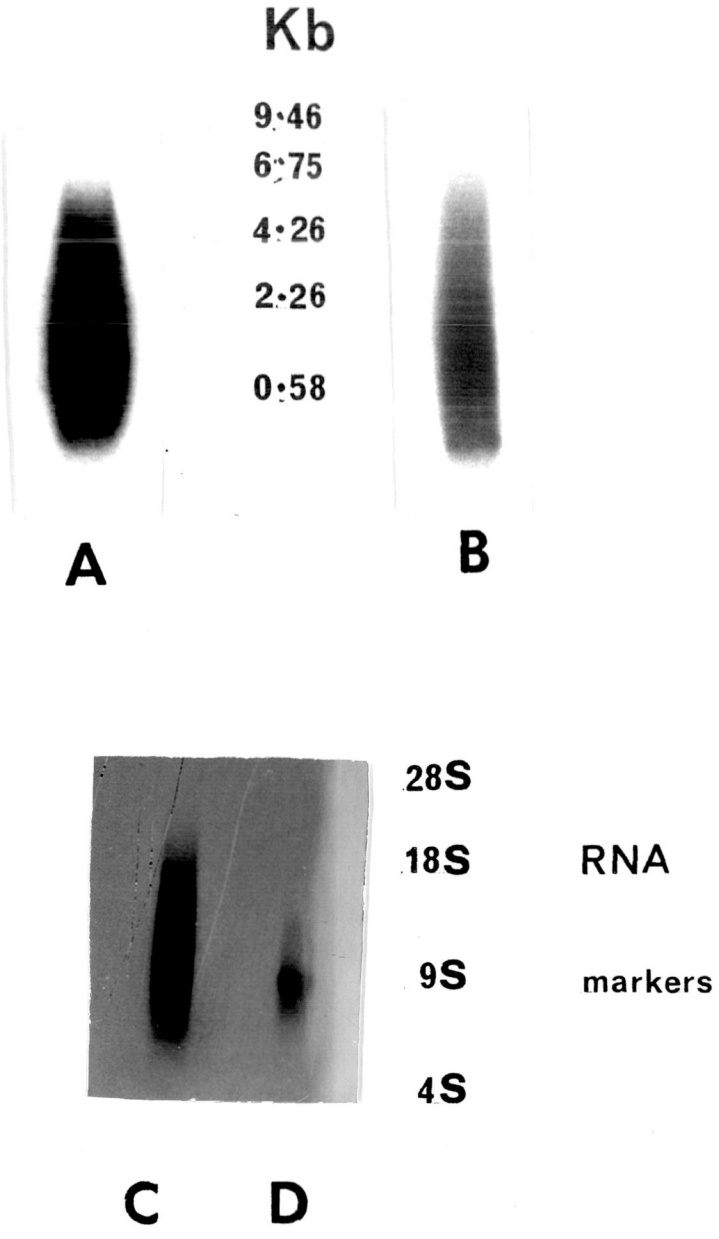


Figure 3.3.1. 1% agarose gels of ^{32}P single stranded cDNA.

Track A Fibroblast ss cDNA

Track B CF fibroblast ss cDNA

Track C Lymphocyte cDNA

Track D Mouse globin cDNA

inhibited the double stranding but 50mM NaCl was satisfactory. The reaction was allowed to proceed for 2 hours at 18°C whereupon a small aliquot was removed and the remainder snap-frozen in dry-ice/ethanol. The extent of the double stranding was then assessed by digesting the aliquot with the single strand specific nuclease S_1 nuclease from Aspergillus oryzae followed by acid precipitation. Double stranding was deemed to be satisfactory when greater than 60% of the radioactivity was resistant to S_1 nuclease digestion.

During the synthesis of the first strand cDNA many workers use Actinomycin D to achieve larger cDNA which is more comparable in length to its mRNA template (Verma,1973). This antibiotic also inhibits the formation of double stranded regions in the cDNA as a result of self priming of the cDNA on the hairpin region which appears at the 3' end of the cDNA during reverse transcription. However, this antibiotic cannot be used during the synthesis of cDNA to be used for cloning since its presence inhibits the double stranding reaction. In my experience cDNA synthesized in the absence of Actinomycin D was not appreciably smaller but the extent of double stranding in some cDNA preparations as monitored by S_1 nuclease resistance was high (25-30% in some instances). When these preparations were double stranded the extent of the double stranding was accordingly higher (75-80%). In the case of the lymphocyte

library, when double stranding was complete the sample was thawed, the reaction stopped by addition of EDTA to 10mM and desalted on Sephadex G-50 in S_1 buffer as described in section 2.4.4. A high concentration of S_1 nuclease was then added to remove the hairpin loops formed during the first strand synthesis. These reaction conditions were shown in test experiments to result in complete degradation of labelled single stranded cDNA without significant breakdown of double stranded DNA. In the case of the fibroblast libraries, the double stranded material was thawed and the reactions stopped with EDTA. The volume of the sample was then increased by dilution with water, 1/10 volume of 10xS_1 buffer was added and a high concentration of S_1 nuclease added as for the lymphocyte cDNA. Again this reaction was monitored by digestion of single stranded cDNA under the same conditions. After incubation for 1 hour at 37°C all samples were thoroughly extracted with phenol/chloroform. Size analysis on 1% agarose gels shows that there is a decrease in the mean size of the double stranded (ds) cDNA as compared to the single stranded cDNA, indicating that the double stranding reaction does not give full length ds cDNA (Fig.3.4.1). In order to remove low M.Wt. material the S_1 treated ds cDNA was fractionated on 15-30% sucrose gradients. The fractions were then sized either by agarose gel electrophoresis or by comparison to the profile of radioactive markers on parallel gradients. Fig. 3.4.2. shows the profiles obtained for the

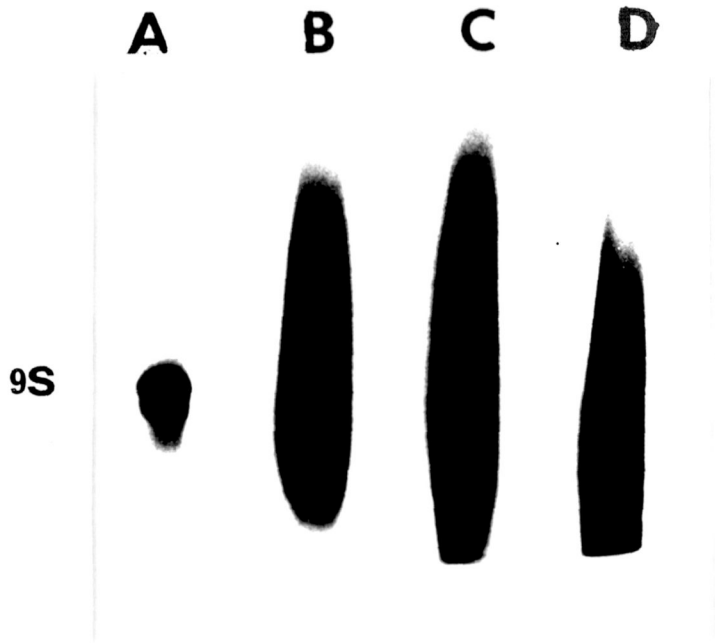


Figure 3.4.1 Analysis of double stranded lymphocyte cDNA on 1% agarose gels.

Track a: ^{32}P labelled mouse globin cDNA.

Track b: ^{32}P labelled single stranded lymphocyte cDNA.

Track c: ^{32}P labelled double stranded lymphocyte cDNA before S_1.

Track d: ^{32}P labelled double stranded lymphocyte cDNA after S_1 digestion.

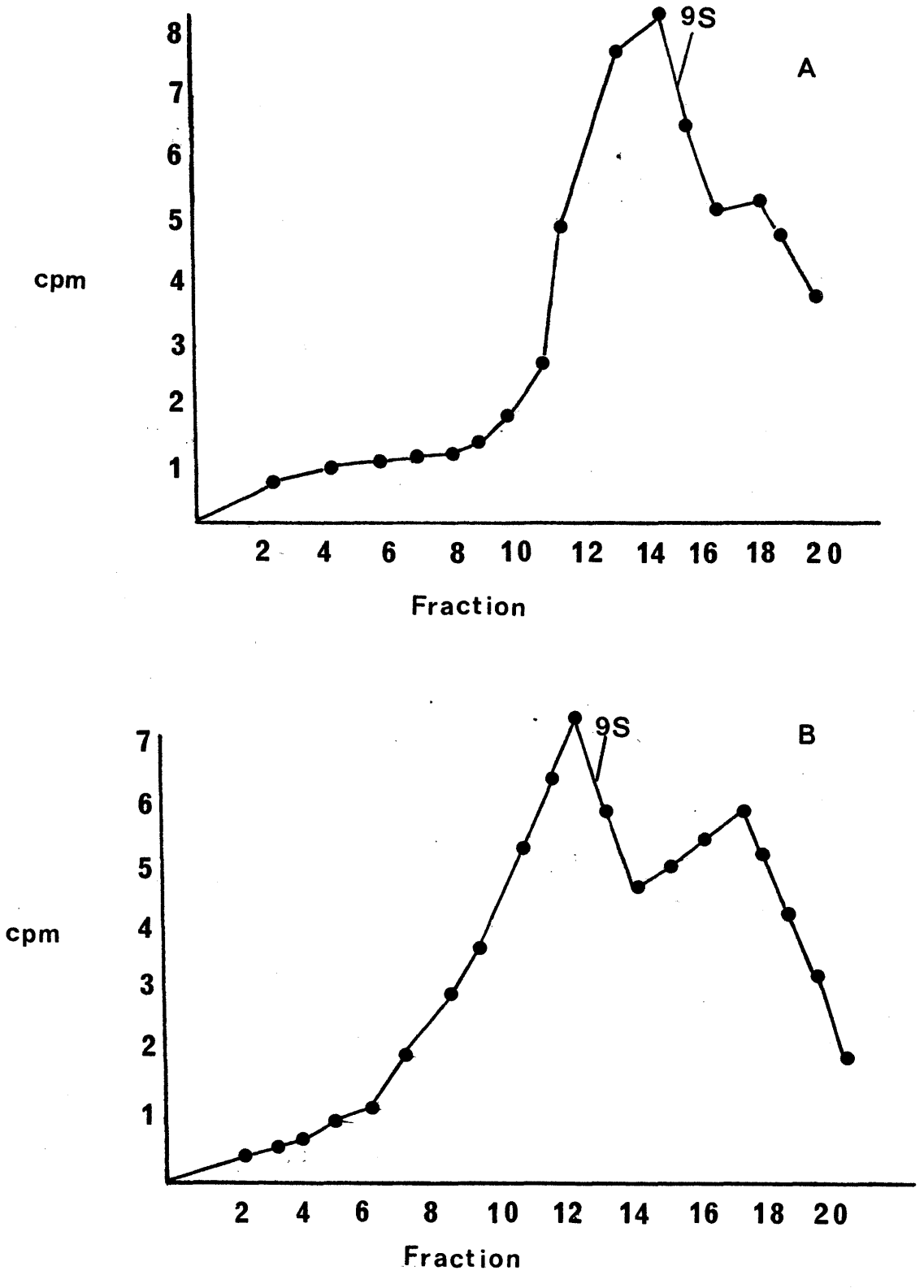


Figure 3.4.2 Sucrose gradient profiles of normal and CF fibroblast ds cDNA.

A) Normal fibroblast ds cDNA fractionated on a 15-30% sucrose gradient at 15°C for 13 hrs. at 40,000 r.p.m. in an SW40 rotor and B) CF ds cDNA similarly for 17 hrs.

fibroblast cDNAs and Fig. 3.4.3 shows sizing of the lymphocyte cDNA by agarose gel electrophoresis. In all cases cDNA greater than or equal to 9S mouse globin cDNA markers were retained. The presence of a peak near the top of the gradient profile indicates the nucleotides digested by the S_1 nuclease. The indicated fractions were pooled, ethanol precipitated and harvested by ultracentrifugation, washed thoroughly with 70% ethanol and recentrifuged at 40,000 r.p.m. for 30 minutes. Samples were then vacuum dried, resuspended in sterile distilled water and again sized on agarose gels. Figs. 3.4.4 and 3.4.5 show the final ds cDNA from each cell type which was used to clone the libraries. Fig. 3.4.4 shows some smearing to the top of the gel but clearly indicates the presence of discrete bands. Fig 3.4.5 shows that the CF fibroblast cDNA is on average larger than the normal fibroblast but the discrete bands observed in the RNA are not as evident.

3.5 Cloning in the host/vector system E.coli HB101/pAT153

The guidelines controlling genetic manipulation in this country (Williams,1976) require certain levels of physical and biological containment dependent on the nature of the experiment to be carried out. In 1979, after revision of these guidelines a new vector was developed. This vector, pAT153 (Twigg & Sherratt, 1979), a derivative of pBR322, used in conjunction with the rec A^-

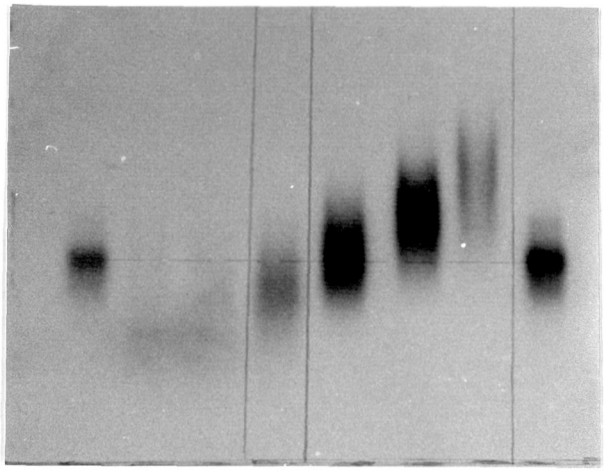


Figure 3.4.3. Autoradiograph of sucrose gradient fractions of lymphocyte double stranded cDNA analysed on 1% agarose gel.

DNA was denatured prior to electrophoresis. The radioactive markers on the gel are 9S mouse globin cDNA.

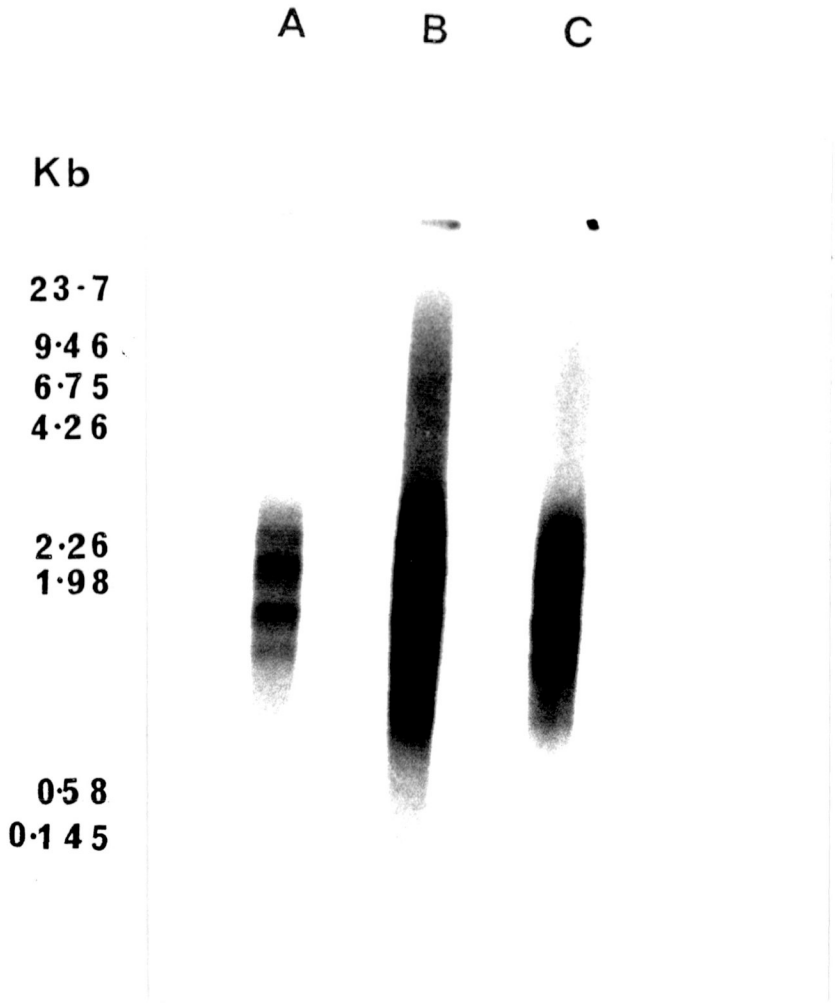


Figure 3.4.4. Analysis of double stranded lymphocyte cDNA on 1% agarose gel.

Track a: 1,000 c.p.m. double stranded lymphocyte cDNA.

Track b: 10,000 c.p.m. " " " ".

Track c: 4,500 c.p.m. " " " ".

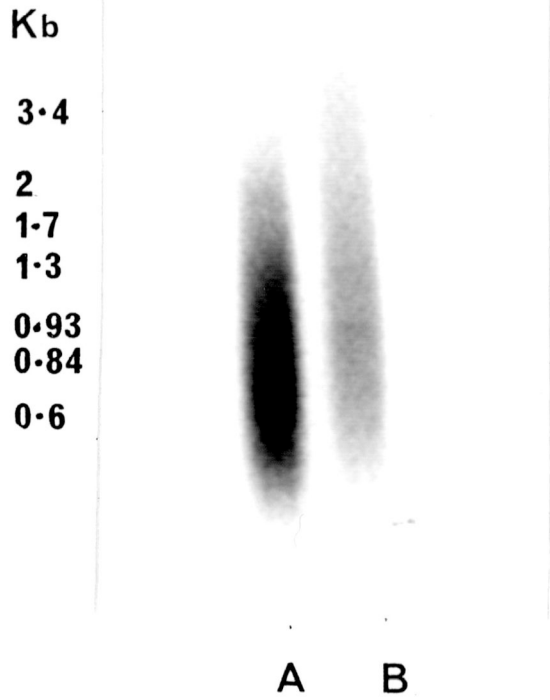


Figure 3.4.5 Analysis of fibroblast and CF fibroblast double stranded cDNA on 1% agarose gel.

Track a: 2,000 c.p.m. normal fibroblast cDNA.

Track b: 2,000 c.p.m. CF fibroblast cDNA.

host E.coli HB101 (Boyer & Roulland-Dussoix, 1969) satisfies the biological containment requirements.

The vector pAT153 contains genes for resistance to several antibiotics eg. tetracycline and ampicillin. It exists within its host as a covalently closed supercoiled circular DNA of about 3700 bases in length. Located within the genes for antibiotic resistance are single sites for several restriction endonucleases. Insertion of donor DNA into one of these sites inactivates the gene so making the bacterium which recieves such a hybrid plasmid sensitive to that antibiotic. After linearization of the original vector by the particular restriction endonuclease there are basically three ways of joining the ds cDNA to the plasmid (Old & Primrose,1980):-

a) The direct ligation of "blunt ended" double stranded cDNA into a blunt ended linearized plasmid (Cochet et al.,1979).

b) The blunt end ligation of synthetic restriction enzyme recognition sites or linkers to the cDNA followed by ligation to the termini of the vector cleaned with the same enzyme (Ullrich et al.,1977).

c) The addition of homopolymer tracts to ds cDNA using terminal deoxynucleotidyl transferase and annealing to tracts of the complementary homopolymer, similarly added to the vector. (Jackson et al.,1972).

During the course of this study all three methods were

assessed and the results are now presented.

3.6 Cloning cDNA libraries by blunt end ligation

The use of blunt end ligation to clone a cDNA library is potentially the most simple method available. Basically it entails the generation of blunt ends at the termini of the vector and the cDNA followed by ligation of the fragments together using T4 DNA ligase. One of the major problems in the construction of cDNA libraries is the presence after transformation of background parental plasmids containing no inserted DNA. This problem can be overcome by inserting the donor DNA into the BamHI site of pAT153. This restriction enzyme recognition site is located in the gene conferring resistance to tetracycline. Tetracycline (Tet) is a bacteriostatic antibiotic and so if bacteria containing recombinant plasmids at the BamHI site are placed on Tet. plates they cannot grow, but neither are they destroyed. However, any bacteria harbouring parental pAT153 plasmids can still grow. If a second antibiotic, cycloserine, is present in the same media it is only taken up by those bacteria which are actively growing ie. non recombinants. Cycloserine is toxic to these bacteria so they are all killed while those bacteria which do contain recombinants survive. When the bacteria are subsequently selected on media containing ampicillin the population should theoretically be 100% recombinant. The BamHI site was therefore the site used for this study.

3.6.1 Preparation of the plasmid

40 μg of pAT153, the majority of which was supercoiled DNA as judged by agarose gel electrophoresis, was linearized by digestion with BamHI as described in 2.2.1. Fig.3.6.1.1 shows the digested DNA to be entirely in the linear form. Supercoiled DNA transforms at 2-3 orders of magnitude more efficiently than linear DNA so the presence of any contaminating supercoiled DNA not visible by gel electrophoresis increases the transformation background. The linearized DNA was therefore used to transform HB101 as in 2.3.3. Transformation efficiencies of supercoiled DNA were routinely 10^6- $5x10^6$ transformants/μg of supercoiled DNA and values of 10^3/μg or less for linear DNA were acceptable. Samples which gave unacceptably high backgrounds were redigested with BamHI and again tested for transformation efficiency. Linearized plasmids were phenol extracted, ethanol precipitated and dissolved in sterile water.

The generation of blunt ends was carried out essentially as described by Seeburg et al.(1977) except that the incubation was allowed to proceed for 1 hour at 37°C at a plasmid concentration of 250 g/ml. The recognition site for BamHI is $\frac{GGATCC}{CCTAGG}$ generating "sticky ends" of $\frac{G}{CTAG}$ after digestion. This reaction was assayed by addition of a known amount of ^{32}P or ^{3}H-dCTP to the incubation. Incorporation of the label into the plasmid indicated that the enzyme had 'filled in' up to the terminal G and so generated

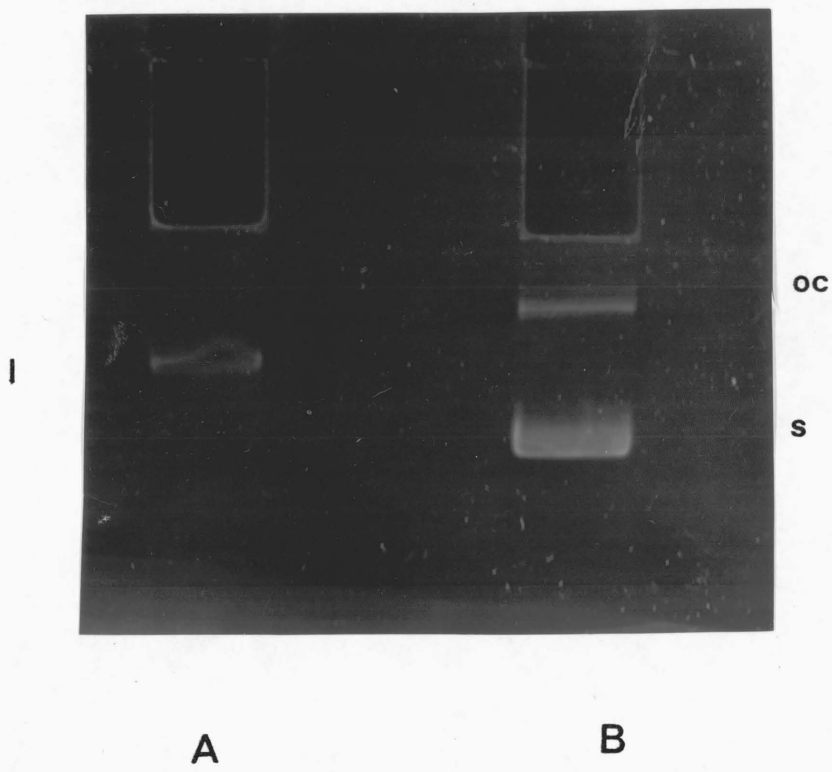


Figure 3.6.1.1. 1% agarose gel of pAT153 DNA.

Track a: 1μg of pAT153 DNA digested by BamH1.

Track b: 2μg of undigested pAT 153 DNA.

s, oc and l indicate supercoiled, open circular and linear DNA respectively.

true blunt ends.

In general, ligation reactions with plasmid vectors must be carried out at high vector:donor ratios in order to promote the joining of a high proportion of vector molecules onto the donor DNA. Even under these conditions a high percentage of the vector molecules recyclise and so increase the background of non-recombinants. An approach has been developed which circumvents these problems caused by self-cyclisation of the vector DNA. The linear plasmid molecule is treated with bacterial alkaline phosphate (BAP) or calf intestinal phosphatase (CIP) before the ligation reaction. The vector DNA can then no longer cyclise since 5' phosphate groups will have been removed from each end of the molecule. The ds cDNA is not treated with the phosphatase, however, and so DNA ligase can join vector and donor DNA molecules in one of the two strands through the 5' phosphoryl groups still on the donor DNA and the 3' OH groups on the vector. Such a molecule can be transformed and introduced into E.coli where it will be repaired and replicate.

