THE MOLECULAR BIOLOGY OF THE INTERFERON RESPONSE: THE 1-8/9-27 GENE FAMILY

Adrian H Brasnett

Thesis presented in partial fulfillment of the degree of
Doctor of Philosophy to the University of London

ICRF Laboratories
44 Lincoln's Inn Fields
London WC2A 3PX

Department of Biochemistry
Imperial College of Science
Technology and Medicine
University of London

September 1990

To my parents.

ABSTRACT

The molecular biology of the interferon response: the 1-8/9-27 gene family.

The treatment of mammalian cells with Type I interferons leads to a number of changes in the pattern of gene expression in the cells. At least ten or more genes have been identified whose level of expression is increased by 3 to more than 50 times by interferon.

One gene activated produces a small RNA of about 800 nucleotides in length. Southern blot analysis of human DNA suggests there are about 20 or more copies of this gene, known as 1-8, present in the genome. Cosmids containing 1-8 sequences were analyzed by transfection into mouse cells and examination of the transfectants for the expression of human 1-8 against a background of cross hybridizing mouse 1-8 sequences. No evidence of expression of human 1-8 was found. An oligonucleotide probe which spans two exons of the related 9-27 gene was used to examine these cosmids and it was concluded that they contained processed 1-8 pseudogenes.

A genomic copy of the related 9-27 gene which is present in a single copy in the human genome, and a full length cDNA, were examined and the complete sequence of the gene and its 5' flanking region determined.

Analysis of the 5' flanking region indicated the presence of a consensus sequence found in the 5' regions of other genes activated by interferon.

Oligonucleotide probes corresponding to this consensus sequence were used to seek to identify any unidentified genes which also have such a sequence and are thus potentially interferon inducible.

ACKNOWLEDGEMENTS

There are any number of colleagues throughout the ICRF and in particular in the Molecular Biology Laboratory who I would like to thank for their help and support towards my research, and it would be impossible to even start to attempt to list them individually. Special thanks must however go to Katy Smith for her invaluable support and friendship both during and since my three years at the ICRF. I would also like to thank my present employers, the partners of J A Kemp and Co, for their moral and material support which has enabled me to complete this thesis. Finally, I am enormously indebted to George Stark who I would like to thank for providing me with the opportunity of working in his laboratory and for his supervision, enthusiasm and support.

CONTENTS

LIST OF FIGURES 10

1. INTRODUCTION. 12

A. Interferons. 13
i) The immune system.
ii) Interferons.
iii) Clinical use of interferon.
iv) Interferon inducible genes
v) Summary

B. Genes and Genomes. 26
i) Complexity of the genome.
ii) Globin genes as a model for gene families.
iii) Other gene families.

C. Oligonucleotides 30
i) Probes.
ii) Primers.

D. Introduction of cloned sequences into cells. 32
i) Transfection techniques.
ii) Study of gene expression using transfection experiments.

E. Objectives. 36

Supplement to the Introduction 37a

2. MATERIALS AND METHODS. 38

A. Cell culture. 39
i) Cell culture.
ii) Interferon preparations.

B. Transfection of cells. 40
i) Transfection proceedure.
ii) Harvest of cells for analysis.

C. CAT assays. 41

D. Bacterial cell culture and preparation. 42
i) Cells.
ii) Preparation of bacteria competent for transfection.

E. Molecular cloning and analysis of DNA. 43
i) General considerations.
ii) Transformation of competent bacteria by plasmid DNA.
iii) Small and medium scale preparation of plasmid DNA.
iv) Large scale preparation of plasmid DNA.
v) Preparation of bacteriophage lambda DNA.
vi) Sequencing DNA.
vii) Isolation of DNA fragments from agarose gels.
viii) Neutral agarose gel electrophoresis.
ix) Labelling DNA by random priming.
x) Kinase labelling oligonucleotides.
xi) Immobilisation of DNA to membranes.
xii) Hybridization of labelled probes to filters.

F. Isolation and analysis of RNA. 53
i) Isolation of total cellular RNA.
ii) Glyoxal agarose gel electrophoresis.

Appendix: Formulations of commonly used solutions. 56

3. RESULTS--CONSENSUS SEQUENCE EXPERIMENTS. 58

A. Introduction. 59

B. Transfection of MT//TK constructs. 59
i) Introduction.
ii) Transfection of the constructs into mammalian cells.
iii) Conclusions.

C. Consensus sequence oligonucletide experiments. 69
i) Introduction.
ii) Probing 6-16 and 1-8 cosmids.
iii) Conclusions.

4. RESULTS--ISOLATION AND ANALYSIS OF A 9-27 cDNA. 82

A. Introduction. 83

B. Screening and Sequencing. 83
i) Screening a lambda-gt 10 library.
ii) Ratio of 9-27 to 1-8 in the library.
ii) Sequencing the cDNAs.
iii) Confirming that the 5' end of clone 2 is from 9-27.
iv) Translation of the open reading frame.
v) Database searches.

5. RESULTS--1-8 GENOMIC COSMIDS. 104

A. Introduction. 105

B. Transfection experiments. 105

C. The cosmids contain processed pseudogenes. 112

i) Introduction
ii) An exon spanning probe

6. RESULTS.--9-27 GENOMIC WORK. 119

A. Introduction. 120

B. Screening for a genomic clone of 9-27. 120
i) A cosmid library.
ii) A lambda phage library.
iii) Conclusions.

C. Studies on a phage genomic clone of 9-27. 125
i) Introduction
ii) Lambda-g9-27 has undergone rearrangements during propagation.
iii) The 14kb insert of lambda-g9-27 could not be subcloned.
iv) Lambda-g9-27 contains 1-8-related sequences over a 5kb region.
v) The 0.5 kb BamH1 fragment contains the 9-27 3' end.
vi) Lambda-g9-27 contains two 1-8-like 3' ends.
vii) No 5' region of 9-27 was detected in the phage clone.
viii) Conclusions

D. Structure of a functional 9-27 genomic clone. 142
i) Introduction.
ii) 9-27Cos3 contains 3' and 5' 9-27 sequences.
iii) Two 1-8 genes are also within Cos3.
iv) The 9-27 gene is located between the two 1-8 genes.
v) The 8 kb Not1fragment could not be subcloned.
vi) A 2 kb PstI fragment contains a functional 9-27 gene.
vii) Cos 3 contains the same 3' end as lambda-g9-27.
viii) All remaining 9-27 mRNA sequences are contiguous.

ix) 9-27 has sequence homologies to other interferon inducible genes.
x) CAT assays with the 5' region.
xi) 9-27 has multiple start sites.
xii) Sequence homologies to other genes.
xiii) Searching the database.

7. DISCUSSION. 163

A. Overview. 164

B. The 9-27/1-8 gene family. 166
i) Genes and pseudogenes.
ii) Total number of 1-8 genes.

C. The 9-27 genomic clone. 170

D. Use of consensus sequence probes. 172

E. Conclusions 176

REFERENCES 178

LIST OF FIGURES

Figure 3.1 Homologies between the upstream regions of interferon inducible genes.

Figure 3.2 Fusions of the metallothionein-IIa gene to the 5' end of the HSV thymidine kinase gene.

Figure 3.3 Assay of mouse Ltk$^-$ cells for their interferon response.

Figure 3.4 Oligonucleotides used in experiments described in chapter 3, and their homologies to the interferon consensus sequence.

Figure 3.5 Experiment using oligonucleotide probe OCT 351 on a cosmid containing the 6-16 gene.

Figure 3.6 Experiment using oligonucleotide probe OCT 351 on cosmid 1-8.8.

Figure 3.7 Oligonucleotide probes used on a panel of plasmids and cosmids.

Figure 4.1 Recombinant lambda-gt10 clones containing 9-27 cDNAs.

Figure 4.2 Ratio of 9-27 to 1-8 in the lgt10 library.

Figure 4.3 Sequence comparisons between clone 2 from the lgt10 library and previous clones 7-8 and 2-1.

Figure 4.4 Probing HeLa cytoplasmic polyA$^+$ RNA with a 24mer probe of the 5' end of 9-27 cDNA.

Figure 4.5 Complete sequence of the longest 9-27 cDNA with a translation of its longest open reading frame.

Figure 4.6 Hydrophobicity profile of the putative 9—27 gene product.

Figure 4.7 Short homology between 9-27 and mouse EGF precursor.

Figure 5.1 Eco R1 digests of six 1-8 cosmids.

Figure 5.2 Northern blots of RNA from Ltk$^-$ cells transfected with 1-8 cosmids.

Figure 5.3	Experiment to determine the probable origin of 1-8 genes found in cosmids 2 - 10.
Figure 5.4	Result of exon-spanning oligonucleotide experiment.
Figure 6.1	Summary of the EMBL 3 phage containing 9—27 genomic sequences.
Figure 6.2	Alignment of 9-27 cDNA and genomic 9-27 sequence derived from the 0.7 kb Bam H1 fragment of lg9—27.
Figure 6.3	The 3' portion of the 1-8 sequence aligned with genomic sequence obtained from lg9—27.
Figure 6.4	Alignment of 1-8 cDNA sequence with sequence derived from lg9-27.
Figure 6.5	Experiment using oligonucleotide probes to 3' and 5' sequence of 9-27.
Figure 6.6	Experiment showing presence of two 1-8 genes in 9-27Cos3.
Figure 6.7	Map of 9-27Cos3.
Figure 6.8	Sequence of the 9-27 gene from its 5' flanking region through to the end of the first exon.
Figure 6.9	Homologies between sequences flanking human interferon inducible genes.
Figure 6.10	CAT gene expression under the control of a 230 bp region of the 9-27 promoter.
Figure 6.11	Homologies between the 9-27 gene and two other interferon inducible genes.

CHAPTER 1. INTRODUCTION

A. INTERFERONS.

i). The immune system.

The evolution of vertebrates would not have been possible without the concomitant development of a sophisticated immune system, capable of defending the body against infection and disease. This notion is reflected in the large number of genes already known to play a role in the defence of the body. These include not only the immunoglobulin genes produced by B-cells, the T-cell receptor genes, histocompatibility and complement genes, but also a number of cytokine genes. Cytokines are conveniently thought of as the hormones of the immune system in that they are responsible for triggering responses in cells, rather than acting or interacting directly upon invading pathogenic agents. Cytokines include the interleukins, tumour necrosis factor, colony stimulating factor and the interferons. The biology, history and uses of cytokines have been considered by Balkwill (1988) and Ruddon (1987).

ii).Interferons.

The interferons were the first members of the cytokine family to be observed when Isaacs and Lindenmann (1957) first reported a virus-induced protein produced by cells, capable of transferring resistance to viral infection to other cells. Subsequent research has identified three

subtypes of interferons, of different cellular origins. There are at least 15 alpha-interferon genes, which are expressed by leukocytes. In addition, there are also 6 related pseudogenes (reviewed by Burke and Shuttleworth, 1984). The α-interferons bind to the same receptor as the single β-interferon (fibroblast interferon) molecule, and the two types are referred to as Type I interferons. There are reports of an immunologically related "β-2" interferon (Friedman-Einat et al, 1982) which has growth inhibitory properties (Resnitzky et al, 1986). There is a single gamma-interferon gene (Gray et al, 1982), which binds to a different receptor and has less sequence homology to the α-interferons than β-interferon does.

The reason for the existence of several different alpha-interferon genes is not fully understood. However it has been found that the various subtypes of α-interferons have different binding affinities for their common receptor (Aguet et al, 1984), and that the binding affinities correlate with the specific activity of the interferons. Although the specific activity of interferon α-1 is much lower than that of α-2, mixtures of the two types have been found to behave synergistically (Orchansky et al, 1983). Since all the alpha-interferon subtypes bind to the same receptor, the synergism must result from a process occurring after binding to the receptor.

Besides being inducible by viral infection of cells, it is also know that α- and β- interferons are induced when

cells are treated with viral or synthetic double stranded RNA (dsRNA). Interferon gamma is induced by mitogenic or antigenic stimulation of the immune system. It has also been reported that some cells produce endogenous interferon, which may be involved in autocrine growth regulation and differentiation (Creasy et al, 1983; Yarden et al, 1984).

The interferons have been extensively reviewed by several authors including Lengyel (1982), Burke (1985), Clemens and McNurlan (1985) and McMahon (1985).

iii). Clinical use of interferon.

The discovery of the anti-growth effects of interferon aroused considerable interest in the possibility that interferon would prove a useful clinical tool in the fight against disease, especially cancer. Early hopes that interferon was a biological "magic bullet" produced inflated expectations of its properties, especially in the popular press. A main problem of early clinical research was the limited supply of interferon, which was produced from large scale culture of human leukocytes. The subsequent cloning and production of interferons using recombinant DNA technology allowed large scale clinical trials, using a larger range of dosages, to be undertaken. The results of the more extensive trials generally confirmed the observations of the earlier tests that interferon is potentially useful in the treatment of a

limited number of specific types of cancer. In a summary of clinical trials using interferons, Goldstein and Laszlo (1986) divided various types of malignancies into three categories: highly responsive, somewhat responsive and rarely or non-responsive.

The most well known responsive malignancy is hairy cell leukemia, which a number of studies have shown undergoes a complete or partial remission (CR+PR) in over 70% of patients treated with either partially purified leukocyte or recombinant interferons (Quesada et al, 1984; Gastl et al, 1985). Lymphomas show a median response rate (CR+PR) of about 25%, although this figure is higher when patients with favorable histology only are considered (Foon et al, 1984).

The "somewhat responsive" malignancies include melanomas and renal cell carcinomas, which have both shown a CR+PR of between 0 and 30% in a number of different clinical trials. A trial of three hundred and seventy one patients with renal cell carcinoma in Japan using both recombinant or lymphoblastoid interferon showed a CR+PR of 17.7% (Umeda and Nijima, 1986), which is comparable to the median CR+PR of the trials reviewed by Goldstein and Laszlo (1986). The Japanese study also examined the effect of interferon on the metastases in the various organs of the body and found that these secondary tumours only had an average response rate (CR+PR) of 10%. This is an analogous situation to the treatment of tumours by chemotherapy, where secondary

growths are often observed to be less responsive to treatment with the same agent than the original growth.

Cancers which do not respond well (if at all) to interferon include some of the most commonly occurring malignancies such as breast, lung and colon cancers. A clinical trial involving the treatment of twelve patients with small cell lung carcinoma using gamma-interferon failed to produce any remission of the disease (Newman et al, 1987). However, it is possible that interferon may still prove of some benefit in these common cancers when used in conjunction with other agents, for example cis-platinum (Fergusson et al, 1987).

Interferons have also been used to treat a variety of viral infections, both singly and in combination with other agents (see Strander, 1986). As with the treatment of cancers, the effectiveness of treatment varies. For example, interferon has proved effective in treating some patients with chronic Hepatitis B infection (Thomas and Scully, 1985).

A major problem associated with the clinical use of interferon is the numerous side effects that occur in patients. The induction of a high fever is commonly observed, while gastrointestinal symptoms are not uncommon. For these reasons, it seems unlikely that interferon will be widely used for the treatment of non life-threatening diseases.

iv). Interferon inducible genes.

In order to improve the clinical effectiveness of interferons and develop improved therapeutic strategies for their use with or without other agents, an understanding of how interferon exerts its effects at a cellular and molecular level is desirable.

The binding of interferon to a specific receptor on the cell surface is believed to lead to the production, activation or release of one or more "second messengers" which in turn activates or modulates the expression of the genes required to bring about the observed anti-growth, anti-viral and immunomodulatory effects of interferon. Receptor bound interferon is also taken up by the cells by receptor mediated endocytosis (Zoon et al, 1983; Branca et al, 1982). It is possible that some of the biological effects of interferon such as the transcriptional induction of a set or subset of genes may be activated directly by the interferon/interferon-receptor complex, or alternatively this internalization process may be connected with the down-regulation of the interferon receptor (Branca and Baglioni, 1982).

There are estimated to be fifteen to twenty different genes known to be inducible by interferon (Revel and Chebath, 1986), at least 12 of which are induced by α-, β- and γ- interferon, while the others are only or primarily induced by either the Class I interferons or γ-interferon. Genes known to be induced by interferon include 2-5 A

synthetase and the dsRNA protein kinase, (reviewed by Lengyel, 1982), the major histocompatibility antigens (Collins et al, 1984; Rosa and Fellous, 1984), guanine nucleotide binding proteins (Cheng et al, 1983), and the mouse Mx gene (Staeheli et al, 1986a) A number of cDNAs of interferon inducible genes have also been cloned (Friedman et al, 1984; Larner et al, 1984; Luster et al, 1985) and some of these cDNAs identified as known genes including metallothionein IIa, thymosin $\beta 4$ and a gene related to platelet factor-4.

Of the genes mentioned above, the role of the mouse Mx gene is the most clearly defined. The Mx locus was identified by genetic methods as determining the resistance of mice to infection with influenza virus (reviewed by Haller, 1981). The Mx^+ allele is responsible for the production of a 75 kd protein in response to type I but not gamma-interferon (Staeheli et al, 1984). The response to interferon is both rapid and large, with Mx RNA rising from undetectable levels to 0.1% of polyadenylated RNA within 5 hours of interferon treatment (Staeheli et al, 1986b). Although it is possible that other interferon inducible genes may be required to mediate a full response to influenza virus infection, Mx^- cells (which do not produce the 75 kd protein), when transformed with a vector expressing Mx become resistant to viral infection, even in the absence of interferon treatment (Staeheli et al, 1986a). Therefore, even though the way in which the Mx

protein interacts with the influenza virus remains to be determined, the Mx gene is a good example of a single interferon inducible gene with a specific function.

The other interferon inducible genes may not have any specific functions in the way that the Mx gene has, but probably play a more general role in the interferon response. Cytotoxic T cells and helper T cells only recognize antigens in conjunction with MHC antigen expression (Hood et al, 1985) and therefore the enhanced expression of MHC molecules on a cell surface in response to interferon may facilitate the cellular immune response. It is likely that given the widespread importance of tissue antigens, their induction by interferon is significant for more than one single reason.

Thymosin $\beta 4$ (reviewed by Horecker and Morgan, 1985) is believed to be involved in the maturation of T lymphocytes. The enhancement of thymosin $\beta 4$ activity would therefore appear to be desirable as part of an enhanced immune response, when the increased production of T cells is required.

A cDNA clone of a γ-interferon inducible gene, pIFNg-31 was found to have significant homology to both platelet factor 4 and β-thromboglobulin, which are both important mediators of inflammation and wound healing (Luster et al, 1984). The induction of a gene with similar properties at a time when the body is likely to be launching a general immune response also appears desirable.

The reasons for the induction of the metallothionein IIa (MT-IIa) gene in response to interferon is less clear. Metallothioneins are cysteine-rich metal-binding proteins which play a major role in the metabolism and the detoxification of the essential metals zinc and copper, and also other metals including cadmium and mercury (Vasak and Kaji, 1983). Besides interferon, metals and glucocorticoid hormones are also inducers of metallothionein (Karin et al, 1980, 1981). The metallothionein genes have separate 5' flanking regions responsible for the metal and hormone responses (Karin et al, 1984a; Stuart et al, 1984), and the MT-IIa also has a more distal 5' sequence which shares homology to a sequence found 5' to three HLA genes and is believed to play a role in mediating the interferon response (Friedman and Stark, 1985, also see Discussion). Therefore, metallothioneins are induced under a number of circumstances, and probably serve different functions depending upon their location and the reason for their induction. The role that metallothionein plays in the interferon response can only be speculated upon at present. One conjecture is that by raising the levels of metallothionein in virally infected cells, metal ions required by enzymes for the synthesis of new virus particles are sequestered and therefore viral growth is inhibited. It is possible that the sequestering of metal ions may also inhibit the normal growth of cells, and Leone et al (1985) consider that a similar explanation may

contribute to the deficiencies in copper-dependent enzymes seen in the cells of patients with Menke's disease. This condition, which manifests itself in a wide range of clinical symptoms such as skeletal abnormalities and neurological defects, appears to result from a continuous over-expression of metallothionein genes, which is due to the induction of these genes by unusually low concentrations of copper. The result of this is an intracellular build up of metallothionein-copper complexes in the cells (and the concomitant lack of copper-dependent enzymes) which leads to the clinical conditions observed. Therefore, although the continuous over-expression of metallothionein is very detrimental to cells, temporarily enhanced levels may prove of benefit. Although the relatively modest induction of metallothionein by interferon (3-5 fold, (Friedman et al, 1984)) makes the sequestering of metal ions unlikely to account for more than a proportion of the anti-growth and anti-viral effects of interferon, these responses may well be the cumulation of several subtle metabolic and physiological changes such as an increase in metallothionein levels, rather than the result of a single Mx-type gene which produces a specific response.

The 2',5'-A synthetase and dsRNA dependent protein kinase system has been extensively studied and its role in the inhibition of protein synthesis characterised (reviewed by Clemens and McNurlan, 1985). Interferon induces the

production of the synthetase and kinase, which require the presence of dsRNA for their activation. Large amounts of dsRNA can be generated during the replicative cycle of several types of DNA and RNA viruses. Activation of the synthetase results in the production of oligomers of 2' to 5' linked adenylates, of the general formula pppA2'p5'(A2'p5')nA (p = phosphate moiety, n is typically 2 - 4), generally abbreviated to 2',5'-A. The 2',5'-A molecules activate a latent ribonuclease, RNase L, which degrades mRNA and rRNA. The protein kinase molecule is also activated by dsRNA, and in its active form phosphorylates the alpha subunit of initiation factor eIF-2, which results in the inhibition of the factor in catalyzing the initiation of protein synthesis.

The identity and function of other interferon inducible genes, presently characterized as either cDNAs or proteins, remains to be determined. These include a 15 kd protein (human and bovine) induced by both γ-interferon and more strongly by Type I interferons (Korant et al, 1984; Blomstrom et al, 1986), two γ-interferon induced HeLa mRNAs of about 1600 and 2900 nucleotides (Larner et al, 1984), the murine Ly-6 antigen expressed by T-cells (Dumont et al, 1986) and the human 1-8, 9-27 and 6-16 genes (Friedman et al, 1984).

The 6-16 gene encodes a mRNA of about 1000 nucleotides, which is strongly induced by Type I, but not γ-interferons in a variety of cell lines (Kelly et al, 1985). A cosmid

containing human genomic 6-16 was transfected into mouse Ltk^- cells and the resulting transfectants showed mouse-interferon-inducible human 6-16 expression. This indicated that not only did the cosmid contain all (or most) of the 6-16 gene, including its regulatory elements, but also that the signals responsible for the induction of at least one interferon-inducible gene are conserved between man and mouse (Porter et al, 1988). The 6-16 gene was independently cloned by isolating polysomes using an antibody to an interferon inducible membrane protein (see Kelly et al, 1986) which was first identified by Burrone and Milstein (1982).

Hybridization of the 1-8 cDNA to Southern blots of human genomic DNA revealed between 10 and 30 bands which hybridized to the probe (Friedman et al, 1984), indicating that 1-8 is a member of a multigene family. There is only a single 9-27 gene, though two overlapping cDNA clones suggested that sequences 5' to the 220 bp of the original 9-27 cDNA indicated that a region of 9-27 shares strong homology to 1-8 (McMahon, 1985). The 1-8 and 9-27 genes are both strongly inducible by all three types of interferon, and produce small mRNAs of about 800-900 nucleotides in length. The uninduced levels of these genes are generally very low and often undetectable, although the levels of 1-8 often appear to rise significantly as cells become more confluent (M. McMahon, pers. comm., and own observations). After induction by interferon, the levels

of 1-8 mRNA in cells remains high even after the rate of transcription of the gene has declined, which may be due to post-transcriptional stabilization (Friedman et al, 1984).

v). Summary.

The wide number of different genes known to be induced by interferon are a reflection of the wide range of responses that interferon produces in cells. It is possible that the binding of interferon to its cell-surface receptor activates more than one responsive pathway. Forti et al (1984) showed that although oxyphebutazone (a nonsteroidal anti-inflammatory agent) could be used to inhibit the anti-viral effects of interferon, the anti-growth properties of interferon were unaffected by the drug. McMahon (1985) studied the expression of a range of interferon inducible genes in a line of Daudi cells which are resistant to the anti-growth (but not anti-viral) effects of interferon. Although metallothionein, class I HLA, thymosin β4, β-2 microglobulin and 2',5'-A synthetase were still induced by interferon treatment of the cells, the 1-8, 9-27 and 6-16 genes were not. Moreover, the expression of c-myc, which is normally down-regulated in response to α- and β- interferons (Jonak and Knight, 1984; Kelly et al, 1985), was also unaffected.

These results strongly suggest that interferon mediates its effects through at least two distinct pathways after binding to its receptor. The use of mutant cell lines

which only partially respond to interferon or methods of inhibiting or over-expressing subsets of or individual interferon-inducible genes may prove useful in dissecting further the mechanism of the interferon response.

B. GENES AND GENOMES.

i). Complexity of the genome.

The control of gene expression has been a major area of interest in molecular and cellular biology in recent years. The development of increasingly sophisticated techniques with which to analyze processes of cell growth, regulation and development has allowed a greater understanding of how cells are controlled, and cast new light on the origins of and relationships between genes, both within and across species.

One theme that has emerged from the research which has taken place is one of an underlying unity of many of these processes. During the 1950s and 1960s, studies of cell biology by two types of techniques, protein electrophoresis and DNA reannealing revealed a vast complexity in the composition of cells. By counting the number of spots on two dimensional protein gels or dividing the amount of single and low copy number DNA by the size of an "average gene", it was estimated that the number of genes in a human or other mammalian cell is in the order of 10,000 to 100,000. From this jigsaw of many individual pieces,

clusters of genes which show common features have been gradually brought together. Initially these links were demonstrated between genes at a structural level in terms of size and function of polypeptides and subsequently at a molecular level in the homologies between the DNA sequences of different genes. One example of organization between a group of genes is found in the study of human haemoglobin genes which show many of the features which provide a model for the structure and function of other gene families.

ii). Globin genes as a model for gene families.