Blunt ended linear plasmid was dephosphorylated by treatment with BAP (0.1 units/μg DNA) at 65^{o}C for 1 hour in 10 mM TRIS, 1 mM EDTA pH 8.0, at a DNA concentration of 15μg/ml, followed by exhaustive phenol and ether extractions. After ethanol precipitation, this plasmid was washed with 70% ethanol and resuspended in 10 mM TRIS pH 7.5 at 100μg/ml.

3.6.2. Ligation of linearized pAT153 and ds cDNA for cloning

As described in 3.4 the double stranded cDNA had been treated with S_1 nuclease prior to size fractionation on sucrose gradient. Although S_1 nuclease is a single strand specific nuclease it was possible that it may have digested into the ends of the ds cDNA and so perfect blunt ends would not be present. In order to overcome this possibility protocols for blunt end ligation (Seeburg et al.,1977; Affara et al.,1981) incorporate a reaction to increase the number of blunt ends. This involves a short incubation of ds cDNA with either E.coli DNA polymerase I or T4 DNA polymerase in the presence of deoxynucleoside triphosphates (dNTPs). The reaction conditions used were as described by Sippel et al.,(1978) using E.coli DNA polymerase I, 250 μM dNTPs and 25 μg/ml ds cDNA. The buffer used in this reaction was also suitable for subsequent ligation of plasmid and cDNA so the blunt ended dephosphorylated plasmid could be added directly.

20 ng of cDNA were made blunt ended and then ligated with a 25 fold excess by weight (500 ng) of plasmid in 100 μl reaction volume as described by Fantoni et al.,(1979). The reaction was allowed to proceed for 16 hours at 12°C (Sugino et al.,1977) in the presence of a high concentration of T4 DNA ligase (Sgaramella & Ehrlich,1978). Ten 10μl aliquots were then removed from the incubation, each was diluted with 90 μl of TEN buffer and the transformation carried out as described in 2.3.3. After transformation

bacteria were filtered onto nitrocellulose filters and selection carried out on ampicillin. The results are summarized in Table 3.6.2.

Linear plasmid(ng.)	cDNA(ng.)	No. of colonies	Efficiency (per μg plasmid)
500	20	2500	5×10^3
50(unligated)	--	50	10^3

Efficiency with supercoiled pAT153 : 5×10^6

TABLE 3.6.2

In view of the transformation efficiency achieved with supercoiled pAT153 the efficiency for the hybrid DNA is low. However the efficiency is such that a library of 125,000 recombinants could be achieved from only 1μg of ds cDNA. In order to remove any parental non-recombinant bacteria, as discussed previously, a cycloserine selection procedure was carried out. Standard methods of cycloserine selection involve growing the cells in liquid culture in the presence of tetracycline and cycloserine, harvesting the cells, washing off the antibiotic and plating out the surviving cells on ampicillin plates. This approach has the disadvantage that any given recombinant molecule will be present in many colonies when the liquid culture is plated out. To overcome this problem replicas were made of the original filters containing the 2500 putative recombinants as described by Hanahan & Meselson (1980).

These replicas were incubated on tetracycline plates for 3 hours in order for non-recombinants to be growing optimally when cycloserine was introduced to the medium. They were then transferred to plates containing tetracycline and cycloserine (100 μg/ml) and incubated for 6 hours. These filters were then laid on Whatman paper soaked with L-broth and sucked dry on a vacuum manifold. Several cycles of this treatment ensured the removal of cycloserine from the filters which were then grown overnight on ampicillin plates. Approximately 80% ie. 2000 of the colonies survived the selection and several of these recombinants were picked and the plasmid isolated as described in 2.3.2. When analysed by electrophoresis on 1% agarose gels (Fig.3.6.2) most of these clones appeared to be either non-recombinants or to have extremely small inserts. Non-recombinant bacteria should, therefore, have been eliminated by cycloserine selection (see discussion). The absence of inserts was confirmed by restriction analysis of the clones using the enzymes Sau3A and BglII (K.Davies, C.Coutelle, personal communication).

The most likely explanation of this was that the phosphatase reaction had been incomplete so the plasmid molecules had recyclised. To confirm this the phosphatased vector was ligated and transformed in the absence of ds cDNA and indeed transformants were produced which behaved as non-recombinants. When assayed by removal of ^{32}P from end-labelled DNA the phosphatase was found

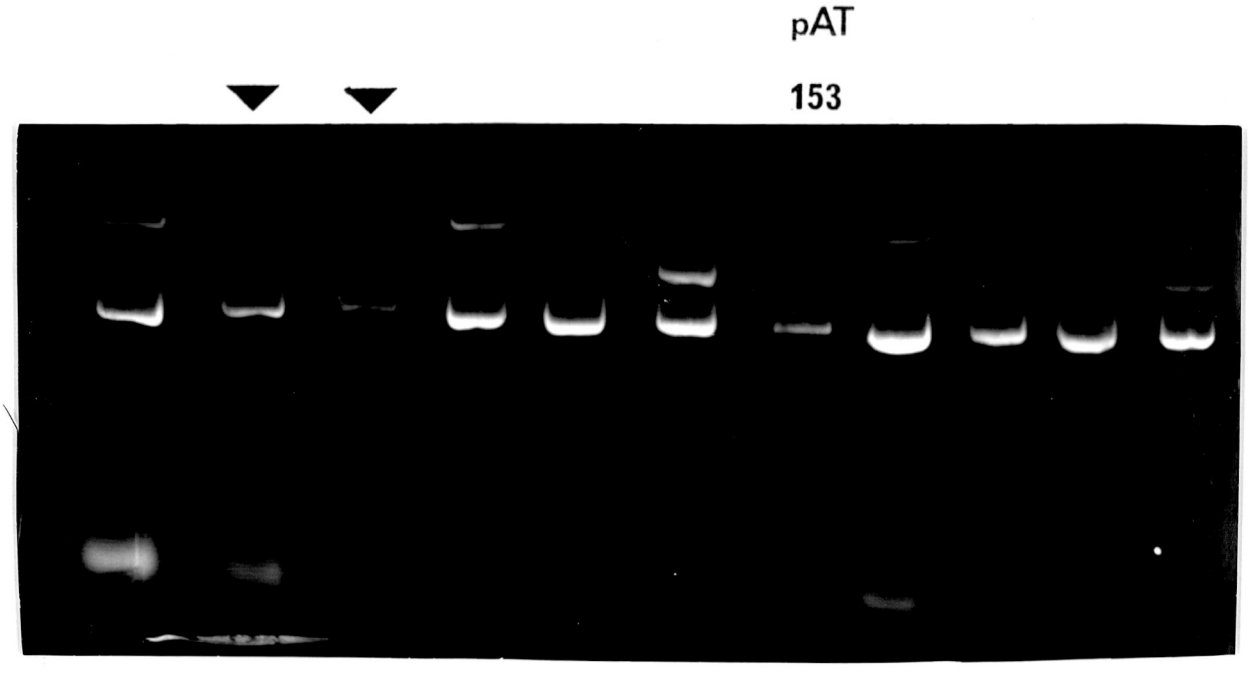


Figure 3.6.2 Size analysis of blunt end ligation clones on 1% agarose gel.

Ten recombinants produced by blunt end ligation were isolated and plasmid prepared from each. 1μg was electrophoresed on a 1% agarose gel with 0.5μg pAT153 as size marker. Thetwo most obvious true recombinants are indicated.

to have low activity. The plasmid vector was therefore re-phosphatased using other batches of enzyme and after ligation used to transform bacteria as usual. When the vector had been satisfactorily dephosphorylated the cloning experiments were repeated exactly as before but the transformation efficiency was markedly reduced. In view of these problems attempts to clone the library by this method were discontinued.

3.7 Cloning cDNA libraries using synthetic restriction enzyme recognition sites.

This method of cloning has been applied in several situations (Ullrich et al.,1977; Heyneker et al.,1976; Seeburg et al.,1977) usually for cloning specific cDNAs from highly enriched populations. A number of chemically synthesized double stranded DNA oligomers containing one or more restriction endonuclease cleavage sites are available. A large molar excess of these linkers is normally ligated to the double stranded cDNA using T4 DNA ligase and the product is then restricted with the particular restriction enzyme. This generates a single stranded terminus of four nucleotides in length at each end of the ds cDNA molecule. These termini can then be ligated with very high efficiency to the termini of a cloning vector cleaved with any enzyme which generates the same four nucleotide sticky ends (Heyneker et al,1976).

One problem of this procedure is that some of the ds cDNA

molecules will have naturally occurring restriction sites for the particular enzyme used and so partial clones may be obtained as a result of cleavage during the digestion necessary to cleave the linkers. This can be overcome if the recognition sites in the cDNA can be methylated before addition of linkers so that they will not be cleaved during the digestion. This is possible when using EcoR1 linkers since the enzyme EcoR1 methylase is available to perform this function. For this reason EcoR1 linkers were used in this study.

3.7.1. Methylation and blunt ending of cDNA

In order to ligate the blunt ended linker molecules to ds cDNA it was necessary to ensure that the ds cDNA was blunt ended. This was carried out by a short incubation of the cDNA with E.coli DNA polymerase I exactly as was described in 3.6.2. for cloning by blunt end ligation. 200 ng of blunt ended ds cDNA were then methylated exactly as described by Maniatis et al., (1978) using 6 μM S-adenosyl methionine as methyl donor and 0.05 units EcoR1 methylase in a 100 μl reaction volume. Two 5 μl aliquots were taken before and after addition of the enzyme and mixed with 0.5 μg lambda DNA. After incubation at 37°C for 1 hour these aliquots were adjusted to reaction conditions for EcoR1, and digested with this enzyme for a further hour at 37°C. Products were analyzed on 1% agarose gels and are shown in Fig.3.7.1.1.

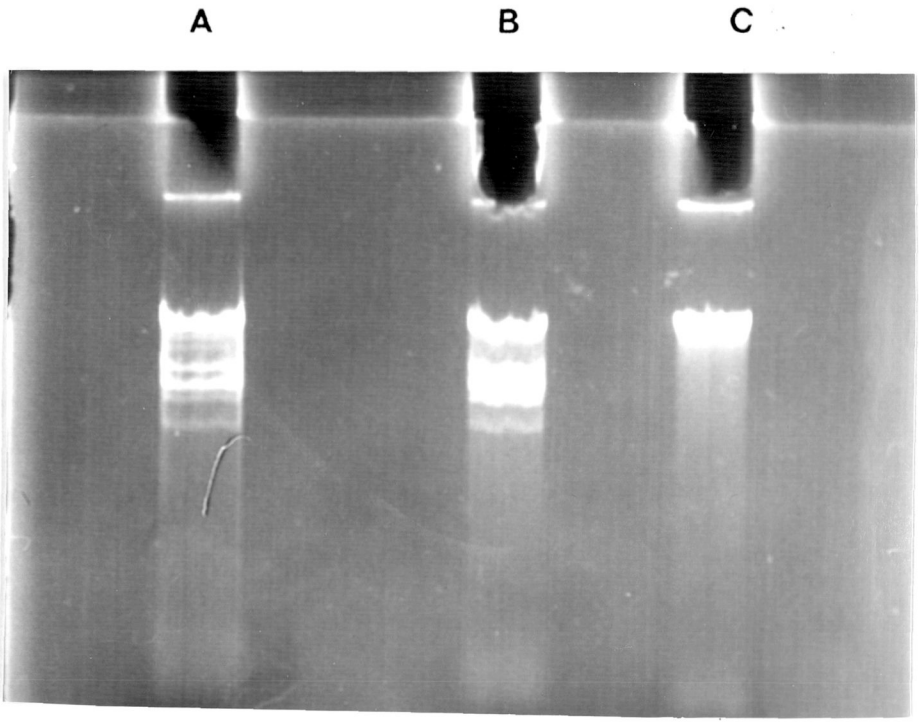


Figure 3.7.1.1 Use of EcoR1 digestion of lambda DNA to monitor methylation of cDNA.

Track a: 0.5μg EcoR1 digested lambda marker DNA.

Track b: 0.5μg lambda DNA + cDNA aliquot before addition of methlase, followed by EcoR1 digestion.

Track c: As for b, after addition of methylase(see text).

Clearly the aliquot removed after addition of methylase has not been digested by EcoR1 so the methylation appears to have been successful.

In view of the small amounts of ds cDNA available pilot experiments were carried out on DNA fragments having known blunt ends. The fragments used were obtained by digestion of lambda DNA with the enzyme Sma1 which restricts the DNA into four flush ended fragments of 20, 12, 8.6 and 8.4 Kb. respectively. However, when lambda DNA was methylated at the EcoR1 site the DNA could not be digested by Sma1. For this reason the DNA was digested with Sma1 primarily amd subsequently methylated, whereupon the EcoR1 sites were protected from EcoR1 digestion (Fig.3.7.1.2). After methylation both blunt ended ds cDNA and lambda fragments were thoroughly phenol extracted and ethanol precipitated.

3.7.2 Phosphorylation of linkers and ligation to DNA

The synthesis of the linkers produces molecules with 5' hydroxyl ends (Scheller et al.,1977). Since the ligase requires 5' phosphate ends the first step in the joining reaction is to phosphorylate the linker. Conditions for phosphorylation and ligation were as described by Maniatis et al.,(1978). 500 ng. of EcoR1 linkers were phosphorylated using T4 kinase for 1 hour at 37°C and 50ng. of these linkers were directly ligated to 200ng. of methylated ds cDNA for 3 hours at room temperature. This reaction

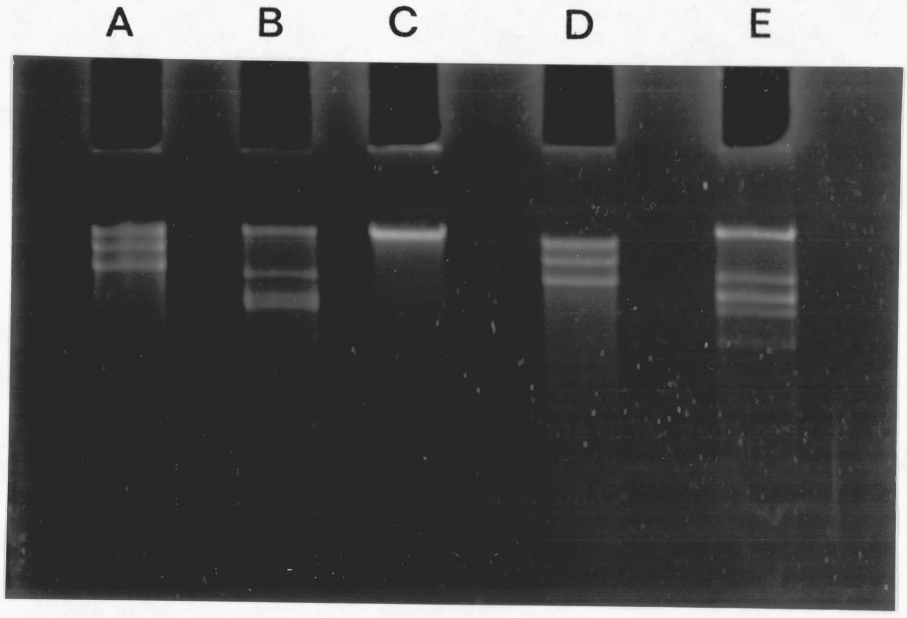


Figure 3.7.1.2 Protection of EcoR1 sites in SmaI digested lambda DNA by methylation.

Track a: 1µg SmaI digested lambda DNA,methylated and redigested with EcoRI.

Track b: 1µg SmaI digested lambda DNA redigested with EcoRI.

Track c: 1µg undigested lambda DNA.

Track d: 1µg SmaI digested DNA.

Track e: 1µg EcoRI digested DNA.

was indirectly monitored by removal from the reaction of small aliquots before and after addition of ligase. These aliquots were incubated with 0.5μg of linear pAT153 and analysed by gel electrophoresis on 1% agarose gels. Fig.3.7.2.1 shows the presence of multiple bands indicative of dimers, trimers etc. of pAT153 as evidence of ligation.

The cDNA-linker ligations were phenol extracted and ethanol precipitated. This precipitate was then taken up in 50μl of EcoR1 buffer and digested with EcoR1 for 1 hour at 37°C in order to generate cohesive ends for cloning. Most of the unligated linker molecules should have been eliminated in the ethanol precipitation but any which were not removed eg. linker oligomers, would compete for digestion by EcoR1. A small aliquot was therefore removed from the incubation, added to 0.1μg of lambda DNA and restricted in parallel. Complete digestion of linkers was evidenced by a limit digestion pattern of the lambda DNA (Fig.3.7.2.2). After digestion of the linker molecules with EcoR1 there would be a large number of "sticky ends" in the incubation which would interfere with the cloning. These were removed by gel filtration on Sephadex G-100 in 5mM TRIS pH 7.5 and the peak of radioactivity containing cDNA-linkers was pooled.

3.7.3 Preparation of the vector, ligation and transformation

10μg of pAT153 was linearised by digestion with EcoR1 as in 2.2.1, phenol extracted and ethanol precipitated. As described in

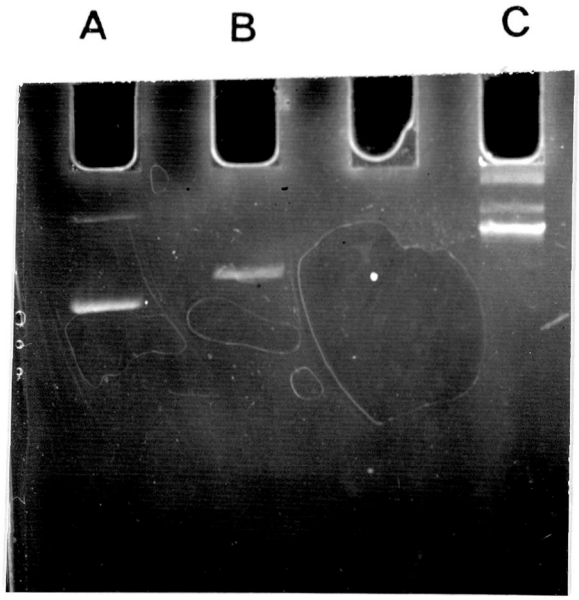


Figure 3.7.2.1 Ligation of EcoRI digested pAT153.

As described in text linker ligation reactions were monitored by parallel ligations of EcoRI digested pAT153.

Track a: 0.3μg undigested pAT153.

Track b: 0.25μg EcoRI pAT153 + aliquot from linker ligation before addition of ligase.

Track c: 0.5μg as b after addition of ligase.



Figure 3.7.2.2 Digestion of lambda DNA with EcoRI in the presence of EcoRI linkers.

Track a: 50ng. undigested lambda DNA.

Track b: 100ng. EcoRI digested lambda DNA + aliquot of linker reaction.

section 3.6.1 unless dephosphorylated this vector will recyclise on ligation and so increase the number of background colonies on transformation. Phosphatase treatment of the linearised plasmid was exactly as described in 3.6.1 and was assayed by ligation of the phosphatased plasmid and electrophoresis on an agarose gel. Fig. 3.7.3.1 shows that the dephosphorylated plasmid remains in the linear form while the untreated plasmid ligates to form open circular molecules as well as higher molecular weight oligomers. After dephosphorylation the plasmid was exhaustively phenol and ether extracted and ethanol precipitated.

To ensure ligation of most of the cDNA molecules to plasmid DNA a molar excess of the plasmid DNA is necessary. From the initial 200 ng. of ds cDNA only 100ng. remained after linker ligation etc. as calculated from its specific activity. This particular cDNA had an average size of approximately 1Kb. and since the plasmid is 3.7Kb. in length, ligation of 1.5 g plasmid to 100ng cDNA should give a 5 fold molar excess. Ligation was carried out at room temperature overnight (Heyneker et al,1976) in a reaction volume of 40µl, the incubate was diluted with dilution buffer and transformed as described(2.3.3). Transformations of supercoiled pAT153, linear pAT153 and dephosphorylated pAT153 after ligation were carried out in parallel and the results are shown in Table 3.7.3.

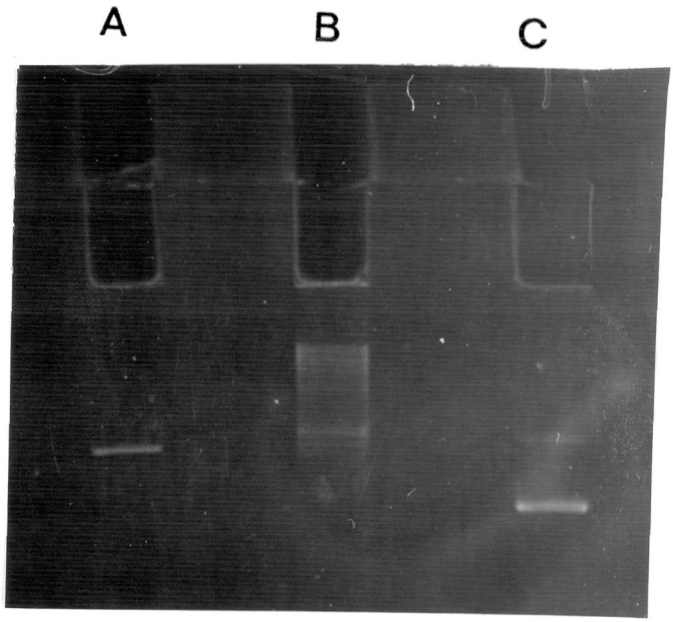


Figure 3.7.3.1 1% agarose gel showing ligation of EcoRI pAT153 before and after alkaline phosphatase treatment.

Track a: 100 ng. dephosphorylated EcoRI pAT153 after ligation.

Track b: As a but not dephosphorylated.

Track c: 100 ng. undigested pAT153.

Linear plasmid(ng)	cDNA(ng)	No. of colonies	Efficiency (/μg plasmid)
1500	100	3000	2×10^3
50	-	48	10^3
50(dephos. & relig.)	-	44	10^3

Efficiency of supercoiled pAT153 = 2×10^6/μg.

TABLE 3.7.3.

Although this method could potentially yield 30,000 recombinants from 1 μg of cDNA the overall efficiency is poor. Clearly a background of parental colonies would be as high as 50% in this case and since there is no selection system for the EcoR1 site in pAT153 this is unacceptable. There are several other disadvantages with the use of linkers to clone heterogeneous populations of cDNA and this method was not pursued further.

3.8 Cloning by homopolymer tailing

This method of cloning relies on the enzyme terminal deoxynucleotidyl transferase(TdT) which catalyses the addition of nucleotides to a 3' terminus of DNA so creating a homopolymer tract or "tail" on the cDNA. If vector and ds cDNA are tailed with complementary deoxynucleotides they may be annealed together to form recombinant circularised molecules for transformation(Jackson et al. 1972; Lobban and Kaiser,1973).

3.8.1 Tailing the vector

The mechanics of the tailing reaction have been well characterised (reviewed by Nelson & Brutlag,1979) and it is apparent that the use of poly dG-poly dC tailing as opposed to poly dA-poly dT results in a 10-100 fold higher efficiency of cloning. The G-C tailing method is particularly useful because by G-C cloning into the Pst1 site of pAT153 the restriction site is regenerated on each side of the insert allowing subsequent purification of insert from vector (Otsuka,1981; Roychoudhury et al.,1976). The Pst1 site is situated within the ampicillinase gene of pAT153 conferring resistance to ampicillin, thus screening for ampicillin sensitive clones is a simple test for recombinants.

50µg of pAT153 was restricted with Pst1 until only linear molecules were present as judged by gel electrophoresis. This plasmid was then extracted with chloroform/octanol and ethanol precipitated. The success of the method for regenerating restriction sites on either sideof the DNA depends on the elimination of 3' exonuclease from the restriction endonuclease. The removal of even a single nucleotide from the 3' end of the vector prior to the terminal transferase reaction will prevent resynthesis of the site. This was assayed by ligation of the linear plasmid by T4 ligase and analysis by gel electrophoresis. Fig.3.8.1.1 shows that a large proportion of the linear molecules do religate to form circular molecules, dimers, trimers etc. indicating that the site is still intact(Track C). When redigested with Pst1(Track D) the plasmid

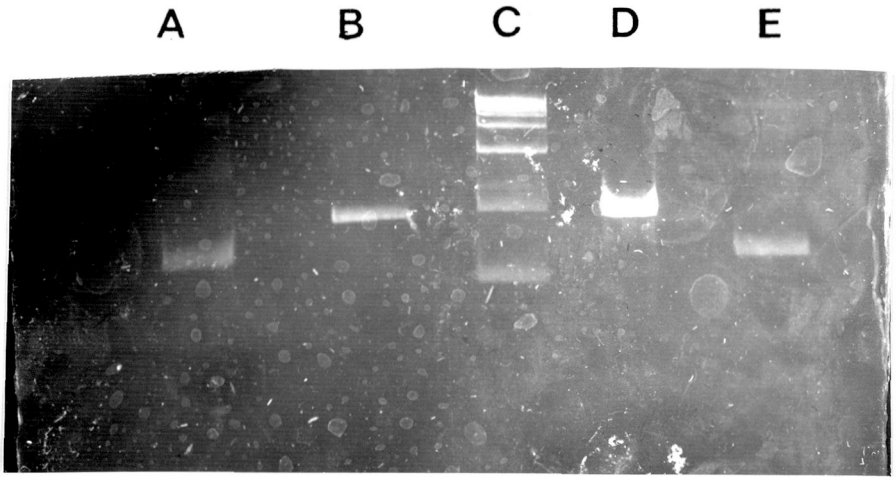


Figure 3.8.1.1 1% agarose gel of ligated and redigested PstI pAT153.