Normal human adult haemoglobin (HbA) consists of two alpha and two beta chain subunits. There is almost 50% homology between the two chains, which strongly implies that they share a common evolutionary origin. However, as the α- and β- genes are on separate chromosomes, it is apparent that the two subunit genes must have diverged many millions of years ago. This can also be inferred from the observation that the degree of homology of the β-chain between species is greater than that between the α- and β-chains within a single species. There are other α- and β-like genes, of more recent origin, found near the chromosomal location of the α- and β- genes. Some of these genes are expressed instead of the α- and β- genes during the early stages of development, to make embryonic and fetal forms of haemoglobin. During the development of the fetus and neonate, expression of the embryonic and fetal

forms of haemoglobin declines as the expression of HbA increases. Even from this descriptive outline of haemoglobin gene expression, there are many mechanisms of gene expression which may be inferred. The fact that haemoglobin genes are only expressed in erythropoietic cells demonstrates a form of control mechanism that must tightly repress the expression of haemoglobin genes in other cell types and/or activate their expression in blood cells. However, although all the haemoglobin genes must share some common feature to allow this cell-specific expression, there is also temporal control of gene expression to allow the appropriate form of haemoglobin to be expressed. Finally, there is also coordinated expression between the two subunit chain genes in order that they are expressed in stoichiometric proportions.

The globin genes have been extensively characterized, and they each show many of the characteristics of an archetypal gene, with "TATA" and "CAAT" boxes upstream to the transcription start site of the genes, in addition to specific DNA sequences which control the expression of individual globin genes. For example, the Aγ gene on the β-gene cluster, which is expressed during fetal development, has the sequence YYTTGA (Y=pyrimidine) immediately upstream of its "CAAT" box in man, goat, rabbit and mouse. This sequence is not found in the adult β-genes and a mutation within this motif has been found in individuals with hereditary persistence of fetal

haemoglobin (HPFH), when the Aγ gene is expressed during adult life (Karlsson and Nienhuis, 1985, and references therein). Another more distal point mutation in the Aγ gene also results in HPFH (Giglioni et al, 1984), indicating that the interactions between regulatory factors and the DNA sequences of the promoters of globin genes are probably varied and complex. In addition there are further lines of evidence (reviewed by Karlsson and Nienhuis) which indicate that sequences downstream of the start of transcription may also play a role in controlling haemoglobin gene expression.

iii). Other gene families.

There are a large number of gene families known to be present in all types of living organisms. Many of them have been shown to have features of spatial, temporal and coordinately or differentially regulated expression similar to those found in the haemoglobin gene family. The sizes of the families range from those with a few members, for example α-tubulin (4 separate genes in chickens (Lopata et al, 1983)) and β-crystallins (3 genes in rat, 4 in humans (Aarts et al, 1987)) through to the immunoglobulin "supergene family" (Hood et al, 1985). Beyond this is the hypervariable "minisatellite" regions of DNA (Jeffreys et al, 1985) present in numerous copies per cell.

C. OLIGONUCLEOTIDES.

i). Probes.

The ability to synthesize and radioactively label short specific sequences of DNA for use as high specific activity single stranded probes has been of great benefit to the study of molecular biology (see for example Gait et al, 1982). One main area of use has been in isolating cDNA or genomic clones of genes whose protein sequence (or partial sequence) has been determined (Wood et al, 1984; Ullrich et al, 1984a; Jaye et al, 1983; and Anderson and Kingston, 1983). There are a number of theoretical and practical factors to consider when using oligonucleotide probes (reviewed by Lathe, 1985). In designing a probe for a gene whose protein sequence has been determined it is possible to use either a short (15 to 20 nucleotides) mixed probe, covering most or all possible codon usages, or a single long probe (30 to 70 nucleotides) based upon preferred codon usage, which may prove to be 80-90% homologous to the actual target sequence. The advantages and disadvantages of each approach are considered by Lathe (1985). Short oligonucleotide probes only require a single base-pair mismatch to their target sequence to be far less stable than a perfectly matched equivalent (Wallace et al, 1979), especially if the mismatch is within, rather than at the end of the probe. Therefore, if a short mixed probe is used which covers most, but not all of the possible codons, the target sequence may not be detected at all, due to just

a single base mismatch. On the other hand, since gene libraries are not simply random collections of nucleotides but instead contain some sequence motifs which occur in apparently unrelated coding regions (Lathe, 1985), and furthermore, many genes are members of multigene families, some long oligonucleotide probes may detect non-target genes, depending upon the exact hybridization conditions employed.

It is also important to consider the size of the library to be probed when considering the use of oligonucleotide probes. Obviously, a human cDNA library, representing approximately 10^7 base pairs of sequence is far less complex than a human genomic DNA library of over 10^9 bp, and therefore, shorter probes may be used with a higher expectation of success on the former. Hybridization conditions for oligonucleotide probes have been reviewed by Meinkoth and Wahl (1984); also see Materials and Methods and Discussion.

ii). Primers.

Oligonucleotides are also used as primers for dideoxy sequencing and primer extension experiments. Traditionally, dideoxy sequencing was accomplished using specific M13 primers. This often required time-consuming manipulation of small regions of cloned DNA fragments, especially in cases where regions of secondary structure of the DNA made sequencing through the region difficult unless

it was, for example, immediately adjacent to the primer, or in cases where a lack of suitable restriction enzyme recognition sites made the sequencing of overlapping regions of DNA impractical. The increasing availability of low cost, custom-made oligonucleotides for use as specific primers, designed to correspond to a known sequence of DNA adjacent to an unknown region has facilitated the accurate determination of DNA sequences. The use of specific primers avoids the need to divide the DNA being sequenced into short (200-400 bp) lengths which have them to be sub-cloned into an M13 vector. Instead, longer lengths of DNA may be subcloned and the first 200 or 300 nucleotides sequenced in the normal way, followed by the next 200 -300 nucleotides from a primer determined by the 3' end of the first region sequenced.

D. INTRODUCTION OF CLONED SEQUENCES INTO CELLS.

i). Transfection techniques.

The introduction of cloned DNA sequences into mammalian cells in culture is widely used for the study of gene expression and function. The most widely used technique is transfection by co-precipitation with calcium phosphate (Graham and Van Der Eb, 1973; Wigler *et al*, 1979), although other techniques including transfection with polybrene-DMSO (Kawai and Nishizawa, 1984), microinjection, electroporation and infection with recombinant viruses are

also available. Various modifications of the calcium phosphate precipitation technique have been proposed including the use of the antibiotic amphotericin B which was reported to enhance the efficiency of transfection of DNA into CHO and L-cells (Hidaka et al, 1985). However, since the efficiency of each technique appears to vary not only according to the particular protocol used but also according to the recipient cell line, such modifications may not be universally applicable. The high efficiency transfer of DNA into mammalian cells has been reviewed by Gorman (1985).

ii). Study of gene expression using transfection experiments.

The use of a variety of different vectors for the manipulation of cloned DNA sequences allows a range of questions to be addressed. By using heterologous promoters, genes may be under or over-expressed in a cell line in which they are normally expressed at high or low levels (e.g. oncogenes) or expressed in cell lines from which they are normally absent (e.g. the Mx gene in Mx^- cell lines). These approaches may prove useful in determining the role of a number of genes that are inducible by interferon. Conversely, by linking the promoter of a gene to a heterologous marker gene, the function of the promoter and regions within it may be studied. This technique is widely used in the study of

gene expression and has been used to identify promoter and enhancer elements, as well as the cellular elements which bind to these regions of DNA (reviewed by Dynan and Tjian, 1985). A variety of marker genes are available, including the herpes simplex virus thymidine kinase (tk) gene (McKnight, 1980) and the bacterial chloramphenicol acetyl transferase (CAT) gene (Gorman et al, 1982a).

The tk gene has been fused to the promoter regions of both the metallothionein I (Stuart et al, 1984) and the metallothionein II-a genes (Karin et al, 1984a,b) in order to characterize sequences within the promoters which are responsible for the induction of the native metallothionein genes by both metals and glucocorticoid hormones. Karin et al (1984a,b) made a series of deletions from the 5' region of the MT-IIa promoter which were stably transfected into Rat 2 cells and assayed by northern blot analysis. They identified a region of the promoter immediately 3' to nucleotide -50 (+1 = start of transcription) that conferred metal inducibility on the fused gene. However, a homologous sequence in the MT-I gene did not by itself render the MT-I/TK fusion inducible by metals (Stuart et al, 1984). However, the assay used for the MT-I/TK fusion was a transient expression assay following the microinjection of the DNA constructs into mouse oocytes. Subsequent work by Karin and Holtgreve (1984c) revealed that their "-50" construct was not inducible in a transient assay, and that a sequence element located further upstream

was also required under these conditions.

This difference between transient and stable transfection assays demonstrates that the results of these types of experiments should be treated with caution. Since the normal state of cellular genes is as part of the cellular genome, the analysis of gene expression in stably transfected cell lines is likely to be more meaningful than transient assays (Karin and Holtgreve, 1984c), although even with stable transfectants the actual site of integration within the genome will influence the level of gene expression. This is an important consideration when the efficiency of transfection is low, so that the resulting population of cells may only be descended from a small number of individual transformants, or when a single colony is isolated for examination.

Both the herpes virus tk gene and the CAT gene are useful markers because they are not normally present in mammalian cells and therefore the assays for the markers will not have any naturally occurring background levels of expression. The levels of tk expression are commonly measured by either northern blot analysis or S1 analysis, providing a direct measurement of mRNA levels. CAT activity can be rapidly measured by a protein activity assay (see Materials and Methods) and this is generally favoured for rapid determination of results and for ease of use when handling multiple samples. Although measuring CAT activity is a less direct method of measuring levels of

expression than an assay for mRNA, it has been shown that in general, CAT activity does correlate with steady state levels of CAT mRNA levels (Gorman et al, 1982b).

E. OBJECTIVES.

The aim of my research was to examine one or more of the interferon-inducible genes previously identified by Friedman et al (1984). Although the cDNA clones obtained by Friedman et al had been sequenced, these clones were in general much smaller than the size of their corresponding mRNAs as detected on northerns. Obtaining and sequencing longer cDNAs of these genes would provide information about potentially translated reading frames of these genes. Identification of the 5' end of a cDNA would lead to the identification of the 5' regulatory sequences of the corresponding gene. This information could be correlated to the observation of putative interferon consensus sequences (Friedman and Stark, 1985).

Of particular interest was the 1-8 gene family. One aim of my work was to seek to examine how individual members of the family were related, and whether or not all or just some of the family members were expressed and whether this expression was, for example, dependent on interferon to the same degree for different family members.

A further area of interest was to examine the putative interferon responsive consensus sequence and attempt to discover whether this did infact play a role in the

interferon response and whether it was present in any other interferon induced gene.

In the long term, identification and characterization of a set of related genes would permit a fuller understanding of the mechanisms by which these genes work. Common features in the DNA or amino acid sequences of related genes imply that these common features may serve a function in the regulation and expression of these genes. This offers a route to the identification of factors which regulate the expression of these genes and of factors (which may be the same or different) which mediate this regulation in response to interferon.

Supplement to the Introduction

The work presented in this thesis was completed in November 1987 and the above introduction was written shortly afterwards in early 1988. At that time, research on interferon inducible genes was largely confined to studies of the kinetics of their induction, the function of the gene products and their primary genomic structure. Little was known about the mechanisms by which expression of these genes was brought about after the binding of interferons to their receptors on the cell surface.

The work presented in Chapter 6.D below and other contemporary research led to the identification of a DNA sequence element which is termed the Interferon Stimulated Response Element (ISRE). Over the last two to three years, a number of protein factors which bind to the ISRE have been identified (reviewed by B.R.G.Williams, Eur.J.Biochem (1991), in press). A number of techniques have been used to identify these factors, although much initial characterization was performed using electrophoretic mobility shift assays using DNA fragments corresponding to native and mutant ISRE sequences (eg A.M.Ali Imam *et al*, Nucl. Acids Res., **18**, 6573-6580 (1990)).

The best characterized of these protein factors is factor E (Imam *et al*, ibid), or ISGF3, for *I*nterferon-*S*timulated *G*ene *F*actor 3 (D.Levy and J.E.Darnell, The New Biologist **2**, 923-928 (1990)). This factor requires the 5' margin and the central core of the ISRE for binding. The factor is activated rapidly after binding of interferon

alpha or gamma to their receptors, and requires no new protein synthesis. ISGF3 has been characterized as a multimeric protein complex comprising a 48-kD component termed ISGF3γ, the level of which can be increased in the cytoplasm of cells by treatment with IFNγ, and a second component, ISFG3α. This latter component is made up of at least three sub units. Treatment of cells with IFNα results in the activation of ISGF3α, which combines with ISGF3γ to form an active complex in the cytoplasm. The complex enters the nucleus, initiating transcription of a family of interferon-inducible genes.

Two other factors, ISGF1 (or C1-C2) and ISGF2 (or M) have also been identified although less well characterized. It appears that ISGF2 is a positive acting factor which plays a role in the maintenance of transcription after treatment of cells by interferon. Both ISGF1 and 2 appear to bind to the core region of the ISRE.

There are interesting parallels between the ISGF1-3 factors and the transcription factor NF-κB, which is activated in the cytoplasm in response to a number of agents by release from a complex with an inhibitory subunit, I B. There is however no evidence to date of an inhibitory factor involved in ISGF3 activation.

The characterization of the specific factors discussed above has fueled a wider debate about the mechanisms of signal transduction in transcriptional activation of genes (Levy and Darnell, ibid; I.Kerr and G.Stark, (1991), in press). There are a number of second messenger pathways which have been investigated in the

activation of transcription of genes. The binding of a polypeptide ligand to its receptor on a cell surface will often result in a number of global changes to the levels of intracellular messengers such as cAMP, diacylglycerol and calcium ions. These global changes in turn result in a cascade of phosphorylations (or dephosphorylations) of cytoplasmic proteins, and it has been widely conjectured that these changes are responsible for variations in the rate of transcription of specific sets of genes.

For example, L.M. Pfeffer et al (Proc. Natl. Acad. Sci., 87, 6537-6541 (1990)) have reported that activation of protein kinase C, and in particular, its β isoform, by α-interferon appears to be a necessary element in the antiproliferative and antiviral actions of α-interferon, while G.E. Hannigan and B.R.G. Williams (Science, 251, 204-207 (1991)) report that arachidonic acid metabolites also appear to be required in such actions.

However, the discovery of specific protein factors which bind to specific DNA sequences suggests that there may be other "direct" signalling pathways which provide for transcription of one or more target genes without a requirement for classical second messengers. It may be that the relatively few number of second messenger pathways, which are greatly outnumbered by specific receptor-ligand pairs, could be involved in the maintenance of one or more components of direct pathways. The likely future recognition of further specific transcription factors should allow the mechanisms of cell signalling and gene activation to be understood more fully.

CHAPTER 2. MATERIALS AND METHODS.

A.CELL CULTURE.

i). Cell Culture.

Human HeLa and mouse Ltk$^-$ cells were grown on sterile plastic tissue culture plates (Nunc) at 37°C in a humidified atmosphere of 90% air, 10% carbon dioxide. The cells were grown as monolayer cultures in Dulbecco's modified Eagle's medium (E4) supplemented with 10% foetal calf serum. When the cells reached confluency they were passaged by trypsinization. The monolayers were rinsed once with PBS and detached with a small volume of trypsin-versene (0.05% w/v trypsin, 0.02% w/v EDTA) in tris-saline. The trypsin was inactivated by the addition of complete medium and the cells seeded at appropriate dilutions.

ii). Interferon preparations.

Human cells were treated with Wellferon, a human lymphoblastoid interferon preparation which is a mixture of alpha interferons of specific activity if greater than 10^8 units per mg of protein, supplied by Wellcome (Allen et al, 1982). Mouse cells were treated with recombinant A/D hybrid interferon of specific activity of greater than $8x10^7$ units per mg of protein (Rehberg et al, 1982) or recombinant E.coli derived gamma interferon at $1-2x10^7$ units per mg of protein, supplied by Genentech Inc. Interferon preparations were diluted in E4 medium prior to

use to achieve appropriate concentrations.

B. TRANSFECTION OF CELLS.

i). Transfection procedure.

The protocol used was a modification of the technique of Graham and van der Eb (1973). Exponentially growing cells were plated at 10^6 cells per 90 mm tissue culture plate and grown overnight.

Plasmid DNA was prepared for transfection into cells by centrifugation through a caesium chloride gradient (see section E(iv) below) followed by dialysis against TE buffer and finally ethanol precipitation in clean sterile microfuge tubes of the required amount of DNA. Carrier salmon sperm DNA and/or plasmid DNA carrying a selectable marker gene was also added at this point if required. The DNA was resuspended in 40μl of TE pH7.4 and mixed with 440μl hepes buffered saline (HBS). 400μl of 260mM calcium chloride solution was added drop by drop whilst air was bubbled through the solution. The mixture was left for 30 minutes to allow the formation of a precipitate which was added to 7ml of freshly changed E4/10% FCS over the cell monolayer. After 8-10 hours of incubation the cells were treated with medium containing 25% v/v glycerol for 1 minute (Parker and Stark, 1979), rinsed twice in E4 and left in normal growth medium. For transient assays, interferon was added at this time if required and the cells

grown for a further 24 hours before being harvested. If the cells were being selected for the stable integration of plasmid DNA, they were allowed to grow for 24-48 hours in non-selective medium before being trypsinized and replated in selective medium. After 2-3 weeks colonies were visible and they were either pooled and cultured as mixed populations or individual colonies were picked using cloning rings. Transfected cells were maintained in selective medium.

Transfections for transient expression assays were performed using smaller 30mm diameter dishes, with the volumes of reagents scaled down accordingly.

ii). Harvest of cells for analysis.

Cells were harvested 24 hours after the addition of interferon in transient expression transfection experiments. The culture medium was removed, the cells rinsed with cold PBS followed by the addition of 100μl lysis buffer (see appendix) to each 30mm plate. After the cells had lysed, the plates were tilted to allow removal of the supernatant which was either analyzed immediately or stored at -20°C.

C. CAT ASSAYS.

Chloramphenicol acetyl transferase (CAT) was assayed non-chromatographically by the method of Sleigh (1986). [^{14}C]-acetyl coenzyme A (Amersham International, 50-60

mCi/mol) was mixed 1 in 10 with 0.5 M non radioactive acetyl coenzyme A. 30μl of the diluted coenzyme was added to 30μl of the supernatant being assayed and made up to 100μl to provide a final concentration of 1.6mM chloramphenicol and 50mM tris pH7.6. After incubation for 1 hour at 37°C, the reaction mixture was extracted twice against an equal volume of ethyl acetate. Unreacted acetyl coenzyme A remains in the aqueous phase, leaving the organic phase to be assayed by liquid scintillation counting to measure the incorporation of ^{14}C into chloramphenicol. This provides a measure of CAT activity of the supernatant.

D. BACTERIAL CELL CULTURE AND PREPARATION.

i). Cells.

The genotypes of *E.coli* strains HB101, C600, ED8767 used for growth of plasmid DNA are provided in Maniatis *et al*. Details of *E.coli* JM101, used for colour selection of recombinant bacteriophage M13 or pUC plasmids, are provided in the handbook supplied by Amersham International plc (1984).

All *E.coli* strains were grown in L-broth except for JM101 which was grown in 2x TY medium (see appendix).

ii). Preparation of bacteria competent for transformation.

A culture of bacterial cells were grown overnight from a single colony in 5ml of L-broth at 37°C. 1ml of the culture was added to 100ml of fresh L-broth and grown at 37°C to and O.D. at 500nm of 0.3, whereupon it was chilled in an ice water bath for 5 minutes. The culture was centrifuged in 25ml aliquots in pre-cooled 30ml Corex tubes at 7000 rpm for 5 minutes at 0°C. The pellet of cells obtained were resuspended in 12.5ml 50mM calcium chloride and left on ice for a further 20 minutes. The suspension was centrifuged and the cells resuspended in 0.5ml of 50mM $CaCl_2$ containing 15% v/v glycerol, and either used immediately or stored at -70°C in 200μl aliquots. When E.coli JM101 were prepared for transformation, a solution of 50mM $CaCl_2$/50mM $MgCl_2$ was used in place of the calcium chloride alone, and this increased the efficiency of transformation of this strain.

E. MOLECULAR CLONING AND ANALYSIS OF DNA.

i). General considerations.

Maniatis et al (1982) describe a wide range of protocols for the manipulation and analysis of DNA and RNA. The protocols used from this source which were modified in any way are described below.

Protocols for the sequencing of DNA by the Sanger chain termination method (Sanger, 1977) are thoroughly described

in the handbook published by Amersham International plc (1984) and were used with only minor modifications, also described below.

The catalogue published by New England Biolabs (1987) is one of many commercially available sources which provide details of the activities and specificities of DNA modifying enzymes, especially restriction endonucleases, used in the work described here. These catalogues also provide detailed restriction maps of commonly used cloning vectors such as pBR322 and pUC18/19.

The composition of the buffers and solutions used are described in the appendix to this chapter.

ii). Transformation of competent bacteria by plasmid DNA.

Aliquots of competent cells were transformed by plasmid DNA using standard protocols. Heat shock was at 42°C for 2 minutes, and cells were then incubated at 37°C for 30 minutes before being plated onto 9cm agar plates containing appropriate antibiotics for selection.

When plasmids allowing the colour selection of recombinants by the expression of β-galactosidase were used, the plates were first overlaid with a Pall Biodyne nylon membrane which was coated with 1mg of both IPTG and X-gal (isopropyl β-D-thiogalactoside and 5-bromo-4-chloro-3-indolyl β-D-galactoside).

The plates were incubated at 37°C overnight for the growth of colonies.

iii). Small and medium scale preparation of plasmid DNA.

A single colony of bacteria containing plasmid DNA was used to inoculate 2ml of L-broth containing a selective antibiotic to maintain plasmid growth. The medium was grown at 37°C for 7-14 hours with constant shaking.

1.5ml of the culture was used to prepare plasmid DNA using the procedure of Birnboim and Doly (1979), using 2-5mg/ml lysozyme (Boehringer) in the first step of the procedure.

The DNA prepared was extracted with phenol, phenol-chloroform (1:1 v/v) and chloroform, and ethanol precipitated. The precipitate was centrifuged and the pellet of DNA prepared resuspended in 50μl TE buffer containing 20-50μg/ml RNase A (Sigma). 5μl of the preparation was usually sufficient for analysis of the DNA by restriction enzyme digest and agarose gel electrophoresis.

The procedure was scaled up to produce larger quantities of DNA, for example 50-500μg. This can be done by seeding up to 400ml of L-broth with 1-5ml of a starter culture of bacteria and growing the medium overnight. The DNA prepared is pure enough for digestion by most of the common commercially available restriction enzymes and hence useful for preliminary restriction analysis and the preparation of linear fragments for subcloning or labelling by random priming (section ix).

iv). Large scale preparation of plasmid DNA.

To prepare 1mg or more of plasmid DNA, overnight cultures of 1-2 litres of bacteria were used to prepare a crude pellet of plasmid DNA using a scaled up version of the Birnboim and Doly protocol described above.

The pellet of DNA obtained was resuspended in about 6ml TE and 1.05 g of caesium chloride per ml TE was added, and the solution transferred into a 16x76mm polyallomer centrifuge tube (Beckman). 200μl ethidium bromide (10mg/ml) was added and the tubes filled with CsCl in TE at the same (1.05:1 w/v) concentration. The tubes were sealed and centrifuged at 55,000 rpm for 18 hours in a Beckman Ty65 ultracentrifuge rotor. The red band of supercoiled plasmid DNA in the middle of the tube was removed using a 21 gauge needle and the DNA prepared by dialysis against TE after butan-1-ol extraction of the ethidium bromide.

The DNA could be further concentrated by ethanol precipitation, although with most small plasmids (eg less than 5-6 kb) which grow at high copy numbers this was not usually necessary.

v). Preparation of bacteriophage lambda DNA.

An overnight culture of the appropriate bacterial strain was centrifuged into a pellet and resuspended in half its original volume in 10mM magnesium sulphate. 50,000 plaque forming units (pfu) of the phage were added to 100μl

aliquots of the bacteria, incubated for 15 minutes and plated in an agarose overlay on to pre-warmed 90mm agar plates. The plates were incubated overnight to allow the growth of a confluent lawn of plaques, which was overlaid with 3ml of lambda dilution (lambda-dil) buffer. The plates were left at 4°C for a few hours, the buffer drawn off into 15ml Corex tubes and the plates rinsed in a further 3ml of buffer which was also collected. The bacterial debris and loose pieces of agarose were pelleted by centrifugation. The supernatant was transferred to clean tubes and treated for 30 minutes at 37°C with 10μg/ml RNase A and 1μg/ml DNase I. Phage particles were precipitated by the addition of an equal volume of cold 20% polyethylene glycol (PEG) 6000 in 2M sodium chloride. A pellet of particles obtained by centrifugation was resuspended in 0.5ml TE, and SDS and EDTA added to final concentrations of 0.1% w/v and 50mM respectively. After a 15 minute incubation at 65°C to lyse the phage particles DNA was recovered by phenol extraction and ethanol precipitation.

The use of agarose rather than agar for growing the lawn of phage is important since DNA prepared from phage grown on agar was often found to be difficult to cut with restriction enzymes.

vi). Sequencing DNA.

All sequencing was performed using variations of the Sanger dideoxy nucleotide chain termination method. M13mp18 and mp19 were used as cloning vehicles.

Sequencing reactions were performed either using a standard 17mer primer or specific oligonucleotides from 18 to 24 nucleotides in length which were known to be complementary to a known region of the cloned DNA.

The annealing of oligonucleotides to M13 DNA was effected by heating the DNAs together at 90°C for 2 minutes and allowing them to cool to room temperature before initiating the reaction.

The chain termination reaction was performed at 20-42°C using α-35S dAPT αS at >600 Ci/mmol (Amersham) as the radioactive nucleotide, and the reaction analyzed on 6, 8 or 10% w/v polyacrylamide gels, depending upon the range of sequence information required.

Sequencing was also performed on denatured double stranded DNA plasmids using the method of Hattori and Sakaki (1986). In general, this produced less clear results than sequencing single stranded DNA, but was useful for rapid preliminary analysis of DNA.

A further variation of standard protocols used was the use of the bacteriophage T7 DNA polymerase ("Sequenaze"- USB Corp.) in place of the E.coli large fragment polymerase ("Klenow"). This produced significantly superior results, especially when the sequencing reactions were performed in

two stages, initially using α-35S dAPT αS as the sole source of dATP, and then using excess concentrations of unlabelled nucleotide/dd-nucleotide mixtures. This produces more uniform labelling of bands throughout the resulting sequencing gel, since each band is labelled to essentially the same extent.

vii). Isolation of DNA fragments from agarose gels.

The band of DNA required was excised from an agarose gel using a clean sterile scalpel, whilst being visualised under a long wave ultraviolet light. The fragment of agarose was transferred into a length of Spectropor dialysis tubing which had been soaked in the gel running buffer. The tubing was closed off at both ends using clips and placed in the gel tank buffer. 100 volts was passed across the tank for 1 hour, the current reversed for a few seconds and the buffer inside the tubing drawn off into a microfuge tube. The buffer was extracted once against butan-1-ol to remove ethidium bromide, ethanol precipitated and resuspended in a small volume of TE. A fraction of the recovered DNA was run on an agarose "mini gel" to estimate the amount recovered and check its purity.

vii). Neutral agarose gel electrophoresis.

Gels were prepared with an agarose concentration of from 0.6 to 1.8% in either tris-acetate (TAE) or tris-borate (TBE) buffers in the presence of 50μg/ml ethidium bromide.

Samples of DNA to be electrophoresed were mixed with 0.2 volumes of loading buffer and loaded into wells of the gels. Electrophoresis was performed for 1-12 hours at 20-100 volts. The gels were examined under ultraviolet light (280nm), and photographed at this wavelength using an orange filter and Polaroid 667 Professional film.

DNA markers of known sizes were used to determine the sizes of samples being analyzed. For the size range 200bp to 12kb the 1kb ladder from BRL was found to be especially useful, since this contains multimers of a 1018bp band present in approximately equal amounts by weight, plus pBR322 digested with Hinf II. For larger markers restriction digests of lambda DNA were used, most usually Hind III (to provide a 23kb band) or Kpn I (17 and 30kb).

ix). Labelling of DNA by random priming.

DNA was labelled using my own adaptation of the protocol described by Feinberg and Vogelstein (1983, 1984). 50-200ng of linear DNA (either a cut plasmid or an isolated fragment) in a volume of 2μl was added to 10μl of a hepes, tris and magnesium chloride (HTM) buffer and 2μl of random hexanucleotide primers (Pharmacia) at a stock concentration of 90 O.D. units per ml. This mixture was boiled for 2 minutes in a microfuge tube which was then centrifuged to bring down any condensation from the walls of the tube and effect cooling. 3μl of the deoxynucleotide dATP, dGTP and dTTP were added to a final concentration of 20μM along with

7μl (70μCi of α-^{32}P dCTP (3000 Ci/mmol, Amersham). The reaction was initiated by the addition of 1μl (5 units) of Klenow fragment E.coli DNA polymerase (Anglian Biotechnology) and the reaction mixture incubated for 3-16 hours at room temperature. DNA was separated from unincorporated nucleotides by passing the mixture through a Sephadex G-50 spun column. The specific activity of the labelled DNA obtained was usually in the range of $5x10^8$ to $1x10^9$ cpm/μg.

x). Kinase labelling oligonucleotides.

Oligonucleotides were labelled by the addition of a ^{32}P phosphate group to their 5' end. 50-100ng of oligonucleotide in a volume of 1μl was added to 1μl of 10x kinase buffer, 7μl (70μCi, 5000Ci/mmol, Amersham) [τ^{32}P]-ATP and 1μl (1-10 units) T4 DNA kinase (Anglian Biotechnology). The reaction mixture was incubated at 37°C for 1 hour and the incorporation of the label into DNA measured by the percentage of an aliquot of the reaction mixture precipitable in 10% w/v trichloroacetic acid. Measured incorporation varied from 5 to 50% and to some extent appeared to depend upon the particular oligonucleotide used. Separation of the labelled oligonucleotides was difficult and not usually necessary, although some enrichment of the ration of incorporated to unincorporated label could be obtained by ethanol precipitation of the reaction mixture using 0.1 volumes 4M

ammonium acetate and 3 volumes of ethanol.

xi). Immobilization of DNA to membranes.

DNA from agarose gels was transferred to either nitrocellulose (Millipore) or Biodyne nylon (Pall Ultrafine Filtration Corporation) using the standard method of Southern (1975). Transfer of DNA was usually allowed to proceed for 12-18 hours, unless two blots of one gel were required, in which case the first filter was only left on the gel for about 1-2 hours.

Filters were dried at room temperature and baked at 80°C under vacuum for an hour.

Plaque lifts from agar or agarose plates were made onto nylon filters, and the DNA from the phage particles which attached to the filters was denatured by rinsing the filters in a mixture of 1.5M NaCl and 0.5M NaOH for 2 minutes. The filters were neutralized for a further 2 minutes in a mixture of 1.5M NaCl and 0.5M tris-HCl pH7.5, and finally rinsed in 2xSSC, air dried and baked at 80°C under vacuum for 1 hour.

xii). Hybridization of labelled DNA probes to filters.

Baked filters with immobilized DNA were first wetted in either water or 0.2X SSC. Prehybridization and hybridization were preformed in disposable heat sealed bags placed on a rocking platform inside an incubator. Hybridization was performed at 42°C except in the case of

short oligonucleotides, which were hybridized at their T_m (calculated according to Meinkoth and Wahl, 1984) minus 15°C. The prehybridization buffer used comprised 50% v/v formamide, 5X Denhardts solution (Denhardt, 1966), 5X SSC, 50mM sodium phosphate, 0.2% SDS and 100μg/ml sheared herring sperm DNA. Filters were prehybridized for 0.5-16 hours. The hybridization buffer was the same as the prehybridization buffer, with the addition of denatured labelled probe, and the hybridization reaction was performed for from 0.5 to 16 hours.

After hybridization the filters were rinsed and washed in a mixture of SSC and 0.1% SDS. The concentration of the SSC varied according to the nature of the probe; oligonucleotide probes were generally washed in 6X SSC at up to their T_m minus 5°C, while random primed DNA was washed initially at 2X SSC at room temperature followed by further washes at 60°C in 0.2X SSC.

The filters were blotted dry on Whatmann 3MM paper, wrapped in cling film and exposed against Kodak XAR5 or Fuji X-Omat X-ray film using intensifying screens.

F.ISOLATION AND ANALYSIS OF RNA.

i). Isolation of total cellular RNA.

Cellular RNA from mammalian cells was isolated and prepared according to the method of Chirgwin *et al* (1979). Cells grown in 90mm diameter dishes in a monolayer culture

were drained of medium, washed once with cold PBS and lysed with 1.5ml of a buffer comprising 4M guanidinium thiocyanate, 25mM trisodium citrate pH7.0, 0.5% v/v N-lauryl sarcosine and 100 mM 2-mercaptoethanol. The lysed cells were scraped off the dish, a further 0.5 ml of buffer added, the plates rescraped and the total lysate passed 10-12 times through a 21 gauge needle to shear high molecular weight DNA.

The extract was layered onto a cushion of 2ml 5.7M CsCl (Koch-Light) in 0.1M EDTA pH7.0 in a polyallomer tube and centrifuged at 35,000 rpm at 20°C in a Beckman SW55 swing-out rotor for 17-20 hours. After centrifugation, the RNA is pelleted at the bottom of the tubes, and the supernatant containing DNA and cellular debris is carefully discarded. The pellet of RNA was dissolved in 0.3M ammonium acetate plus 10mM EDTA and precipitated by the addition of 2.5 volumes of ethanol. The RNA was repelleted by centrifugation in microfuge tubes and resuspended in TE and stored at -80°C.

Recovered RNA was quantified by measuring the absorbance at 260nm with 1 A_{260} unit equivalent to 40μg/ml of RNA.

ii). Glyoxal agarose gel electrophoresis.

RNA samples were prepared for electrophoresis by denaturation using the method of Thomas (1980). 5-20μg of total cellular RNA was precipitated in ethanol and resuspended in a buffer comprising 10mM sodium phosphate

pH7.0, 50% v/v DMSO and 1M deionised glyoxal. The RNA was denatured by incubation at 50°C for 1 hour. 0.2 volumes of loading buffer (100mM sodium phosphate pH7.0, 0.3% w/v bromophenol blue, 50% glycerol) was added and the RNA electrophoresed at 70 volts for 3-6 hours through a 1.4% agarose, 10mM sodium phosphate gel using a 10mM sodium phosphate buffer which was recirculated. RNA was transferred immediately after electrophoresis to nylon membranes using a 20X SSC buffer. Parallel tracks of RNA run of the same gel for use as markers were cut from the gel before transfer and stained using ethidium bromide in order to observe the position of the 18 and 28S ribosomal bands.

Appendix: Formulation of commonly used solutions.

TAE	40mM tris-acetate pH7.5 2mM EDTA
TBE	0.089M tris-borate pH8.3 2mM EDTA
TE	10mM tris-base 1mM EDTA adjust to desired pH with HCl
DNA loading buffer	10x TAE 0.25% w/v bromophenol blue 50% v/v glycerol
SSC	0.15M sodium chloride 0.15M sodium citrate pH adjusted to 7.5 with NaOH
SSPE	0.15M sodium chloride 10mM sodium dihydrogen phosphate 1mM EDTA pH adjusted to 7.5 with sodium hydroxide
HTM	0.3M Hepes 0.125M tris-base 12.5mM magnesium chloride 20mM 2-mercaptoethanol - make from stock solutions of hepes pH6.6 and tris pH8.0.

PBS	0.8% w/v sodium chloride
	0.02% w/v potassium chloride
	0.11% w/v disodium hydrogen phosphate
	0.02% w/v potassium dihydrogen phosphate
	0.013% w/v calcium chloride ($CaCl_2.2H_2O$)
	0.01% w/v magnesium chloride ($MgCl_2.6H_2O$)
L-Broth	1% w/v Bacto-tryptone
	1% w/v sodium chloride
	0.55% w/v Bacto-yeast extract
2x TY	1.6 w/v Bacto-tryptone
	0.5% w/v sodium chloride
	1% w/v Bacto-yeast extract
Lysis buffer	0.15M sodium chloride
	10mM tris-Cl, pH 7.8
	1.5mM magnesium chloride
	0.65% v/v NP40
10x Kinase buffer	0.5M tris-CL, pH 7.6
	0.1M magnesium chloride
	50mM dithiothreitol
	1mM spermidine
	1mM EDTA

CHAPTER 3. CONSENSUS SEQUENCE EXPERIMENTS.

A. INTRODUCTION

Following the identification of metallothionein-IIa (MT-IIa) as an interferon inducible gene, a homology between a sequence in its 5' flanking region and the flanking regions of three HLA genes was noted by Friedman and Stark (1985). It was proposed that this sequence played a role in the interferon response of these and possibly other genes. Experiments presented in this chapter describe approaches used to test this hypothesis. Work in other laboratories demonstrated that this sequence is also found upstream of mouse H-2 genes (Fig. 3.1) and does play a role in the induction of genes by interferon (Israel et al, 1986; Korber et al, 1987).

B. MT-IIa/TK FUSIONS.

i). Introduction.

Human metallothionein genes are also inducible by heavy metals and glucocorticoid hormones (Karin et al, 1980). This regulation has been studied by fusing selected regions of the 5' ends of these genes onto heterologous marker genes which can be assayed by northern or protein analysis (Stuart et al, 1984; Karin et al, 1984b). One such series of deletions of the MT-IIa promoter region was fused to the herpes simplex virus thymidine kinase (tk) gene by M. Karin and these deletions were used to define regions of the

Figure 3.1.

Homologies between the upstream regions of a class II HLA gene, two class I genes (*one unidentified), a mouse H2 gene and the human metallothionein II-a gene.

HLA DR	**ACCTC-AGAGTTTCTCCTCT-CA**
HLA A3	**AACTCCGCAGTTTCTTTTCT-CT**
HLA*	**ACCTCCGCAGTTTCTCTTCTTCT**
H2 K^b	**ATCTCCACAGTTTCACTTCTGCA**
MT-IIa	**ACCT--GCAGTTTCTCCTCT-CT**
Consensus	**ACCTCCGCAGTTTCTYYTCTNCT**

promoter involved in the transcriptional response to dexamethasone and cadmium (Karin et al, 1984a). Four such constructs were generously supplied to me by M. Karin. These included plasmid 1'4, containing 5' sequence to -652 bp upstream of the cap site and containing the consensus sequence, and plasmid 1'3, to -582 and lacking the consensus (Fig. 3.2).

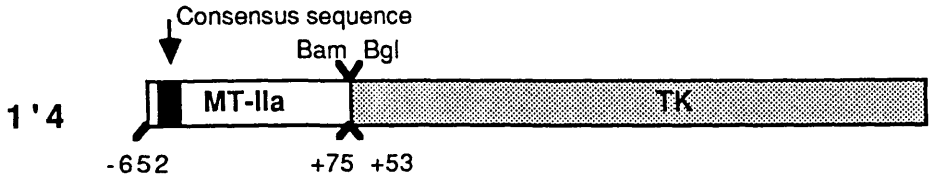
ii). Transfection of the constructs into mammalian cells.

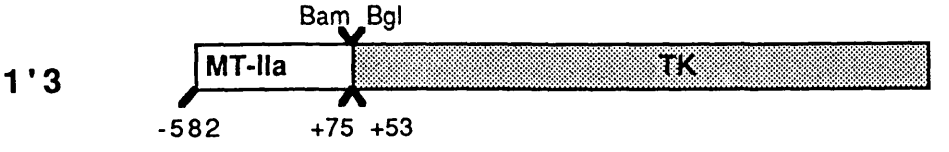
Initial attempts to transfect these constructs into a human cell line, HeLa Heb7A, were only partially successful. Heb7A is a tk^- cell line (Tunacliffe 1983) which allowed selection of transfected cells in HAT media. The cells were first assayed for sensitivity to alpha-interferon and found to be growth inhibited and resistant to EMC virus when treated. Although 20 μg of each plasmid was transfected into 2 x 10^7 cells, only 3 colonies were obtained in total, all for the 1'3 construct. A second similar attempt was also unsuccessful. Tunacliffe (1983) also reports a low transfection frequency of this cell line.

The plasmids 1'3 and 1'4 were successfully transfected into mouse Ltk^-cells. These cells had been used to express and induce a genomic clone of human 6-16, another interferon inducible gene (Porter et al, 1988), indicating conservation of the interferon signalling pathway between species. For these transfections pSV2neo was used as a

Figure 3.2.

Fusions of the Metallothionein-IIa gene to the 5' end of the HSV thymidine kinase gene. These fusions are cloned into the Hind III site of pBR322.

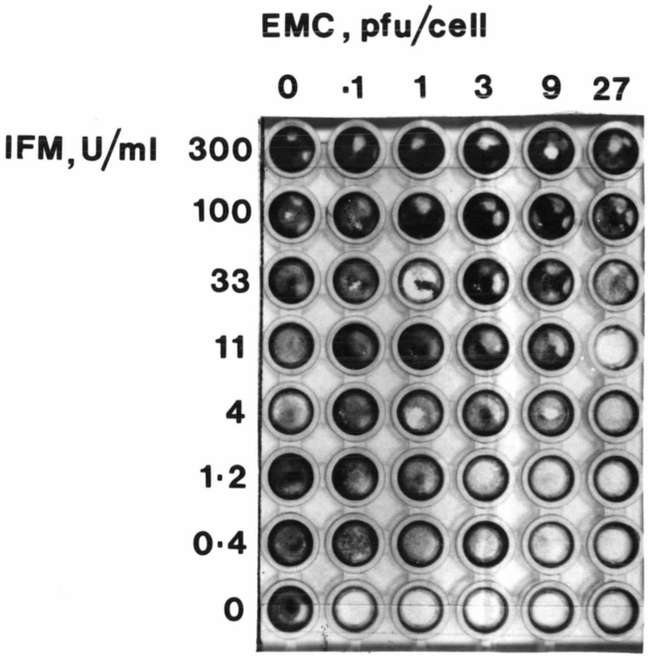

Consensus sequence
Bam Bgl
1'4
MT-IIa
TK
-652
+75 +53


Bam Bgl
1'3
MT-IIa
TK
-582
+75 +53

selectable marker instead of selecting the cells in HAT to try and avoid any potential bias for transfectants with high levels of tk expression, as this was the gene being assayed. The Ltk$^-$ cells used were tested for an interferon response by their resistance to EMC virus upon treatment with mouse interferon (Fig. 3.3) and were found to be positive. (This also demonstrated the activity of the batch of interferon used.) 60 μg of each plasmid was transfected into 3 x 10^6 cells with 10 μg of pSV2neo and after two weeks in selection over 1000 colonies on each 90 mm dish were visible. The colonies were pooled and expanded for experimentation. For each construct, a plate of cells was either treated with 1000 units per ml of mouse interferon or left untreated for 8 hours. Total RNA was prepared using GTC and aliquots glyoxalated and run out on a 10 mM sodium phosphate agarose gel. The gel was blotted and probed with nick translated plasmid 4'7 which contains the tk gene fused to the smallest amount of flanking MT-IIa sequence. The result (not shown) showed a message of the expected size in all four tracks containing RNA from the transfected lines (and no signal in the untransfected controls). There was little apparent difference in the intensity of the four signals, though when measured on a Joyce-Lobel Chromoscan 3 a small (1.5 fold) increase between the intensity of the signal from 1'4-transfected cells treated with interferon over the 1'4-transfectants minus interferon bands was observed, with the other two

Figure 3.3.

Assay of mouse Ltk^- cells for their interferon response. $5x10^4$ cells were plated out in the wells of a 96 well microtitre dish, left to settle for 8 hours and treated with interferon for 24 hours. EMC virus was added overnight, the cells were then stained with crystal violet/20% ethanol for 30 minutes and rinsed in water. The results show that even very low levels of interferon provide a degree of protection from EMC infection.


EMC, pfu/cell
0
·1
1
3
9
27
IFM, U/ml
300
100
33
11
4
1·2
0·4
0

having similar levels to the 1'4-. However, on a single experiment this can not be considered a significant finding.

iii). Conclusions.

The results described in this section, though inconclusive, illustrate some of the difficulties involved with transfection experiments. MT-IIa is only induced 3 to 5 fold by interferon (Friedman et al, 1984) and there are a number of major differences between the native gene and the system used to test the promoter which could have reduced the maximum possible induction observable. These include the use of mouse instead of human cells in which to test the constructs, although it appears that for other human interferon induced genes tested such as 6-16 (Porter et al, 1988) and 9-27 (see Ch. 6.D.vi) a response is obtained. Israel et al (1986) also reports the expression and induction of mouse H-2 genes in human cells which provides further evidence that the response mechanism is conserved. The induction observed by Israel of fusions of the H-2Kb gene to the chloramphenicol acetyl transferase gene is reported to be between 1.9 and 2.8 fold, which is less than the induction observed for native H-2 genes and may be partly due to the absence of post-transcriptional events thought to contribute towards their stabilization in interferon treated cells. Similar considerations may apply

to MT-IIa. The consensus sequence in MT-IIa is also much further away from the start of transcription than the equivalent sequences in other genes. Between it and the cap site lie sequences responsible for the induction of the gene by heavy metals and glucocorticoid hormones, which may have contributed to the relatively high basal levels of transcription seen in the experiment described above. In conclusion, it would probably be difficult to demonstrate the role of the consensus sequence found in MT-IIa without further modifications of the approach described in this chapter, such as removing other response elements from the gene and possibly using more than one consensus sequence element in any construct under test.

C. OLIGONUCLEOTIDE PROBES TO THE CONSENSUS SEQUENCE.

i). Introduction.

Experiments devised to examine the possible presence of the consensus sequence in the flanking regions of other interferon inducible genes are presented in this section. These were performed using oligonucleotide probes representing the consensus sequence and annealing them to potential target sequences at low stringency. This was followed by increasingly more stringent washes in order to observe any regions of DNA that bind the probe with any degree of specificity. By using controls of either totally

homologous or partially matched sequences, a qualitative impression of the degree of hybridization to other regions of DNA could be obtained.

Conditions for the hybridization and washing of oligonucleotide probes were derived from Wallace et al (1979) and Meinkoth and Wahl (1984) and are discussed in Ch. 2.E.(xii). Single base mismatches between an oligonucleotide probe and a target sequence reduces the melting temperature (Tm) of the hybrid formed, though the extent of this reduction will vary according to the location of the mismatch.

ii). Probing 6-16 and 1-8 cosmids.

Oligonucleotide probes of the consensus sequence and two variants (Fig. 3.4a) were synthesised and supplied by Celltech and UMIST. The first two, OCT 351 and 352 are overlapping 23mers that encompass the consensus and a few downstream nucleotides found in MT-IIa. These were annealed together in Klenow buffer at 55°C for 1 hour and extended to produce a double stranded fragment of DNA by the addition of nucleotides and Klenow enzyme. After kinasing the 5' ends the fragments were cloned into the Sma1 site of M13mp8 and recombinants sequenced to confirm the presence of the cloned sequence. An initial control experiment to use one of the constructs as a probe on the MT-IIa flanking region was not successful and all future

Figure 3.4.

a). The four oligonucleotides used in experiments described in this chapter. OCT 351 and 352 were supplied by Celltech Ltd whilst Oligos 303 and 304 were synthesised at UMIST and supplied through Genetic Research Instrumentation. OCT 351 contains the two highly conserved regions (underlined) of the consensus sequence and overlaps OCT 352 by 9 base pairs at it's 3' end. An Eco R1 recognition sequence (underlined) was included in OCT 352 for cloning purposes.

b). Alignment of OCT 351 to the consensus sequence found in human metallothionein-IIa. There is a central contiguous match of 10 nucleotides flanked by gaps and two further short matches.

c). Alignment of the consensus sequence with sequence found flanking the 6-16 gene. There is a perfect match with the conserved 'ACCT' motif and a 7/9 match with the centrally conserved (GCAGTTTCT) region.

OCT 351 5' CACCTCCGCAGTTTCTCTTCTCT 3'

OCT 352 3' GAGAAGAGATGTTACTCTTAAGC 5'

OLIGO 303 5' A(A/C)CT(G/C)G(A/G)AG(C/A)TTCT(C/T)T 3'

OLIGO 304 5' A(A/C)CTC(G/C)G(A/G)AG(C/A)TTCT(C/T)T 3'

3.4a

```
OCT 351    5' CACCTCCGCAGTTTCTCTTCTCT 3'
                ||||  ||||||||||| |||
MT-IIa     5' GACCT--GCAGTTTCTCCTCTAC 3'
```

3.4b

```
Consensus  5' ACCTCCGCAGTTTCTYY 3'
              ||||  | || ||||||
6-16       5' ACCT-GGAAGCTTCTCT 3'
```

3.4c

experiments were performed using γ-[^{32}P]ATP labelled oligonucleotides as the probes, with the cloned sequences acting as controls.

An initial experiment using the 23mer OCT 351 is shown in Fig. 3.5. The labelled oligonucleotide was used to probe a positive control, a cosmid containing the 6-16 gene and an MT-IIa/tk fusion constructs (sec. B above), containing its version of the consensus sequence (not an exact match to the probe, Fig. 3.4b). When washed at a low temperature, some, but not all, bands of DNA are detected by OCT 351, including four of the cosmid bands and the twelve multimeric units of the 1 kb ladder marker, believed to contain yeast 2 micron plasmid DNA (B.R.L.-Gibco, pers. comm.). The band containing the consensus sequence in MT-IIa is also detected, as is the positive control, which produces the strongest signal. Neither of the other bands in these tracks or either band of the negative control is seen. After a series of washes at higher temperatures only two bands of hybridization in addition to the positive control were seen. The first one was the MT-IIa flanking region which contains the consensus sequence and the second was the 1.8 kb Eco R1 fragment of 6-16, which at the time was believed to contain the 5' region of the gene. A similar experiment with a 1-8 cosmid, 1-8.8, revealed a sequence with specificity for the probe in a region of DNA within 4 kb upstream of the 1-8 gene (Fig. 3.6).

Sequence obtained by J. Kelly from the 1.8 kb Eco R1

Figure 3.5.

Experiment using oligonucleotide probe OCT 351 on a cosmid containing a 6-16 gene. Washed at 24 and 38 °C in 6XSSC. 'MT' = track containing plasmid 1'4 supplied by M. Karin, cut with EcoR1; '6-16' = cosmid containg the human 6-16 gene supplied by J.Kelly (pers comm) cut with EcoR1; '-' = M13mp13 containing OCT 351 cloned into the SmaI site, cut with ClaI; '+' = plasmid 1'3(see Fig. 3.2), supplied by M.Karin, cut with ClaI and HindIII. See text for further discussion.

At 38°C, the probe only hybridizes to 3 bands - a 3.1 kb fragment containing the metallothionein consensus sequence, a 1.8 kb fragment from the 6-16 cosmid and the positive control.