Track a: 100 ng. supercoiled pAT153.

Track b: 250 ng. PstI pAT153.

Track c: 500 ng. PstI pAT153 religated.

Track d: 500ng. as c redigested with PstI.

Track e: 100 ng. supercoiled pAT153.

returns to the linear form, also indicating correct regeneration of the site. Interestingly a band appears in the ligated track to run faster than supercoiled pAT153 markers and disappears on redigestion with Pst1. No obvious explanation of this band, which is apparently exclusive to Pst1 digestion and religation, can be given.

Tailing conditions had been developed in this laboratory (D. Woods,personal comm.) and were as described in 2.4.5. The enzyme activity was found to be extremely variable from day today and standard assays were performed on the same day as preparative reactions. The extent of the reaction was best controlled by varying the time of reaction, concentration of nucleotides and temperature but not by altering the concentration of enzyme since its activity declines rapidly with over-dilution. Analytical tailing reactions of the plasmid with ^{32}P-dGTP were carried out at various nucleotide concentrations, termini concentrations and temps. (see Fig.3.8.1.2). With reference to these standard reactions 50μg of linear pAT153 at an end concentration of 160nM were incubated for 2 mins. at 31^{0}C with 500μM dGTP and 50μCi ^{32}P dGTP. An aliquot of the reaction was removed and the tail length calculated from the amount of acid precipitable radioactivity. The tail length was found to be 16 nucleotides. Unincorporated nucleotides were removed by filtration on Sephadex G-50 after phenol/chloroform extraction and the peak of radioactivity precipitated by the addition of ethanol.

The calculation of tail length is based on the premise that

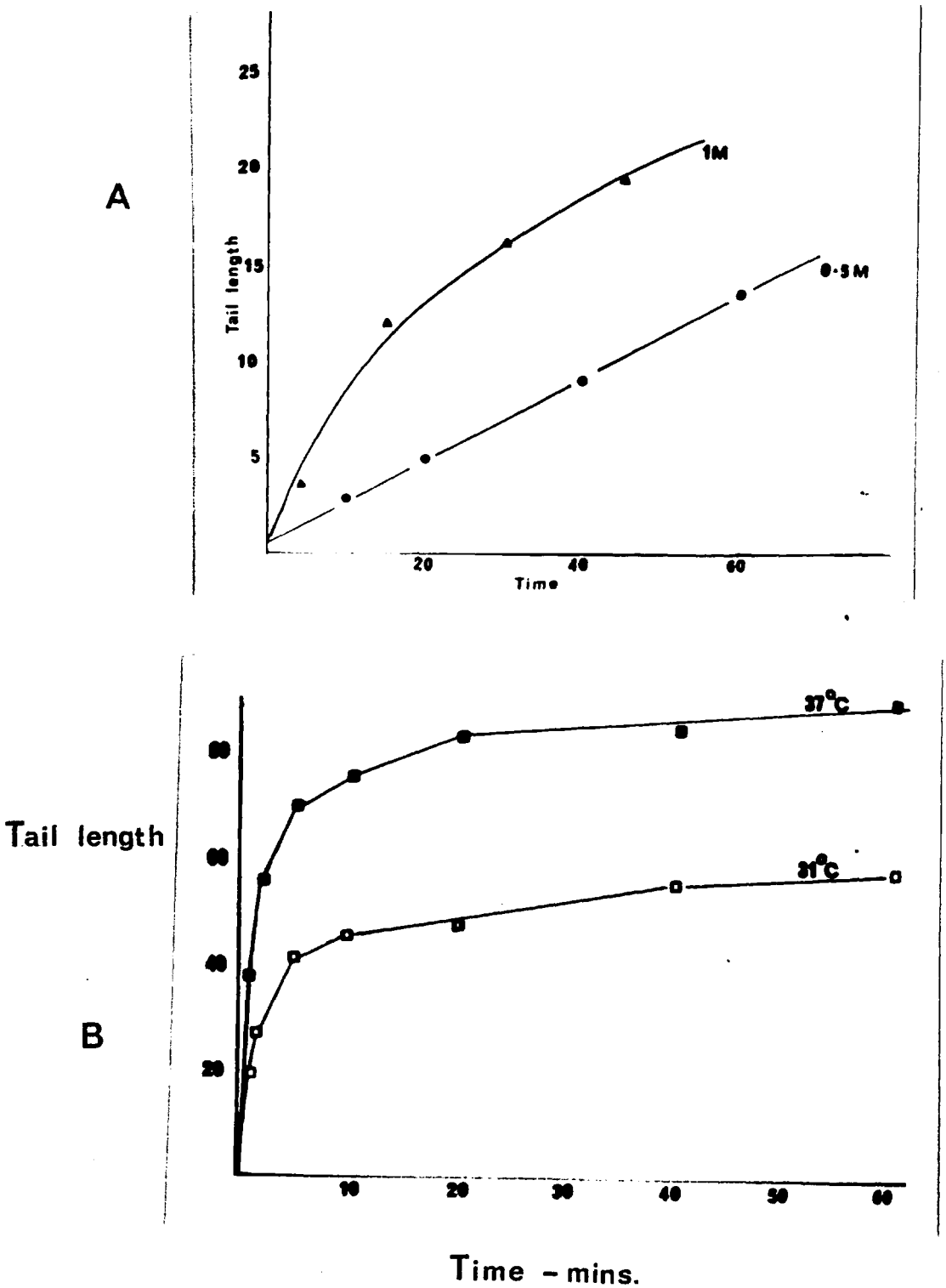


Figure 3.8.1.2 Varying parameters of tailing reaction.

A)Nucleotide concentration. Reactions were carried out at 37°C with 40 nM ends concentration.

B)Temperature . Reactions carried out at 500µM dGTP with 80 nM ends concentration.

the incorporation of labelled deoxynucleotides occurs only at the termini. In practice the DNA may contain some single stranded nicks which also act as a primer for the tailing reaction(Jackson et al., 1972). If the nicking is excessive this may result in a serious error in the calculation of tail length. To confirm that the tailing was only at the Pst1 sites two approaches were taken. Firstly, the radioactively labelled plasmid was denatured by alkali, electrophoresed on a 1% agarose gel and autoradiographed. Fig. 3.8.1.3(a) shows that the denatured plasmid runs mainly as a distinct single band with little evidence of smaller molecular weight(i.e. nicked) material. Secondly, the tailed plasmid was restricted with the enzyme Taq1 which generates 8 fragments from 1 to 0.14Kb. in length from Pst1 digested pAT153. If tailing has only occurred at the termini the only bands which will be radioactively labelled are the bands representing molecules of 1038 and 406 nucleotides in length. Fig.3.8.1.3(b) shows the autoradiograph of the tailed plasmid after Taq1 digestion and clearly most of the radioactivity is present in the two fragments representing the termini of the linear molecule. Fig.3.8.1.3(c) shows a similar autoradiograph of another batch of plasmid under the same conditions which has obviously been tailed at nicks. The good plasmid was therefore suitable for cloning and was used for all the libraries.

3.8.2 Tailing and cloning lymphocyte cDNA

Since the vector was tailed with tracts of dG it was necessary

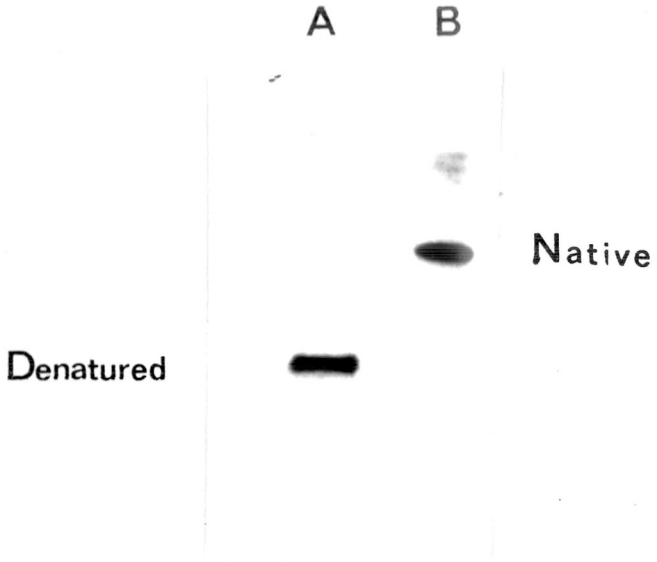


Figure 3.8.1.3a) Autoradiograph of 1% agarose gel of ^{32}P dGTP tailed pAT153.

Track a: 10,000 c.p.m. tailed pAT153 denatured in 0.2M NaOH.

Track b: 10,000 c.p.m. tailed pAT153.

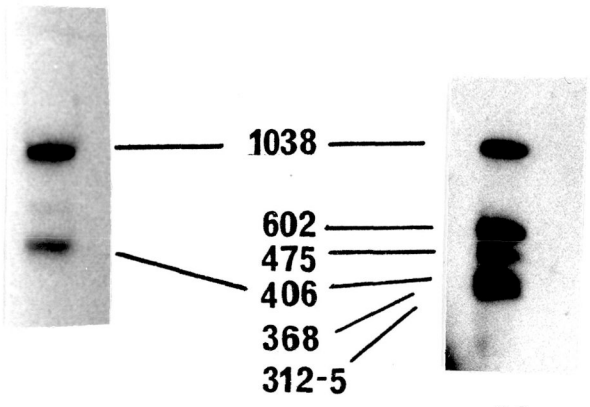


Figure 3.8.1.3 b) and c). TaqI digestion of ^{32}P tailed PstI pAT153.

10,000 c.p.m. electrophoresed on 1% agarose gels in each case.

to tail the ds cDNA with dC tracts. The ds cDNA was labelled with ^{32}P and so it was impossible to perform analytical reactions because the uptake of ^{32}P dCTP could not be followed. Tailing conditions for dCTP were therefore established by using linear plasmid DNA and, with reference to these, ds cDNA was incubated at 5μg/ml. in 100μM dCTP for 5 minutes at 37°C.(Section 2.4.5). Because of the cDNA size heterogeneity it was impossible to accurately determine the concentration of DNA termini. However under these conditions it was estimated that 10-30 dC residues would be added per DNA terminus. After tailing the cDNA was phenol extracted, precipitated with ethanol and dissolved in TEN buffer.

The final step in the cloning procedure involves annealing approximately equimolar amounts of double stranded cDNA and plasmid DNA by slow cooling from an elevated temperature(Wensink et al., 1974) followed by transformation of HB101 with the circular DNA. The plasmid is approximately 3700 nucleotides in length and the lymphocyte cDNA was estimated to be an average 1000 nucleotides. A 5:1 weight to weight ratio of plasmid to cDNA should therefore be approximately equimolar. 320 ng of tailed cDNA were annealed with 2.4μg of tailed plasmid for 2 hours at 60°C and left to cool to room temperature overnight (see 2.4.6) at a plasmid concentration of 2.5μg/ml. This mixture was then split into 50 μl aliquots and diluted with TEN buffer. Following transformation

the cells were filtered directly onto nitrocellulose filters and selected on L-agar plates containing 10 μg/ml tetracycline. The results of this transformation are shown in Table 3.8.2.

Plasmid(ng)	cDNA(ng)	No. of colonies	Efficiency (/μg plasmid)
2400	320	25,000	10^4
300	-	46	2×10^2

Efficiency with supercoiled plasmid = 5×10^5/μg.

TABLE 3.8.2.

The result of this analytical cloning was encouraging in that the apparent background of non-recombinants was only 2% and the recombinant transformation efficiency was high. However, supercoiled transformation efficiency was low so it was possible to generate an even better yield of recombinants during a good transformation. This experiment was scaled up to use 1 μg of cDNA and 5 μg of vector and a library of 200,000 recombinants was obtained,ie. a recombinant transformation efficiency of 5×10^4 transformants/ μg of plasmid, with a non-recombinant background of 4%.

3.8.3. Tailing and cloning normal and CF fibroblast ds cDNA.

The fibroblast libraries were constructed in basically the same way as the lymphocyte except that the terminal transferase used was much more active. The reaction,in fact, proceeded so

quickly that the temperature of reaction was reduced to 10°C and the concentration of ends reduced, 1 μg of normal fibroblast and 500 ng of CF fibroblast ds cDNA were tailed for 5 minutes at 10°C at a dCTP concentration of 5 μM and a ds cDNA concentration of 5 μg/ml. This was estimated to give tail lengths of approximately 20 dC residues per DNA terminus. In each case the reaction was stopped by the addition of EDTA, the incubate extracted with phenol/chloroform and desalted on a small Sephadex G-50 column in one tenth TEN buffer. The peak of radioactivity was pooled, lyophilised and brought to TEN buffer by addition of one tenth the original volume of sterile distilled water. This cDNA was then annealed to the same batch of tailed plasmid which was used for the lymphocyte library. The host used for these transformations was E.coli MC1061 (a gift from Dr. R. Dalgleish) since transformation efficiencies were reportedly higher. Initial transformations were carried out after annealing cDNA to plasmid at a ratio of 2.5:1 and 5:1 and the results are shown in Table 3.8.3. Again, both of these tailed ds cDNA preparations appeared to clone well with background levels of 1.5 and 2.8% for the normal and CF libraries respectively. Increasing the ratio of plasmid to cDNA did not increase the yield of recombinants significantly but resulted in higher backgrounds of non-recombinants. These cloning experiments were again scaled up using a plasmid:cDNA ratio of 2.5:1 and ultimately libraries of 70,000 and 20,000 recombinants were obtained for normal and CF fibroblast cDNAs respectively. The overall

transformation efficiencies remained almost exactly as shown in Table3.8.3.

Plasmid(ng)	cDNA(ng)	No.of colonies	Efficiency (per μg plasmid)
25	10 FIB	1134	4.5×10^4
50	10 FIB	1526	3×10^4
50	-	40	8×10^2
25	-	13	5.2×10^2
25	10 CF	512	2×10^4
50	10 CF	604	1.2×10^4
50	-	35	7×10^2
25	-	14	5.6×10^2

Supercoiled transformation efficiency = $3.5 \times 10^6/\mu$g.

TABLE 3.8.3.

3.9 Storage and analysis of the libraries.

Little is known about the stability of complex populations of recombinant clones. For this reason it was decided to store the cDNA libraries in three ways; as clones on nitrocellulose filter replicas of the original transformations; as purified supercoiled plasmid DNA and asglycerol stocks. The storage of the clones on nitrocellulose filters was performed as described

by Hanahan and Meselson (1980). Glycerol stocks and plasmid DNA were obtained as follows: cloned recombinants on the nitrocellulose filters were shaken into 1 litre of L broth and small aliquots (10 mls) were removed and stored as 50% glycerol stocks at -20°C. From the remaining 900 mls the recombinant plasmid was isolated by the standard cleared lysate method after amplification of the bacteria. Plasmid was then separated from chromosomal DNA and RNA by caesium chloride/ethidium bromide centrifugation (section 2.3.2.). Typical separation is shown in Fig. 3.9.1.

During the pilot experiments in order to demonstrate directly that recombinants did indeed contain inserted sequences, lysates were prepared from single colonies of bacteria and analysed on agarose gels as described in 2.3.5. Colonies were picked separately into a lysis mix containing SDS and proteinase K and incubated for 2 hours. Samples were usually split in two and to one half was added 50 ng of marker supercoiled pAT153 so that the presence of a doublet indicated an inserted sequence (Fig. 3.9.2). There is no plasmid DNA in track A and this is probably due to poor amplification. In other cases it is difficult to discern a doublet indicating either a small insert or a non-recombinant.

The size range of the sequences in the complete libraries was analysed by the Southern blot technique (Southern, 1975) as described in 2.4.13. 5 μg of lymphocyte library plasmid undigested and 10 μg of Pst I digested plasmid were electrophoresed on a 1%

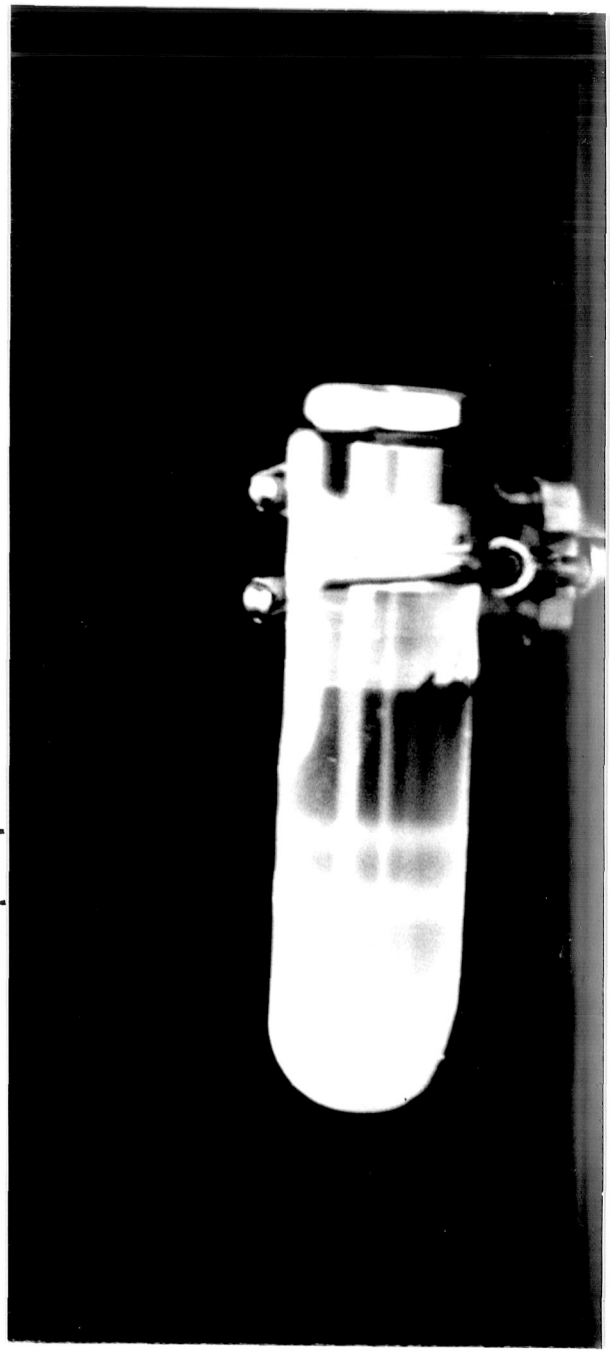


Figure 3.9.1 Purification of plasmid DNA on caesium chloride gradient.

Plasmid prepared as in 2.3.2 was centrifuged at 20°C for 68 hrs. at 40,000 r.p.m. in a TV-865B rotor.

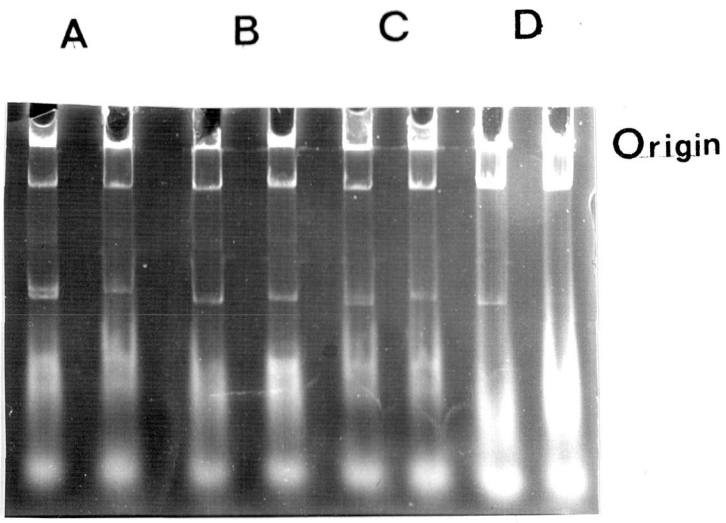


Figure 3.9.2 1% agarose gel electrophoresis of DNA from single colony bacterial lysates.

Samples are run in pairs with the left hand sample containing 50 ng. of supercoiled pAT153 DNA. Recombinants are identified by the presence of doublets in the left hand tracks.

agarose gel, transferred to nitrocellulose paper, hybridised with 30 x 10^6 cpm of ^{32}P Kinase labelled lymphocyte poly $(A)^+$ RNA (Van de Sande et al.,1973) and autoradiographed. The result (Fig.3.9.3) shows, in the Pst I digested DNA, no detectable nicked circular or linear plasmid carrying sequences which hybridise to poly $(A)^+$ RNA showing that the Pst I sites have been regenerated. The average size of the inserted sequences is approximately 600 base pairs but inserts of up to 1300 base pairs are evident.

Similarly 10μg of Pst I digested CF fibroblast library plasmid was analysed by hybridisation with ^{32}P CF fibroblast RNA. (Fig. 3.9.4). In this case there was some hybridisation to the linearized vector probably indicating some recombinants which had only reconstructed the correct Pst I site on one side of the insert or that the digestion was incomplete. Again there was a wide distribution of insert sizes but more interesting, was the presence of intense bands representing molecules of approximately 1.2, 0.95, 0.8, 0.65 and 0.45 Kb in length. Whether these are artefacts of the cloning procedure or genuine products will be discussed subsequently. The normal fibroblast library was not analysed by this method.

3.10. Stability of the libraries.

As previously discussed the non-recombinant, ie. ampicillin resistant, background of the lymphocyte library was 4%. However, whenthe glycerol stocks of the library were plated onto tetracycline

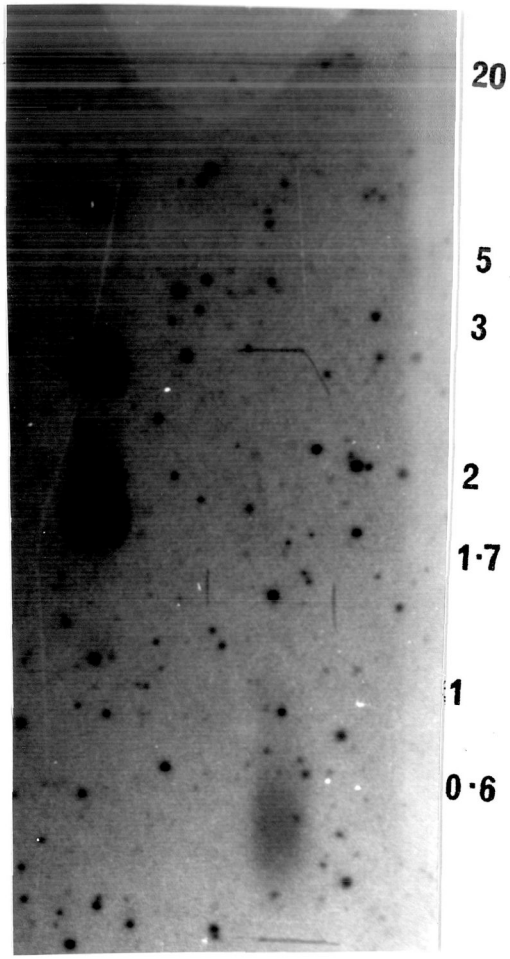


Figure 3.9.3. Southern transfer of lymphocyte cDNA library plasmid.

a) Undigested and b) PstI digested DNA probed with 3 x 10^7 c.p.m. ^{32}P kinase labelled lymphocyte poly A^+ RNA.



Figure 3.9.4 Southern transfer of CF fibroblast library plasmid.

10 µg of CF library plasmid digested with PstI was analysed on a 1% agarose gel and hybridised with 5 x 10^6 c.p.m. CF poly A^+ RNA. Hind III digested lambda DNA was used as markers.

and ampicillin plates a higher background of 20% was observed and most of the clones seemed to be very small. There seemed to be three possible explanations for this:-

a) The ampicillin was breaking down allowing breakthrough of the true recombinant molecules.

b) The plasmids were discarding their inserts and reverting to the parental type.

c) Despite the presence of inserts in the β-lactamase gene, recombinant plasmids were expressing resistance to ampicillin at a slower rate, hence the smaller sized colonies.

In order to test these possibilities 8 of the small colonies were isolated and small preparations made. These plasmids were then analysed on an agarose gel and as shown in Fig.3.10.1. they all clearly did contain inserts. These clones were therefore grown in liquid culture and on agar plates in the presence of ampicillin using mild type HB101 (ampicillin sensitive) as a parallel control. All 8 colonies were resistant to ampicillin. Similarly when the fibroblast libraries were tested for ampicillin sensitivity from their glycerol stocks the apparent background had increased from 2-3% to about 15%.