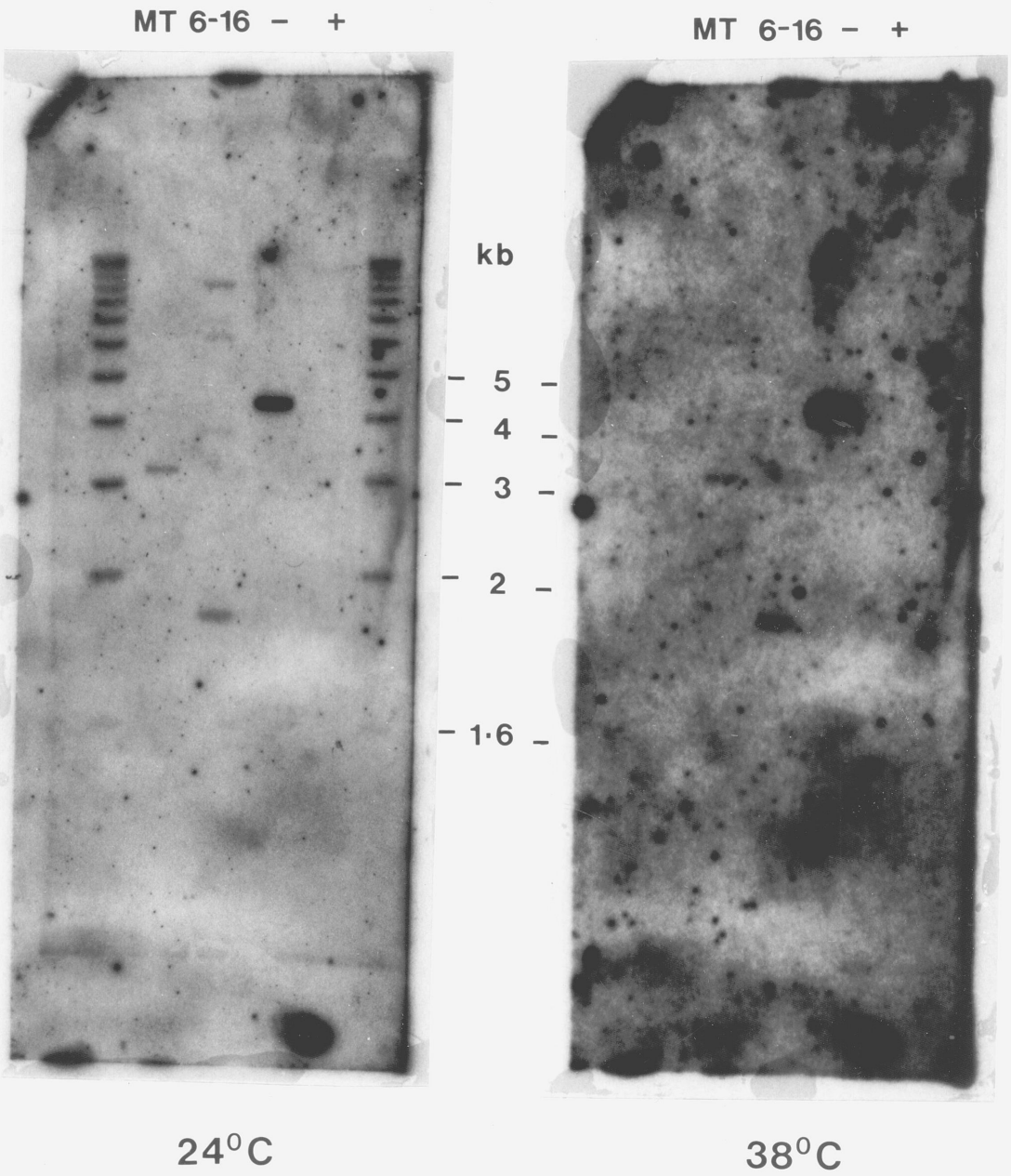
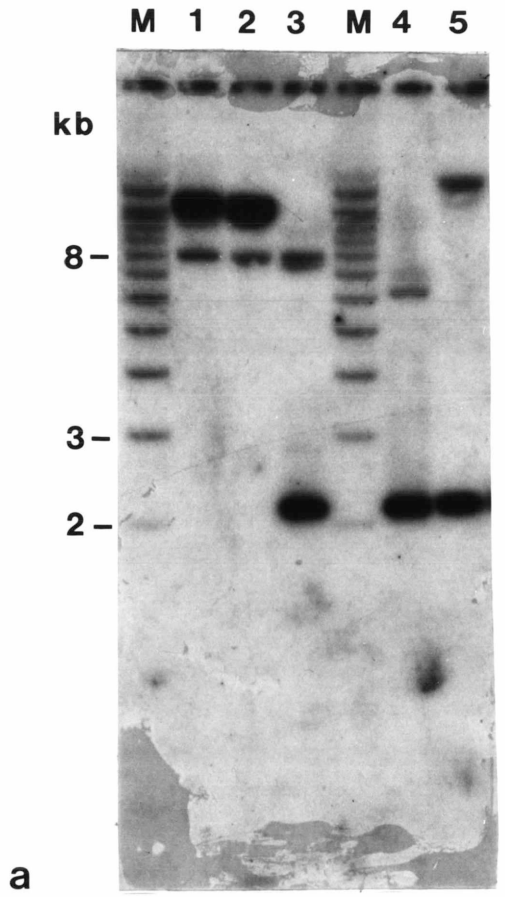

MT 6-16 - +
MT 6-16 - +
kb
5
4
3
2
1·6
24°C
38°C

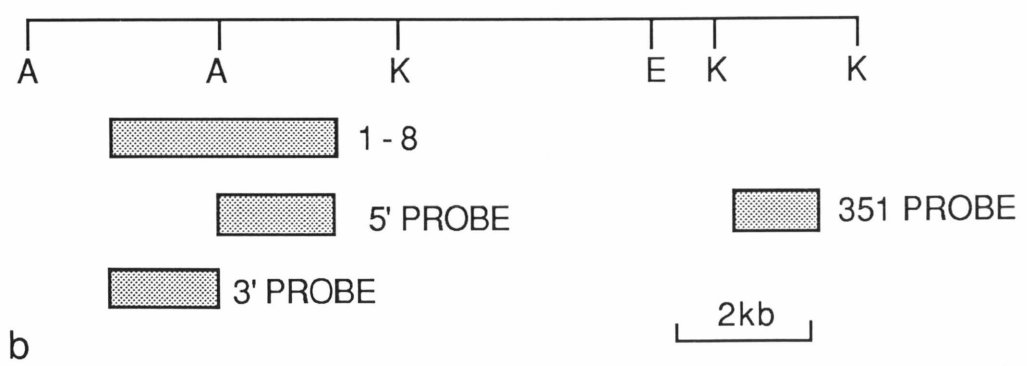
Figure 3.6.

(a) Experiment using oligonucleotide probe OCT 351 on different digests of cosmid 1-8.8. M = marker, 1 - 5 = cosmid digested with: 1) ApaI + SalI, 2) ApaI, 3) ApaI + KpnI, 4) EcoRI + KpnI, 5) KpnI. The blot was probed with kinase labelled OCT 351, washed at 31°C in 6XSSC and exposed overnight. The probe hybridized relatively strongly to a 2.1 kb KpnI fragment.

(b) Previous experiments with the same blot had indicated a 1-8 cross-hybridizing region between an ApaI and KpnI site could be cut at a second ApaI site into 5' and 3' cross-hybridizing regions, as indicated. (A=ApaI, K=KpnI, E=EcoRI).



a



b

fragment of the 6-16 cosmid revealed a homology to the consensus sequence (Fig 3.4c). Two further oligonucleotides of mixed composition were obtained (Fig. 3.4a) and used to probe a panel of 1-8 cosmids (Fig. 3.7). Both these mixed probes contained sequences with Tms that varied over an 8°C range under standard (6xSSC) conditions, though by performing a series of washes in tetramethylammonium chloride (TMA) which eliminates the difference in stability between A:T and G:C base pairs, the degree of hybridization was entirely dependent on the length of matched sequences alone (Wood et al 1985). Equimolar amounts of each DNA were loaded onto slot blots, hybridized overnight and initially washed in 6xSSC at room temperature. Following this, a series of washes in TMA were performed at 35, 45 and 49°C. Except for the perfectly matched controls, only the 23mer and mixed 16mer on the 6-16 cosmid produced a detectable signal at 45°C or above. The range of hybridization to the 1-8 cosmids varied, though none showed any degree of specificity above that of a control (non 1-8) cosmid. There is only a faint signal from the 23mer on the MT-IIa plasmid, which is probably below the level of the other slots due to the smaller total amount of DNA loaded, reducing the degree of non specific cross hybridization at the lower temperatures.

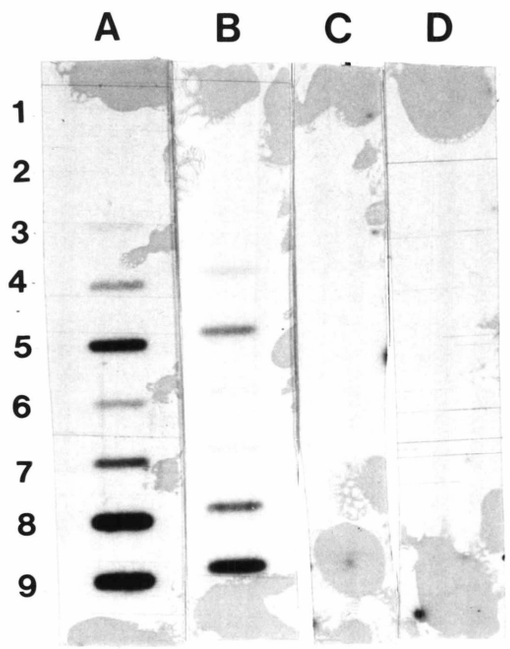
Figure 3.7.

Oligonucleotide probes used on a panel of plasmids and cosmids. 16mer probe = oligo 303, 17 mer = oligo 304, 23mer = OCT 351. Wash conditions: A = room temperature, 6X SSC; B = 35°C, C= 45°C, D = 49°C, all in TMA buffer (3.0 M Me_4NCl/50 mM Tris.HCl pH8.0/2 mM EDTA in 0.1% SDS)

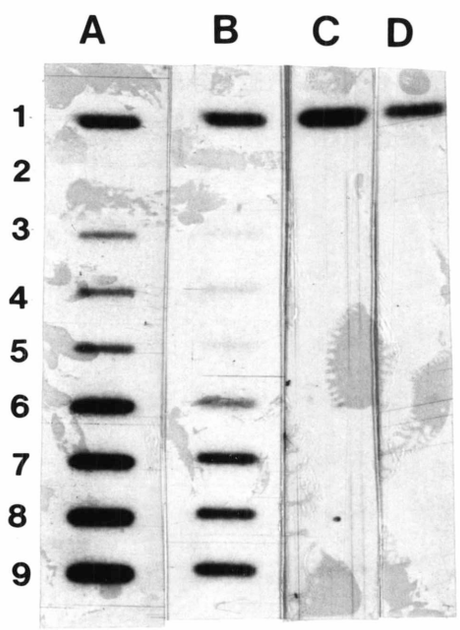
DNA samples: 1 = M13mp13 containing OCT 351 (40ng); 2 = plasmid 1'4 containing the metallothionein consensus sequence (30 ng); 3-7 = 1-8 cosmids nos. 2,3,4,8 & 10 respectively (230 ng); 8 = control cosmid (230 ng); 9 = cosmid containing 6-16 gene (230 ng).

See text for a discussion of the results.

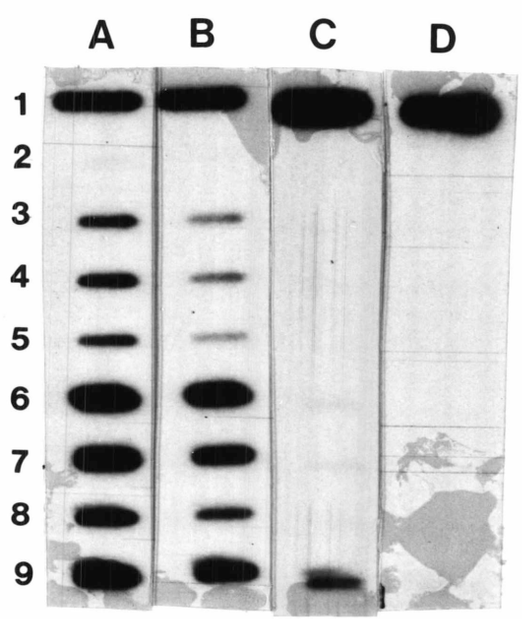
16mer



17mer



23mer



iii). Conclusions.

Experiments described in chapter 5 indicate that the 1-8 cosmids used for the experiment in section (ii) above are probably unexpressed pseudogenes and hence unlikely to have any 5' flanking control sequences. The 6-16 "consensus sequence" to which the mixed probes were made was subsequently found not to be in the 5' flanking region of this gene and hence unlikely to play a part in any interferon response. However, further work based upon these initial experiments and taking into account the information from the mouse H-2 genes was carried out by K. Smith, resulting in the isolation of six genomic lambda phage clones which are positive for other variants of the consensus sequence. However, preliminary analysis of these clones failed to suggest that any of them contained sequences related to interferon-inducible genes. The prospects for the wider application of this technique are considered further in Chap. 7.

CHAPTER 4. ISOLATION AND ANALYSIS OF A 9-27 cDNA.

A. INTRODUCTION.

The original 9-27 cDNA clone was a short sequence of 240 nucleotides, isolated from the T98G cell line. A longer cDNA, 2-1, isolated by M. McMahon from a Daudi cell cDNA library first revealed the homology of 9-27 to 1-8. A further isolate from the same library, 7-8, overlapped with the first 110 bases of 2-1 where there is also the close homology to the 1-8 cDNA , though because no sequence 3' of this clone was obtained, it was not clear if 7-8 represented further example of the 9-27 sequence or was another member of the 1-8 gene family. In order to resolve this and to obtain a potential full open reading frame of 9-27 the isolation of a longer cDNA of the gene was undertaken.

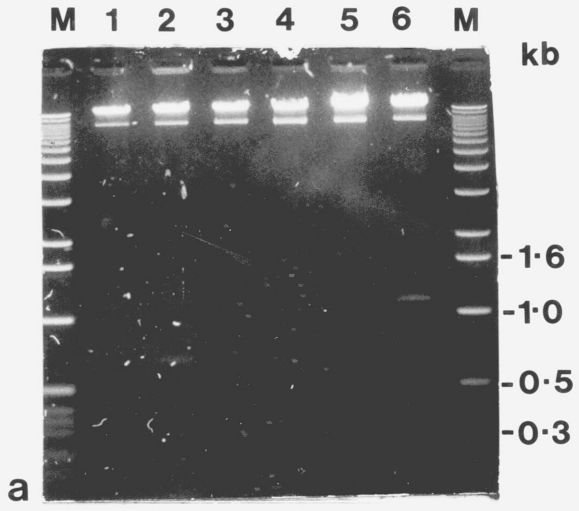
B. SCREENING AND SEQUENCING.

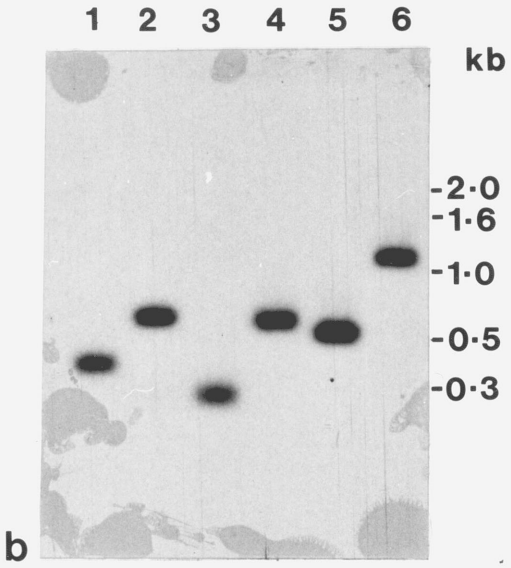
i). Screening a lambda gt10 library.

A cDNA library was constructed by D. Gewert in lambda gt10 (lgt10) using RNA from HeLa cells treated with α-interferon for 24 hours. The cDNA synthesis was performed using RNaseH and DNA polymerase to construct the second strand which produces more complete cDNAs than the original self priming method. Eight agar dishes were each plated with approximately 50,000 p.f.u. and grown overnight. Single lifts onto nitrocellulose filters were taken off each plate, probed with 9-27 unique sequences,

Figure 4.1.

Recombinant lgt10 clones containing 9-27 cDNAs. EcoRI digests were run out on an agarose gel where insert bands were only just visible (a). Probing a blot of the gel with a 9-27 specific cDNA probe confirmed the presence of 9-27 cDNA clones (b). Hybridization under standard conditions for 6 hours; exposure of autoradigraph = 15 minutes.


M 1 2 3 4 5 6 M
kb
1·6
1·0
0·5
0·3
a


1 2 3 4 5 6
kb
2·0
1·6
1·0
0·5
0·3
b

washed at high stringency and exposed against Kodak XAR film overnight. Between 100 to 200 positives over a low background were seen on each filter. Twelve plugs of agar were taken from areas of the plates corresponding to positive signals and each replated on a 90 mm plate at a concentration of 50 - 500 p.f.u per dish. Duplicate lifts onto nylon filters were taken and probed and washed in the same manner to the primary screen. Between 1 and 13 duplicate positive were detected on ten of the plates and upon replating the remaining two primary screen positives at a higher density, secondary positives from them were also obtained. Each positive signal represented an area of the plate containing between 1 and 5 plaques, allowing one positive from each plate to be picked and replated at a low density for a final round of screening. Following this, six independent and well isolated positives were picked, grown to confluence on 90 mm agarose dishes and the phage DNA prepared. The DNA was digested with EcoR1 which excises the cloned sequences, run on an agarose gel and blotted onto a nylon membrane. Bands of between 300 and 1100 bp were visible on the gel (Fig. 4.1a) and they were all positive for 9-27 (Fig. 4.1b). Number 6 is larger than the actual 9-27 mRNA and may have resulted from the erroneous ligation of 9-27 to another message at some stage in the construction of the library. The next two longest clones, numbers 2 and 4 were both between 600 - 700 bp in size and were selected for further examination.

ii). Ratio of 9-27 to 1-8 in the library.

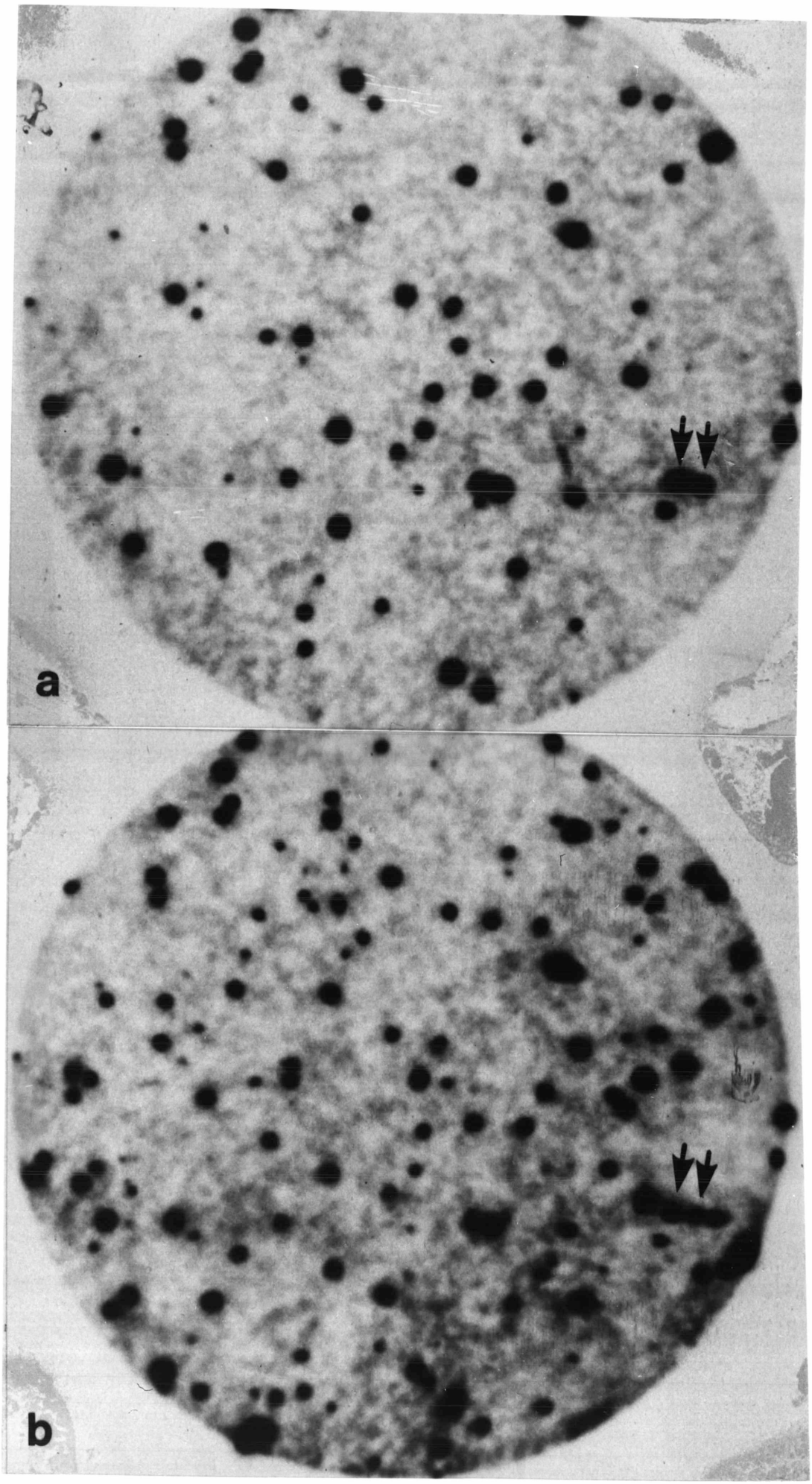
In order to estimate the frequency of 9-27 mRNA in interferon treated cells as a percentage of total poly A+ RNA, and to compare the ratio of 9-27 to 1-8, the titre of the library was redetermined and a further 60,000 p.f.u. plated onto three 15 cm plate. Duplicate nylon filters from each plate were screened firstly with 9-27 specific cDNA and then again (without stripping the filters) with the 1-8 cDNA (Fig. 4.2). A total of 196 duplicate positives for 9-27 were identified and a further 188 1-8s seen when the filters were probed for the second time. This gives a 9-27:1-8 ratio of almost 1:1 and an estimate of 0.3% as the amount of 9-27 mRNA in treated cells, assuming that the cDNA library is representative of the total polyA$^+$ RNA. These results differ slightly from the figures of 2:3 and 0.2% reported previously (Friedman et al 1984). As it is now known that at least two 1-8 genes are expressed, this indicates that 9-27 is either more strongly induced or its mRNA is more stable than 1-8 genes.

iii). Sequencing the cDNAs.

Clones 2 and 4 were both subcloned into the EcoR1 site of pUC13 by mixing 1 - 5 μg of EcoR1 cut phage DNA with 20 ng of cut and phosphatased vector, ligating overnight at 12°C with T4 DNA ligase, transforming competent E.coli JM101 and selecting colonies on indicator plates.

Figure 4.2.

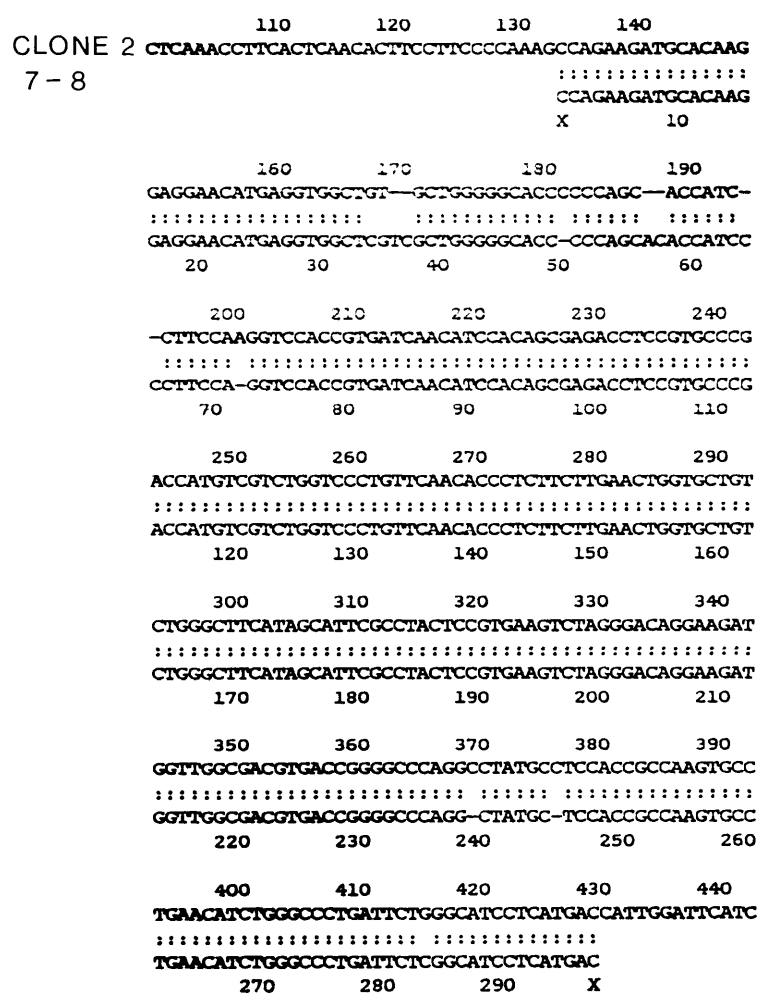
Ratio of 9-27 to 1-8 DNA in the lgt10 library. (a) = Filter containing approximately 10,000 plaques probed with 9-27 specific cDNA. (b) = the same filter re-probed with 1-8. The arrows indicate the same pair of 9-27 positive signals, flanked in (b) by two additional positives.

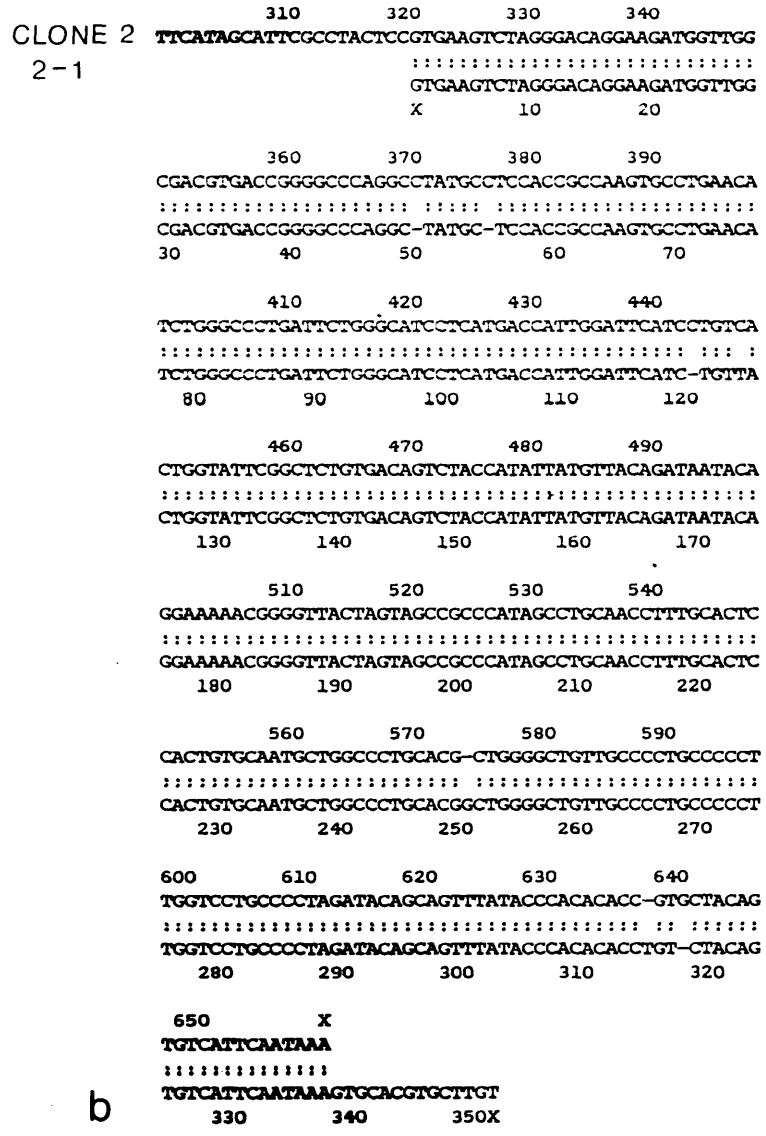

a
b

The recombinant plasmids were grown on a large scale and DNA was prepared over caesium chloride gradients. The EcoR1 inserts were isolated and subcloned into the EcoR1 site of M13 mp18 for sequencing. Only preliminary sequence of the 5' region of clone 4 was obtained; this was found to be in agreement with the equivalent part of clone 2, which was analyzed in detail. The alignments of the final clone 2 sequence against the cDNAs 7-8 and 2-1 are presented in Fig 4.3. It is apparent that 7-8 is almost certainly a partial 9-27 clone rather than a distinct member of the 1-8 family, even though there are several minor differences between the two sequences. Two of these differences, both in the potential open reading frame of the gene, were probably the result of ambiguous data from the sequencing reactions being incorrectly interpreted. The first occurs in and downstream of a run of six C residues, which follow the sequence 5' GGGGGCA 3'. This run of Gs are the probable cause (through the formation of secondary structure) of the stacking and compression seen on sequencing gels of the Cs which were previously interpreted as only being five in number. This compression also causes artifacts in the succeeding sequence. The correct and unambiguous reading of this sequence was only possible by sequencing the opposite strand which was achieved by subcloning the 5' EcoR1 to Sau3A fragment (residues 0-213 of the Figure 4.5 sequence) of this clone into the EcoR1 and BamH1 sites of M13 mp18. It is worth noting that when

Figure 4.3.

Sequence comparisons between clone 2 from the lgt10 library and previous clones (a) 7-8 and (b) 2-1. Clone 7-8 has a 3' end which terminates in sequence common to both 9-27 and 1-8, though the comparison shows that it is very similar to 9-27 at its 5' end. The sequence of clone 2-1 agrees well with that of clone 2.


CLONE 2
7 - 8
110 120 130 140
CTCAAACCTTCACTCAACACTTCCTTCCCCAAAGCCAGAAGATGCACAAG
CCAGAAGATGCACAAG
X 10
160 170 180 190
GAGGAACATGAGGTGGCTGT—GCTGGGGGCACCCCCCAGC—ACCATC-
GAGGAACATGAGGTGGCTCGTCGCTGGGGGCACC-CCCAGCACACCATCC
20 30 40 50 60
200 210 220 230 240
-CTTCCAAGGTCCACCGTGATCAACATCCACAGCGAGACCTCCGTGCCCG
CCTTCCA-GGTCCACCGTGATCAACATCCACAGCGAGACCTCCGTGCCCG
70 80 90 100 110
250 260 270 280 290
ACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCTTGAACTGGTGCTGT
ACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCTTGAACTGGTGCTGT
120 130 140 150 160
300 310 320 330 340
CTGGGCTTCATAGCATTCGCCTACTCCGTGAAGTCTAGGGACAGGAAGAT
CTGGGCTTCATAGCATTCGCCTACTCCGTGAAGTCTAGGGACAGGAAGAT
170 180 190 200 210
350 360 370 380 390
GGTTGGCGACGTGACCGGGGCCCAGGCCTATGCCTCCACCGCCAAGTGCC
GGTTGGCGACGTGACCGGGGCCCAGG-CTATGC-TCCACCGCCAAGTGCC
220 230 240 250 260
400 410 420 430 440
TGAACATCTGGGCCCTGATTCTGGGCATCCTCATGACCATTGGATTCATC
TGAACATCTGGGCCCTGATTCTCGGCATCCTCATGAC
270 280 290 X
a


CLONE 2
2 - 1
310 320 330 340
TTCATAGCATTCGCCTACTCCGTGAAGTCTAGGGACAGGAAGATGGTTGG
GTGAAGTCTAGGGACAGGAAGATGGTTGG
X 10 20
360 370 380 390
CGACGTGACCGGGGCCCAGGCCTATGCCTCCACCGCCAAGTGCCTGAACA
CGACGTGACCGGGGCCCAGGC-TATGC-TCCACCGCCAAGTGCCTGAACA
30 40 50 60 70
410 420 430 440
TCTGGGCCCTGATTCTGGGCATCCTCATGACCATTGGATTCATCCTGTCA
TCTGGGCCCTGATTCTGGGCATCCTCATGACCATTGGATTCATC-TGTTA
80 90 100 110 120
460 470 480 490
CTGGTATTCGGCTCTGTGACAGTCTACCATATTATGTTACAGATAATACA
CTGGTATTCGGCTCTGTGACAGTCTACCATATTATGTTACAGATAATACA
130 140 150 160 170
510 520 530 540
GGAAAAACGGGGTTACTAGTAGCCGCCCATAGCCTGCAACCTTTGCACTC
GGAAAAACGGGGTTACTAGTAGCCGCCCATAGCCTGCAACCTTTGCACTC
180 190 200 210 220
560 570 580 590
CACTGTGCAATGCTGGCCCTGCACG-CTGGGGCTGTTGCCCCTGCCCCCT
CACTGTGCAATGCTGGCCCTGCACGGCTGGGGCTGTTGCCCCTGCCCCCT
230 240 250 260 270
600 610 620 630 640
TGGTCCTGCCCCTAGATACAGCAGTTTATACCCACACACC-GTGCTACAG
TGGTCCTGCCCCTAGATACAGCAGTTTATACCCACACACCTGT-CTACAG
280 290 300 310 320
650 X
TGTCATTCAATAAA
TGTCATTCAATAAAGTGCACGTGCTTGT
330 340 350X
b

the same region of the genomic clone of 9-27 (Chap. 6) was sequenced using T7 DNA polymerase ("Sequenase") instead of Klenow, this difficulty did not occur.

The second discrepancy is a result of a similar G/C compression, at bases 379 to 362 of the clone 2 sequence. Sequencing the coding strand of the cDNA gives the apparent reading 5' GGC 3', though the opposite strand also gives this reading which would mean that the coding sequence is 5' GCC 3'. The correct interpretation is obviously 5' GGCC 3'. Two other extra C residues not found in cDNA2-1 occur within the next 70 nucleotides, restoring both sequences to the same open reading frame. Clone 2 was sequenced on both strands through these problematic regions and the final sequence derived was later found to be in agreement with the genomic sequence described in chapter 6.

iv). Confirming that the 5' end of clone 2 is from 9-27.

An oligonucleotide corresponding to the opposite strand of clone 2 from nucleotides 50 to 27 was obtained for use as a probe for a 9-27 genomic clone. In order to confirm that this sequence was authentic and not the result of a cloning artefact, cytoplasmic polyA$^+$ RNA from HeLa cells treated or untreated with interferon was glyoxalated, run on a 10 mM sodium phosphate gel and blotted onto nitrocellulose. The oligonucleotide was end-labelled with γ-[32P]ATP and hybridized to the filter at 42°C in

6xSSPE/5xDenhardts overnight. After washing, the filter was exposed to Kodak XAR film overnight and a faint inducible signal observed (Fig. 4.4) at the expected position for 9-27. This was good (though not conclusive) evidence for the 5' end of the cDNA being true 9-27 sequence.

v). Translation of the ORF.

There is only one long open reading frame in the DNA sequence of 9-27. A translation of this frame into amino acid sequence is presented in Figure 4.5, and the hydrophobicity profile of the translated sequence is shown in Figure 4.6. These analyses predict a protein of 15 kd in size which is largely hydrophobic, with a possible membrane spanning region near the C-terminus. There is a large hydrophilic peak. A peptide which corresponds to the sequence spanning this peak, when inoculated into rabbits caused the production of an anti-serum which was capable of detecting an interferon inducible protein of approximately 15 kd in size (L. Reid, pers comm).

vi). Database searches.

Release number 50 of Genbank was searched with the full cDNA sequence of clone 2. Besides the 1-8 and previous 9-27 sequences already in the database, no homologies of significance were identified.

The NBRF database was searched with the predicted

Figure 4.4.

Probing HeLa cytoplasmic polyA$^+$ RNA with a 24mer probe to the 5' end (complement of residues 50 to 27 of Fig. 4.5) of 9-27 cDNA clone 2. 0.5mg of RNA from cells either untreated (-) or treated (+) with interferon was run out on an agarose gel and northern blotted onto nitrocellulose. The blot was probed at 30°C overnight and washed at 55°C in 2XSSC. An overnight exposure indicated a faint induced band of about 800 nucleotides in length.

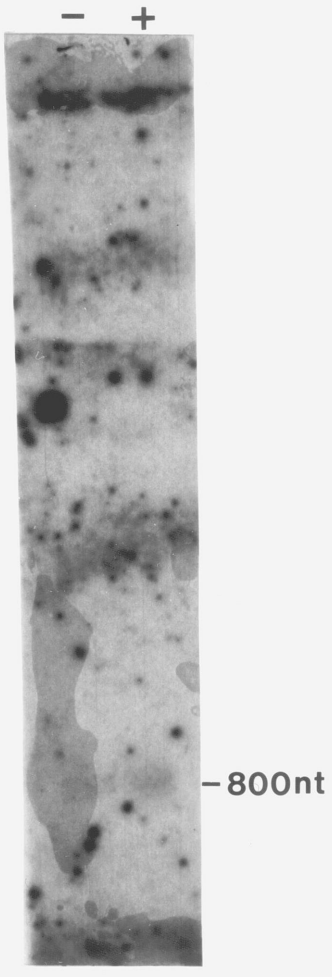

−
+
800nt

Figure 4.5.

Complete sequence of the longest 9-27 cDNA (Clone 2) with the translation of its longest open reading frame.

```
                10        20        30        40
      TCGAGAAACTGAAACGACAGGGGAAAGGAGGTCTCACTGAGCAC

     50        60        70        80        90
CGTCCCAGCATCCGGACACCACAGCGGCCCTTCGCTCCACGCAGAAAACC

     100       110       120       130       140
ACACTTCTCAAACCTTCACTCAACACTTCCTTCCCCAAAGCCAGAAGATG
                                                Met

     150       160       170       180       190
CACAAGGAGGAACATGAGGTGGCTGTGCTGGGGGCACCCCCCAGCACCAT
HisLysGluGluHisGluValAlaValLeuGlyAlaProProSerThrIl

     200       210       220       230       240
CCTTCCAAGGTCCACCGTGATCAACATCCACAGCGAGACCTCCGTGCCCG
eLeuProArgSerThrValIleAsnIleHisSerGluThrSerValProA

     250       260       270       280       290
ACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCTTGAACTGGTGCTGT
spHisValValTrpSerLeuPheAsnThrLeuPheLeuAsnTrpCysCys

     300       310       320       330       340
CTGGGCTTCATAGCATTCGCCTACTCCGTGAAGTCTAGGGACAGGAAGAT
LeuGlyPheIleAlaPheAlaTyrSerValLysSerArgAspArgLysMe

     350       360       370       380       390
GGTTGGCGACGTGACCGGGGCCCAGGCCTATGCCTCCACCGCCAAGTGCC
tValGlyAspValThrGlyAlaGlnAlaTyrAlaSerThrAlaLysCysL

     400       410       420       430       440
TGAACATCTGGGCCCTGATTCTGGGCATCCTCATGACCATTGGATTCATC
euAsnIleTrpAlaLeuIleLeuGlyIleLeuMetThrIleGlyPheIle

     450       460       470       480       490
CTGTCACTGGTATTCGGCTCTGTGACAGTCTACCATATTATGTTACAGAT
LeuSerLeuValPheGlySerValThrValTyrHisIleMetLeuGlnIl

     500       510       520       530       540
AATACAGGAAAAACGGGGTTACTAGTAGCCGCCCATAGCCTGCAACCTTT
eIleGlnGluLysArgGlyTyrENDEND

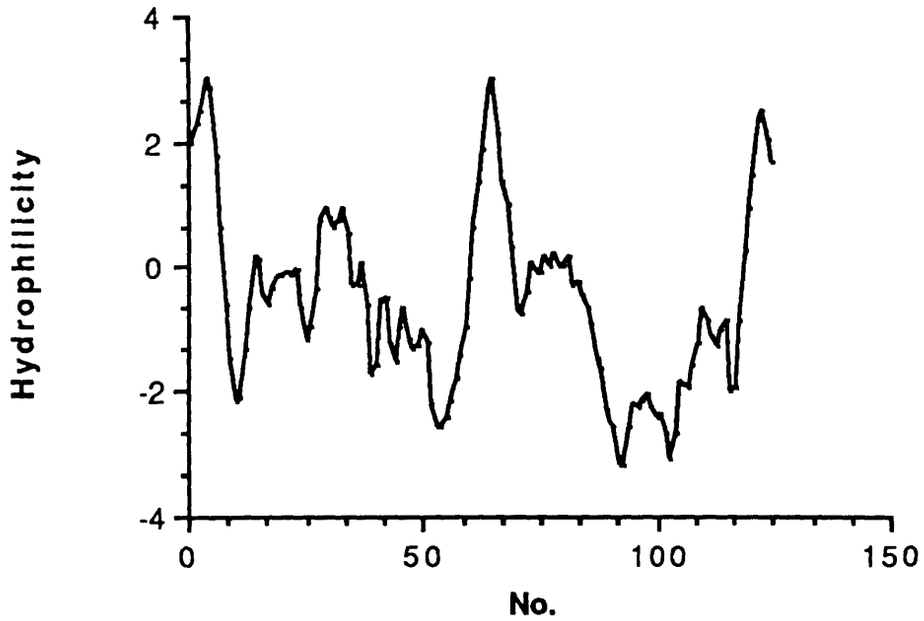
     550       560       570       580       590
GCACTCCACTGTGCAATGCTGGCCCTGCACGCTGGGGCTGTTGCCCCTGC

     600       610       620       630       640
CCCCTTGGTCCTGCCCCTAGATACAGCAGTTTATACCCACACACCGTGCT

     650
ACAGTGTCATTCAATAAA
```

Figure 4.6.

Hydrophobicity profile of the putative 9-27 gene product using the algorithm of Kyte and Dolittle (1982). Residue number (x-axis) plotted against hydrophilicity value; positive numbers represent hydrophilic reisdues.


Hydrophilicity
4
2
0
-2
-4
0
50
100
150
No.

protein sequence for 9-27. No significant homologies were found, with the possible exception of a short homology to the mouse EGF precursor (Fig. 4.7). Although the homology is only of seven out of ten residues, it may be of note that it occurs in the region of the hydrophobic peak of 9-27. Without any further information about the possible role of 9-27, it would be wrong however at this stage to place any emphasis on what may be nothing more than a coincidence.

Figure 4.7.

Homology between residues 63 to 72 of the longest open reading frame of 9-27 and mouse epidermal growth factor precursor. The homology was detected by a search of the NBRF database.

```
        60                                  70
9-27    SerValLysSerArgAspArgLysMetValGlyAspValThrGlyAla
                  *   *   *   *   *           *   *
EGF     TyrThrLeuSerArgAspArgLysTyrCysGluAspValAsnGluCys
        351                                 361
```

CHAPTER 5. 1-8 GENOMIC COSMIDS.

A. INTRODUCTION.

A series of cosmids containing 1-8 sequences were isolated by McMahon from the cosmid library supplied by Collins (see Chapter 6.B). The aim of the work described in this chapter was to identify which, if any, of the cosmids contained a complete and functional gene. The principal method used was transfection of the cosmids into mouse Ltk^- cells followed by northern blot analysis. McMahon (1985) had shown that the 1-8 cDNA probe was able to detect homologous sequences in both mouse DNA and RNA. Because of this, in addition to the cDNA probe used, probes were made from the 1-8 cross-hybridizing regions of each cosmid and used to probe the RNA from cells in which the parental cosmid had been transfected. This would allow more stringent washing of northern blots prepared from transfected cells probed with such fragments, since these probes would be expected to provide an exact match to the human 1-8 sequence.

B. TRANSFECTION EXPERIMENTS.

i). Summary of experiments.

The six cosmids used for these experiments were named pCos1-8.2, 3, 4, 5, 8 and 10. They were initially analyzed in collaboration with Martin McMahon who has presented some of the preliminary mapping data (McMahon, 1985). All six

have different pattern of restriction fragments when digested with a range of enzymes, including Eco R1 (Figure 5.1).

Cosmid DNA was prepared over caesium chloride gradients and 10 μg of each was transfected into 10^6 mouse Ltk$^-$ cells. The cosmids contain the herpes simplex virus tk gene, enabling the selection of transfectants in HAT media (see methods). After 2-3 weeks, between 100 and 1000 colonies were visible on each plate and they were pooled and expanded in mixed cultures. 3-5x10^6 cells were plated out into 90 mm dishes, treated for 24 hours with 300 units/ml of mouse interferon and RNA prepared. Typically 300 μg of total RNA was recovered. 10-20 μg (a constant amount for each experiment) was glyoxalated, run out on a 1.4% agarose gel, blotted and probed. Figure 5.2 shows a typical result in which an interferon inducible signal is seen in both transfected cell lines and untransfected controls when the blots were washed at a medium stringency. When washed at a higher stringency, no message was detected in any of the cell lines. Although numerous experiments were performed, using a variety of probes (genomic, 5'cDNA and 3'cDNA) and wash conditions, the result in all cases was that either no message could be seen in both transfected or untransfected mouse cells, or a 1-8 signal was detected in transfected cell lines at the same level as the mouse control.

Figure 5.1.

EcoR1 digests of six 1-8 containing cosmids, numbered 2, 3, 4, 5, 10^3, and 8. Cosmid 10^3 is referred to as cosmid 10 since the cosmid in the track labelled 10^1 did not contain 1-8 sequences. Since it was isolated from the same library as the 1-8 cosmids it was used as a control in some experiments.

"M" = 1 kb ladder marker (BRL).

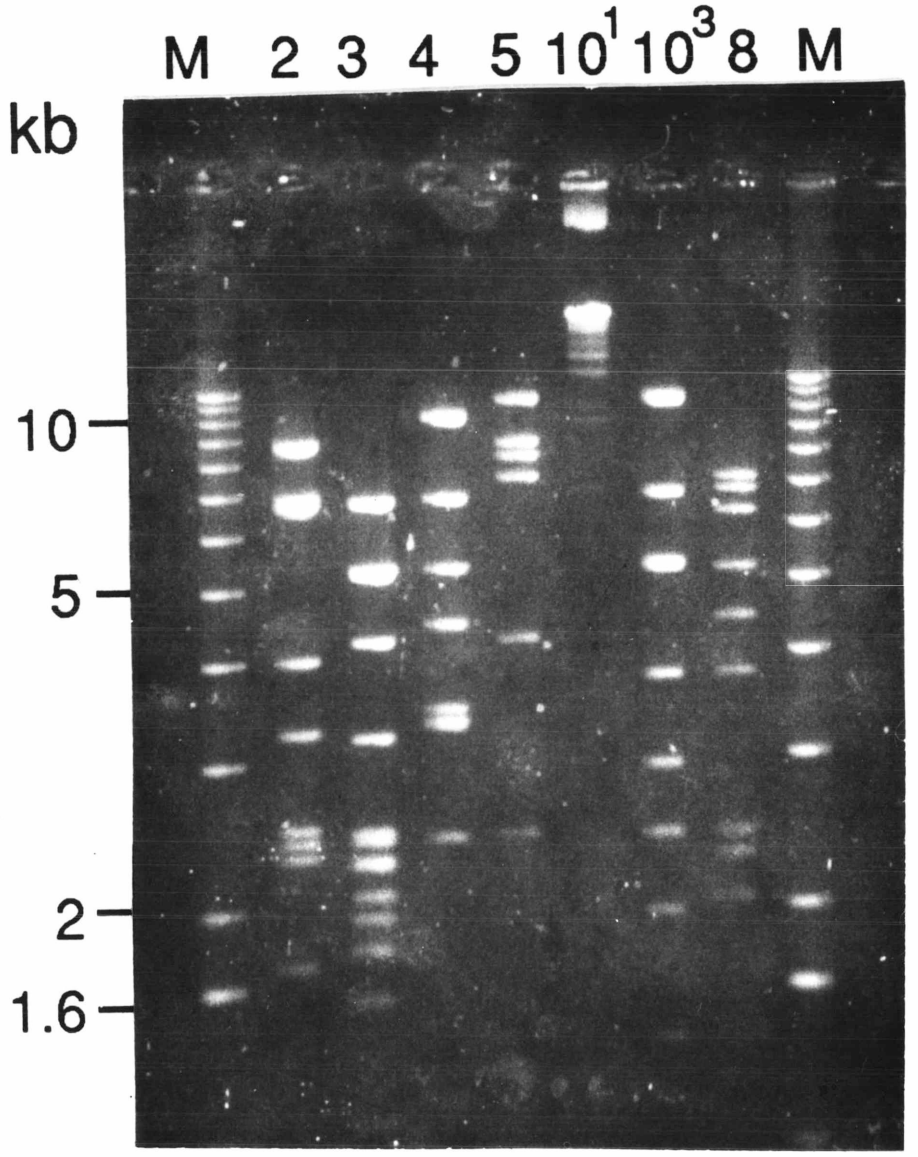
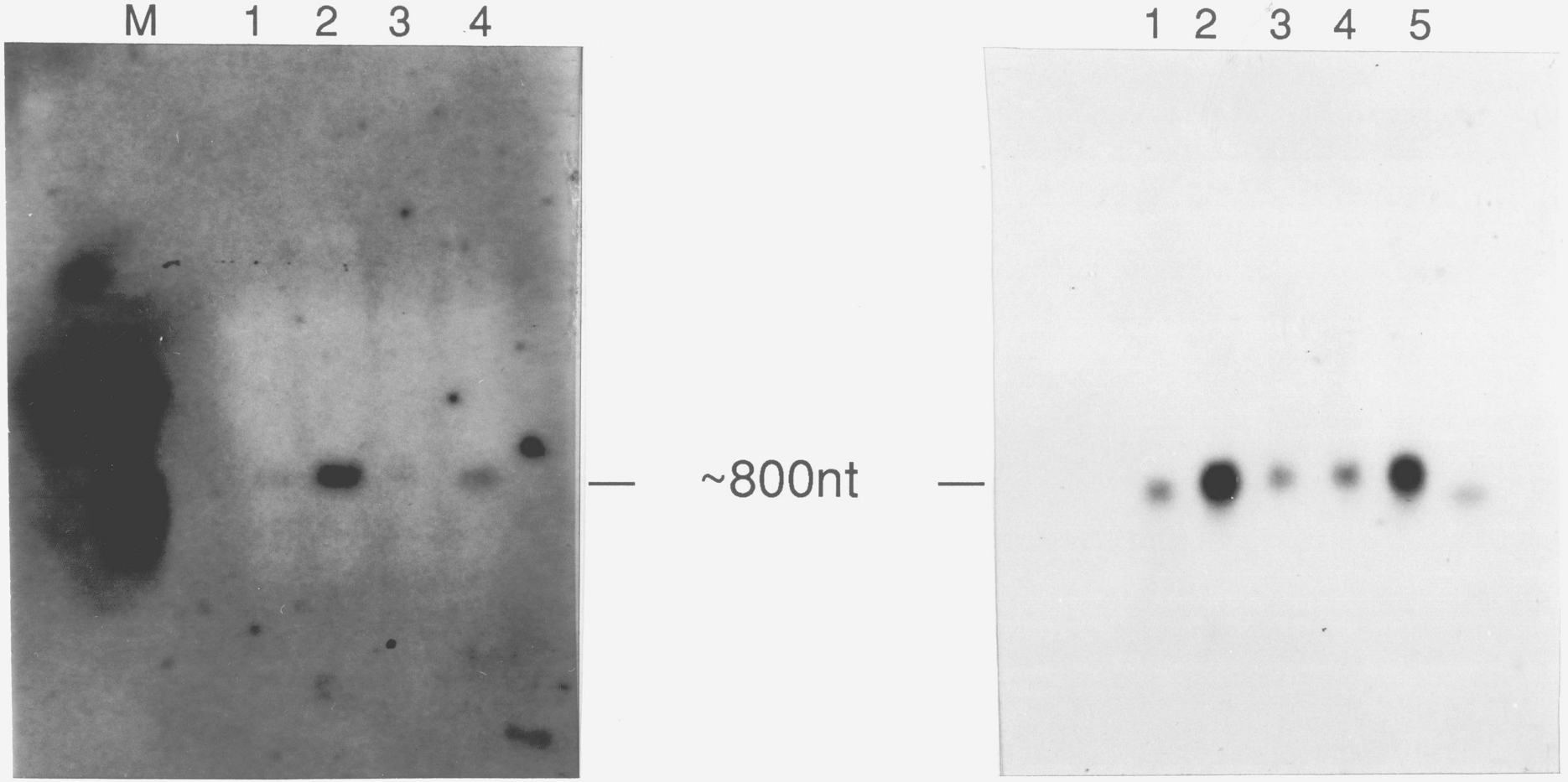

M
2
3
4
5
10^1
10^3
8
M
kb
10
5
2
1.6

Figure 5.2.