As previously mentioned the libraries were also grown up amplified and plasmid extracted for storage as DNA. When this DNA was used for transformation the apparent background of ampicillin resistant colonies was as high as 60-70% and when

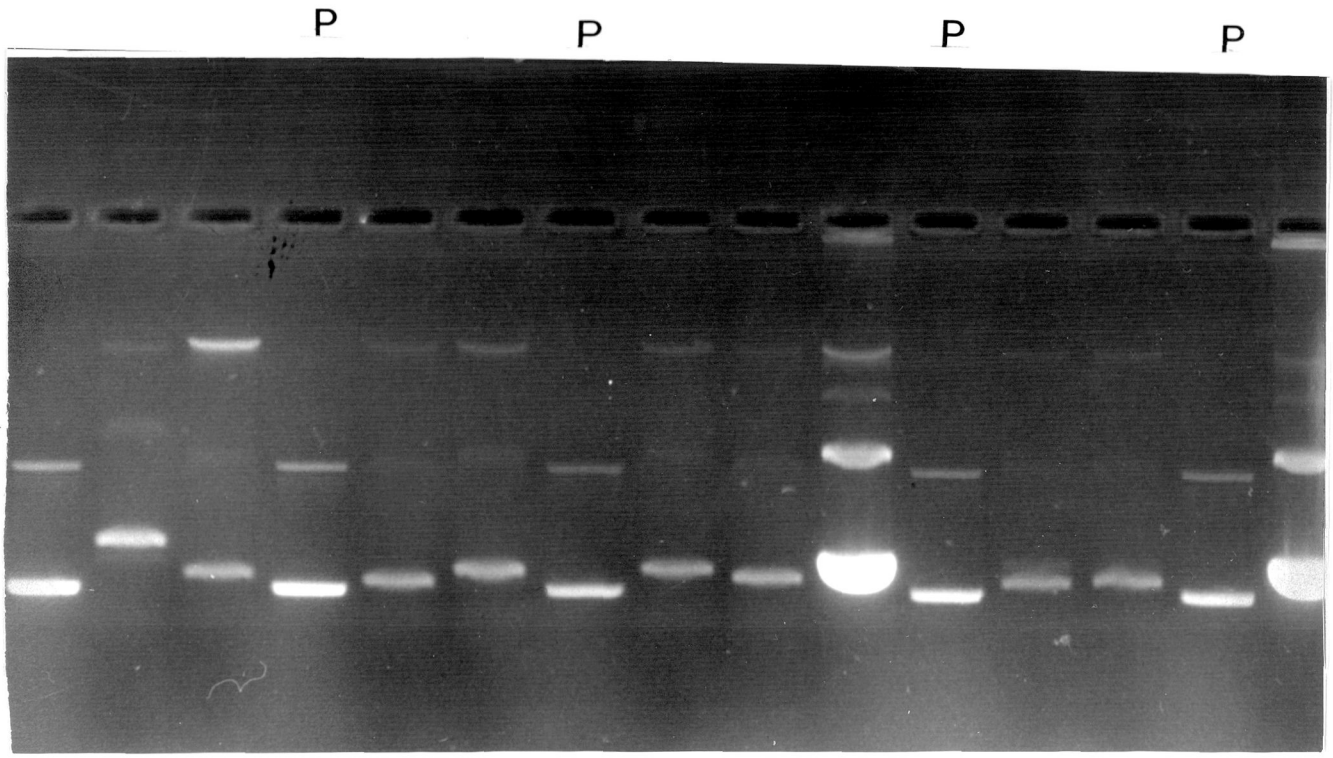


Figure 3.10.1 1% agarose gel of ampicillin resistant recombinants.

Each track contains 500 ng. of recombinant plasmid. pAT153 markers were also run as indicated(P).

individual clones were analysed in this case the background was due to non-recombinants. Since the original background was 20% (including the ampicillin resistant recombinants) an increase of 40-50% background had occurred during the plasmid preparations. One possible explanation for this was that non-recombinant plasmid containing bacteria grew quicker than recombinants. To assess this, several clones with various insert sizes were grown individually with non-recombinants in parallel and their growth rate monitored spectrophotometrically. As can be seen from Figure 3.10.2 there was no significant difference in growth rate of any of the clones.

3.11 DISCUSSION

3.11.1 Synthesis of double stranded cDNA.

Since the original conditions for reverse transcription of mRNA (Kacian et al.,1972; Ross et al.,1972; Verma et al.,1972) were reported, numerous reports have been published describing the optimal conditions for the reaction (Monahan et al.,1976; Efstratiadis et al.,1975; Kacian & Myers,1976). Two recent papers (Buell et al.,1978; Retzel et al.,1980) have attempted to rationalize these conditions and in most cases differences were due to ribonuclease contamination of the enzyme or to the method of preparation of the RNA. The conditions used in this study were based on those of Efstratiadis et al (1975) and included high

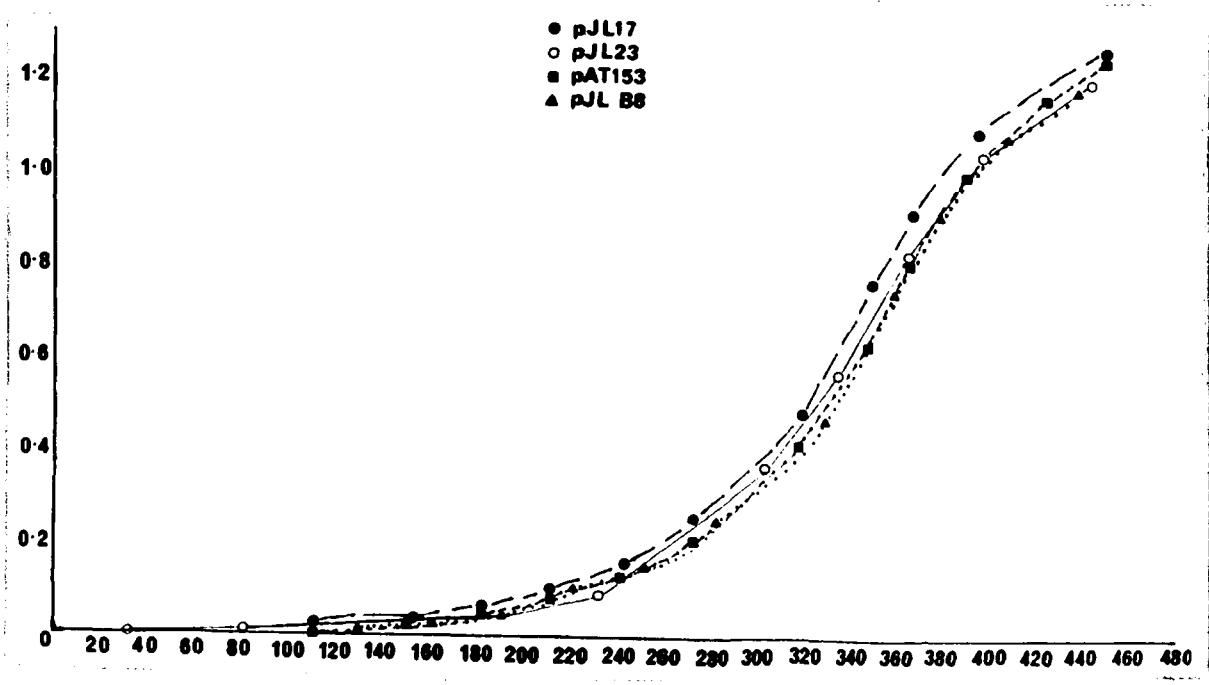


Figure 3.10.2 Growth curves of individual clones.

Clones containing the four plasmids indicated were grown in L broth and their growth rates measured spectrophotometrically. Insert sizes for pJL23, pJL17, pJL B8 and pAT153 were 1.9, 1.05, 0.335 and 0 Kb. respectively.

concentrations of precursor nucleotides and high enzyme concentrations (Friedman & Rosbash,1977). These conditions correlate reasonably well with the "ideal" conditions of Retzel et al.,(1980) apart from the temperature of the incubation (42°C) and the monovalent cation concentration. Each poly A^+ RNA molecule is thought to act as a template for the synthesis of at most one cDNA molecule which remains hydrogen bonded to its template and renders it unable to serve as a template for further synthesis (Friedman & Rosbash,1980). In view of the fact that the yields of cDNA correlated well with the estimated purity of the poly A^+ RNA preparations this suggests that reverse transcription under these conditions was efficient. Further evidence of this is shown by the similarities in size distribution of the poly A^+ RNA and cDNA populations.

The presence of discrete bands in the poly A^+ RNA is interesting but may have been caused by aggregation. However, if this were true discrete bands should not be seen in the cDNA preparations. The presence of bands in cDNA has also been described elsewhere (Efstratiadis et al.,1975) and are thought to indicate partial transcripts. It seems logical that if an abundant poly A^+ RNA contains sequences causing premature termination of reverse transcription then discrete bands would be observed. In the case of fibroblast libraries one particular poly A^+ RNA, that coding for procollagen, is notoriously difficult to reverse transcribe (P. Tolstoshev, personal communication) which may well be caused by such sequences towards the 3' end of the RNA. Of course,

theoretically, any full length cDNA molecules representing highly abundent poly A^+ RNA molecules should also give discrete bands.

The double stranding reaction relies on the existance of a region of self complementarity at the 3' end of a cDNA molecule which results in the formation of a "snapback" or hairpin structure (Leis & Hurwitz,1972). This hairpin can then act as a primer for elongation of the cDNA using DNA polymerase I (Efstratiadis et al., 1976) or reverse transcriptase (Rougeon & Mach,1976). There are potentially important differences between the enzymes. When reverse transcriptase is used there is a virtual absence of intermediates between full length cDNA and full length ds cDNA (Monahan et al.,1976; Kay et al.,1980). In contrast, with pol I, large amounts of partially double stranded cDNA are present (Wickens et al.,1978). This implies that a higher proportion of full length double stranded cDNA molecules would be produced by using reverse transcriptase. However, Rougeon and Mach (1976) have shown that pol I will copy cDNA prepared from mouse Ig light chains while reverse transcriptase will not. In this study it was more important to have a complete representation of RNA in the library than to have full length copies so therefore it was decided to use pol. I in this study. Conditions were as described by Humphries et al.,(1977) and, as expected, there was a decrease in the average size of the cDNA indicating incomplete double

stranding. Because of the accumulation of a significant proportion of less than full length ds cDNA the sucrose gradient size fractionation was performed.

3.11.2 Relative merits of cloning methods

There are arguably three main criteria when assessing different cloning methods:- 1) Simplicity 2) Selection method 3) Reisolation of cloned sequences. The use of blunt end ligation is a relatively untried method of cloning but it has been used successfully for cloning chicken conalbumin (Cochet et al.,1979), mouse embryonic globin (Fantoni et al.,1979) and mouse foetal β-globin (Affara et al.,1981). Blunt end ligation into the BamHI site fits all three criteria well. Insertion of DNA into the BamHI site of pAT153 inactivates the gene for tetracycline resistance so that cycloserine selection is possible if background non-recombinants are present. In this study, however, cycloserine selection did not eliminate the background. The explanation of this is as follows: digestionof pAT153 by BamHI occurs at the sequence $\frac{GGATCC}{CCTAGG}$ to generate ends of $\frac{G}{CCTAG}$ and $\frac{GATCC}{\ \ G}$ which when "blunt ended"become $\frac{GGATC}{CCTAG}$ and $\frac{GATCC}{CCTAG}$. Phosphatase treatment of the blunt ended molecule should prevent recircularization on ligation but if the phosphatase is not efficient then a large number of recircularized molecules will form. Clearly the sequence so generated $\frac{GGATCGATCC}{CCTAGCTAGG}$ is now a different sequence to that of the original pAT153. In effect

there is an insertion of the tetranucleotide sequence TCGA causing a frame shift during expression of the tetracycline resistance gene and so these colonies behave as "pseudo" recombinants. During the study on blunt end ligation exhaustive dephosphorylation was required to inhibit this recircularization and when satisfactorily dephosphorylated the plasmid was not efficient for cloning. It has been suggested (N. Lewis, personal communication) that this is due to the enzyme removing the terminal nucleotide as well as the 5' phosphate so destroying the true blunt ends. This particular system also enables inserts to be removed easily since blunt ending regenerates sites for the enzymes MboI and Sau3A on either side of the insert. (Backman et al.,1976). There is a way of overcoming the pseudo-recombinant problems and developing this procedure into a potentially very efficient cloning method. The enzyme BalI cuts pAT153 at a single restriction site close to the BamHI site ie. within the tetracycline resistance gene, This enzyme however cuts the plasmid to give flush ends and therefore pseudo recombinants cannot form nor is any filling in step necessary, yet cycloserine selection is still possible. Once again, however, a good phosphatase preparation is required so that blunt ends are left intact once the phosphate is removed.

Duplex DNA linker molecules bearing restriction enzyme recognition sites have been chemically synthesised (Bahl et al., 1977; Scheller et al.,1977) and used to insert a number of

different DNA molecules into plasmids (Heyneker et al.,1976; Fiddes & Goodman,1979; McLeod et al., 1981; Bozzonu et al.,1981). The use of linkers has several advantages over other systems in that it will always be possible to resect the inserted DNA and hence purify it. A less obvious advantage is that having obtained long double stranded cDNA all cDNA clones should contain equally long inserts. When using tailing methods, however, if tails are added to internal nicks then shorter cloned inserts may be formed through annealing at these nicks and repair within the bacterium. One final advantage in using linkers is the higher transformation efficiency obtained with ligated DNA as compared with annealed DNA. Efficiencies as high as 40 recombinants per ng of ds cDNA have been reported (Fiddes & Goodman,1979). This figure has itself been superceded by the use of two different linkers eg. EcoR1 and SalI, where figures as high as 500-1400 recombinants per ng of ds cDNA have been reported (Kurtz & Nicodemus,1981; Lewis et al., 1981).

In spite of this the use of linkers has limited applicability to cloning of heterogeneous populations. This is mainly because of the necessity to check the double stranded cDNA for the presence of particular restriction enzyme cleavage sites which is impossible when preparing a complete cDNA library. The use of EcoR1 linkers can overcome this to some extent due to the availability of EcoR1 methylase. This in turn has the disadvantage that use of the EcoR1

site in pAT153 is not selectable, although there is the possibility of blunt end ligating EcoR1 linkers into another site (eg. BalI) of EcoR1 methylated pAT153 and selecting by cycloserine etc. A further disadvantage of using linkers is purely in the number of reactions involved. If yields are not optimal at each stage, although the ultimate transformation efficiency will be high, a larger amount of starting material is required.

In the study described here the ultimate transformation efficiency was not as high as would be expected for linkers. This may be explained by the observation of Kurtz & Nicodemus (1981) that the transformation efficiency of DNA methylated by EcoR1 methylase is 10-fold decreased. On the other hand inefficiency at one of the reaction steps may have contributed to this. Although each step was monitored indirectly it would have been preferable to label the linkers with γ-^{32}P ATP durimg the phosphorylation. If the cDNA had then been made with ^{3}H instead of ^{32}P the actual ligation of the linkers to the cDNA could have been observed. It has been suggested (K. Davies, personal communication) that the most inefficient part of the procedure is the generation of true blunt ends on the cDNA which would obviously reduce the efficiency of ligation of linkers and the ultimate yield of recombinants. It is unlikely that the phosphatase would affect the efficiency, as in blunt end ligation, even if the terminal 5' nucleotide was removed since one strand could still

ligate through the"sticky ends". The observation that lambda DNA which is methylated at EcoR1 sites cannot be digested by SmaI is interesting. The recognition sequences of EcoR1 and SmaI are GAATTC and CCCGGG respectively and show no similarity. It may be argued therefore that the SmaI may bind anywhere along the DNA and then move rapidly along the DNA until the correct site is reached as has been described for EcoR1 (Woodhead & Malcolm,1980). The presence of methyl groups may inhibit the SmaI by an effect of steric hindrance, stopping the enzyme moving to its recognition site. Conversely, SmaI may bind to DNA at sites distant from the cleavage site which may include EcoR1 recognition sites and hence methylation there will inhibit SmaI. In summary then, cloning using linkers, although an excellent method for specific cDNA, is not the method of choice for heterogeneous cDNA populations.

Terminal deoxynucleotidyl transferase (TdT) isolated from calf thymus was first used by Jackson et al., (1972) to add homopolymer tails to the termini of DNA molecules and has subsequently been used commonly in the cloning procedures. It has been observed that, under certain conditions, terminal transferase will catalyse the polymerisation of homopolymer tracts in the absence of a primer. These tracts would be acid precipitable and would result in an artificially high estimation of tail length. To minimise this effect the reactions were carried

out in cacodylate buffer (Kato et al.,1967). Originally terminal transferase only tailed on exposed 3' termini but subsequently conditions were developed which allowed tailing on blunt ended or even 3' recessed termini (Brutlag et al.,1977; Roychoudhury et al., 1976). This involved destabilizing the termini by including Co^{++} in the reaction. In the presence of cobalt, however, any internal nicks in the plasmid or cDNA can also be tailed and this may lead to deletion of DNA sequences after bacterial transformation. Thus it is reasonably common to find recombinant plasmid DNA molecules which are shorter than the parental plasmid (Humphries et al.,1978; Williams & Lloyd,1979). Although cobalt was used in cloning these libraries no individual clones have been identified which are shorter than pAT153. This is probably because care was taken to ensure that nicked plasmid was not used in the cloning procedures. However nicks may have been present in the cDNA which would not have been observed.

The choice of PstI site as the cloning site had several advantages. Firstly, it was possible to use G-C tailing to regenerate PstI sites on each side of the insert. The use of G-C tailing in itself is preferable to A-T tailing since G-C tails are more stable, so shorter tails are required and a higher cloning efficiency results. A second advantage of the PstI site is that a large proportion of inserts at this site appear to be expressed (Villa Komaroff et al.,1978; Keshet et al.,1981;

Auster et al.,1980). The PstI site is situated in the ampicillinase gene of pAT153 and inserts are often expressed as hybrid peptides fused to the N terminal region of the ampicillinase protein. If the inserted sequence is in the correct reading frame then it is possible that correct antigenic determinants may be produced and the clone identified by hybridization with radioactively labelled antibody to a specific protein. This may be useful for picking specific clones from the libraries. It would also be interesting to screen the CF and normal fibroblast libraries with the antibody developed against the CF protein (Manson & Brock, 1980).

3.11.3 Cloning and characterisation of the libraries

The major problems associated with cloning by tailing apart from the "nicks" appears to be the problem of background transformants. This is amplified by the fact that viable recombinant molecules are formed with a relatively low efficiency, so it is necessary to cleave the DNA such that the transformation efficiency with linear plasmid DNA is extremely low. This has been such a problem in some situations that elaborate methods for separating linear molecules from contaminating closed circles have been reported (Norgard et al.,1980). In this study no plasmid preparation was used until exhaustive PstI digestion reduced its transformation efficiency to a low level. As a result of this

the initial background levels in the libraries were never more than 5-10% as calculated from the number of colonies generated by parallel transformation using linear pAT153. However, when the recombinants were plated onto ampicillin 15-20% were resistant. This has been observed in other situations (Liebscher et al., 1980; Villa-Komaroff et al., 1978) where as high as 60% or more of the true recombinants are resistant to ampicillin. The reason for this is that the PstI site is situated in the region encoding aminoacids 181-182 of the ampicillinase while the active site of the protein is much nearer the N-terminus of the protein. This in effect, means that the ampicillinase is still active in the form of a hybrid protein.

The increase in the background after amplification and plasmid preparation cannot be explained by these expression clones. Fig.3.10.2. shows that the increase in the background is not due to the parental pAT153 containing bacteria dividing faster than the recombinants. It is possible that the plasmids were throwing out their inserts and reverting to the parental type since this would necessitate a recombination event occurring. The host bacterium HB101 is a rec A^- bacteria so this seems unlikely. However several reports (Brutlag et al.,1977; Hershey et al.,1977) have suggested that cloning using tailing does result in loss of some inserts particularly if the tails are too long. Indeed C.Coutelle (personal communication) observed a background

increase from 9 to 84% as a particular library constructed with G-C tails of 200 in length, was grown. Several individual clones from the lymphocyte libraries have been studied (D. Woods personal communication) and there is no evidence of any clones excising their inserts. The most likely explanation therefore is that the background increases as a result of amplification during the plasmid preparation. The PstI site itself is not situated in the region which regulates plasmid copy number (Covarrubias et al.,1981) but it is concievable that the increased size of the recombinant plasmid reduces the rate of amplification if not the actual copy number.

The cloning efficiency in the construction of the libraries varied from 40 per ng of ds cDNA for CF fibroblasts, to 200 per ng for lymphocytes. It has been suggested (Liebscher et al.,1980) that differences in transformation procedure can account for these differences. The basis of the transformation procedure is that E.coli cells when treated with calcium chloride solutions become competent to take up DNA (Mandel & Higa,1970; Cohen et al., 1973). Various modifications of the procedure have been developed (Norgard et al.,1978; Dagert & Ehrlich,1979) and also a procedure using $RbCl_2$ in place of $CaCl_2$ (Kushner,1978). During the course of this study the transformation procedure was optimised using modifications of several of these methods. In general it seems that the quality of the DNA being used is more important than

the actual method used and this has been suggested by D.Sherratt (personal communication). The fibroblast libraries were cloned using a different strain of host, E.coli MC1061. This strain had been reported (R.Dalgleish, personal communication) as consistently transforming at 5 x 10^7 transformants/μ g of super-coiled plasmid. In my hands, efficiencies this high were not obtained but efficiences were reproducibly higher than HB101.

Figures 3.7.3 and 3.9.4 illustrate the sizes of the inserts in the lymphocyte and CF fibroblast libraries respectively, and also shows the extent of regeneration of PstI sites on either side of the insert. The average sizes of the inserts seem to be larger in the CF library than the lymphocyte and this correlates with the similar increased size of the original ds cDNA preparations. The presence of bands in the CF library may represent an artefact of cloning such aspremature termination by reverse transcriptase or polymerase I. Conversely it is possible that the bands may represent discrete species of cDNA present in high abundance, particularly since discrete bands have been observed in poly A^+ RNA populations. It is also possible that prolonged exposure of the lymphocyte library autoradiograph may have shown that such bands are not exclusive to the fibroblasts.

3.11.4 Sequence representation in the libraries

The nucleotide complexity of the original lymphocyte cDNA

preparation was determined by hybridisation to lymphocyte poly A^+ RNA. From analyses of the components of the Rot curve so generated (data supplied by J.M.Crampton) it can be calculated that there are approximately 8000 discrete poly A^+ RNAs present in normal human lymphocytes(Table 3.11.3.). This figure agrees with that of Ostrow et al.,(1979) for human lymphoblastic leukaemic cells.

Component	% cDNA hybrid.	$Rot_{\frac{1}{2}}$	$Rot_{\frac{1}{2}}$ corrected	No.of 1000 sequences	Relative copy no.
1	54	0.22	0.119	35	260
2	46	56	26	7800	1

TABLE 3.11.3.(see Woods et al.,1980)

From these figures it is possible to calculate the number of clones required to represent every expressed sequence at least once. From Table 3.11.3 it can be seen that 46% of the cDNA hybridizes to the 1800 least abundant mRNA sequences. The minimum number of individual clones required to generate a complete library is therefore 7800/0.46 = 16956. However, due to the inclusion of some clones containing some low abundance sequences and to the absence of clones containing others, a much larger number of clones must be generated to guarantee obtaining a given sequence. This number can be deduced using the formula which Clarke & Carbon (1976) applied to genomic

cloning, $N = \frac{\ln (1 - P)}{\ln (1 - 1/n)}$

where P is the probability of obtaining a given sequence

n is the minimum no. of individual clones required

For a 99% probability of obtaining any given sequence in this population only 78,052 colonies need be generated.

The fraction of the lymphocyte cDNA complexity represented in the cDNA library can be determined by comparing the hybridisation kinetics and saturation levels of the cDNA/poly A^+ RNA reaction with the hybridisation of the same cDNA preparation with plasmid DNA isolated from the library. This has been carried out (J. Crampton, personal communication) and the kinetics of the two reactions show a similar distribution of abundance classes in the cloned sequences to those represented in the poly $(A)^+$ RNA population and similar saturation levels indicating complete representation in the library.

No hybridisation data is available for the original fibroblast cDNA preparations so similar calculations cannot be performed. However, if one assumes a similar distribution of sequences in normal and CF fibroblasts as lymphocytes, then 70,000 normal fibroblast clones and 20,000 CF clones represents probabilities of 98.5% and 70% respectively of representing any one poly A^+ RNA sequence.

4. COMPARISON OF POLY A^+ RNA SEQUENCES IN CF AND NORMAL CELLS

4.1 Introduction

Having constructed the cDNA libraries it was then possible to compare the sequences represented in each. The basis of the comparison involved the use of colony hybridisation as developed by Grunstein and Hogness (1975) whereby a very large number of colonies carrying different hybrid plasmids can be readily screened to determine which colonies contain particular sequences.

Two approaches were initially adopted to detect any sequences which were differentially expressed in CF or normal cells. Firstly the lymphocyte library was hybridised with ^{32}P labelled CF and normal lymphocyte RNA and the patterns of hybridisation compared. Secondly the normal fibroblast library was hybridised with ^{32}P labelled normal fibroblast RNA in the presence of an excess of plasmid from the CF fibroblast library. Screening experiments such as these are limited by the inability to detect clones containing sequences present in low copy number in the probe. Thus only clones containing sequences which code for the middle and highly abundant RNAs can be identified.