Typical northern blot experiments using polyA$^+$ RNA prepared from Ltk$^-$ cells transfected with 1-8 cosmids.

a). 1 & 2 = RNA from cells transfected with 1-8cos10, minus (1) or plus (2) interferon. 3 & 4 = untransfected controls minus (3) or plus (4) interferon. Probe = 3' fragment (approx 250 bp) of 1-8cDNA; blot washed at 55°C, 0.2xSSC. M = marker lane (1 kb ladder).

b). 1 & 2 = Ltk$^-$ control, minus (1) and plus (2) interferon. 3 & 4 = RNA from Ltk$^-$ cells transfected with 1-8cos8, minus interferon, 5 = as 3/4 plus interferon. Probe = PstI - ApaI fragment of 1-8cos8 (this is a subfragment of the 5' PstI - KpnI fragment illustrated in Fig. 3.6b); blot washed at 60°C in 0.2xSSC.


M 1 2 3 4
~800nt
1 2 3 4 5
a
b

ii). Conclusions.

Subsequent experiments such as those discussed below in part C indicate that the cosmids examined in these experiments are unlikely to contain functional genes. However, since the results of the subsequent experiments were not available at the time of initial attempts to characterize these cosmids, it is believed that the approach described above was a priori the most rational way to proceed.

The problems which arose with these experiments are common to situations where although a positive answer would be unequivocal, a negative one requires further experimentation. In this case, very high stringency hybridization and washing conditions revealed no signal from RNA prepared from transfected cells. Although this could mean that no expression of human 1-8 was occurring, it could also be a result of other factors, such as a low level of expression or less than perfect stringency with the probe. However, by relaxing these conditions, murine 1-8 was detected. In some circumstances a host background level of expression may be acceptable, e.g. if the transfected gene was being over-expressed at a high level.

The experiments undertaken attempted to balance these conflicting considerations. With hindsight, it is clear that their lack of success was due to the absence of functional 1-8 genes in the cosmids tested, rather than a failure in the experimental protocols.

C. THE COSMIDS CONTAIN PROCESSED PSEUDOGENES.

i). Introduction.

Pseudogenes are found throughout the human genome. They are non functional copies of genes and are thought to have arisen in at least two different ways.

In one scenario, duplication of a gene provides the genome with two copies of the gene. A mutation in one copy resulting in loss of function would not have any deleterious effects and would not be selected against during the course of evolution. Any further mutations would be effectively neutral, leading to the divergence of this copy from its functional homologue. (Gene conversion is a further complication to this.) Such a gene may still be expressed, depending upon the type of mutation.

Alternatively, retroviral insertion of mRNA back into the genome can give rise to a pseudogene which would be a DNA copy of the transcribed and processed sequence of the gene without any flanking control sequences.

The experiments described below sought to distinguish between a functional gene with intact introns and a spliced or processed counterpart by the use of a specific probe.

ii). An exon spanning probe.

The 9-27 gene and at least one 1-8 gene have a common splice junction at the start of their final exon (Chapter

6.C). This occurs in a region where on both sides of this junction where there is 100% homology between the two genes. This implies that this sequence is likely to be conserved in other 1-8 genes. A 34mer oligonucleotide probe that corresponds to the cDNA sequences 17 nucleotides upstream (5') and 17 nucleotides downstream (3') of the splice site was obtained. At a low stringency of hybridization the 17 bp match from either exon would allow hybridization of the oligonucleotide. At a higher stringency, only a longer match, such as that found in the cDNA of 9-27 or a processed pseudogene would suffice to produce continued hybridization of the probe (see Fig. 5.3).

When this 34mer was used to probe the 1-8 cosmids and appropriate controls (Fig. 5.4), hybridization to 5 of the 6 cosmids was seen above the temperature that resulted in the loss of signal from known functional genes. The sixth, cosmid 5, did not hybridize at low stringency as strongly as the negative control. The remaining five, plus an additional 1-8 cosmid isolated subsequently, are all potentially processed pseudogenes which would explain the failure to detect expression of human 1-8 from these cosmids in transfected mouse cell lines. The 1-8 sequences in cosmid 5 are obviously the most diverged from the conserved 1-8/9-27 region. Since 1-8 appears not to be expressed from this cosmid, it also is likely to be a pseudogene. However, the result leaves open the

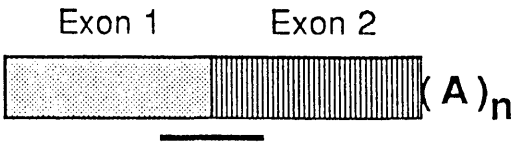
possibility that the pseudogene in this cosmid is of the first type (a genomic duplication) described above.

Figure 5.3.

Experiment to determine the probable origin of 1-8 genes found in cosmids 2-10. A 34mer probe which spans the exons of the known 1-8 gene will hybridize along a contiguous length of the cDNA (top) but only along one or other of its halves to the genomic gene from which the cDNA is transcribed (centre).

If however the probe is used on a processed pseudogene, hybridization will occur over the entire sequence (ignoring pseudogene mutations which produce mismatches), bottom figure.

cDNA



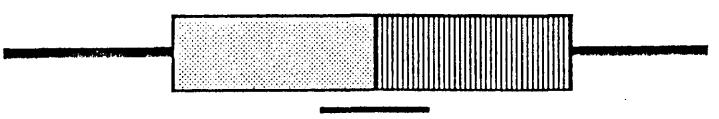
34mer oligo that
spans the exons.
(Common to 1-8 & 9-27)

TRANSCRIBED GENE



17/34 bases hybridize

PROCESSED PSEUDOGENE



Complete hybridization
of the 34mer oligo.

Figure 5.4.

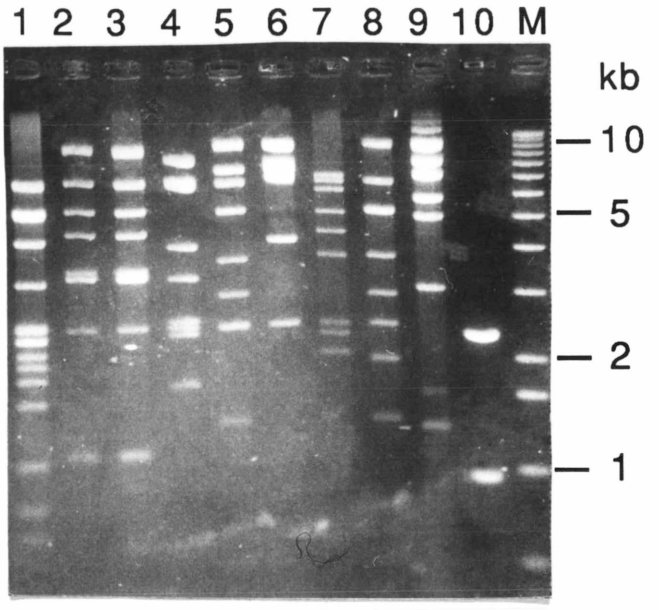
Result of exon-spanning oligonucleotide experiment.

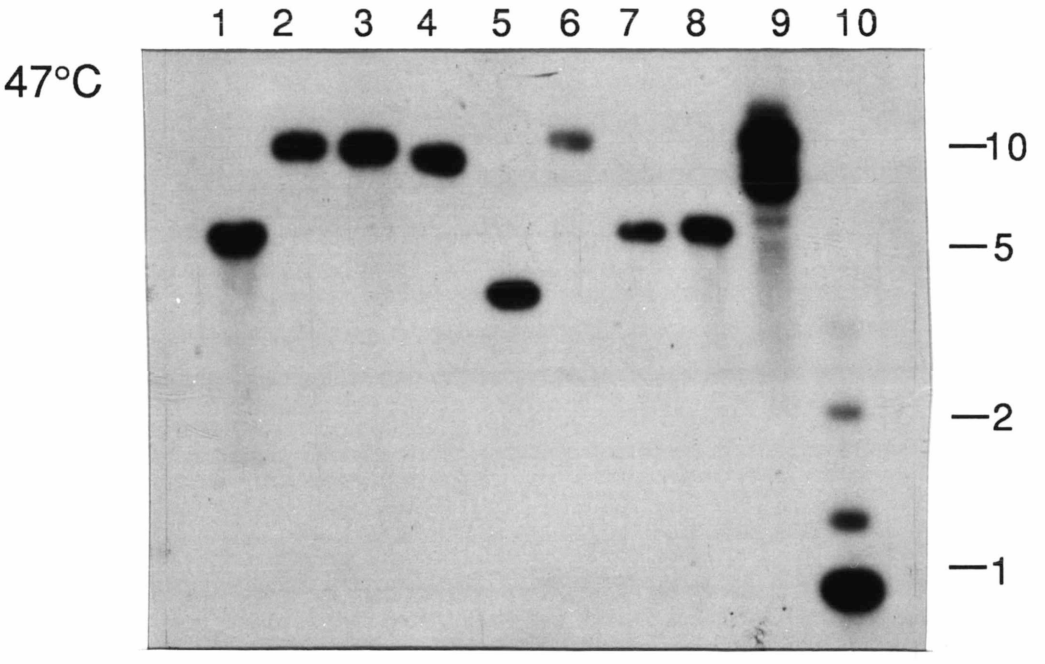
Lanes 1 to 8 contain EcoR1 digests of 1-8 cosmids as follows: 1 = cos3, 2 = cos4, 3 = cos4 (independent isolate), 4 = cos2, 5 = cos19 (additional isolate from the Collins library), 6 = cos5, 7 = cos8, 8 = cos10. Lane 9 = 9-27cos3, which contains 9-27 and two functional 1-8 genes (Hind III digest). Lane 10 = 9-27 cDNA in pUC18, digested with PvuII.

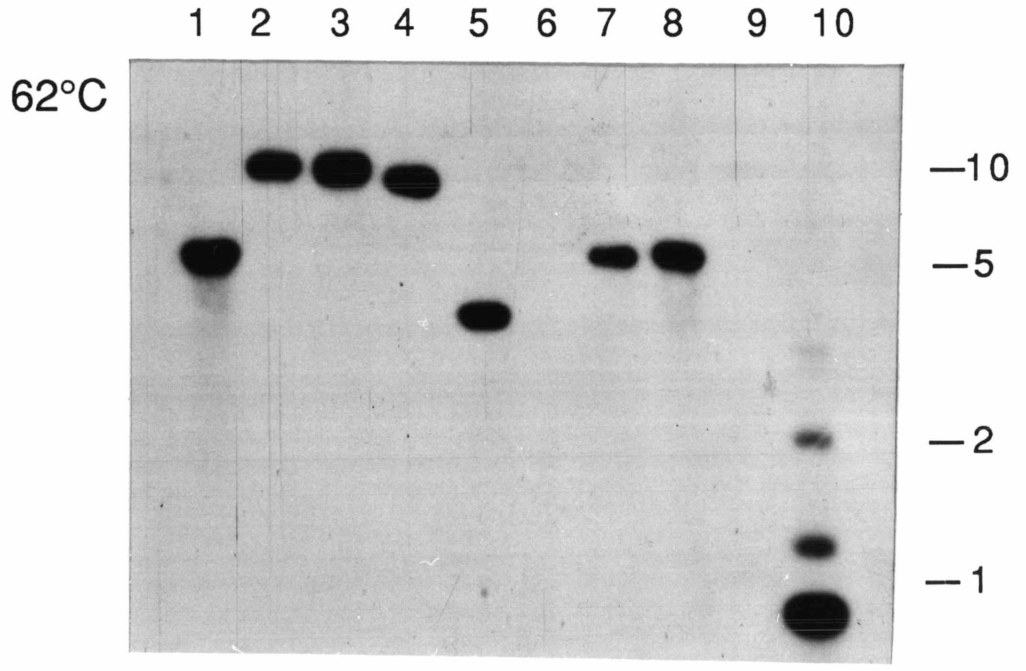
Top photograph shows DNAs after digestion run out on an agarose gel.

Centre is an autoradiograph of a southern blot of the digests, probed with a ^{32}P labelled exon-spanning oligonucleotide (a 34mer corresponding to (ie. same strand) nucleotides 311-344 of the Fig. 4.7 sequence) washed at 47°C in 4xSSC.

Bottom photograph is an autoradiograph of the same blot rewashed at 62°C in 4xSSC. Both autoradiographs were exposed for approximately the same length of time, 50-60 minutes. The strong signal from 3 bands in lane 9 have completely disappeared.


1 2 3 4 5 6 7 8 9 10 M
kb
10
5
2
1


47°C
1 2 3 4 5 6 7 8 9 10
10
5
2
1


62°C
1 2 3 4 5 6 7 8 9 10
10
5
2
1

CHAPTER 6. ANALYSIS OF THE GENOMIC STRUCTURE OF 9-27.

A. INTRODUCTION

Following unsuccessful attempts to demonstrate expression of 1-8 genes from any of the cosmids cloned and previously described, it was decided to concentrate upon obtaining a genomic clone of 9-27. This would be advantageous over further work on the 1-8s as 9-27 is both a unique and expressed gene and so any genomic clone containing a complete gene would also contain a promoter that was known to be both functional and interferon inducible. Furthermore, analysis of mouse L cells show no detectable counterpart of 9-27 mRNA, which would allow a functional analysis of the human gene in these cells in a similar manner to the 6-16 gene (Porter et al, 1988). As previously described, a low background of murine 1-8 was an added complication in the analysis of the 1-8 cosmids.

B. SCREENING FOR A GENOMIC CLONE OF 9-27.

i). A cosmid library.

In collaboration with S.Manly (Primary and secondary screens by S.M., DNA preparations and final analysis by myself), a human genomic cosmid library supplied by John Collins was screened for 9-27. This library had already been used successfully to isolate eight independent 1-8 clones and a functional 6-16 gene (Kelly et al, 1986).

Over 60,000 colonies were grown on three 22 x 22 cm agar plates and duplicate filters screened for 9-27. The probe used was a fragment of 9-27 cDNA which did not contain any homology to 1-8. As this fragment was isolated from a pBR plasmid it is inevitable that some vector sequences will be co-purified and therefore present in the probe. This will cause a background signal during the screening process due to hybridization to pBR derived sequences in the cosmids, which under some circumstances will result in false positives being picked. In order to minimize this possibility, an excess of unlabelled of pBR322 DNA was added to the probe and this mixture was boiled together before being added to the hybridization. The labelled pBR sequences in the probe should rapidly anneal to the unlabelled DNA and therefore be unavailable for hybridization to sequences on the filters.

Twenty two potential positives for 9-27 were identified. The autoradiographs of the filters were lined up with the original plates and plugs of agar containing 10 - 100 colonies pulled from the area of the positive signal. The bacteria from each plug were replated onto 15 cm agar plates at about 1000 colonies per plate. Duplicate lifts onto Biodyne nylon filters were taken from each plate, probed for 9-27, washed at high stringency and exposed for 24 - 48 hours. Positive signals of varying strength were seen and 37 non-independent positives were picked from 19 plates. As most of the positives picked were either single

colonies or a mixture of only 2 or 3 colonies it was decided to grow small cultures of each for a rapid DNA preparation, perform restriction digests of the DNA and run analytical agarose gels that could be blotted and probed. This would allow rapid identification of independent positives and reduce or eliminate the requirement for a tertiary screen.

Initial BamH1 digests of the preparations show that many of the clones picked were identical or related. Eight of them were chosen for further characterization and digested with EcoR1. Two of them, numbers 2 and 19 appeared to be identical, and their 1.5 and 4.1 kb bands were also seen in numbers 5 and 0.9. It was also noted that whilst all the clones contained a vector derived band of 2.4 kb from the thymidine kinase gene of the construct, the larger vector only band was either 6.8 or 8.1 kb. This appears to result from the original vector (Lindenmaier 1982) having two cos sites 1.3 kb apart. Erroneous packaging of both cos sites would result in the recombinant cosmid having the 8.1 kb EcoR1 fragment whilst either correct packaging or deletion of one of the repeated 1.3 kb units which contain the cos sites would result in the smaller fragment. Upon probing a southern blot of these digests, it was revealed that the 4.1 and 1.5 kb bands both cross hybridize to the 9-27 probe. However, the signal was weak when compared to authentic 9-27 (i.e. the probe upon itself). When the filters were stripped and probed with an oligonucleotide

specific for 9-27, no signal was seen from any of the clones. It is possible that the sequences which do cross hybridize to 9-27 are in some way related to the gene, though not as a member of the 1-8 gene family, as these cosmids are not positive for 1-8. The specificity of the cross hybridization to 9-27 was confirmed when during a subsequent independent screen of a fresh pool of the library the only positives obtained were cosmids with apparently identical 4.1 and 1.5 kb cross hybridizing bands (L. Reid, pers. comm.).

ii). A lambda phage library.

A human placental DNA library in the vector Charon 4A that originated in the laboratory of L. Hood was supplied by C. Dickson. Eight 15 cm agar plates were each plated with 50,000 p.f.u. and duplicate lifts probed with both an oligonucleotide (226-AB) which corresponds to residues 50-27 of the opposite strand of the sequence of Fig. 4.5. It was also probed with the longest cDNA of 9-27 labelled using the random priming protocol but replacing the random hexamers with 226-AB. 226-AB corresponds to the non-coding strand of 9-27, with its 3' end at the 25th nucleotide of the cDNA. By using it as a primer on an isolated fragment of the cDNA, the probe produced should be biased in favour of the first 48 nucleotides of this fragment, even though control experiments indicated that non-specific priming events produced a probe containing 1-8 cross-hybridizing

sequences.

Of eight potential positives from the initial screening with the oligonucleotide, none were substantiated as 9-27 positive in the more stringent wash conditions of the secondary screen. The primary screening with the cDNA probe produced fifteen duplicate positives, the thirteen strongest of which were picked. However, of the eight carried through the second round of screening, none were positive for the 9-27 unique (i.e. 3') probe. An earlier independent screening of this library by M. McMahon (pers. com.) also failed to produce a 9-27 positive.

iii). Conclusions.

It is apparent that the 9-27 gene is under-represented in at least two different genomic DNA libraries. Both were independently screened twice and it seems unlikely that a positive clone would have been missed. Both libraries have been used successfully to isolate other genes, including 6-16 (Kelly et al, 1986) and several 1-8s (McMahon 1985) in the case of the cosmid library, and the human homologue to the mouse int-1 gene (C. Dickson, pers com) from the phage. Results presented in sections C and D describe features of the genome in the region of 9-27 that are apparently unstable and which may account for the absence of the gene from these libraries.

C. STUDIES ON A PHAGE GENOMIC CLONE OF 9-27.

i). Introduction.

A single EMBL3 phage clone containing some 9-27 genomic DNA sequences had been isolated in 1984 by R. Friedman, from a human placental DNA library constructed and supplied by J. Nathans at Stanford University (S. Manly, pers comm). This phage, named lambda-g9-27 (lg9-27) contains about 14 kb of human DNA cloned into the BamH1 sites of the vector. These sites are flanked by Sal1 sites, which are rare in human DNA and can therefore be used to excise the cloned sequences from the phage arms as a single or limited number of fragments. Transfection of the phage DNA into mouse L cells (by L. Reid) failed to show any detectable expression of 9-27 mRNA which strongly suggested the absence of a complete and functional gene within these cloned sequences. Results presented below and in section D show that this was due to rearrangements and deletions of the cloned insert, although information obtained from this phage, summarized in Fig. 6.1, was still of significance to these other studies.

ii). Lg9-27 has undergone rearrangements during propagation.

Initial analysis of lg9-27 was performed by restriction digestion using BamH1 and Sal1 singly and together. The BamH1 digestion produces various size fragments of cloned

DNA plus the two phage arms of 9 and 30 kb attached to the cloned DNA up to the first internal BamH1 sites at each end of the insert (Fig. 6.1). These junction fragments are released by the double digestion with Sal1. lg9-27 had 3 BamH1 sites between 1 and 4 kb in size visible on an agarose gel and two other bands of 0.8 and 0.5 kb, detected by their hybridization to a 1-8 cDNA probe. The 3.3 kb band and the right hand junction fragment also hybridized to 1-8. A further band of 0.2 kb was found upon subcloning the BamH1 fragments of this phage.

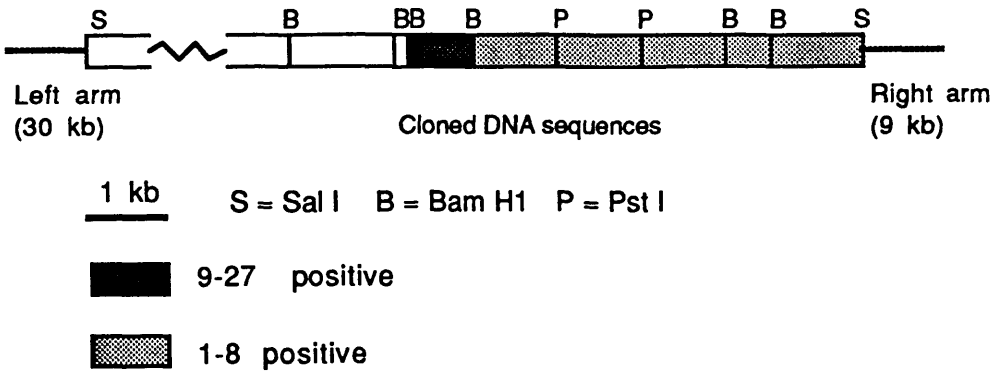
Careful analysis of the phage DNA digested with Sal1 and run out on a low percentage agarose gel revealed that this isolate was a heterogeneous mixture of four different size insert fragments, between about 13 and 17 kb in size. All the isolates appeared to contain the small BamH1 fragments mentioned above, but a different left-hand junction (SalI-BamH1) piece. The most likely explanation is that all four forms originate from a common isolate which underwent rearrangement during propagation.

The phage stock of lg9-27 was replated at low density and individual plaques were picked and grown up on a large scale for the preparation of DNA. Two separate isolates examined showed a left hand Sal1 to BamH1 junction fragment of either 3.1 or 4.1 kb. No apparent heterogeneity was observed in either isolate, which may indicate that both forms of this phage represent stable rearrangements of an unstable parental form.

Figure 6.1.

Summary of the EMBL 3 phage containing 9-27 genomic sequences.

The 9-27 positive region contains the 3' exon of the gene. The 1-8 cross-hybridizing sequences examined both contained 3' ends of 1-8 genes (section C.vi). The left hand junction fragment was found to have undergone rearrangements into four apparently stable forms (section C.ii). The only Pst I sites shown are those within the 3.3 kb Bam H1 fragment.


S
B
BB
B
P
P
B
B
S
Left arm
(30 kb)
Cloned DNA sequences
Right arm
(9 kb)
1 kb
S = Sal I B = Bam H1 P = Pst I
9-27 positive
1-8 positive

iii). The 14 kb insert of lg9-27 could not be subcloned as a single piece.

In order to facilitate further analysis it was attempted to subclone the entire Sal1 insert of lg9-27 into a plasmid vector. This would allow larger quantities of DNA to be easily obtained and more accurate restriction mapping to be performed. Subcloning large pieces of DNA is generally more difficult than subcloning small ones, usually because the efficiency of transformation is reduced (Hardy, 1981), though it is still not uncommon to be able to construct and transform cells with plasmids in the 10 to 20 kb size range. Repeated attempts at subcloning the Sal1 insert of lg9-27 failed to produce any recombinant clones. The methods used were standard DNA preparation and purification procedures which are normally successful. pBR322, pBR328 and pUC18 vectors were cut at the Sal1 site with or without the presence of calf intestinal phosphatase and then gel purified. These were mixed with either total Sal1 cut lg9-27 or the purified insert and ligated under standard conditions. After transformation of competent bacteria fewer colonies were observed than would have been expected from the relevant control experiments. None were ever due to the presence of the required recombinant construct.

Finally it was attempted to subclone a preparatior of the purified Sal1 cut insert into a Charomid vector, Ch 9-34. This vector was developed by I. Saito as a cosmid

type vector containing a variable size stuffer fragment which could be adapted for cloning target DNA sequences of a wide range of sizes (Saito and Stark, 1986). This was also unsuccessful on two occasions.

It was possible to subclone the BamH1 fragments of lg9-27 into pUC18. This was done by ligating total BamH1 digested phage with the cut vector and picking colonies at random. The same was done for the junction BamH1 to Sal1 fragments. This first revealed the existence of the 0.2 kb BamH1 fragment, which was often found as a multiple insert with other size fragments. The main fragments of interest were 0.5, 0.8 and 3.3 kb in size as well as the 1.1 kb BamH1 to Sal1 junction fragment (see Fig. 6.1). All of these were obtained, though the 0.8 kb fragment only after a second screen of colonies of these 'shotgun' clones. The 2.3 kb fragment was not seen in any of the colonies screened, although no rigorous search was performed for it.

iv). lg9-27 has 1-8 related sequences spread over a 5 kb region.

In order to map the BamH1 fragments of lg9-27 a series of BamH1 partial digests were performed upon the phage which had been first digested with an excess of Sal1. These digests were probed with the 0.2 kb BamH1 fragment and the 9-27 cDNA probe which is unique to 9-27, and only hybridizes to the 0.8 kb BamH1 fragment. Information from

these digests produced the map shown in Fig. 6.1, indicating a 5.7 kb region at the right hand end of the cloned DNA containing 1-8 related sequences. This region was separated by nearly 4 kb from the apparently unstable end of the cloned sequences and so was considered worthy of further examination. The large spread of homology to 1-8 indicated an unusual structure of this region of DNA. A full length 9-27 cDNA (Chap. 4.B) indicated that stringent homology to the 1-8 cDNA was limited to a 250 bp region. As there are no BamH1 sites in the 9-27 cDNA this implies that the 250 bp region was spread over possibly four exons (i.e. within the 0.8, 3.3, 0.5 kb BamH1 fragment and the 1.1 kb BamH1 to Sal1 fragment assuming that lg9-27 only contains a single gene. Alternatively, some of the 9-27 sequences may have been duplicated during the rearrangements of the phage, or, a 1-8 gene was also present in this region. This latter hypothesis was to prove correct.

v). The 0.8 kb BamH1 fragment contains the 9-27 3' end.