4.2 ^{32}P labelling of RNA with polynucleotide kinase

The enzyme polynucleotide kinase isolated from T4 infected

E.coli catalyses the transfer of the γ -phosphate group from ATP to the 5' hydroxyl group of double or single stranded nucleic acid (Richardson,1965). If the ATP is labelled with ^{32}P in the position then the nucliec acid becomes radioactively labelled. The 5' end of eukaryotic mRNAs are usually "capped" which means that no free 5' hydroxyl group exists and so labelling is not possible. Even if the 5' end of the RNA was available for labelling the resulting specific activity would be inadequate for screening experiments since only one ^{32}P would be added per RNA molecule. Prior to labelling , the mRNA was subjected to mild alkali hydrolysis to generate more 5' ends. Hydrolysis in 0.1 N NaOH for 15 minutes was found to generate RNA with average length of 100 - 200 bases which could be labelled to specific activity of 10^7 cpm/μg as described by Van de Sande et al., (1973). Unincorporated nucleotides were removed by gel filtration and this RNA was used directly for colony hybridisation.

4.3 Colony hybridisation to lymphocyte cDNA library

Conditions for colony hybridisation were basically as described by Grunstein and Wallis, (1979) with modifications as described in 2.3.7 and 2.4.9/10. The lymphocyte library was diluted from glycerol stocks and filtered onto nitrocellulose filters to give a colony density of about 1500 colonies/9cm filter. When hybridised with ^{32}P poly A^+ lymphocyte RNA, only about 500 colonies

ie. 30% of the total clones, could be detected by autoradiography (see Figure 4.3.1). Gergen et al.,(1979) have shown that using reasonable probe concentrations, hybridisation and autoradiographic exposure times hybrids will be detected when a clone represents a sequence present in greater than 0.5% of the probe. However, Wahl et al.(1979) have reported that by incorporating 10% dextran sulphate into the hybridisations, signals of equal intensity are generated about 100 times faster. Replica filters were therefore hybridised in the presence and absence of 10% dextran sulphate in order to assess whether greater than 30% of the clones gave signals. As can be seen in Figure 4.3.2 the dextran sulphate did increase the signal to a small degree but was not as effective as expected. Also it appeared that the background signal was increased significantly using dextran sulphate so the resolution of the positively hybridising colonies was more difficult. Dextran sulphate was therefore not used in subsequent experiments.

4.4 Comparison of expressed sequences in lymphocyte RNA from different normal individuals.

Since the ultimate aim of this series of experiments was to identify any differences in expressed sequences from normal and CF individuals it was important to eliminate the possibility that any differences observed were due to normal individual

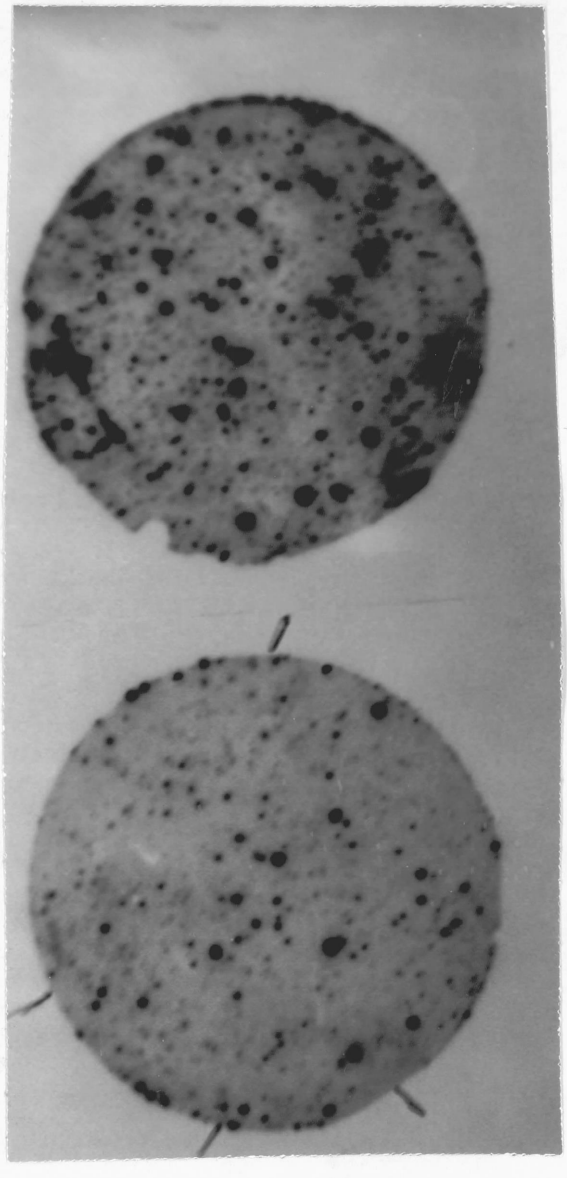


Figure 4.3.1 Grunstein-Hogness hybridisation of lymphocyte cDNA lib.

Filters containing approx. 1500 colonies were hybridised with 4×10^6 c.p.m. ^{32}P- lymphocyte RNA.

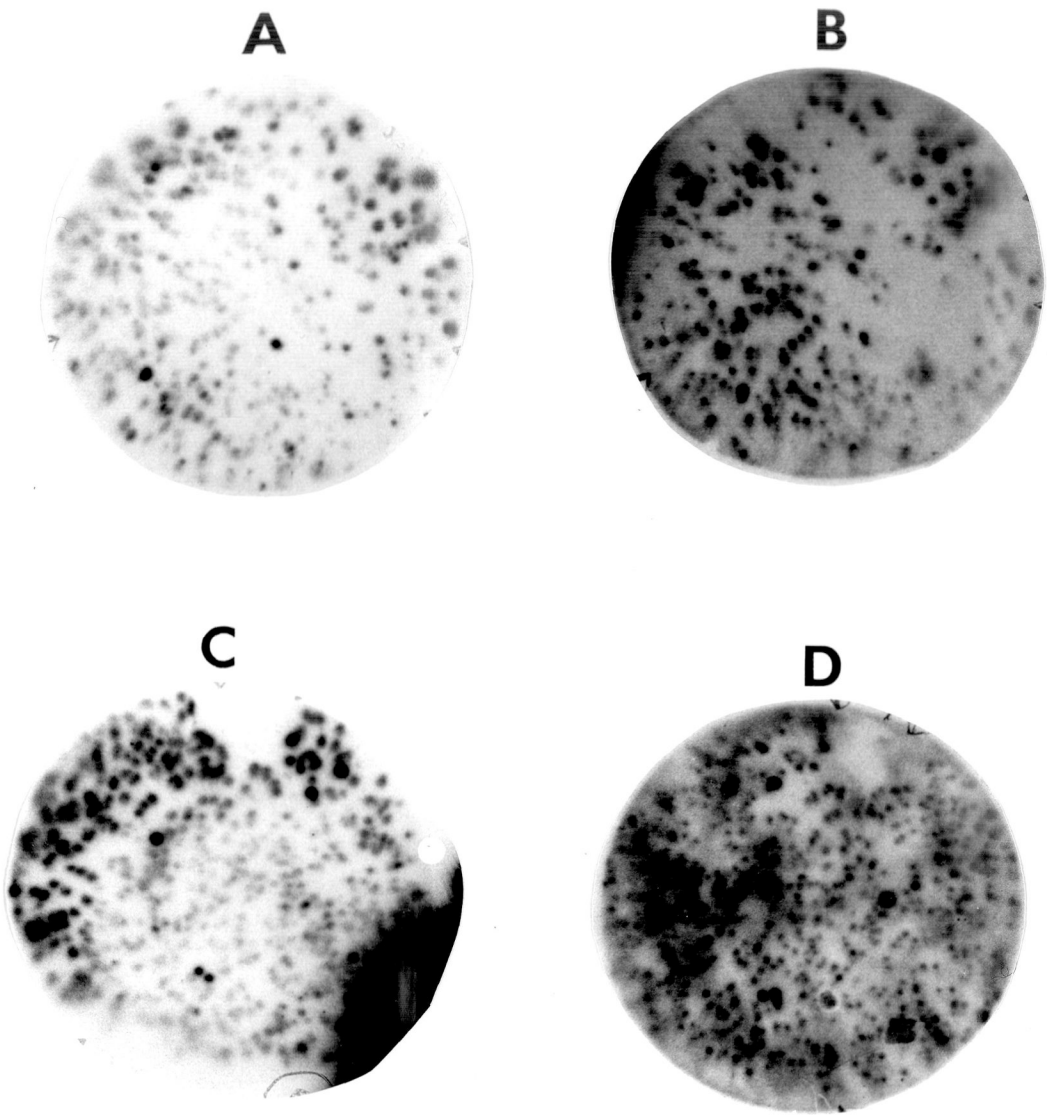


Figure 4.3.2 Use of Grunstein-Hogness hybridisation in the presence of 10% dextran sulphate.

Clones from the lymphocyte library were hybridised with ^{32}P lymphocyte RNA in the presence (B and D) and absence (A and C) of dextran sulphate. A and B are replica plates while C and D are mirror images.

variations. It was clear from Figure 4.3.1 that comparisons of RNA populations by differential hybridisation to random spreads of the library would be difficult. It was therefore decided to pick individual clones into micro titre dishes so that they could be grown as ordered arrays. These ordered arrays could then be transferred onto filters for hybridisation using a home made replicating tool. Twenty colonies were picked at random from the lymphocyte library, grown in ordered arrays and replica filters made. In order to compare expressed sequences from normal individuals, buffy coat residues were obtained from the National Blood Transfusion centre from four individuals. In order to cover possible differences between blood groups or sex the phenotypes of the four individuals selected were:-

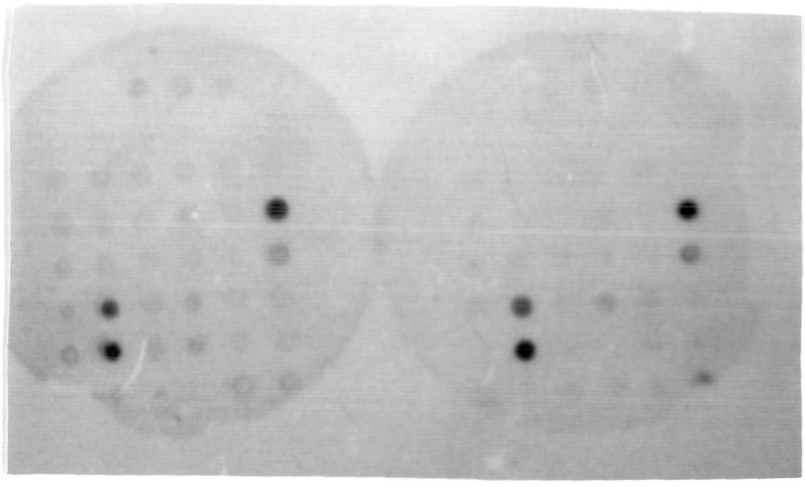
Sample	Sex	Blood Group
A	Male	O -ve
B	Female	O -ve
C	Male	A -ve
D	Male	O +ve

Lymphocytes were fractionated from the buffy coats of these individuals and RNA extracted as described in 2.4.1. Typically 500- 600 μgms. of total RNA were obtained from each pack of buffy coats. After oligo dT cellulose chromatography poly A^+ RNA was ^{32}P labelled as previously described. Each individual

RNA (4 x 10^6 cpm) was then hybridised to a replica of the ordered array, the unhybridised probe washed off and autoradiographed. When screened with RNA from the male samples the pattern of hybridisation was virtually identical (Fig. 4.4.1A) regardless of blood group. However when screened with the single female RNA one clone was consistently reduced in intensity (Fig. 4.4.1B). It is possible that other clones may have shown differences in intensity if all combinations of phenotype were used as probe and if more clones were screened. The difference in intensity of signal for this clone suggested that great care should be taken when interpreting any subsequent CF differences as being related to the basic defect.

It was possible that the nature of this difference lay either in the fact that the subject was female or that the difference was related to another factor, such as recent infection leading to stimulation in expression of a particular antibody. This was easily examined by hybridising this clone to DNA from a male and female by the procedure of Southern (1975). If the clone was present only on the Y-chromosome then no hybridisation to female DNA would be observed. 15μg of DNA from two males and two females were digested with EcoR1 and electrophoresed on 1% agarose gels. The DNA was then transferred to nitrocellulose and hybridised with plasmid DNA from this clone labelled to high specific activity by nick translation (Rigby et al.,1977).

A



B

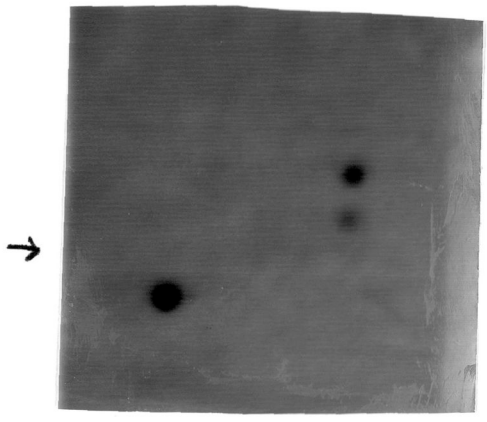


Figure 4.4.1 Hybridisation of lymphocyte cDNA clones with ^{32}P RNA from normal individuals.

A) RNA derived from male individuals.

B) RNA derived from a female individual.

Figure 4.4.2 shows the result of this hybridisation after autoradiography. In the ♂ males an intensely hybridising band was observed at approximately 8Kb. In the female samples a band was also observed at 8Kb but was much less intense. A faint band was also observed in the female samples at 3.6Kb but whether this band was also present in the male DNA was hidden by background hybridisation. Clearly, then, this clone could not be 'Y' specific.

4.5 Preselection of highly expressed lymphocyte clones.

As previously mentioned Gergen et al.,(1979) had shown that the limit of detection of Grunstein - Hogness screening was those clones whose inserts were complementary to greater than 0.5% of the probe. During the course of this study an increased sensitivity of 0.2% was reported (Crampton et al., 1980). From Table 3.11.3. it can be seen that 54% of lymphocyte cDNA hybridises to approximately 35 different highly abundant sequences. From this it can be calculated that the minimum number of clones required to generate a complete library of highly abundant lymphocyte sequences is 3.5/0.54 = 65 clones. Using the formula of Clarke and Carbon (1976) ,

$$N = \frac{\ln (1 - P)}{\ln (1 - 1/n)}$$

where N = number of clones required to identify a given highly

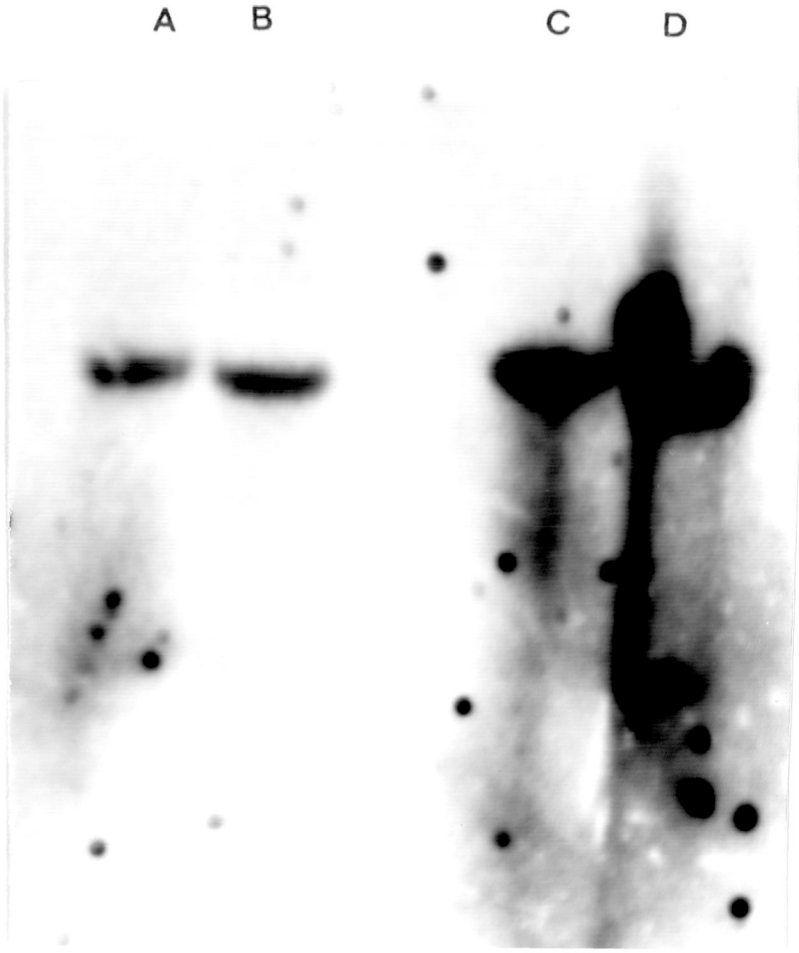


Figure 4.4.2. Southern transfer of male and female DNA with a putative 'Y' specific clone.

DNA from two females (A and B) and two males (C and D) were digested with EcoR1 electrophoresed on a 1% agarose gel, transfered to nitrocellulose and hybridised with the nick-translated plasmid.

abundant clone.

P = Probability of obtaining a given highly abundant clone.

n = minimum number of clones required to generate a complete bank of highly abundant sequences, it can be calculated that to obtain a 99% probability of obtaining any given highly abundant clone only about 300 clones need to be screened. Also from Table 3.11.3. it can be calculated that the total number of expressed sequences in lymphocytes is in the order of 16,900, of which 35 x 260 ie. 9100 are represented by the 35 most highly abundant sequences. This means that each highly abundant sequence is on average 1 - 5% of the total poly A^+ RNA. Similarly it may be calculated that a clone present at 0. 2% of the total poly A^+ RNA will have a relative copy number of 0.2/100 x 16900 = 33.8. Therefore the ratio of highly abundant sequences to those present at 0.2% is 260:33.8 or 7:1. By screening 7 times the number of clones calculated to obtain a 99% probability of obtaining a highly abundant clone ie. 7 x 300 = 2100 the same probability of obtaining a clone represented at 0.2% or more results.

As previously mentioned, it is much easier to perform comparative hybridisations on ordered arrays of clones than on random spreads. Since only approximately 2100 clones were needed to represent all those colonies detectable by Grunstein-Hogness hybridisation the use of ordered arrays seemed feasible. Such a statistical analysis cannot be absolutely accurate so

to ensure screening of a representative sample 5,000 clones from the lymphocyte library were hybridised with ^{32}P lymphocyte poly A^+ RNA. Typical examples of autoradiographs obtained are shown in Figure 4.5.1. By alignment with the master filters from which each was made 600 clones giving autoradiographic signals were individually picked into micro titre dishes and grown as ordered arrays. These clones were then used for the comparative studies.

4.6. Hybridisation of ordered arrays with CF and normal lymphocyte RNA.

In collaboration with the Brompton Hospital it was arranged for three relatively healthy male CF patients to visit the laboratory whereupon 200 mls. of blood was obtained from each with their consent. This blood was pooled, the lymphocytes fractionated by Ficoll-Hypaque centrifugation (2.4.1) and the RNA extracted. The yield of total RNA from the 600mls. was 310μgms. Previous experiments (data not shown) had shown that it was possible to screen filters with 20μgms. of total RNA instead of the usual 1μgm. of poly A^+ RNA, thus eliminating the purification step of the poly A^+ RNA using oligo (dT) cellulose chromatography. Although rRNA in the probe would be labelled by the kinase the absence of rRNA clones in the library meant that this was not a problem.

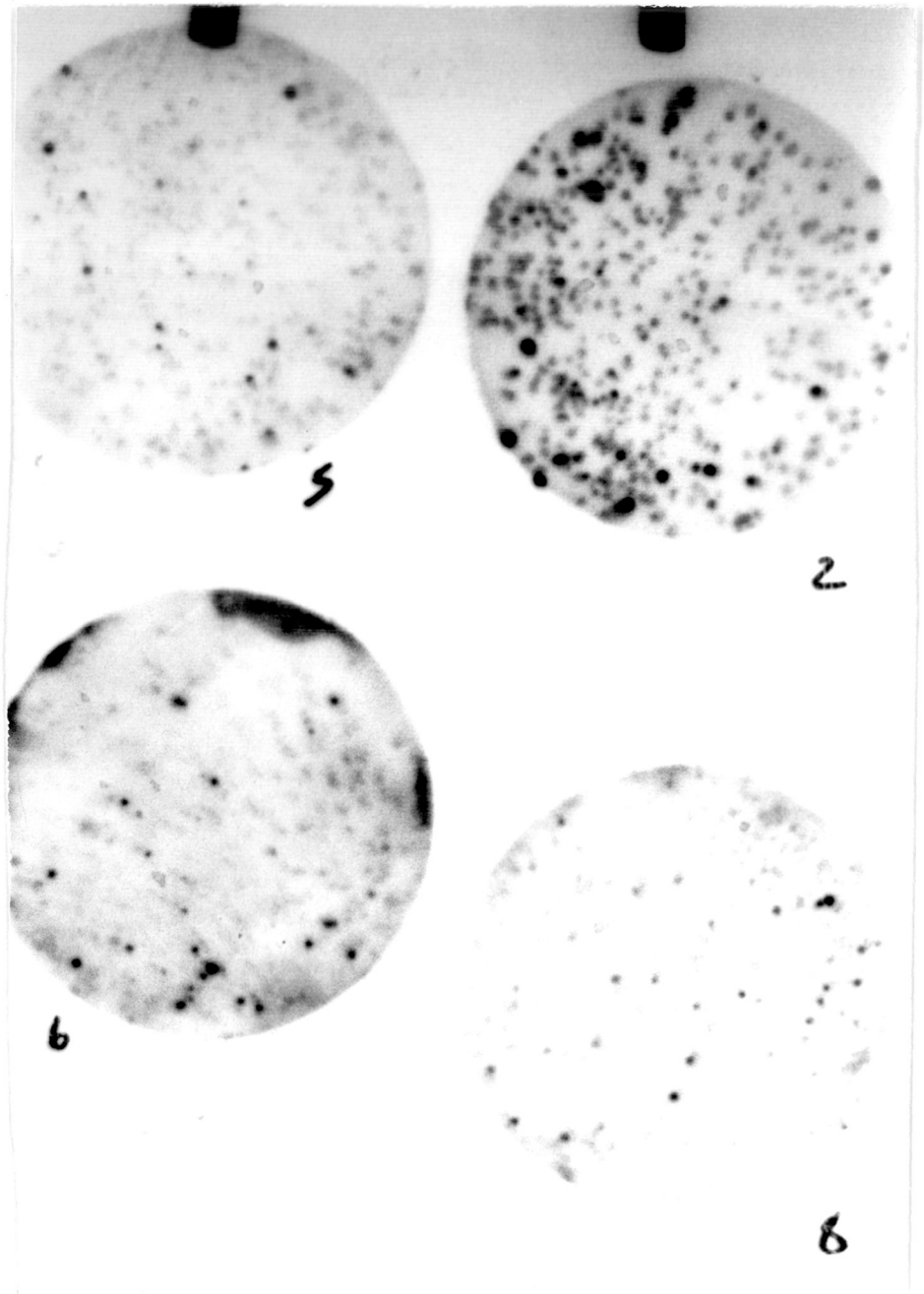
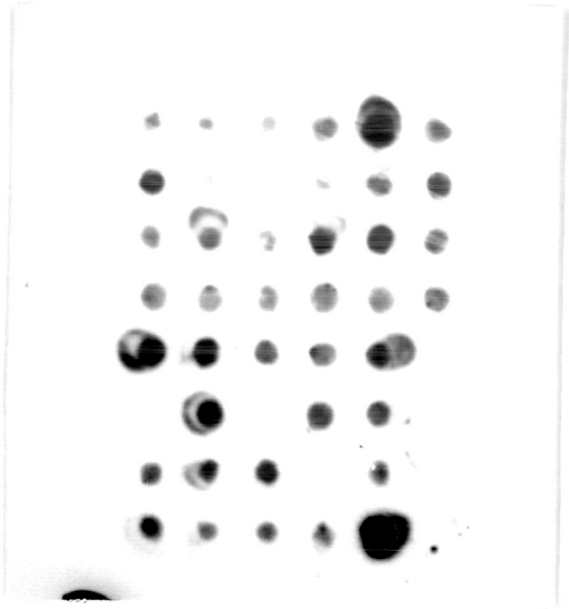


Figure 4.5.1 Grunstein-Hogness hybridisation of lymphocyte cDNA library.

Random spreads of clones (5000 in total) were hybridised with ^{32}P lymphocyte RNA.

20μgms of CF and normal RNA were alkali treated, ethanol precipitated and kinase labelled to a specific activity of 3×10^7 cpm/20μg. These preparations were then hybridised to the 600 ordered array clones in 10 mls. of formamide buffer overnight. Random spreads of the lymphocyte library were also screened with both RNAs to check whether any differences could be detected. After washing off unhybridised RNA the filters were autoradiographed and typical results are shown in Figure 4.6.1. The overall pattern of hybridisation was very similar with both types of RNA, as might be expected since they were derived from the same cell type. No qualitative differences were observed in any case, but several clones appeared to be quantitatively different. Clones which appeared to be under and over expressedin CF lymphocyte RNA were observed. Several clones also appeared promising in the comparative hybridisation of the randomly spread clones, In all, 28 clones were picked from the filters and inoculated into the wells of a single micro titre dish. These bacteria were then rescreened with both types of RNA as before and the results are shown in Figure 4.6.2. The autoradiographs shown in Figure 4.6.2 are the result of short exposure times but on longer exposure times as more clones gave signals the two arrowed clones showed the most marked differences. Both these clones represented sequences which appeared to be under expressed in CF lymphocytes compared

NORMAL



CF

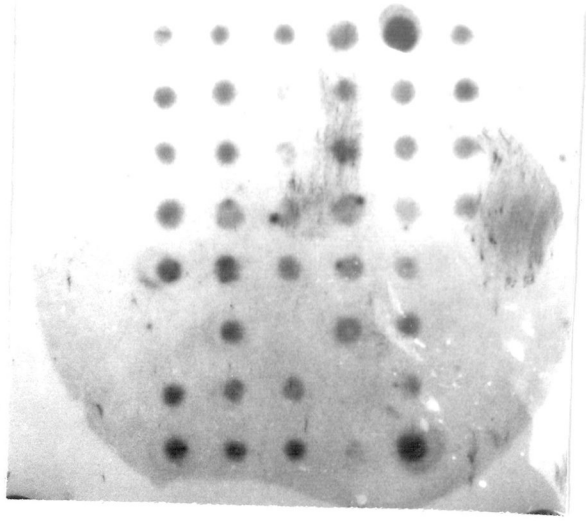
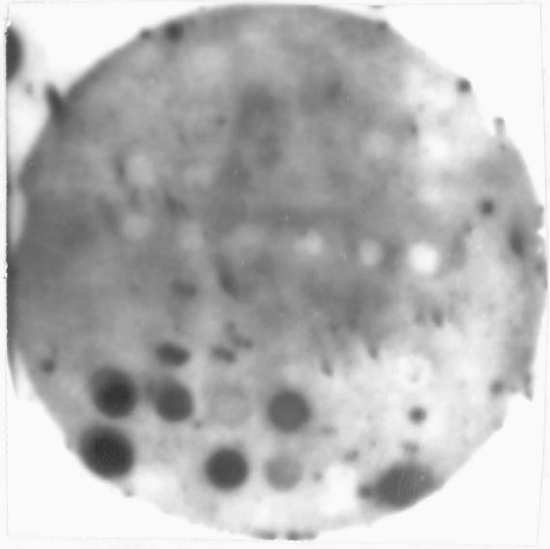


Figure 4.6.1 Grunstein-Hogness hybridisation of preselected lymphocyte clones with CF and normal lymphocyte RNA.

Clones grown in ordered arrays were hybridised in replica to ^{32}P normal and CF lymphocyte RNA. The most obvious differences are indicated and these clones were further analysed.

NORMAL



CF

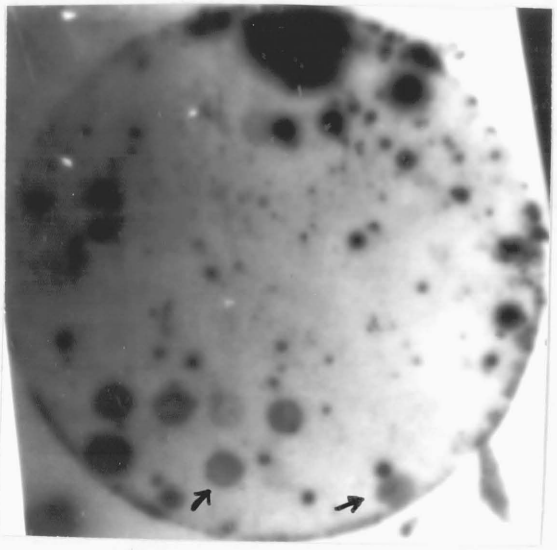


Figure 4.6.2 Rescreen of selected clones with ^{32}P normal and CF lymphocyte RNA.

28 clones picked after initial screening with CF and normal RNA were rescreened similarly. The two clones indicated were used for quantitative examination.

to normals, whereas none were observed which were over expressed. These clones were designated CDL-3 and CDL-6 (Cystic Differential Lymphocyte).

4.7 Quantitation of abundance of CDL-3 and CDL-6 in CF and normal lymphocyte poly A^+ RNA.

There are a number of methods available for estimating the abundance of a particular cloned sequence in poly A^+ RNA. The method used here was a direct hybridisation method originally described by Gillespie and Spiegelman (1965) and modified by Birnsteil et al.,(1972). Constant amounts of plasmid DNA (5µg) were denatured and immobilised to individual nitrocellulose filter discs. The efficiency of binding to the filters was assayed by binding of ^{32}P labelled plasmid DNA in parallel and was routinely 70-75%. Background hybridisation was estimated by including a set of filters loaded with pAT153. All filters were hybridised to ^{32}P labelled RNA (3×10^7 cpm/ 20µg) from CF and normal lymphocytes. Labelled mouse globin RNA was mixed with each probe so that 0.1% of the total counts was globin, and filters loaded with a mouse globin cDNA plasmid were included in the hybridisation reactions. A comparison of the levels of hybridisation to the two clones with this globin kinetic standard gave a figure for the amount of the RNA complementary to each cloned sequence. This figure must be corrected for the

insert size of each clone which was 580 bp and 450 bp for CDL -6 and CDL - 3 respectively as judged by gel electrophoresis. Taking these insert sizes into account an accurate figure for the amount of complementary RNA to each clone was determined for CF and normal lymphocyte RNA and these are shown in Table 4.7.1.

Immobilised plasmid	% of total RNA complementary to cloned sequence	
DNA	NORMAL	CF
CDL - 3	1.13	0.68
CDL - 6	1.6	0.93

Table 4.7.1

These figures show that the level of transcripts complementary to CDL - 3 and CDL - 6 in CF lymphocyte RNA are 58 and 60% respectively of the levels in normal lymphocyte RNA.

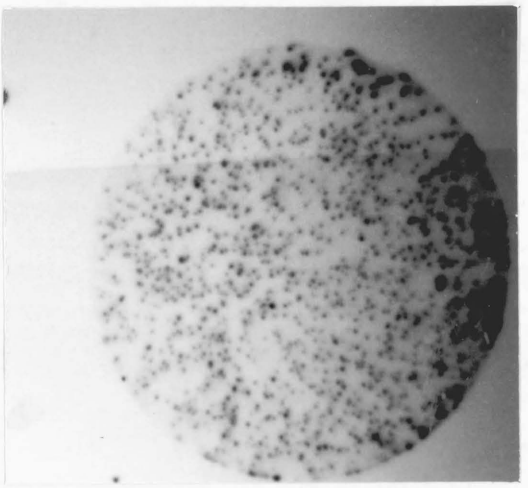
4.8 Use of fibroblast libraries in competition hybridisation.

As previously mentioned, the second approach to identifying any differntial expression was to use competitive hybridisation. The basis of this technique is that by hybridising ^{32}P labelled normal mRNA sequences in the presence of a vast excess of cloned CF sequences only those sequences which are absent or reduced in CF will be available for hybridisation to immobilised cloned

sequences during colony hybridisation. This experiment was initially impossible since insufficient RNA or DNA was available from CF tissue to enable a vast excess hybridisation to be performed. However, when libraries from CF and normal fibroblasts were constructed this problem was overcome since unlimited amounts of DNA representing the expressed sequences from CF cells were available. These experiments were performed at a high colony density since no comparison was necessary. Initially 3×10^6 cpm. of kinase labelled fibroblast RNA were hybridised to filters containing 5000 normal fibroblast recombinants and a 50 fold excess by weight of sonicated CF fibroblast library plasmid was included in the hybridisation. This corresponds to an insert excess of 10-20 fold by weight over the probe in order to compete out all common sequences. In fact, as can be seen in Figure 4.8.1 the presence of the CF library plasmid made absolutely no difference to the hybridisation pattern.

In a second series of experiments it was decided to prehybridise fibroblast ^{32}P cDNA to an excess of CF fibroblast library plasmid and then use the prehybridised mixture for screening the library. 450 μg of ^{32}P fibroblast cDNA was prehybridised with 25 μg of sheared, denatured CF fibroblast library plasmid (10 fold excess by weight of insert over cDNA). The hybridisation was performed in 3 x SSC at 65° and allowed to proceed to a Cot value of 10(calculated from the concentration

A



B

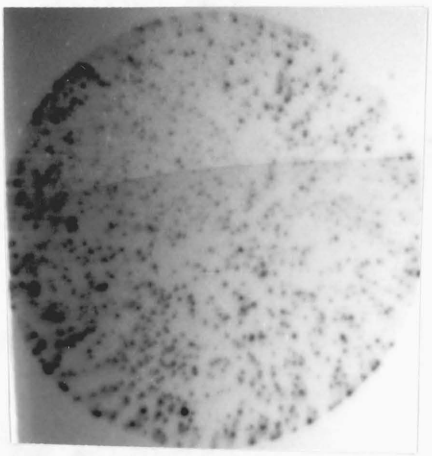


Figure 4.8.1 Use of competition hybridisation.

Grunstein-Hogness hybridisation of 5000 normal fibroblast recombinants with ^{32}P fibroblast RNA in the presence of (A) and absence of (B) unlabelled CF fibroblast cDNA library plasmid.

of the cold DNA and the time of incubation) by which time any sequences capable of giving an autoradiographic signal would have hybridised. Analysis by S_1 digestion showed 65% of the fibroblast cDNA to be double stranded. This mixture was then used directly for Grunstein-Hogness hybridisation and the result is shown in 4.8.2. Although the competitor has reduced the number of clones giving signals a substantial number of clones have not been competed out so, although this may give an enrichment for any missing sequence, a clone representing a sequence absent in CF RNA could not be identified.

4.9. The use of ^{32}P cDNA in Grunstein-Hogness hybridisation.

As previously mentioned the use of Grunstein-Hogness screening with mixed RNA populations as probe is limited by the sensitivity of the method. It is possible to kinase label poly A^+ RNA to about 10^7 cpm/μg and those clones representing 0.2% or more of the probe can be detected, (Crampton et al.,1980). If cDNA is used as the probe it is possible to generate much higher specific activities up to 10^9 cpm/μg. A short study was therefore performed to assess whether the sensitivity of this technique could be increased by using cDNA as the probe. Lymphocyte poly A^+ RNA was reverse transcribed in the presence of radioactive nucleotides as shown in Table 4.9 and each sample of ^{32}P cDNA was hybridised to

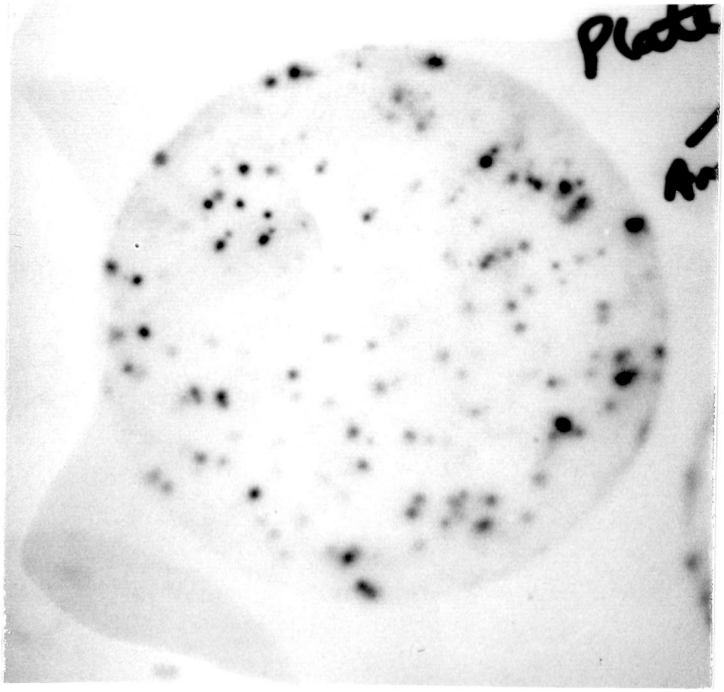


Figure 4.8.2 Grunstein-Hogness hybridisation after prehybridisation of probe and competitor.

Randomly spread normal fibroblast clones were hybridised with ^{32}P fibroblast cDNA which had been prehybridised to a vast excess of CF fibroblast library plasmid.

replica filters of clones from the lymphocyte library.

Sample	^{32}P Input	Specific Activity (cpm/ μg
A	10 μCi dCTP	184×10^6
B	40 μCi dCTP	184×10^6
C	10 μCi each dCTP,TTP	368×10^6
D	10 μCi each dCTP,TTP,dGTP	552×10^6
E	10 μCi each dCTP, TTP,dGTP,dATP	739×10^6

Table 4.9

Figure 4.9.1 shows the resultant autoradiographs after exactly the same exposure time. Replicas A and B show that increasing the amount of a single radioactive nucleotide has little effect on the number of clones which appear. However, as one adds different radioactive nucleotides to the incubation more colonies appear, (C D and E) ie. sensitivity is increased.

In order to compare the sensitivity of the hybridisation replica filters were also hybridised with ^{32}P poly A^+ lymphocyte RNA (10^7 cpm/ μg) and exposed for similar times. The resultant autoradiograph (Replica F) shows that there is an increase in the number of colonies giving signals but also the relative intensities and patterns of hybridisation of some clones vary. The increase in sensitivity has not yet been determined.

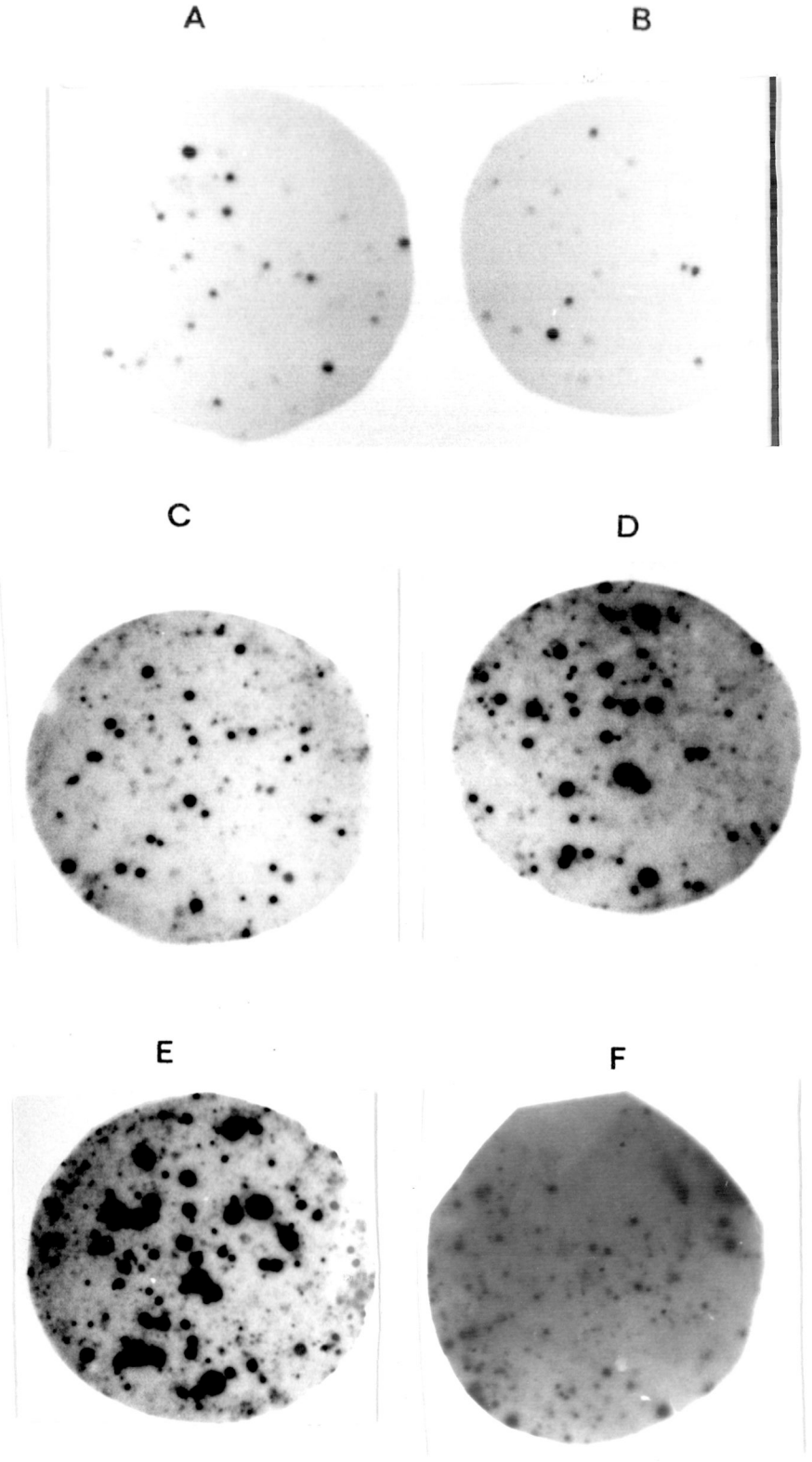


Figure 4.9.1 Grunstein-Hogness hybridisation using cDNA.

Replica filters (A-E) were hybridised with ^{32}P lymphocyte cDNA as described in the text and F with ^{32}P lymphocyte poly A^+ RNA.

4.10 DISCUSSION

4.10.1 Use of comparative hybridisation

These experiments demonstrate a new approach to the study of genetic disease using comparative hybridisation. This type of screening has been used in the past for identifying RNA sequences that are differentially expressed in lymphocytes and fibroblasts (Crampton et al.,1980) developmentally regulated RNA molecules in Xenopus (Golden et al., 1980),Dictyostelium (Williams & Lloyed,1979) and quantitatively for sea urchin development (Lasky et al.,1980). The closest analogy to this study, however, is the molecular analysis of the mutation which prevents the synthesis of many chorionproteins in the silk moth (Latron et al.,1980). In this study cell free translation had shown the absence of specific proteins and differential hybridisation to a cDNA library allowed identification of corresponding clones.

Kinetic analysis and translational studies of mRNA from normal and cystic fibrosis cells shows no gross differences. Neither of these techniques, however, precludes the possibility that a sequence or small family of sequences may be differentially expressed in cystic fibrosis cells.

The method of differential hybridisation is limited by the fact that only those RNA sequences expressed at a level of 0.2% or greater are being examined. Clearly, therefore, this

method will only be of value for a genetic disease which is manifested by a change at this level. Similarly, this method has limited applicability since it would not detect a defect such as a point mutation or a defect in processing of a particular sequence unless the processing resulted in an altered abundance of the sequence. Nevertheless, the identification of the two clones which appear to be under expressed in CF lymphocytes is certainly encouraging. The fact that the relative ratios of expression for both clones are very similar and that their insert sizes are similar may mean that they are different transcripts of the same RNA. Restriction mapping of the two clones is now in progress and will clarify this point. Otherwise it may be speculated that the defect in CF is in a region of the genome which controls co-ordinate expression of a number of genes, and these clones may be different members of this family. Further characterization of these clones including quantitation of their respective RNAs in CF individuals and heterozygotes is necessary before any definite relationship with CF is established. This is particularly true in view of the clone which was shown to vary between normal individuals.

Some misgivings were felt initially regarding the pooling of blood from the three CF patients in order to extract the lymphocytes. If as has been suggested, CF is a heterogeneous

disease then it is possible that the defect in one of these individuals could be masked by the correct function in others. However, in comparison with three normal individuals there would be a difference in the relative concentrations of this defective sequence which is the basis of this screening procedure. Conversely the use of a mixed sample may mean that individual variation may be eliminated and so any detectable differences should be related to CF. Since a mixed population of normal lymphocytes was used to construct the library and to make ^{32}P RNA for screening this approach was felt to be justified.

With the construction of libraries from fibroblast samples it was possible to carry out similar comparative hybridisations for another cell type. Although in CF the defect is thought to be expressed in all cell types the use of two different cell types should reduce the possibility of missing the defect due to cell specific expression. When using fibroblast RNA for screening any problems ofheterogeneity are eliminated since the fibroblasts are grown in tissue culture and are essentially derived from a single individual. It should be emphasised again however, that individual variations may be present between normal fibroblasts from different individuals

Since the defect in CF is thought to be manifest in all cell types it may be possible to detect differences in expression by comparative heterologous hybridisation between the normal

fibroblast library and CF or normal lymphocyte RNA (or vice versa). It has been shown that in some cases 80% of the sequences present in one cell type are also present in other cell types (Ostrow et al.,1979). However the relative abundance of these sequences are different between cell types and many of the sequences present at 0.2% or more of the RNA are cell specific. This means that in heterologous hybridisations the number of clones giving autoradiographic signals is much reduced, so comparison is easier. Of course, this still requires that the basic defect in CF involves a sequence present at 0.2% or more of the RNA.

4.10.2 Use of competitive hybridisation.

The use of competition hybridisation in colony screening has been reported recently (Mangiarotti et al.,1981) although it has been an extremely useful technique in other areas of molecular biology for many years (Skoultchi & Gross,1973; Church & McCarthy,1967). The use of competition hybridisation offers several advantages over conventional colony hybridisation. The introduction of competition a) simplifies the screening procedure since it can be done with a single filter hybridisation and thus b) avoids any problems of unequal amounts of DNA on replica filters which may influence the intensity of the signals; c) reduces the problem of background hybridisation, because of the large

excess of unlabelled DNA which minimizes the non specific sticking of the labelled probe to the filter.

Despite these advantages the competition experiments described here were unsuccessful. Initial experiments showed that unlabelled competitor CF library plasmid did not compete with ^{32}P-cDNA from normal fibroblasts at all. When the CF library plasmid was sheared and denatured single stranded DNA consisting of insert and plasmid sequences were produced. It can be speculated that on addition of this to the filters the plasmid sequences hybridised to plasmid sequences in the clones, regardless of insert sequences, leaving unhybridised tracts of single stranded fibroblast sequences. When ^{32}P-cDNA was added it hybridised to these tracts and so all clones gave autoradiographic signals.

This problem was reduced by prehybridising fibroblast ^{32}P-cDNA to a vast excess of CF library plasmid so that all plasmid sequences could reanneal and so any hybridisation in the subsequent colony hybridisation would be due to unhybridised fibroblast ^{32}P-cDNA. Clearly this did reduce the number of clones giving signals but not as much as would be desired. One possible explanation for this is the fact that a fraction of cDNA populations is non-hybridisable by solution hybridisation (Getz et al.,1975). Whether this cDNA could hybridise to immobilised filter DNA in colony hybridisation is unclear but it is a possible explanation

for the number of clones still showing as positive. These sequences could be removed by initially hybridising fib. cDNA and poly A^+ RNA to completion followed by hydroxyapatite chromatography to isolate the single stranded DNA . Although in this study the CF library plasmid was used as competitor it is, of course, also necessary to perform the reverse experiment to detect any clones which may be over expressed in CF cells.

Modifications of this technique have also been reported (Vitek et al.,1981; Hirsch et al.,1978). These involve enrichment for specific sequences by using cDNA synthesised on oligo (dT) cellulose columns. This could be applied to CF by synthesising CF cDNA using oligo (dT) cellulose as primer, (to which it covalently binds) and then passage through the column of normal fibroblast RNA. Any sequences which are not present in CF will not be retained and the eluate could then be used as a probe enriched in the CF deficient sequence.

4.10.3. Increasing sensitivity of screening.

In an attempt to increase the sensitivity of screening, ^{32}P-cDNA of very high specific activity was used as probe in place of kinase labelled RNA. The sensitivity of the procedure was not increased significantly but the pattern of hybridisation was different for the two probes. cDNA can only be made from

poly A^+ RNA, theoretically, whereas kinase labelled RNA can be poly A^+ or poly A^- RNA. Species of mRNA which occur in poly A^+ and poly A^- forms are known to exist for many proteins (Cann et al.,1974; Ruderman & Pardue,1977; Hunter & Garrels,1977). Clearly, therefore, the signals generated for these clones will differ according to the probe. Further variation in hybridisation pattern may be due to a decreased concentration of probe when using cDNA and differing lengths of sequence in the probe eg. due to alkaline hydrolysis of RNA prior to labelling.

Several methods have been developed for identifying specific low abundance clones (Harpold et al.,1978; Parnes et al.,1981) but these all involve identification of characteristic proteins in cell free translations etc., which is impossible to apply for CF since the defective protein is unknown. Indeed, the absence of a specific product of the abherrant CF gene is the major stumbling block to work in this area. Any partially purified product or sequence data would permit much more direct study of the defective gene.

5.SUMMARY AND PERSPECTIVES.