Sequencing the 0.8 kb fragment revealed 9-27 3' end sequence close to and going away from (i.e. 5' to 3') one end of this subclone. The sequence is in good agreement with the cDNA within the bounds of sequencing error (Fig. 6.2). There is a consensus splice acceptor site near the BamH1 site and although the sequence was not accurately determined further, the sequencing gel indicated that the

Figure 6.2.

Alignment of 9-27 cDNA and genomic 9-27 sequence derived from the 0.8 kb Bam H1 fragment of lg9-27.

The putative splice acceptor site is indicated, and the other differences between the two sequences are marked with asterisks.

```
                10         20         30         40
9-27: TCGAGAAACTGAAACGACAGGGGAAAGGAGGTCTCACTGAGCAC

     50         60         70         80         90
CGTCCCAGCATCCGGACACCACAGCGGCCCTTCGCTCCACGCAGAAAACC

     100        110        120        130        140
ACACTTCTCAAACCTTCACTCAACACTTCCTTCCCCAAAGCCAGAAGATG

     150        160        170        180        190
CACAAGGAGGAACATGAGGTGGCTGTGCTGGGGGCACCCCCCAGCACCAT

     200        210        220        230        240
CCTTCCAAGGTCCACCGTGATCAACATCCACAGCGAGACCTCCGTGCCCG

     250        260        270        280        290
ACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCTTGAACTGGTGCTGT

     300        310        320        330        340
CTGGGCTTCATAGCATTCGCCTACTCCGTGAAGTCTAGGGACAGGAAGAT
Genomic sequence:        TCCCCCAGTCTAGGGACAGGAAGAT
                         *SPLICE*

     350        360        370        380        390
GGTTGGCGACGTGACCGGGGCCCAGGCCTATGCCTCCACCGCCAAGTGCC
GGTTGGCGACGTGACCGGGGCCCAGGC-TATGCCTCCACCGCCAAGTGCC
                           *

     400         410        420        430        440
TGAACATCT-GGGCCCTGATTCTGGGCATCCTCATGACCATTGGATTCAT
TGAACATCTTGGGCCCTGATTCTGGCCATCCTCATGACCATTGGATTCAT
         *                *

      450        460        470        480        490
CCTGTCACTGGTATTCGGCTCTGTGACAGTCTACCATATTATGTTACAGA
CCTGTTACTGGTATTCGGCTCTGTGACAGTCTACCATATTATGTTACAGA
     *

      500        510        520        530        540
TAATACAGGAAAAACGGGGTTACTAGTAGCCGCCCATAGCCTGCAACCTT
TAATACAGGAAAAACGGGGTTACTAGTAGCCGCC-ATAGCCTGCA
                                    *

      550        560        570        580        590
TGCACTCCACTGTGCAATGCTGGCCCTGCACGCTGGGGCTGTTGCCCCTG

      600        610        620        630        640
CCCCCTTGGTCCTGCCCCTAGATACAGCAGTTTATACCCACACACCGTGC

      650
TACAGTGTCATTCAATAAA
```

sequence was consistent with the cDNA at least to the AATAAA polyadenylation signal at the 3' end. In order to orientate this sequence relative to the other 1-8 sequences, the BstX1 sites of the 0.8 and 3.3 kb fragments were mapped and compared to a BstX1 digest of lg9-27. The 0.8 kb fragment is cut into two pieces of 220 and 550 bp whilst the 3.3 kb fragment has 4 BstX1 sites, with end fragments of 320 and 800 bp. Probing the BstX1 digest of the phage with 9-27 produces hybridization to a 870 bp fragment, which can only be a combination of the 550 and 320 bp BstX1 to BamH1 pieces. This demonstrates that 9-27 is downstream of the other 1-8 sequences within lg9-27.

vi). lg9-27 contains two 1-8-like 3' ends.

Probing BstX1 digests of the 3.3 kb BamH1 fragment with 1-8 and the full length 9-27 probes revealed a region of this fragment that was specific to 1-8 . This must be due to sequence outside the common region of the two cDNAs, which is mainly in their 3' ends. After mapping the 3.3 kb fragment, sequence was obtained from the right hand Pst1 site going towards the right arm of the phage. This revealed 3' sequence with very high homology to the 1-8 cDNA (Fig. 6.3)

The 0.5 kb BamH1 fragment was also sequenced. This too had sequence closely related to 1-8, though unlike the cDNA it had no tandem repeat of 68 bases in the 3' region (Fig. 6.4). This sequence also has a consensus splice acceptor

in exactly the equivalent part of its sequence as the 3' exon of 9-27, and differs by only one nucleotide to 1-8 over the first 44 bases of the exon. This illustrates a remarkable degree of conservation between three related but nonetheless distinct genes. The orientation of this second 1-8 exon relative to the other two 3' ends was not determined.

vii). No 5' region of 9-27 was detected upstream of the 3' exon.

The subcloned 3.3 kb BamH1 and 1.1 kb BamH1 to Sal1 fragments from lg9-27 were probed with two different oligonucleotides. 498-AB is a 21mer of the sequence 5' to the start of the 3' exon of 1-8 and 9-27, while 492-AB is a 18mer of 9-27 sequence further upstream, confirmed in two independent cDNA clones but not found in 1-8. Both sequences were used to probe the two fragments that are presumed to be 5' to the final exon of 9-27. Whilst 498-AB was positive for the large BstX1 fragment of the 3.3 kb BamH1 region where homology to both 1-8 and 9-27 had previously been detected, both subclones were negative for the 9-27 specific oligonucleotide (Fig. 6.5)

viii). Conclusions.

The results described in sections vi and vii above about the relative positions of 9-27 and 1-8 genes are almost certainly due to rearrangements within the cloned DNA

Figure 6.3.

The 3' portion of the 1-8 cDNA sequence (upper sequence) aligned with genomic sequence obtained from the 3.3 kb Bam H1 fragment of lg9-27.

```
       360        370        380        390        400
AGATCAGGAGGCATCATTGAGGCCTGGAGCTCT-GCCCGTGACCTGTATC
      :::::::::::::::::::::::::::: ::::::::::::::::
      GGAGGCATCATTGAGGCCTGGAGCTCTTGCCCGTGACCTGTATC
                10         20         30         40

          410        420        430        440
CCAC--GTACTCTATCTTCCATTCCTCGCCCTGCCCCCAGAGGCCAGGA-
::::  :::::::::::::::::: ::::::::::::::::::::::::
CCACACGTACTCTATCTTCCATTC-TCGCCCTGCCCCCAGAGGCCAGGAA
      50         60         70         80         90


 450        460        470        480        490
GCTCTGCCCTTGA-CCTGTATTCCACTTACTCCACCTTCCATTCCTCGCC
:::::::::::::: ::::::::::::::::::::::::::::::::::
GCTCTGCCCTTGAACCTGTATTCCACTTACTCCACCTTCCATTCCTCGCC
       100        110        120        130        140

  500        510        520        530        540
CTGTCCCCACAGCCGAGTCCTGCATCAGCCCTTTATCCTCACACGCTTTT
:::::::::::::::::::::::::::::::: ::::::::::::::::
CTGTCCCCACAGCCGAGTCCTGCATCAGCCCTCTATCCTCACACGCTTTT
       150        160        170        180        190

  550        560
CTACAATGGCATTCAATAAAC
::::::::::::::::::::
CTACAATGGCATTCAATAAAC
       200
```

Figure 6.4.

Alignment of 1-8 cDNA sequence (upper sequence) with the seqence derived from the 0.5 kb Bam H1 fragment of lg9-27.

A consensus splice acceptor site is underlined. The genomic sequence is missing the 68 base-pair double repeat found in the cDNA (compare nucleotides 374-441 with 442-509).

```
       160       170       180       190       200
CTTCATGAACACTGCTGCCTGGGCTTCATAGCATTCGCCTACTCCGTGAA
                :  :     ::    ::    :      :
                GGATCCTGGTCCCCTCACCATCTCCTCTCCCCCA
                          10        20        30

       210       220       230       240       250
GTCTAGGGACAGGAAG-ATGGTTGGCGACGTGACCGGGGCCCAGGCTATG
:::::::::::::: : ::::::::::::::::::::::::::::::::
GTCTAGGGACAGGACGAATGGTTGGCGACGTGACCGGGGCCCAGGCTATG
     40        50        60        70        80

         260       270       280       290
CTCCACCGCCAAGTGCCTGAACATCTGGGCCCTGATTT-TGGGCATCTTC
:::::::::::::::::::::::::::::::::::::  :::::::: ::
CTCCACCGCCAAGTGCCTGAACATCTGGGCCCTGATTCCTGGGCATCCTC
     90        100       110       120       130

         310       320       330       340
ATGACCATTCTGCTCATCATCATCCCAGTGTTGGTCGTCCAGGCCCAGCG
:::::::::::::::::: :::::::::::: :: :: :::::::: :  :
ATGACCATTCTGCTCATCGTCATCCCAGTGCTGATCTTCCAGGCCTATGG
     140       150       160       170       180

         360       370       380       390
ATAGATCAGGAGGCATCATTGAGGCCTGGAGCTCTGCCCGTGACCTGTAT
:::::::::::::::::: ::::::::::::::::::::::::::::::
ATAGATCAGGAGGCATCACTGAGGCCTGGAGCTCTGCCCGTGACCTGTAT
     190       200       210       220       230

         410       420       430       440
CCCACGTACTCTATCTTCCATTCCTCGCCCTGCCCCCAGAGGCCAGGAGC
:::::::::: : ::::::::::::::::::::::: :::
CCCACGTACTCCAACTTCCATTCCTCGCCCTGCCCCCGGAG---------
     240       250       260       270

         460       470       480       490
TCTGCCCTTGACCTGTATTCCACTTACTCCACCTTCCATTCCTCGCCCTG

--------------------------------------------------

         510       520       530       540
TCCCCACAGCCGAGTCCTGCATCAGCCCTTTATCCTCACACGCTTTTCTA
         :::::::::: ::::::::::::::::::::::::::::::
---------CCGAGTCCTGTATCAGCCCTTTATCCTCACACGCTTTTCTA
             280       290       300       310

       560
CAATGGCATTCAATAAAG
:::::::::::::::::
CAATGGCATTCAATAAATGCACGTTTCTGGTGCTGCTGCGACTTCACCTG
   320       330       340       350       360
```

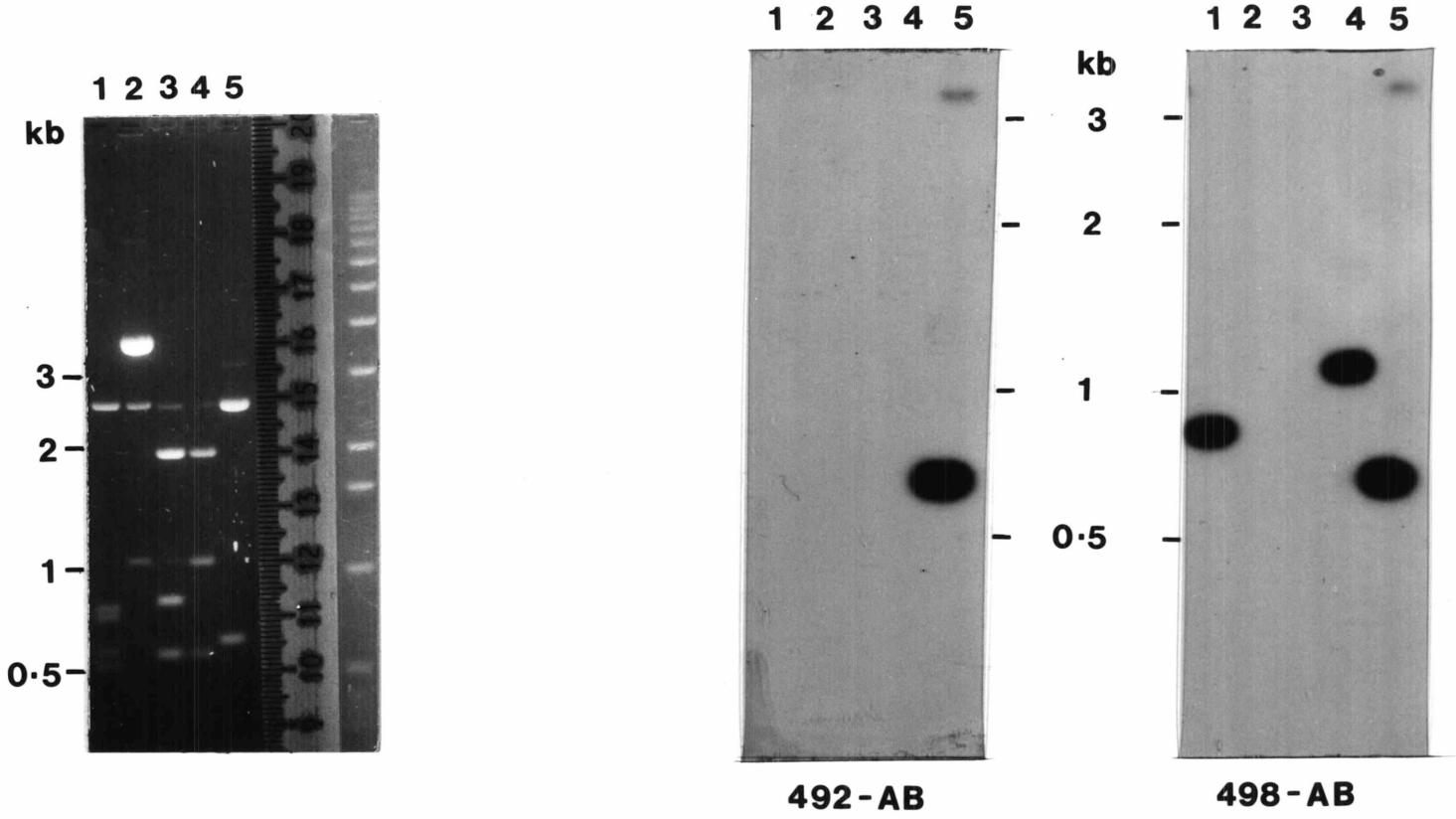
Figure 6.5.

Experiment using oligonucleotide probes to 3' and 5' sequence of 9-27.

Left hand photograph shows agarose gel of electrophoresis of various digests of fragments subcloned from lg9-27.
1 = 3.3 kb Bam H1 fragment in pUC18; Bam H1 + BstX1 digest.
2 = 1.1 kb fragment in pUC18 (upstream of the 9-27 3'end); Bam H1 + SalI digest.
3 & 4 = two different Bam H1-PstI subfragments of the 3.3 kb fragments (see Fig 6.1) in pUC18; Bam H1 + PstI + SspI digest.
5 = 9-27cDNA clone in pUC18; Eco R1 digest.

Middle photograph shows an autoradiograph of a blot of the gel probed with an oligonucleotide specific for 9-27 corresponding to sequence upstream of the final exon.

Right hand photograph shows an autoradiograph of the same blot reprobed with a 3'oligonucleotide.


1 2 3 4 5
kb
3
2
1
0·5
1 2 3 4 5
kb
3
2
1
0·5
492-AB
1 2 3 4 5
498-AB

during its growth as a recombinant phage. However, valid conclusions can be drawn from some of the data shown concerning the conserved sequences and exon structures of 9-27 and 1-8 (at least in their 3' regions) as well as establishing that 9-27 is in close proximity to at least one 1-8 gene within the human genome. These were both relevant pieces of information for subsequent studies.

D. STRUCTURE OF A FUNCTIONAL 9-27 GENOMIC CLONE.

i). Introduction.

A cosmid containing 9-27 genomic sequences within a 40 kb cloned insert was obtained from a library screened by L. Reid. The library was supplied and constructed by P. Little using a vector, pLORIST B, which contains a lambda origin of replication instead of the more usual plasmid ColE1 origin. The variability of growth of cosmids containing this plasmid origin was noted during the preparation of DNA from six cosmids containing 1-8 genes when yields varied over a 10 fold range. In some cases this could lead to the elimination of some cloned sequences during the growth and amplification of the library, while the lambda origin produces a cosmid copy number in each bacterial cell that is not affected by the cloned sequences it carries (Cross and Little 1986, Little and Cross, 1985, 1987). This cosmid, designated 9-27Cos3, was the only independent positive for 9-27 found in the library.

ii). 9-27Cos3 contains 3' and 5' 9-27 sequences.

Cos 3 was obtained by screening the library with the 9-27 specific cDNA probe which identifies the 9-27 3' end. To determine whether this cosmid also contained sequences from the 5' region of 9-27 a series of restriction digests were performed and blotted onto nitrocellulose. They were probed with 226-AB, a 24mer of sequence from the 5' end of the longest 9-27 cDNA. A single band in each track was positive for the oligonucleotide (Fig. 6.6a), indicating the presence of sequence close to the 5' end of the transcribed gene.

iii). Two 1-8 genes are also within Cos3.

Oligo 226-AB also indicated the presence of two regions of DNA with closely related sequences, which were detected under conditions of relatively high stringency for this probe (Fig. 6.6b). There is some homology between the most 5' sequences of the 1-8 cDNA and 9-27, so it is possible that this extends further upstream and results in the bands observed being due to the presence of two nearby 1-8 genes. Probing Pst1 and Hind III digests of Cos3 with a series of oligonucleotide probes established the presence of two 1-8 genes with both 5', middle (i.e. common to 9-27) and 3' sequences.

Figure 6.6.

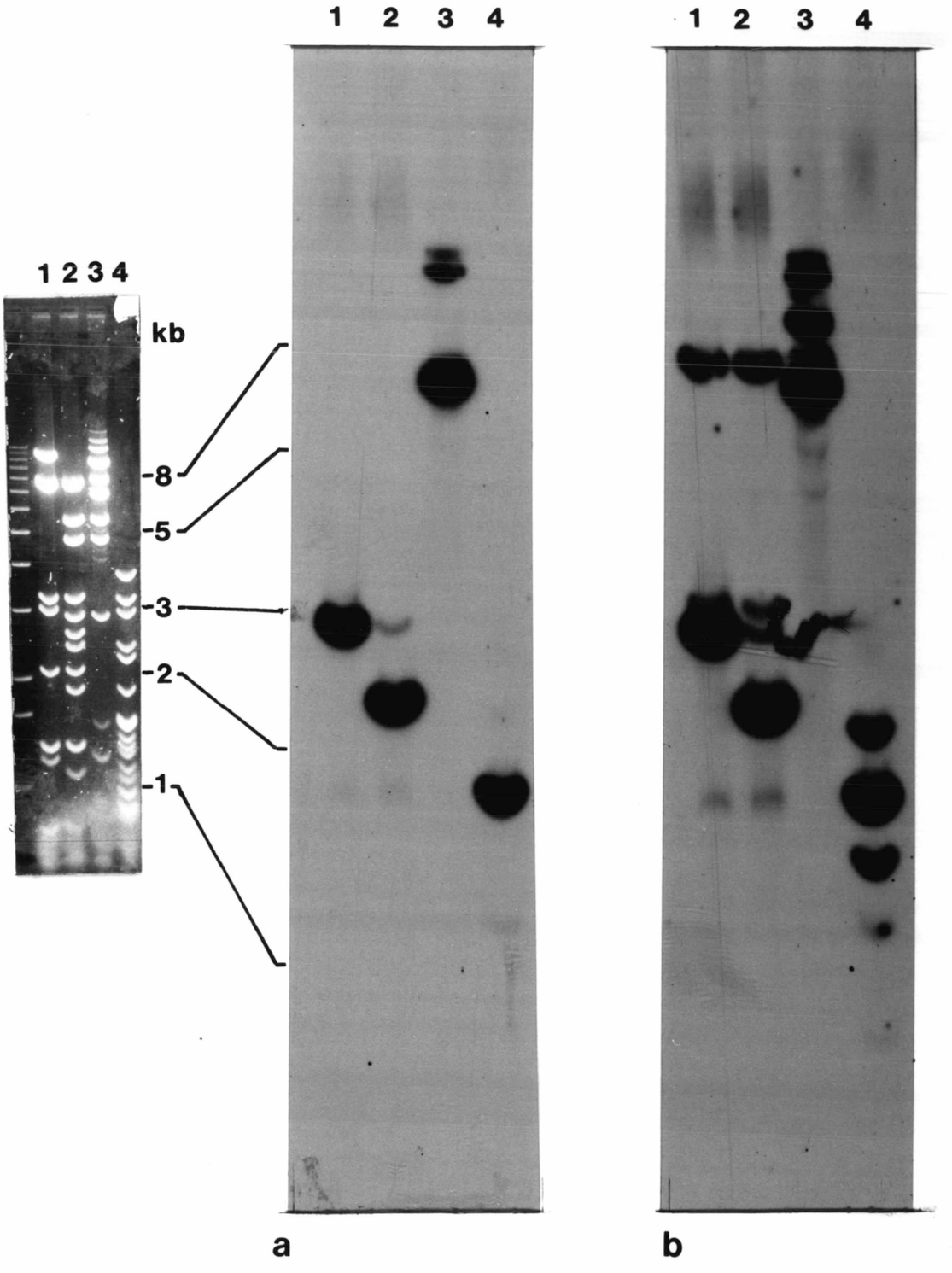
Experiment showing presence of two 1-8 genes in 9—27Cos3.

Left hand photograph shows an agarose gel of electrophoresis of 9-27Cos3 DNA digests.

1 = Bam H1; 2 = HindIII + Bam H1; 3 = HindIII; 4 = Pst1.

a). Autoradiograph of blot of the gel probed with 226-AB, a 9-27 5'-end probe, high stringency wash (65°C, 2xSSC).

b). Low stringency was of the blot (50°C, 2xSSC) revealing cross-hybridization of the probe to two other regions of Cos3.


1 2 3 4
kb
8
5
3
2
1
1 2 3 4
a
1 2 3 4
b

iv). The 9-27 gene is located between the two 1-8 genes.

An initial map of the cosmid was obtained by digests with enzymes that cut infrequently in human DNA. Not1 and Nru1 sites contain CpG dinucleotides, which are under-represented in human DNA (Bird 1985), while Sfi1, like Not1 has an 8 base recognition sequence and so also cuts rarely. A single Nru1 site (plus one in the vector), two Not1 and three Sfi1 sites were found in the cloned sequences. Cutting, blotting and probing with 9-27 unique sequences digests of Cos3 with these enzymes and Hind III produced an initial map shown in Fig. 6.7. The 9-27 gene is situated between a Hind III and Nru1 site within an 8.9 kb Hind III fragment. This is flanked by the other two large Hind III fragments of 7.7 and 11.1 kb which each contain a 1-8 gene, designated 1-8.left and 1-8.right respectively.

The observation of two Not1 sites spaced only 8 kb apart is worth noting. The frequency of Not1 sites in the human genome is of the order of one every one hundred kb, though they are not randomly distributed, with the majority of sites occurring in htf islands (A. Bird, pers. comm.). These are CpG rich regions of DNA found near many actively transcribed genes, and the presence of the Not1 sites maybe indicative of two htf islands within Cos3.

v). The 8 kb Not1 fragment could not be subcloned.

Two attempts were made to subclone the 8 kb Not1

Figure 6.7.

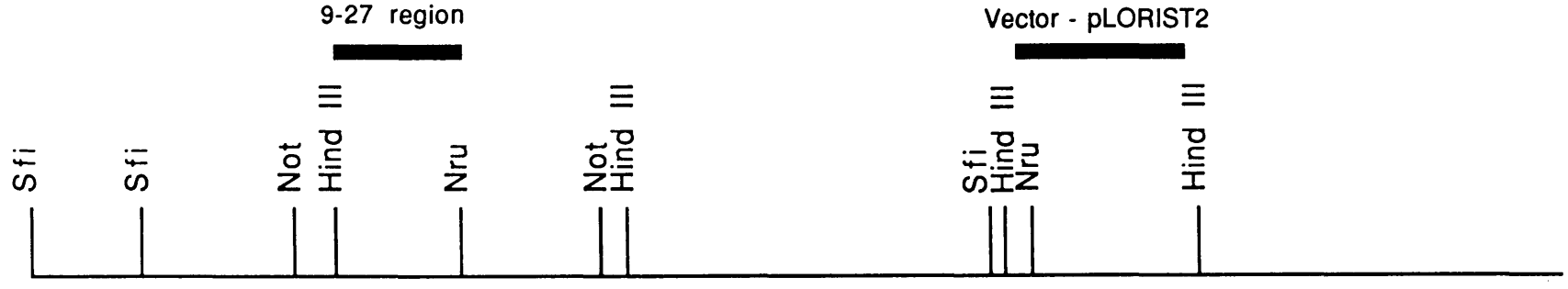
Map (only partial for HindIII) of 9-27Cos3. The 9-27 region indicated contains sequences corresponding to both 5' and 3' regions of 9-27 cDNA.

9-27 Cos 3

44 Kb

(Only partial Hind III map)

1cm = 2 kb



fragment containing 9-27. The first experiment was a blunt-end cloning reaction into the Sma1 site of pUC13 with the Not1 ends of the 8 kb fragment being filled in with reverse transcriptase. No positive results were obtained by this method, and a second attempt using the vector pKSN13+ (Stratagene) which has a Not1 site in its polylinker (enabling a 'sticky end' ligation reaction) was also unsuccessful.

These results were reminiscent of the lack of success in the attempts to subclone the 14 kb Sal1 fragment of lg9-27, and strongly suggests the existence of an 'unclonable' or poison sequence near the 9-27 gene. This sequence is obviously viable in Cos3, which may be due to the phage origin of replication or to a more subtle positional effect, either relative to the origin (at least 12 kb away) or to other sequences within the clone.

vi). A 2 kb Pst1 fragment contains a functional 9-27 gene.

A 2 kb Pst1 band that contains both 9-27 5' and 3' sequences (identified using techniques described in sections ii and iii) was cut out of a gel, electroeluted and subcloned into the Pst1 site of pUC19. Surprisingly, when the resulting plasmid was prepared and transfected into mouse L-cells (in collaboration with L. Reid) it was found to produce an alpha interferon inducible 9-27 mRNA of the correct size.

vii). Cos3 contains the same 9-27 3' end as lg9-27.