This study has described the construction of recombinant cDNA libraries from normal human lymphocytes, normal human fibroblasts and cystic fibrosis fibroblasts. These libraries have been used in preliminary comparative experiments to try to elucidate the nature of the basic defect in cystic fibrosis. So far two clones have been identified whose expression is reduced to 60 and 58% in CF lymphocytes as compared to normals. These clones are now being analysed by study of RNA from CF parents and homozygotes to assess their relationship to the basic defect. Lack of knowledge of the nature of the protein product which is defective in CF makes identification of the clones coding for these sequences difficult. However, the availability of the libraries may allow some of the speculated defects to be verified or discounted. For instance it would be interesting to observe whether plasmid from the CF library could cause hybrid arrested translation of a protein which co-migrates with arginine esterase which has been linked to the basic defect.

The libraries themselves will obviously be useful for a wide range of studies. Already putative sequences for actin, collagen and elastin have been tentatively identified in the normal fibroblast library. Both lymphocyte and normal fibroblast libraries have been studied with regard to the presence of repeated sequences adjacent to the coding sequences. It may also

be possible to detect precursor hnRNA clones in some cases whereby some insight into mechanisms of post transcriptional processing may arise. Arguably the most important use, however, will be in the study of genetic disease.

The method described in this thesis is limited in sensitivity and will only detect significant differences in expression. Defects involving point mutations, translational or post-translational control will not necessarily be detected. If the defect is due to a point mutation then the most promising approach is to identify a restriction fragment polymorphism which is linked to the defect in CF similar to that shown for β-thalassemia and sickle cell disease (Kan et al., 1980; Kan & Dozy, 1978). In order to study linked polymorphisims it is necessary to have recombinant DNA probes for the particular chromosomal region concerned. It has been reported that CF may be loosely linked to several human chromosomes particularly chromosome 4. It is relatively easy to isolate those sequences from the libraries which are expressed by chromosome 4 and so generate a large number of probes for linked polymorphisim studies on CF families.

Once the defective gene(s) has been identified the development of a highly accurate antenatal diagnosis (AND) procedure will be possible. This may not be absolutely essential if Nadler's procedure for AND proves to be efficient and accurate. However, ultimately, the possibility of correcting the defect in CF as

suggested by French Anderson and Diacumakos (1981) will require the defective gene to be positively identified and the development of techniques for introducing genes into mammalian cells in vivo.

Affara,N.A.,Goldfarb,P.S.,Vass,K.,Lyons,A.and Harrison,P.R. (1981) Nuc.Acids Res. 9 3061-3076

Allison,A.C.(1954) Brit.Med.J. 1 290-294

Altman,S.(1981) Cell 23 3-4

Amster,O.,Salomon,D.,Zemel,O.,Zamir,A.,Zeelon,E.P.,Kantor,F.and Schechter,I.(1980) Nuc.Acids Res. 8 2055-2065

Anderson,C.M.and Goodchild,M.C. in Cystic fibrosis-Manual of diagnosis and management, Blackwell Scientific Publications.

Andersen,D.H.(1959) J.Chron.Dis. 7 58

Aviv,H.and Leder,P.(1972) Proc.Natl.Acad.Sci.U.S.A. 69 1408-1412

Avvedimento,V.E.,Vogeli,G.,Yamada,Y.,Maizel,J.V.,Pastan,I.and de Crombrugghe,B.(1980) Cell 21 689-696

Backman,K.,Ptashne,M.and Gilbert,W.(1976) Proc.Natl.Acad.Sci. U.S.A. 73 4174-4178

Bahl,C.P.,Marians,K.J.,Wu,R.,Stawinsky,J.and Narang,S.(1977) Gene 1 81-92

Baldacci,P.,Royal,A.,Bregegere,F.,Abastado,J.P.,Cami,B.,Daniel, F.and Kourilsky,P.(1981) Nuc.Acids Res. 9 3575

Baltimore D.(1970) Nature 226 1209-1210

Barnes,W.M.(1977) Science 195 393-394

Bell,G.I.,Pictet,R.and Rutter,W.J.(1980) Nuc.Acids Res. 8 4091-4099

Benoist.C,O`Hare,K.,Breathnach,R.and Chambon,P.(1980) Nuc.Acids Res. 8 127-142

Berk,A.J.and Sharp,P.A.(1977) Cell 12 721-732

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

Bernheim,M.,Monnet,P.,Jeune,M.,Robert,J.M.,Comby,J.(1961) Pediatrie 16 17-38

Berry,H.K.,Kellogg,F.W.,Lichstein,S.R.and Ingberg,R.C.(1980) Amer.J.Dis.Child. 134 930-934

Besley,G.T.N.,Patrick,A.D.and Norman,A.P.(1969) J.Med.Genet. 6 278

Biggar,W.D.,Holmes,B.and Good,R.A.(1971) Proc.Natl.Acad.Sci. U.S.A. 68 1716

Birkenmeir,E.H.,Brown,D.D.and Jordan E.(1978) Cell 15 1077-1086

Birnie,G.D.,Macphail,E.,Young,B.D.,Getz,M.H.and Paul,J.(1974) Cell Diff. 3 221-232

Birnsteil,M.L.,Sells,B.H.and Purdom,I.F.(1972) J.Mol.Biol. 63 21-39

Bishop,J.O.,Morton,J.G.,Rosbash,M.and Richardson,M.(1974) Nature 250 199-204

Blomfield,J.,Rush,A.R.,Allars,H.M.and Brown,J.(1976) Ped.Res. 10 574-584

Bogenhagen,D.F.,Sakonju,S.and Brown,D.D.(1980) Cell 19 27-35

Bogenhagen,D.F.and Brown,D.D.(1981) Cell 24 261-270

Bolivar,F.and Backman,K.(1979) in Methods in Enzymology 68 245-280 Eds. Colowick,P.and Kaplan,N.O. Academic Press

Botelho,S.Y.,Goldstein,A.M.and Rosenlund,M.L.(1973) J.Ped. 83 601-607

Bowman,B.H.,Lockhart,L.H.and Mc.Combs,M.L.(1969) Science 164 325

Bowman,B.H.,Barnett,D.R.,Matalon,R.,Danes,B.S.and Bearn,A.G.(1973) Proc.Natl.Acad.Sci.U.S.A. 70 548-551

Bowman,B.H.,Lockhart,L.H.and Mc.Combs,M.L.(1969) Science 164 325-326

Boxerbaum,B.,Kagumba,M.and Matthews,L.W.(1973) Amer.Rev.Resp. Dis. 108 777

Boyer,H.W.and Roulland-Dussoix,D.(1969) J.Mol.Biol. 41 459-472

Bozzoni,I.,Beccari,E.,Luo,Z.X.,Amaldi,F.,Pierandrei-Amaldi,P.and Campioni,N.(1981) Nuc.Acids Res. 9 1069-1086

Brackenridge,C.J.(1978) Ann.Hum.Genet. 42 197-201

Brawerman,G.(1981) Crit.Rev.Biochem. 10 1-38

Breathnach,R.,Mantei,N.and Chambon,P.(1980) Proc.Natl.Acad.Sci. U.S.A. 77 740-744

Breslow,J.L.and Epstein,J.(1980) Science 207 1007

Breslow,J.L.,Epstein,J.,Fontaine,J.and Forbes,G.B.(1978) Science 201 180-182

Breslow,J.L.,Mc.Pherson,J.and Epstein J.(1981) New Engl.J.Med. 304 1-5

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

Britten,R.J.and Davidson,E.H.(1971) Quart.Rev.Biol.46 111-138

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

Brock D.J.H.and Hayward,C.(1979) Lancet 1 1244

Brunecky,Z.(1972) J.Med.Genet.9 33-37

Brutlag,D.,Fry,K.,Nelson,T.and Hung,P,(1977) Cell 10 509-519

Buell,G.N.,Wickens,M.P.,Paynor,F.and Schimke,R.T.(1978) J.Biol.Chem 253 2471-2482

Bütterworth,J.(1974) Clin.Chim.Acta 56 159

Campo,M.S.,Bishop,J.O.(1974) J.Mol.Biol. 90 649-663

Cann,A.,Gamino,R.,Banks,J.and Bank,A.(1974) J.Biol.Chem. 249 7536-7540

Carter,C.O.(1967) Mod.Probl.Pediat. 10 372

Chan,L.,Means,A.R.and O'Malley,B.W.(1973) Proc.Natl.Acad.Sci. U.S.A. 70 1870-1874

Chien,Y-H.and Thompson,E.B.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 4583-4587

Chung,S.(1981) Cell 24 785-798

Church,R.and Mc.Carthy,B.J.(1967) Proc.Natl.Acad.Sci.U.S.A. 58 1548-1555

Clarke,L.and Carbon J.(1976) Cell 9 91-99

Clemens,M.J.(1979) Nature 279 673-674

Clewell,D.B.and Helinski,D.(1969) Proc.Natl.Acad.Sci.U.S.A. 62 1159-1166

Cochet,M.,Perrin,F.,Gannon,F.,Krust,A.,Chambon,P.,Mc.Knight,G., Lee,D.C.,Mayo,K.E.and Palmiter,R.(1979) Nuc.Acids Res. 6 2435-2451

Cohen,F.L.and Daniel,W.L.(1974) J.Med.Genet. 11 253

Cohen,S.N.,Chang,A.C.Y.and Hsu,L(1973) Proc.Natl.Acad.Sci.U.S.A 70 3240-3244

Colten,H.,Lapey,A.and Mark,E.J.(1981) New Engl.J.Med. 304 831-836

Conneally,P.M.,Merritt,A.D.and Yu,P.(1973) Tex.Rep.Biol.Med. 31 639-650

Conover,J.H.,Conod,E.J.and Hirschorn,K.(1974) Life Sci. 14 253-266

Conover,J.H.,Conod,E.J.and Hirschorn,K.(1973) Lancet 2 1501

Conover,J.H.,Bonforte,R.J.,Hathaway,P.,Paciuc,S.,Conod,E.J., Ginbender,E.and Hirschorn,K.(1973) Ped.Res. 7 224-28

Conover,j.h.,Conod,E.J.and Hirschorn,K.(1976) Tex.Rep.Biol. Med. 34 45

Cooke,H.J.and Hindley,J.(1979) Nuc.Acids Res. 6 3177-3197

Corden,J.,Wasylyk,B.,Buchwalder,A.,Sassone-Corsi,P.,Kedlinger,C and Chambon,P.(1980) Science 209 1406_1414

Costantini,F.D.,Scheller,R.H.,Britten,R.J.and Davidson,E.H. (1978) Cell 15 173-187

Covarrubias,L.,Cervantes,L.,Covarrubias,A.,Soberon,X.,Vichido,I Kupersztoch-Portnoy,Y.M.and Bolivar,F.(1981) Gene 13 25-35

Cox,R.F.(1977) Biochemistry 16 3433-3443

Crampton,J.M.,Humphries,S.,Woods,D.and Williamson,R.(1980) Nuc.Acids Res. 8 6007-6017

Crawfurd,M.D.(1975) Lancet 1 167

Crawfurd,M.D.(1972) Heredity 29 126

Crick,F.(1979) Science 204 264-271

Crossley,J.R.,Elliot,R.B.and Smith,P.A.(1979) Lancet 472-474

Dagert,M.and Ehrlich,S.D.(1979) Gene 6 23-28

Dandona,P.,Littlewood,J.M.,Ramdial,L.and Evans,R.(1981a) Lancet 380

Dandona,P.,Hodson,M.E.,Bell,J.,Ramdial,L.,Beldon,J.and Batten,J.C.(1981b) Thorax 36 60-62

Dandona,P.,Hodson,M.E.,Ramdial,L.,Bell,J.M.and Batten,J.C. (1979) Lancet 1032

Danes,B.S.and Bearn,A.G.(1972) J.Exp.Med. 136 1313

Danes,B.S.and Bearn,A.G.(1968) Lancet 1061-1063

Danks,D.M.,Allan,J.and Anderson,C.M.(1965) Ann.Hum.Genet.28 323

Dann,L.G.and Blau,K.(1979) Lancet 2 907

Darnell,J.(1978) Science 202 1257-1260

Davidson,E.H.,Hough,B.R.,Klein,W.H.and Britten,R.J.(1975) Cell 4 217-238

Davidson,E.H.and Britten,R.J.(1979) Science 204 1052-1059

Davidson,E.H.,Klein,W.H.and Britten,R.J.(1977) Dev.Biol. 55 69-84

Davis,M.M.,Calame,K.,Early,P.W.,Livant,D.L.,Joho,R.,Weissman,I.L.and Hood,L.(1980a) Nature 283 733-739

Davis,M.M.,Kim,S.K.and Hood,L.(1980b) Cell 22 1-3

Davis,P.B.,Hubbard,V.S.and di Sant'Agnese,P.A.(1980a) Amer.J.Med. 69 643

Davis,P.B.,Shelhamer,J.R.and Kaliner,M.(1980b) New Engl.J.Med. 302 1453

Deeley,R.G.,Gordon,J.I.,Burns,A.T.H.,Mullinix,K.P.,Binastein,M. and Goldberger,R.F.(1977) J.Biol.Chem. 252 8310-8319

Defranco,D.,Schmidt,O.and Soll,D.(1980) Proc.Natl.Acad.Sci. U.S.A. 77 3365-3368

Deininger,P.L.and Schmid,C.W.(1976) J.Mol.Biol. 106 773-790

Deininger,P.L.and Schmid,C.W.(1979) J.Mol.Biol. 127 437-460

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

Denning,C.R.,Huang,N.N.,Cuasay,L.R.,Schwachman,H.,Tocci,P., Warwick,W.J.and Gibson,L.E.(1980) Ped. 66 752

Denning,C.R.,Sommers,S.C.and Quigley,H.J.(1968) Ped. 41 7

De Noto,F.M.,Moore,D.D.and Goodman,H.M.(1981) Nuc.Acids Res.9 3719-3730

Derman,E.,Krauter,K.,Walling,L.,Weinberger,C.,Ray,M.and Darnell J.E.Jr.(1981) Cell 23 731-739

Dierks,P.,Van Ooyen,A.,Meyer,F.,Mantei,N.and Weissman,C.(1980) Experientia 36 742

Dierks,P.,Van Ooyen,A.,Mantei,N.and Weissman,C.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 1411-1415

di Sant'Agnese,P.A.,Darling,R.C..Perera,G.A.(1953) Am.J.Med. 15 777

di Sant'Agnese,P.A.and Talamo,R.C.(1967) New Engl.J.Med. 277 1287-1294

di Sant'Agnese,P.A.and Davis,P.B.(1976) New Engl.J.Med. 295 481-485

Doershuk,C.F.,Matthews,L.W.,Tucker,A.S.and Spector,S.(1965) Ped. 36 675

Donald,L.J.,Hamerton,J.L.(1978) Cytogen.Cell.Genet. 22 5-11

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

Dugaiczyk,A.,Woo,S.L.C.,Colbert,D.A.,Lai,E.C.,Mace,M.L.and O'Malley,B.W.(1979) Proc.Natl.Acad.Sci.U.S.A. 76 2253-2257

Duncan,C.H.,Jagadeeswaran,P.,Wang,R.R.C.and Weissman,S.M.(1981) Gene 13 185-196

Duncan,C.,Biro,P.A.,Choudary,P.V.,Elder,J.T.,Wang,R.C.,Forget,B DeRiel,J.K.and Weissman,S.M.(1979) Proc.Natl.Acad.Sci.U.S.A. 76 5095-5099

Early,P.,Rogers,J.,Davis,M.,Calame,K.,Bond,M.,Wall,R.and Hood,L.(1980b) Cell 20 313-319

Early,P.,Huang,H.,Davis,M.,Calame,K.and Hood,L.(1980b) Cell 19 981-992

Efstratiadis,A.,Maniatis,T.,Kafatos,F.C.,Jeffrey,A.and Vournakis,J.K.(1975) Cell 4 367_378

Efstratiadis,A.,Kafatos,F.C.,Maxam,A.and Maniatis,T.(1976) Cell 7 279-288

Elder,J.T.,Pan,J.,Duncan,C.and Weisman,S.M.(1981) Nuc.Acids Res. 9 1171-1189

Engelke,D.R.,Ng,S-Y.,Shatry,B.S.and Roeder,R.G.(1980) Cell 19 717-728

Epstein,J.and Breslow,J.L.(1977) Proc.Natl.Acad.Sci.U.S.A. 74 1676-1679

Evans,M.I.,Hager,L.J.and Mc.Knight,G.S.(1981) Cell 25 187-193

Fantoni,A.,Bozzoni,I.,Ullu,E.and Farace,M.G.(1979) Nuc.Acids Res. 6 3505-3517

Farrell,P.M.,Pallavicini,J.C.and Ulane,M.M.(1975) Proc.Soc.Exp.Biol.Med. 149 340

Feigal,R.J.and Shapiro,B.L.(1979a) Nature 278 276-277

Feigal,R.J.and Shapiro,B.L.(1979b) Proc.Natl.Acad.Sci.U.S.A. 76 2979-2983

Feigal,R.J.and Shapiro,B.L.(1979c) Ped.Res. 13 764-768

Felber,B.K.,Gerber-Huber,S.,Meier,C.,May,F.E.B.,Westley,B., Weber,R.and Ryffel,G.U.(1981) Nuc.Acids Res. 9 2455

Ferrari,S.,Yehle,C.O.,Robertson,H.D.and Dickson,E.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 2395-2399

Fiddes,J.C.and Goodman,H.M.(1979) Nature 281 351-356

Flavell,R.A.,Kooter,J.M.,de Boer,E.,Little,P.F.R.and Williamson,R.(1978) Cell 15 25-41

Foder,B.,Scharff,O.and Tonessen,P.(1980) Clin.Chim.Acta. 104 187

Fraser,N.W.,Nevins,J.R.,Ziff,E.and Darnell,J.E.(1979) J. Mol.Biol. 129 643-666

French Anderson,W.,Killos,L.,Sanders-Haigh,L.,Kretschmer,P.J. and Diacumakos,E.G.(1980) 77 Proc.Natl.Acad.Sci.U.S.A. 5399-5403

French Anderson,W.and Diacumakos,E.G.(1981) Sci.Amer. August 60-93

Friedman,E.Y.and Rosbash,M.(1977) Nuc.Acids Res. 4 3455-3471

Fritsch,E.F.,Lawn,R.M.and Maniatis,T.(1980) Cell 19 959-972

Galau,G.A.,Britten,R.J.and Davidson,E.H.(1974) Cell 2 9-20

Galey,W.R.,Gillon,M.J.and Murphy,S.(1980) Ped.Res. 14 1269-1271

Garapin,A.C.,Cami,B.,Roskam,W.,Kourilsky,P.,LePennec,J.P., Perrin,F.,Gerlinger,P.,Cochet,M.and Chambon,P.(1978) Cell 14 629-639

Gergen,J.P.,Stern,R.H.and Wensink,P.C.(1979) Nuc.Acids Res. 7 2115-2136

Getz,M.J.,Birnie,G.D.,Young,B.D.,Macphail,E.and Paul,J.(1975) Cell 4 121-129

Gibson,L.E.(1980) Am.J.Dis.Child. 134 925-926

Gibson,L.E.and Cooke,R.E.(1959) Ped. 23 545

Gilbert,W.(1978) Nature 271 501

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

Goff,S.P.and Berg,P.(1979) J.Mol.Biol.133 359-382

Goldberg,R.B.,Galau,G.A.,Britten,R.J.and Davidson,E.H.(1973) Proc.Natl.Acad.Sci.U.S.A. 70 3516-3520

Golden,L.,Schafer,U.and Rosbash,M.(1980) Cell 22 835-844

Goodchild,M.C.,Edwards,J.H.,Glenn,K.P.,Grindey,C.,Harris,R., Mackintosh,P.and Wentzel,L.(1976) J.Med.Genet.13 417-449

Goodman,H.O.and Reed,S.C.(1952) Am.J.Hum.Genet.4 59

Gorin,M.B.and Tilghman,S.A.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 1351-1355

Gottleib,L.(1971) J.Ped. 79 930

Green,M.C.,Sidman,R.L.and Pivetta,O.H.(1972)Science 176 800-803

Grosschedl,R.and Birnstiel,M.L.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 1432-1436

Grosveld,G.C.,Shewmaker,C.K.,Jat,P.and Flavell,R.A.(1981a) Cell 25 215-26

Grosveld,G.C.,Koster,A.and Flavell,R.A.(1981b) Cell 23 573-584

Grunstein,M.and Hogness,D.S.(1975) Proc.Natl.Acad.Sci.U.S.A. (1975) 72 3961-3965

Grunstein,M.and Wallis,J.(1979) in Methods in Enzymol. 68 379-389 Academic Press Inc.

Halbreich,A.,Pajot,P.,Foucher,M.,Grandchamp,C.and Slonimski,P.(1980) Cell 19 321-329

Hall,B.D.and Simpkiss,M.J.(1969) J.Med.Genet. 5 262

Hamer,D.H.,Smith,K.D.,Boyer,S.H.and Leder,P.(1979) Cell 17 725-735

Hamer,D.H.and Leder,P.(1979a) Cell 17 737-747

Hamer,D.H.and Leder,P.(1979b) Cell 18 1299-1302

Hanahan,D.and Meselson,M.(1980) Gene 10 63-67

Hanns,S.,Holsclaw,D.S.and Shin,H.S.(1974) Lancet 2 520

Harpold,M.M.,Dobner,P.R.,Evans,R.M.and Bancroft,F.C.(1978) Nuc.Acids Res. 5 2039-2053

Hastie,N.D.and Bishop,J.O.(1976) Cell 9 761-774

Haynes,J.R.,Rosteck,P.and Lingrel,J.B.(1981) In press

Herod,H.G.,Amos,D.B.and Spock,A.(1979) J.Ped. 94 676

Heyneker,H.L.,Shine,J.,Goodman,H.M.,Boyer,H.W.,Rosenberg,J., Dickerson,R.E.,Narang,S.A.,Itakura,K.,Lin,S-Y.and Riggs,A.D. (1976) Nature 263 748-752

Hicks,J.,Strathern,J.N.and Klar,A.J.S.(1980)Nature 282 478-483

Hirsch,F.W.,Nall,K.N.,Spohn,W.H.and Busch,H.(1978) Proc.Natl.Acad.Sci.U.S.A. 757 1736-1739

Hodson,M.E.,Beldon,I.,Towler,R.,Bamber,M.and Baten,J.C.(1980) Brit.J.Dis.Chest 74 419

Hoeijmakers,J.H.J.,Frasch,A.C.C.,Bernards,A.,Borst,P. and Cross,G.A.M.(1980) Nature 284 78-80

Hofer,E.and Darnell,J.E.(1981) Cell 23 585-593

Honda,B.M.and Roeder,R.G.(1980) Cell 22 119-126

Hosli,P.,Erickson,R.P.and Vogt,E.(1976) Biochem.Biophys.Res. Comm. 73 209-216

Hosli,P.and Vogt,E.(1977) Biochem.Biophys.Res.Comm. 79 741

Hosli,P.,Kollberg,H.and Vogt,E.(1978) Acta Pediatr.Scand. 67 617-620

Houck,C.M.,Rinehart,F.P.and Schmid,C.W.(1979) J.Mol.Biol. 132 289-306

Houghton,M.,Jackson,I.J.,Porter,A.G.,Doel,S.M.,Catlin,G.H., Barber,C.and Carey,N.H.(1981) Nuc.Acids Res. 9 247

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

Hu,S-L.and Manley,J.L.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 820-824

Huez,G.,Bruck,C.and Cleuter,Y.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 909-911

Humphries,P.,Cochet,M.,Krust,A.,Gerlinger,P.,Kourilsky,P.and Chambon,P.(1977) Nuc.Acids Res. 4 2389-2406

Hunter,T.and Garrels,J.I.(1977) Cell 12 1097-1108

Hutchinson,C.A.,Phillips,S.,Edgell,M.H.,Gillam,S.,Jahnke,P. and Smith,M.(1978) J.Biol.Chem. 253 6551-6560

Iatrou,K.,Tsitilou,S.G.,Goldsmith,M.R.and Kafatos,F.C.(1980) Cell 20 659-669

Jackson,D.A.,Symons,R.H.and Berg,P.(1972) Proc.Natl.Acad.Sci. U.S.A. 69 2904-2909

Jacobs,H.and Birnie,G.(1980) Nuc.Acids Res. 8 3087-3103

Jagodzinski,L.L.,Castro,C.E.,Sherod,P.,Lee,D.and Sevall,J.S.(1979) J.Biol.Chem. 254 3038-3044

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

Jelinek,W.R.(1977) J.Mol.Biol. 115 591-600

Jelinek,W.R.,Toomey,T.P.,Leinwand,L.,Duncan,C.H.,Biro,P.A., Choudary,P.V.,Weisman,S.M.,Rubin,C.M.,Houck,C.M.,Deininger,P.L. and Schmid,C.W.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 1398-1402

Jelinek,W.and Leinwand,L.(1978) Cell 15 205-214

Jelinek,W.R. and Darnell J.E.(1972) Proc.Natl.Acad.Sci.U.S.A. 69 2537-2541

Johnson,J,D.,Ogden,R.,Johnson,P.,Abelson,J.,Dembeck,P.and Itakura,K.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 2564-2568

Kacian,D.L.,Spiegelman,S.,Banks,A.,Terada,M.,Dow,L.and Marks,P.A.(1972) Nature New Biol 235 167-169

Kacian,D.L.and Myers,J.C.(1976) Proc.Natl.Acad.Sci.U.S.A. 73 2191-2195

Kamalay,J.C.and Goldberg,R.B.(1980) Cell 19 935-946

Kan,Y.W.and Dozy,A.M.(1978) Lancet 910-912

Kan,Y.W.,Lee,K.Y.,Furbetta,M.,Anguis,A.and Cao,A.(1980) New Engl.J.Med. 302 185-188

Kataoka,T.,Miyata,T.and Honjo,T.(1981) Cell 23 357-368

Kataoka,T.,Kawakami,T.,Takahashi,N.and Honjo,T.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 919-923

Katz,S.and Ansah,T.A.(1980) Clin.Chim.Acta 100 245

Kaufman,R.E.,Kretschmer,P.J.,Adams,J.W.,Coon,H.C., Anderson,W.F. and Nienhuis,A.W.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 4229-4233

Kay,R.M.,Harris,R.,Patient,R.K.and Williams,J.G.(1980) Nuc.Acids Res. 12 2691-2708

Keshet,E.,Rosner,A.,Bernstein,Y.,Gorecki,M.and Aviv,H.(1981) Nuc.Acids Res. 9 19-30

Kindle,K.L.and Firtel,R.A.(1979) Nuc.Acids Res. 6 2403

King,C.R.,Udell,D.S.and Deeley,R.G.(1979b) J.Biol.Chem. 254 6781-6787

King,D.N.,Walsh,M.P.,Heeley,A.F.and Kuzemo,J.A.(1979a) Lancet 1217

Kinniburgh,A.J.and Ross,J.(1979) Cell 17 915-921

Kiper,M.(1979) Nature 278 279-280

Klein,W.H.,Murphy,W.,Attardi,G.,Britten,R.J.and Davidson,E.H. (1974) Proc.Natl.Acad.Sci.U.S.A. 71 1785-1789

Klenow,H.,Overjaard-Hansen,K. and Palkar,S.A.(1971) Eur.J.Biochem. 22 371-381

Knapp,G.,Ogden,R.C.,Peebles,C.L.and Abelson,J.(1979) Cell 18 37-45

Knudson,A.G.,Wayne,L.and Hallett,W.Y.(1967) Am.J.Hum.Genet. 19 388

Kominami(1981) Nuc.Acids Res. 9 3219

Kozak,M.(1980) Cell 22 7-8

Kraus,I.,Antonowicz,I.,Harnish,S.,Lazarus,H.and Schwachman,H. (1971) Ped. 47 1010

Kressman,A.,Hofstetter,H.,Dicapua,E.,Grosschedl,R.and Birnstiel,M.L.(1979) Nuc.Acids Res. 7 1749-1763

Kulczyski,L.and Schauf,V.(1974) Amer.J.Dis.Child. 127 64-67

Kuroiwa,A.and Natori,S.(1979) Nuc.Acids Res. 7 751-764

Kurtz,D.T.(1981) Nature 291 629-631

Kurtz,D.T.and Nicodemus,C.F.(1981) Gene 13 145-152

Kurz,J.B.(1979) Science 206 1317-1319

Kushner,S.R. in Boyer,H.B.and Nicosia,S.(Eds.) Genetic Engineering,Elsevier 1978 pp.17-23

Larsen,G.L.,Barron,R.J.,Landay,R.,Cotton,E.K.,Gonzalez,M.A.and Brooks,J.G.(1980) Amer.J.Dis.Child. 134 1143-1148

Lasky,L.,Lev,Z.,Xin,J-H.,Britten,R.J.and Davidson,E.H.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 5317-5321

Lawn,R.M.,Adelman,J.,Francke,A.E.,Houck,C.M.,Gross,M.,Najarian, R.and Goeddel,D.V.(1981) Nuc.Acids Res. 9 1045

Lee,A.S.,Thomas,T.L.,Lev,Z.,Britten,R.J.and Davidson,E.H.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 3259-3263

Leis,J.P.and Hurwitz,J.(1972) Proc.Natl.Acad.Sci.U.S.A. 69 2331-2335

LeMeur,M.,Glanville,N.,Mandel,J.L.,Gerlinger,P.,Palmiter,R. and Chambon,P.(1981) Cell 23 561-571

Lennox,E.S.(1955) Virol. 1 190-206

Lerner,M.R.,Boyle,J.A.,Mount,S.M.,Wollin,S.L.and Steitz,J.A. (1980) Nature 283 220-224

Lewis,J.A.,Kurtz,D.T.and Melera,P.W.(1981) Nuc.Acids Res. 9 1311-1322

Lieberman,J.,Costea,N.,Yakulis,V.and Kaneshiro,W.(1979) Clin.Res. 27 507A

Liebscher,D-H.,Coutelle,C.,Rapoport,T.A.,Hahn,V.,Rosenthal,S., Prehn,S.and Williamson,R.(1980) Gene 9 233-246

Liedtke,C.,Boat,T.F.,Stern,R.C.and Dearborn,D.G.(1981) New Engl.J.Med. 304 975

Lippe,B.M.,Kaplan,S.A.,Neufeld,N.D.,Smith,A.and Scott,M.(1980) Ped. 65 1018

Little,P.F.R.,Flavell,R.A.,Kooter,J.M.,Annison,G.and Williamson,R.(1979) Nature 27-231

Littlefield,J.W.(1981) New Engl.J.Med. 304 44-45

Lobban,P.E.and Kaiser,A.D.(1973) J.Mol.Biol. 78 453-471

Lobeck,C.C.and Huebner,D.D.(1962) Ped. 30 172

Lomedico,P.,Rosenthal,N.,Efstratiadis,A.,Gilbert,W.,Kolodnes,R. and Tizard,R. Cell 18 545-558

Luisot,P.,Leviat,C.and Gilly,R. Clin.Chim.Acta 48 373

Maki,R.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 2138-2142

Mandel,M.and Higa,A.(1970) J.Mol.Biol. 53 159-162

Mangiarotti,G.,Chung,S.,Zuker,C.and Lodish,H.F.(1981) Nuc.Acids Res. 9 947-963

Mangos,J.A.and Mc.Sherry,N.R.(1968) Ped.Res.2 378-384

Mangos,J.A.(1973) Tex.Rep.Biol.Med. 31 651-663

Mangos,J.A.,Mc.Sherry,N.R.and Benke,P.J.(1967) Ped.Res. 1 436

Maniatis,T.,Hardison,R.C.,Lacy,E.,Lauer,J.,O'Connell,C.,Quon,D. Sim,G.K. and Efstratiadis,A.(1978) Cell 15 687-701

Manley,J.L.,Fire,A.,Cano,A.,Sharp,P.A.and Gefter,M.L.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 3855-3859

Manson,J.C.and Brock,D.J.H.(1980) Lancet 330

Mantei,N.,Boll,W.and Weissman,C.(1979) Nature 281 40-46

Marie,J.,Simon,M-P.,Dreyfus,J-C.and Kahn,A.(1981) Nature 292 70

Marino,J.T.(1980) Amer.J.Dis.Child. 134 1179

Marino,J.T.(1981) J.Ped. 98 511

Matalon,R.and Dorfman,A.(1968) Biochem.Biophys.Res.Comm. 33 954-958

Matthews,L.W.,Doershuk,C.F.and Spector,S.(1967) Ped. 39 176

Maxwell,L.H.,Maxwell,F.and Hahn,W.E.(1980) Nuc.Acids Res. 8 5875

Mayo,B.J.,Klebe,R.J.,Lankford,B.J.,Morris,N.R.,Barnet,D.R. and Bowman,B.H.(1976) Tex.Rep.Biol.Med. 34 107-112

Mayo,B.J.,Klebe,R.J.,Barnett,D.R.,Lankford,B.J.and Bowman,B.H. (1980) Clin.Genet. 18 379-386

Max,E.E.,Seidman,J.G.and Leder,P.(1979) Proc.Natl.Acad.Sci. U.S.A 76 3450

MacLeod,A.R.(1981) Nuc.Acids Res. 9 2675-2689

Mc.Neely,M.C.,Awasthi,Y.,Barnett,D.R.,Srivastna,S.and Bowman,B.H.(1981) In press.

Mears,J.G.,Ramirez,F.,Leibowitz,D.and Bank,A.(1978) Cell 15 15-23

Merritt,A.D.,Hanna,B.L.,Todd,C.W.and Myers,T.L.:J.Lab.Clin.Med. 60 998-999

Mertz,J.E.and Gurdon,J.B.(1977) Proc.Natl.Acad.Sci.U.S.A.(1977) Proc.Natl.Acad.Sci.U.S.A. 74 1502-1505

Meyuhas,O.and Perry,R.P.(1979) Cell 16 139-148

Monahan,J.J.,Harris,S.E.and O'Malley,B.W.(1976) J.Biol.Chem. 251 3738-3748

Moore,K.W.,Rogers,J.,Hunkapiller,T.,Early,P.,Nottenburg,C., Weissman,I.,Bazin,H.,Wall,R.and Hood,L.(1981) Proc.Natl.Acad. Sci.U.S.A. 78 1800-1804

Nadler,H.L.and Walsh,M.M.J.(1980a) Lancet 96-97

Nadler,H.L.and Walsh,M.M.J.(1980b) Ped 66 690-692

Nasif,E.G.,Younoszai,M.K.,Weinberger,M.M.and Nasif,C.M.(1981) J.Ped. 98 320-322

Nagata,S.,Mantei,N.and Weissman,C.(1980) Nature 287 401-408

Naveh-Many,T.and Cedar,H.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 4246-4250

Nelson,T.and Brutlag,D.(1979) in Wu,R.(ed.) Methods in Enzymol. 68 Academic Press,NY.pp.41-50

Ng,S.Y.,Parker,C.S.and Roeder,R.G.(1979) Proc.Natl.Acad.Sci. U.S.A. 76 136-140

Nikaido,T.,Nakai,S.and Honjo,T.(1981) Nature 292 845-848

Nordstrom,O.V.,Blomfield,J.and Brown,J.M.(1980) in proc. of 8th. CF congress,Toronto

Norgard,M.V.,Keem,K.and Monahan,J.J.(1978) Gene 3 279-292

Norgard,M.V.,Tocci,M.J.and Monahan,J.J.(1980) J.Biol.Chem. 255 7665-7672

Nunberg,J.H.,Kaufman,R.J.,Chang,A.C.Y.,Cohen,S.N.and Schimke,R.T.(1980) Cell 19 355-364

Ohno,S.(1971) Nature 234 134-137

Old,J.,Clegg,J.B.,Weatherall,D.J.,Ottolenghi,S.,Comi,P., Giglioni,B.,Mitchell,J.,Tolstoshev,P.and Williamson,R.(1976) Cell 8 13-18

Old,R.W.and Primrose,S.B.(1980) in Principles of Gene Manipulation. Blackwell Scientific Publications

Ostrow,R.S.,Woods,W.G.,Vosika,G.J.and Faras,A.J.(1979) Biochem.Biophys.Acta 562 92-102

Otsuka,A.(1981) Gene 13 339_346

Palmiter,R.D.,Moore,P.B.,Mulvihil,E.R.and Emtage,S.(1976) Cell 8 557-572

Parnes,J.R.,Velan,B.,Felsenfeld,A.,Ramanathan,L.,Ferrini,U., Appela,E.and Seidman,J.G.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 2253-2257

Parnes,J.R.,Velan,B.,Felsenfeld,A.,Ramanathan,L.,Ferrini,U., Appela,E.and Seidman,J.G.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 2253-2257

Paterson,B.M.and Bishop,J.O.(1977) Cell 12 751-765

Pearson,W.R.,Wu,J-R.and Bonner,J.(1978) Biochemistry 17 51-59

Pelham,H.R.B.and Brown,D.D.(1980) Proc.Natl.Acad.Sci.U.S.A. 7 4170-4174

Pelham,H.R.B.,Wormington,W.M.and Brown,D.D.(1981) Proc.Natl. Acad.Sci.U.S.A. 78 1760-1764

Pellicer,A.,Wigler,M.,Axel,R.and Silverstein,S.(1978) Cell 14 133-141

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

Peterson,R.C.,Doering,J.L.and Brown,D.D.(1980) Cell 20 131-141

Pivetta,O.H.,Cassino,R.J.J.and Sordelli,D.O.(1980) Life Sci. 26 1349

Pivetta,O.H.,Vaccaro,M.I.and Maillie,A.J.(1981) Life Sci. 28 2207-2214

Pribnow,D.(1975) Proc.Natl.Acad.Sci.U.S.A. 72 784-788

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

Rabbitts,T.H.and Milstein,C.(1975) Eur.J.Biochem. 52 125-13

Radloff,R.,Bauer,W.and Vinograd,J.(1968) Proc.Natl.Acad.Sci. U.S.A. 57 1514-1521

Rao,G.J.S.and Nadler,H.L.(1975) Ped.Res. 9 739

Retzel,E.F.,Collett,M.S.and Faras,A.J.(1980) Biochemistry 19 513-518

Reymond,C.,Agliano,A.and Spohr,G.(1980) Proc.Natl.Acad.Sci. U.S.A. 77 4683

Richardson,C.C.(1965) Proc.Natl.Acad.Sci.U.S.A. 54 158-165

Rigby,P.W.J.,Dieckmann,M.,Rhodes,C.and Berg,P.(1977) J.Mol.Biol. 113 237-251

Robertson,H.D.,Dickson,E.and Jelinek,W.(1977) J.Mol.Biol.115 571-589

Robins,D.M.and Schimke,R.(1978) J.Biol.Chem. 253 8925-8934

Rogers,J.and Wall,R.(1980) Proc.Natl.Acad.Sci.U.S.A. 77 1877-1879

Rogers,J.,Early,P.,Carter,C.,Calame,K.,Bond,M.,Hood,L.and Wall,R.(1980b) Cell 20 303 312

Romeo,G.(1981) Ital.J.Ped. In press

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

Roop,D.R.,Kristo,P.,Stumph,W.E.,Tsai,M-J.and O'Malley,B.W. (1981) Cell 23 671-680

Rosbash,M.and Ford,P.J.(1974) J.Mol.Biol. 85 87-101

Roscher,A.,Schmid,A.J.,Hadorn,B.and Wiesmann,U.N.(1980) Ped.Res.14 261

Ross,J.,Aviv,H.,Scolnick,E.and Leder,P.(1972) Proc.Natl.Acad. Sci.U.S.A. 69 264-268

Rougeon,F.and Mach,B.(1976) Proc.Natl.Acad.Sci.U.S.A. 73 3418-3422

Royal,A.,Garapin,A.,Cami,B.,Perrin,F.,Mandel,J.L.,LeMeur,M., Bregegere,M.,Gannon,F.,LePennec,J.P.,Chambon,P.and Kourilsky,P. (1979) Nature 279 125

Roychoudhury,R.,Jay,E.and Wu,R.(1976) Nuc.Acids Res. 3 101-116

Ruderman,J.V.and Pardue,M.L.(1977) Dev.Biol. 60 48-68

Ryffel,G.U.,Wyler,T.,Muellener,D.B.and Weber,R.(1980) Cell 19 53-61

Sakano,H.,Huppi,K.,Heinrich,G.and Tonegawa,S.(1979a) Nature 280 288

Sakano,H.,Rogers,J.H.,Huppi,K.,Brack,C.,Traunecker,A.,Maki,R., Wall,R.and Tonegawa,S.(1979b) Nature 277 627

Sakonju,S.,Bogenhagen,D.F.and Brown,D.D.(1980) Cell 19 13-25

Sakonju,S.,Brown,D.D.,Engelke,D.,Ng,S.Y.,Shastry,B.S.and Roeder,R.G.(1981) Cell 23 665^669

Schaap,T.and Cohen,M.M.(1976) In CF:Projections into the future p291-302 Stratton Medical Book Corp.,N.York

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

Scheller,R.A.,Costantini,F.D.,Kozlowski,M.R.,Britten,R.J.and Davidson,E.H.(1978) Cell 15 189-203

Scheller,R.H.,Dickerson,R.E.,Boyer,H.W.,Riggs,A.D.and Itakura,K.(1977) Science 196 177-180

Schmid,C.W.and Deininger,P.L.(1975) Cell 6 345-358

Schmoyer,I.R.,Brooks,S.P.and Fischer,J.F.(1972) Life Sci. 11 1037

Schwachman,H.and Gahm,N.(1956) New Engl.J.Med. 255 999-1001

Schwachman,H.,Leubner,H.and Catzel,P.(1955) Adv.Ped. 7 249

Schwartz,R.H.(1966) Amer.J.Dis.Child. 3 408

Seeburg,P.H.,Shine,J.,Martial,J.A.,Baxter,J.D.and Goodman,H.M.(1977) Nature 270 486-500

Seidman,J.G.,Max,E.E.and Leder,P.(1979) Nature 280 370-375

Sgaramella,V.and Ehrlich,S.D.(1978) Eur.J.Biochem. 86 531-537

Sharp,P.A. Cell 23 643-646

Shen,C-K.and Maniatis,T.(1980) Cell 19 379-392

Shortle,D.and Nathans,D.(1978) Proc.Natl.Acad.Sci.U.S.A. 75 2170-2174

Sippel,A.E.,Land,H.,Lindenmaier,W.,Chi Nguyen-Hu,M.,Wurtz,T., Timmis,K.N.,Giesecke,K.and Schutz,G.(1978) Nuc.Acids Res. 5 3275-3294

Skoultchi,A.and Gross,P.R.(1973) Proc.Natl.Acad.Sci.U.S.A. 70 2840-2844

Sorensen,R.U.,Stern,R.C.,Chase,P.and Polmar,S.H.(1979) Infection and Immunol. 23 398-402

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

Spicer,S.S.,di Sant'Agnese,P.A.,Vincent,R.A.and Ulane,M.(1980) Exp.Mol.Path. 33 104

Spock,A.,Heick,H.M.C.,Cress,H.and Logan,W.S.(1967) Ped.Res. 1 173

Stalder,J.,Larsen,A.,Engel,J.D.,Dolan,M.,Groudine,M.and Weintraub,H.(1980a) Cell 20 451-460

Stalder,J.,Groudine,M.,Dodgson,J.B.,Engel,J.D.and Weintraub,H.(1980b) Cell 19 973-980

Steinberg,A.G.and Brown,D.C.(1960) Hum.Genet. 12 416-424

Steinmetz,M.,Hochtl,J.,Schnell,H.,Gebhard,W.and Zachau,H.G. (1980) Nuc.Acids Res. 8 1721-1729

Stern,R.C.,Boat,T.F.,Abramowsky,C.R.,Matthews,L.W.,Wood,R.E.and Doershuk,C.F.(1978) J.Amer.Med.Assn. 239 2676-2680

Strohman,R.C.,Moss,P.S.,Micou-Eastwood,J.,Przybyla,A.and Paterson,B.(1977) Cell 10 265-273

Sugino,A.,Goodman,H.M.,Heyneker,H.L.,Shine,J.,Boyer,H.W.and Cozzarelli,N.R.(1977) J.Biol.Chem. 252 3987-3994

Super,M.(1975) S.Afr.Med.J. 49 818

Tashima,M.,Calabretta,B.,Torelli,G.,Scofield,M.,Maizel,A.and Saunders,G.F.(1981) Proc.Natl.Acad.Sci.U.S.A, 78 1508-1512

Taylor,J.M.(1979) Ann.Rev.Biochem. 48 681

Telford,J.,Boseley,P.,Schaffner,W.and Birnstiel,M.(1977) Science 195 391-393

Temin,H.and Mizutani,S.(1970) Nature 226 1211-1213

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

Tsai,M.J.,Nordstrom,J.L.,Zimmer,W.and O'Malley,B.W.(1980) Cell 22 219-230

Tsai,S.Y.,Tsai,M.J.and O'Malley,B.W.(1981) Proc.Natl.Acad.Sci. U.S.A. 78 879-883

Tuan,D.,Biro,P.A.,deRiel,J.K.,Lazarus,H.and Forget,B.G.(1979) Nuc.Acids Res. 6 2519-2544

Twigg,A.J.and Sherratt,D.(1980) Nature 283 216-218

Tyler,B.and Adams,J.M.(1980) Nuc.Acids Res. 8 5579

Ullrich,A.,Shine,J.,Chirgwin,J.,Pictet,R.,Tischer,E.,Rutter,W.J and Goodman,H.M.(1977) Science 196 1313-1319

Valenzuela,P.,Quirogq,M.,Zaldivar,J.,Rutter,W.J.,Kirschner,M.W. and Cleveland,D.W.(1981) Nature 289 650

Van de Sande,J.H.,Kleppe,K.and Khorana,H.G.(1973) Biochemistry 12 5050-5055

Van der Ploeg,L.H.T.and Flavell,R.A.(1980) Cell 19 947-958

Verma,I.M.,Temple,G.F.,Fan,H.and Baltimore,D.(1973) In Viral Replication and Cancer Proc. 2nd.Duran-Reynab International

Symposium Barcelona

Villa-Komaroff,L.,Efstratiadis,A.,Broome,S.,Lomedico,P.,Tizard, R.,Naber,S.,Chick,W.L.and Gilbert,W.(1978) Proc.Natl.Acad.Sci. U.S.A. 75 3727-3731

Vitek,M.P.,Kreissman,S.G.and Gross,R.H.(1981) Nuc.Acids Res. 9 1191-1202

Vivell,O.,Jacobi,H.and Munchbach,K.(1963) Monatsschr.Kinderheildk 111 62

Wahli,W.,Dawid,I.,Wyler,T.,Weber,R.and Ryffel,G.U.(1980) Cell 20 107-117

Walsh-Platt,M.,Rao,G.J.S.,Simpson,S.and Nadler,H.L.(1978) Ped.Res. 12 129

Ward(1976) Tex.Rep.Biol.Med.

Wasylyk,B.and Chambon,P.(1981) Nuc.Acids Res. 9 1813

Weihofen,D.M.and Pringle,D.J.(1969) J.Amer.Diet.Assn. 54 206

Weil,P.A.,Luse,D.S.,Segall,J.and Roeder,R.G.(1979) Cell 18 469-484

Weiner,A.(1980) Cell 22 209-218

Weintraub,H.and Groudine,M.(1976) Science 193 848-856

Weintraub,H.,Larsen,A.and Groudine,M.(1981) Cell 24 333-344

Weisbrod,S.,Groudine,M.and Weintraub,H.(1980) Cell 19 289-301

Welch,D.W.and Roberys,R.M.(1975) Ped.Res. 9 698-703

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

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

Wickens,M.P.,Buell,G.N.and Schimke,R.T.(1978) J.Biol.Chem. 253 2483-2495

Wickens,M.P.and Laskey,R.A.(1981) in Genetic Engineering I (Williamson,R.ed.) pp103-167 Acad.Press.

Wiesmann,U.and Neufeld,E.F.(1970) J.Ped. 7 685-690

Wigler,M.,Sweet,R.,Sim,G.K.,Wold,B.,Pellicer,A.,Lacy,E., Maniatis,T.,Silverstein,S.and Axel,R.(1979) Cell 16 777-785

Willi,U.V.,Reddish,J.M.and Teele,R.L.(1980) Amer.J.Roent. 134 1005

Williams,J.G.,Hoffman,R.and Penman,S.(1975) Cell 6 197-206

Williams,J.G.and Lloyd,M.M.(1979) J.Mol.Biol. 129 19-35

Williams,J.G.(1980) in Genetic Engineering I (Wiliamson,R.ed.)

Williams,R.E.O.(1976) Report on the working party on the practice of genetic manipulation,Cmnd.6600,HMSO,London

Wilson,G.B.and Fudenberg,H.H.(1977) Nature 266 463

Wilson,G.B.and Fudenberg,H.H.(1976) Ped.Res. 10 87

Wilson,G.B.,Jahn,T.L.and Fonseca,J.R.(1973) Clin.Chim.Acta 49 79

Wold,B.J.,Klein,W.H.,Hough-Evans,B.R.,Britten,R.J.and Davidson,E.H.(1978) Cell 14 941-950

Woodhead,J.L.and Malcolm,A.D.B.(1980) Nuc.Acids Res. 8 389-402

Woods,D.E.,Crampton,J.M.,Clarke,B.E.and Williamson,R.(1980) Nuc.Acids Res. 8 5157-5168

Woolford,J.L.and Rosbash,M.(1979) Nuc.Acids Res. 6 2483-2497

Wormington,W.M.,Bogenhagen,D.F.,Jordan,E.and Brown,D.D.(1981) Cell 24 809-817

Wozney,J.,Hanahan,D.,Morimoto,R.,Boedtker,H.and Doty,P.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 712-716

Wright,S.W.and Morton,N.E.(1968) Amer.J.Hum.Genet.20 157-169

Wu,C.and Gilbert,W.(1981) Proc.Natl.Acad.Sci.U.S.A. 78 1577-1580

Wu,G.(1978) Proc.Natl.Acad.Sci.U.S.A. 75 2175-2178

Young,B.D.,Birnie,G.D.and Paul,J.(1976) Biochemistry 15 2823-2329

Young,R.A.,Hagenbuchle,O.and Schibler,U.(1981) Cell 23 451-458

Zeelon,E.P.,Bothwell,A.L.M.,Kantor,F.and Schechter,I.(1981) Nuc.Acids Res. 9 3809