A single BamH1 site was found near the centre of the 2kb Pst1 9-27 fragment of Cos3. The 3' BamH1 to Pst1 fragment of this subclone was subcloned between the BamH1 and Pst1 sites of M13mp18 and mp19 and sequenced using both M13 and internal primers. The sequence obtained agrees with the 9-27 cDNA sequence and contains the same splice acceptor that was found in lg9-27. This confirms the structure of the final exon of this gene.

viii). All other 9-27 mRNA sequences are contiguous.

The 5' BamH1 to Pst1 region of the 2 kb Pst1 fragment was also subcloned into M13mp18 and mp19. This was sequenced using a range of internal primers and revealed a stretch of sequence contiguous from the most 5' nucleotide of the longest cDNA of 9-27 down to the sequence immediately adjacent to the 5' end of the final (3') exon. This sequence was in exact agreement to the cDNA. The genomic sequence to the start of an intron is presented in Figure 6.8. There is no consensus splice acceptor site in the sequence to the upstream Pst1 site and it is unlikely that there is a further exon in the 9-27 gene.

ix). 9-27 has sequence homologies to other alpha-interferon induced genes.

A 13 bp sequence is found in the 5' flanking regions of

Figure 6.8.

Sequence of the 9-27 gene from its 5' flanking region through to the end of the first exon.

The first of several transcription start sites (*) is designated as +1. Two upstream restriction enzyme sites are indicated, as are two putative "Cat" boxes and a consensus interferon response element (IRE). A second IRE-like element is also present near the start of transcription of the gene.

The sequence immediately following the end of the exon, GTGCG, is a classic consensus splice donor sequence.

```
     -400       -390       -380       -370       -360
CCCGCTCTGAGCAGGACTGCAGCCTGAGGAAAGAGCAAGGATTTCAGGAG
                  PstI
     -350       -340       -330       -320       -310
AGAGGCCTGCGACAAGTGAGGTGAGGGCTTTTGGGGGATTGTCCTGGCGC

     -300       -290       -280       -270       -260
CTGGAGTGTGGGGGCCTGGCAGGGGCCCTGAACCGGGACAGTGAGGTCCT

     -250       -240       -230       -220       -210
GTACTTGCTGGCCTGGGGTGGACAGAACCCAAGCCTGGGTAAAATCCCCA

     -200       -190       -180       -170       -160
GGGCCCCCTAAACCCAACACTTAGGAAGTCACTAGTCCTGACTTCACTTC
                                  SpeI
     -150       -140       -130       -120       -110
TGATGAGGAAGCCTCTCTCCTTAGCCTTCAGCCTTTCCTCCCACCCTGCC

     -100       -90        -80        -70        -60
ATAAAGTAATTTGATCCTCAAGAAGTTAAACCACACCTCATTGGTCCCTG
                                          Cat box
     -50        -40        -30        -20        -10
CTAATTCACCAATTTACAAACAGCAGGAAATAGAAACTTAAGAGAAATAC
       Cat box                 IRE
        *         10         20         30         40
ACACTTCTGAGAAACTGAAACGACAGGGGAAAGGAGGTCTCACTGAGCAC
           IRE-like
     50         60         70         80         90
CGTCCCAGCATCCGGACACCACAGCGGCCCTTCGCTCCACGCAGAAAACC

     100        110        120        130        140
ACACTTCTCAAACCTTCACTCAACACTTCCTTCCCCAAAGCCAGAAGATG
                                                 Met
     150        160        170        180        190
CACAAGGAGGAACATGAGGTGGCTGTGCTGGGGGCACCCCCCAGCACCAT

     200        210        220        230        240
CCTTCCAAGGTCCACCGTGATCAACATCCACAGCGAGACCTCCGTGCCCG

     250        260        270        280        290
ACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCTTGAACTGGTGCTGT

     300        310        320
CTGGGCTTCATAGCATTCGCCTACTCCGTGAAG
```

several alpha-interferon induced genes and is the major sequence element controlling the elevated transcriptional response of these genes (Reich et al,1987; Porter et al,1988). This consensus is also present in 9-27, between 30 and 18 bp upstream of the start of the longest cDNA, and 32 and 12 bp downstream of the two consensus (CCAAT) 'cat' boxes. There is also an eleven bp sequence at residues 6 to 16 of the cDNA that is a match to the central eleven nucleotides of this consensus. These comparisons are presented in figure 6.9.

x). CAT assays with the 5' region.

The 230 bp region downstream from the Spe1 site of the 9-27 5' flanking region (see Fig. 6.8) was cloned into the Hind III site of pSV0CAT (Gorman, 1982a), upstream of the chloramphenicol acetyl transferase (CAT) gene. This region contains 100 bp of sequence upstream of the most 5' of the two consensus CCAAT boxes, the consensus interferon response element (IRE) and 60 bp of sequence downstream of the start of the cDNA. Constructs containing a single copy of this element in either the sense (i.e. 5' to 3' upstream of CAT) and the antisense orientation were transiently transfected into HeLa cells, which were analyzed for CAT activity by the method of Sleigh (1986) 24 hours after the addition of interferon. The results obtained indicated that this 230 bp of DNA (only in the correct orientation) confers both an alpha and a gamma interferon response to

Figure 6.9.

Homologies between sequences flanking human interferon inducible genes.

Gene	I.R.E. Sequence	Position
9-27	**A G G A A A T A G A A A C T**	-31 to -19
6-16	**G** * * * * * **A T** * * * * * *	-112 to -99
ISG 54	**G** * * * * * **G T** * * * * * *	-87 to -100
ISG 15	**G** * * * * * **C C** * * * * * *	-108 to -95
IP 10	**T** * * * * * **G T** * * * * * **C**	-222 to -209
Factor B	* * * * * * **C** * * * * * * *	-127 to -140
Synthetase	* * * * * * **A** - * * * * * **A**	-100 to -87

the CAT gene. However, the level of induction is low compared to that seen for the native gene and there is a basal level of transcription that increases with the quantity of DNA transfected (see Fig 6.10). It is possible that there are elements upstream of the Spe1 site that would reduce this through some form of transcriptional repression. These types of considerations are difficult to rule out, though the results obtained indicate the presence of significant IREs for both alpha and gamma interferon between the Spe1 and BspM II sites. It is difficult to imagine that the consensus IRE described in section ix above does not play a major role in the response to α-interferon.

xi). 9-27 has multiple start sites.

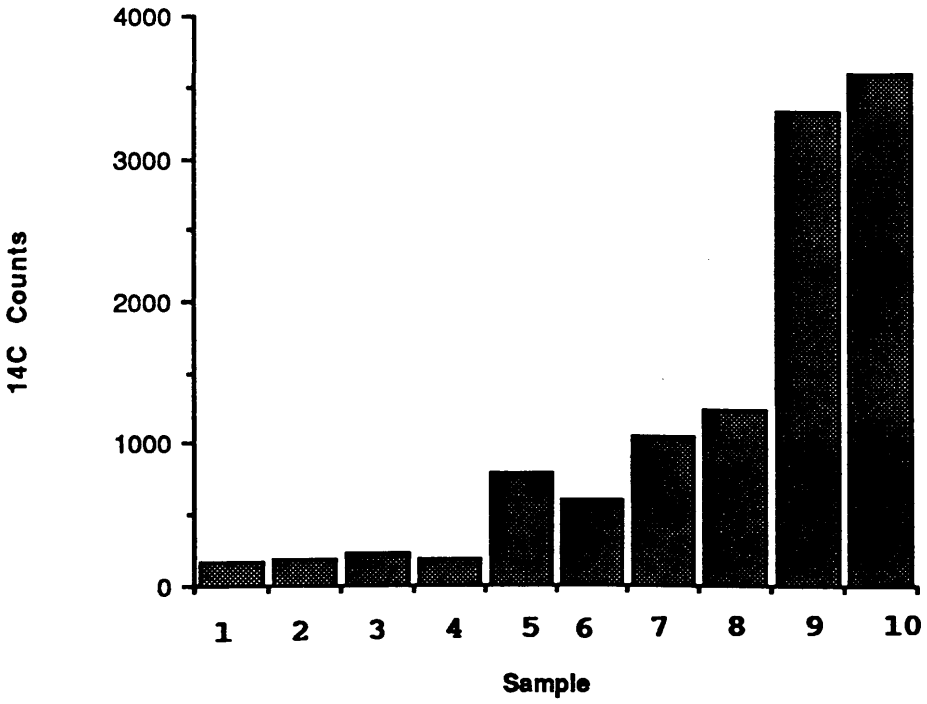
Primer extension analyses (results not shown) indicate the presence of multiple start sites of transcription for 9-27. The primer extension was performed on total HeLa RNA from cells treated with alpha-interferon for 24 hours and an untreated control, using oligonucleotide 492-AB (corresponding to residues 154 to 137 of the opposite strand of the Fig 4.5 sequence) as a primer. A plasmid sequencing reaction using the same oligonucleotide on the 2 kb Pst1 subclone was run in parallel and used as a marker. Two potential major interferon induced bands were seen on the extension analysis which corresponded to start sites 10 and 29 nucleotides downstream of the longest cDNA. This

Figure 6.10.

CAT gene expression under the control of a 230 bp region of the 9-27 promoter.

Lanes 1-2 = control (no DNA).
Lanes 3-6 = 0.5μg pSV0CAT/9-27 promoter plasmid DNA transiently transfected into $2x10^5$ HeLa cells and either untreated (3,4) or treated (5,6) with 300U/ml α-interferon.
Lanes 7-10 = as 3-6, but using 2.5μg of plasmid DNA.

The CAT activity was measured using the method of Sleigh (1987) and provides a relative measure of ^{14}C counts.


4000
3000
2000
1000
0
14C Counts
1
2
3
4
5
6
7
8
9
10
Sample

was confirmed by an S1 analysis which indicated two major (and several minor) start sites in the vicinity (+/- 5 nt) of these residues (L. Reid, pers. comm.).

xii). Sequence homologies to other genes.

Two homologies of possible importance were noted between 9-27 and two other interferon induced genes (Fig. 6.11). Both the homologies described are interesting not only for the direct sequence comparison but also because of they occur in equivalent regions of the genes.

2'-5'A synthetase has multiple start sites of transcription (Revel, pers. comm.) and no definable 'TATA' box. An 18 bp sequence upstream of the majority of these start sites (though containing at least one) bears a striking similarity to the sequence in an equivalent position in 9-27 (One gap in each sequence is required for perfect alignment).

A 14 bp stretch of homology between the two 'CAT' boxes shows a 12 bp identity with sequence immediately downstream of the 'CAT' box of IP 10, a gamma interferon induced gene (Luster and Ravetch 1987). The sequence in IP 10 is orientated in the opposite direction with respect to the start of transcription of the gene, though many defined DNA response elements (such as the alpha-IRE) are effective irrespective of their orientation. The homology is in a form of a block of five, a gap of two (in both cases Gs for Cs) and a block of seven bases, reminiscent of the

structure of the alpha-IRE. The significance of this sequence, if any, is currently being investigated.

No significant homologies were found between 9-27 and the HLA class II sequences published by Miwa et al(1987) or Dorn et al(1987).

xiii). Searching the Genbank Database.

The Genbank DNA sequence database, release number 50, was searched with the 5' genomic sequence of 9-27 from the Pst 1 to the BspMII sites. No significant homologies to either strand were found.

Figure 6.11.

Homologies between the 9-27 gene and two other interferon inducible genes, 2'-5'A synthetase and IP10. The homology found to IP10 is in the region of the "Cat" boxes of each gene, although, as indicated, these boxes are in opposite orientations.

Synthetase	**CCCTTCTGAGGAAAC-GAAACCAACAG**	-106 to -83
	: ::::::: ::::: ::::: ::::	
9-27	**CACTTCTGA-GAAACTGAAACG-ACAG**	-5 to +20

9-27	**TCCCTGGCTAATTCA-CCAAT**	-63 to -44
	::::: ::::::: **X**	
IP 10	**TCCCTCCCTAATTCTGATTGG**	-45 to -65

CHAPTER 7. DISCUSSION.

A. OVERVIEW.

The interferons were initially characterized as proteins with the ability to confer an antiviral state upon treated cells. Subsequently, they were also found to act as antiproliferative and immunomodulatory agents. It is clear that such a diverse range of actions is mediated by the induction of a number of mRNAs (reviewed by Revel and Chebath, 1986) whose translation products are responsible for these effects. Analysis of the function of these proteins, and the analysis of their mechanisms of induction are therefore essential to the understanding of the mechanisms of interferon action, and the cloning of both cDNAs of the induced mRNAs, as well as genomic copies of these genes are important steps to achieving this information. A complete and accurate sequence of these genes, enabling the prediction of the primary protein sequence could be used to over- or under-express normal and mutated forms of proteins to asses the importance of individual proteins (or subregions of them) in the antiviral, antigrowth and immunomodulatory effects of interferon. Antibodies made against peptides of the predicted protein sequences could be used to address similar questions, and reveal the cellular localization and interactions between the gene product under study and other proteins in the cell.

In addition to cDNA copies of these genes, the isolation of genomic clones allows the study of the mechanism of induction of genes. Linking regions of normal and mutated promoter of interferon-inducible genes to a heterologous marker gene enables the regions of the promoter responsible for the induction (or repression) of these genes to be characterized. This in turn should allow the study of factors within cells which bind to or activate these regions. This would be of considerable value in understanding the events which occur within a cell between the moment interferons bind to their receptors on the cell surface and the resulting enhanced transcription of the inducible genes.

The primary aim of the work presented in this thesis was to obtain a complete and accurate sequence of an interferon-inducible gene, namely 9-27. The isolation and initial characterization of a cDNA and a genomic DNA clone was successful and allowed some preliminary work on defining a limited region of the 5' flanking sequence responsible (at least in part) for the inducibility of 9-27. Other work described, involving the use of oligonucleotide probes of potential "consensus sequences", though unsuccessful, provided informative results relating to the feasibility of such schemes in general and provided useful technical experience of the handling and use of such probes for other

purposes. Experiments on a panel of cosmids containing 1-8 related sequences and the discovery of two functional 1-8 genes in close proximity to 9-27 was surprising and serves to demonstrate many salient features of the structure and complexity of the human genome.

B.THE 9-27/1-8 GENE FAMILY.

i). 1-8 Genes and Pseudogenes.

It is apparent that many genes are not distributed at random throughout the eukaryotic genome but occur in clusters which contain several related genes. For example, one of the best characterized clusters is the human β-globin gene family where five functional genes and one pseudogene are found over a 50 kb region of chromosome 16. All six genes are related and almost certainly are the descendants of a single primordial parent, and the study of this system has provided many insights on the mechanisms of evolution, development and gene regulation. There are mow many examples of such clusters including the immunoglobulin, HLA, complement, collagen, actin, tubulin, crystallin and α-interferon gene families. The 9-27/1-8 gene family described in the results here show similarities to all these gene families.

Besides Cos3, the other 1-8 cosmids studied all contained a single 1-8-related sequence. There are two

lines of evidence to suggest that these are non-functional pseudogenes; firstly the absence of any detectable expression of 1-8 from these cosmids in both mouse (Chapter 5.B) and hamster cells, and secondly, the hybridization of the "exon-spanning" oligonucleotide to six of these cosmids. This latter experiment indicates that the cosmids are likely to contain processed pseudogenes, which are be found throughout the genome at random, in contrast to the non-functional intron-containing pseudogenes associated with the globin and histocompatibility gene systems.

The results from Cos3 using the 1-8 3' probe and a range of oligonucleotide probes suggest that the two 1-8 genes found in this cosmid are very similar to each other. The two 1-8 genes flank the 9-27 gene and are transcribed in the same direction, which is the opposite direction from 9-27. This is in contrast to the human globin gene clusters where the expressed genes are all transcribed in the same direction. One possible sequence of events which could have produced such a cluster would be an initial duplication of an ancestral gene to produce two "back-to-back" genes which evolved into the present 9-27 gene and a first 1-8 gene. After the two genes had diverged, a second event gave rise to the second 1-8 gene. It is possible that exchange of information between the two 1-8 genes has occurred by gene conversion in order that their similarity has been maintained.

It is interesting to note that there are apparently no 9-27 pseudogenes. It seems not unreasonable to suggest that the 1-8 pseudogenes arose through the integration of mRNA copies of 1-8 into the genome. This is more likely to have occurred in a cell in which 1-8 was being expressed at high levels, e.g. in response to interferon. Since 9-27 is expressed in a similar manner to 1-8 and with similar kinetics, the absence of 9-27 pseudogenes might be explained by the hypothesis that 1-8 pseudogenes arose before the duplication event that led to the formation of 9-27.

It is also worth noting that there are no good hypotheses to suggest why there is a need for two functional 1-8 genes. There may of course be subtle differences in their mode of induction or action which are not yet discovered. Alternatively, the two genes may be functionally equivalent. If this were true, the 1-8 gene cluster may represent a gene cluster in an early evolutionary stage where, although there is no requirement for two 1-8 genes, there have been no mutational events which have resulted in one or other of these genes becoming a non-functional pseudogene.

ii). Total number of 1-8 genes.

Digests of total human DNA which are blotted onto a filter and probed with the original 1-8 cDNA indicate that there may be 10 to 20 different copies of 1-8 sequences in

the genome. It is difficult to be more precise, given that the range of intensity of the signal from each band on the filter may either represent more than one copy of the gene or merely reflect the degree of hybridization between the probe and the 1-8 homologue detected.

The two 1-8 and single 9-27 genes found in a 14 kb region of the genome may represent the entire expressed family, with all the other members represented by pseudogenes. There are no further 1-8 genes within Cos3, which contains about 10 kb of DNA either side of the gene cluster, though it is possible that there may be other linked family members that are spatially more distant. However, one argument against there being many (or any) more active 1-8 genes comes from the ratio of 9-27 to 1-8 mRNA in interferon-induced cells, as estimated by the number of cDNA copies of these messages in a cDNA library. Friedman et al (1984) present numbers from which the 9-27:1-8 ratio can be calculated as 40:60, while in Chapter 4.B it is calculated as 50:50. With two 1-8 genes, each is only required to produce 50 - 75% of the stable mRNA produced by the 9-27 gene. If there were three functional 1-8 genes, the requirement would be only 30 - 50%. Since it appears that 1-8 and 9-27 are induced in a similar manner to each other, the total number of active 1-8 genes is probably quite small, and may indeed be just two.

C. The 9-27 Genomic Clone.

Obtaining a genomic clone of 9-27 proved to be a more difficult task than might have been expected for a relatively small gene, as the result of Chapter 6 reflect. The results from the Collins library (Chap. 6.B) and the single EMBL3 phage clone (Chap. 6.C), taken together with the difficulties involved in manipulating regions of Cos3 (Chap. 6.D) point towards the presence of an unstable or 'unclonable' (sic) region of DNA near the 9-27 gene. The Collins library contained a complete and functional 6-16 gene (Porter et al, 1988) and an earlier screening of the library indicated the presence of a 9-27 gene (McMahon, 1985). It therefore appears that this library was a statistically complete representation of the human genome and that the 9-27 gene was lost during its subsequent propagation.

There are anecdotal reports of unstable and unclonable sequences (Little and Cross, 1985), including a region of DNA near the CAD gene locus. Genomic CAD DNA was first cloned in a phage, lambda-100 (Padgett et al, 1982a). However, many of the phage clones of this region were found to have a 1.75 kb deletion in a region containing the largest intron of the gene. It has also been calculated that the number of clones positive for CAD obtained from this library was only 20% of the expected number (I. Saito, pers. comm.). A BamH1 subregion of lambda-100, termed 103, spanning the 1.75 kb region could not be subcloned, despite

numerous attempts, including one using the recA$^-$ strain E.coli HB101 (Padgett et al, 1982b). An intact 103 fragment was observed in the cosmid clone cCAD6, which was grown in HB101 (deSaint Vincent et al, 1981), though again, the number of positive clones was about ten-fold lower than expected. However, growth and yield of cCAD6 was apparently normal. A more recent attempt at cloning this region using a cosmid vector and E.coli ED8767 cells was only partially successful, resulting in 20% of the number of expected positives, all of which have only a small region of 103 at either their 3' or 5' end. These results indicate that there is probably a region of DNA near the CAD gene locus that is unstable when cloned. It is possible that there is a region near the 9-27 locus which has similar properties to the 103 region.

There are a variety of reasons proposed to explain the poor growth or loss of some sequences carried by cloning vectors. Direct or inverted repeats of DNA can cause instabilities in the cloned sequences, while the efficiency of the ColE1 origin of replication, used in a majority of non-phage cloning vehicles, is susceptible to the influence of the content and size of the cloned sequences it is replicating. If repeated DNA structures were responsible for the deletion and under-representation of the 9-27 and CAD gene sequences, then the use of a recBC E.coli strain which is reported to maintain palindromic DNA (Boissy and Astell, 1985) may have resulted in the more successful

isolation of these sequences. However, other factors, such as some form of interaction between the plasmid origin of replication and the 'unclonable' sequences may mean that in some cases, the strain of E.coli used would have little effect on the overall result. It is true that the DNA content of vectors has an effect of the growth of that vector; this was observed in the growth of the 1-8 cosmids, where yields of DNA varied over a ten-fold range, and this phenomenon was also reported by Cross and Little (1985) on randomly selected cosmids from another library.

It is also believed, though as yet unproven, that the relative position of the 'unclonable' sequence to the origin of replication may influence the viability and growth of a cosmid (I.Saito, pers. comm.). Whatever, the explanation for the difficulties encountered with the cloning of the 9-27 gene, it may be that the use of the Lorist B vector, which contains a lambda origin of replication (Cross and Little, 1986), may have been the critical factor in the success of their library in maintaining a functional 9-27 clone.

D. Use of Consensus Sequence Probes.

i). Introduction.

The use of oligonucleotide probes of consensus sequences found 5' to a common set of genes in order to seek to identify related family members is a novel method which may

be applicable to other studies of the kind described in Chapter 3. However, it seems unlikely that the technique will find more than a limited number of other applications.

There are two basic scenarios in which this technique could be applied. It could be used for screening whole genomic libraries or alternatively, the technique may find application in a method for secondary screening, to determine whether or not genomic DNA clones containing coding sequences of genes related to the gene family member under study also contained 5' ends which included a known consensus sequence. This latter method was attempted without success in two experiments described in Chapter 3, and possible reasons for this are also discussed below.

Wood et al (1975) derive a formula to calculate the approximate number of sequence matches between an oligonucleotide and a target genome as N (number of matches) = $2C(n-h+1)(p)/4^h$, where C is the complexity of the genome in bp, n the probe length, h the length of contiguous match of the probe to the target sequence and p the pool size, i.e. the number of any individual species of oligonucleotides in the probe if the probe has been synthesized with redundancy at one or more of the bases along its length. This formula is only a first approximation as it does not take into account factors such as the specific base composition of the sequence. For example, CpG dinucleotides are relatively uncommon in mammalian DNA, and therefore any short sequence containing

this motif is likely to be under-represented in the genome.

ii). Screening scenarios.

In two cases presented in Chapter 3, specific cosmid clones were probed with the 23mer oligonucleotide OCT 351 and regions of homology were detected near the 6-16 gene and a 1-8 gene, though later work indicated that neither region is likely to be involved in the interferon response. One reason that these two false positive were detected may be that the probe was actually too long when compared with the subsequently identified consensus sequence, so leading to the identification of sequences that may have contained more homology to regions of the probe either side of the core of the consensus sequence than to the consensus itself. Conversely, when considering other applications of these types of experiments, it would also be possible to miss true positives by making the probe too short and thus missing homologous sequences with only a single base-pair mismatch.

The consensus sequence found in front of several interferon-inducible genes, GGAAANNGAAACT, is thirteen nucleotides long and contains two nucleotides which appear to vary. An oligonucleotide probe of this sequence with all four possible bases at each N would be a mixture of 16 molecules. Using the above formula with the value of C set for a human genome as $C= 3x10^9$, the number of times such a thirteen base sequence would occur in the genome is about

1,600. This is almost certainly well in excess of the number of interferon-inducible genes in humans. Although it is therefore impractical to use such a probe as a means for directly identifying further interferon-inducible genes in a genomic library, the expected frequency of the sequence is still rare enough (once every 2000 kb) to make using an oligonucleotide probe of this type potentially useful as a secondary screening method once potential positives have been suitably identified. Such a probe would also have use as a tool for mapping specific clones to try and identify promoter regions of genes.

If a probe of the sequence described above was used in this manner (as a secondary screen) under stringent wash conditions, a positive result would provide strong evidence for the presence of the 5' end of an interferon-inducible gene. However, since a single base-pair mismatch is very destabalizing to short probes, a negative match would not rule out the presence of such a gene.

It seems less likely that an oligonucleotide probe to a 5' consensus sequences could be used as a tool for primary screening, since the consensus sequence under investigation would need to be long (more than 14-16 nucleotides) and contain no more than one or two variable residues. Since the interferon consensus sequence, at 13 nucleotides, is one of the longer conserved controlling sequences identified to date, it is doubtful that many, if any, suitable candidate sequences will be discovered.

E. CONCLUSIONS.

The various difficulties encountered with the isolation and characterization of a complete and functional 9-27 gene resulted in only a very limited time being available for investigation of its interferon responsive element (IRE). It was perhaps fortunate that because the sequence of the IRE had been identified in other genes, it was possible to perform an initial expression assay using a relatively limited region of the 9-27 promoter in order to confirm a region of significance in the interferon response of this gene. Nonetheless, the identification of an additional IRE provides valuable confirmation that this sequence is an important key to the interferon response, although of course there may well be other factors involved. In addition, the isolation of the 5' end of a gene known to be induced by both type I and II interferons will allow the role of the IRE and other regions of the promoter to be examined in order to characterize the mechanisms responsible for the action of gamma-interferon.

The characterization of members of the 1-8 multigene family through the use of the exon-spanning probe indicates once again the complexity of the human genome, and provides further food for thought to those interested in the origin and evolution of the genome. The apparently functionless 1-8 pseudogenes add several more kilobases to the total amount of DNA known to be carried by man for no apparent

reason. The study of pseudogenes, which are believed to diverge from their parental sequence at an effectively constant - neutral - rate is one possible way in which the evolution of the genome may be examined.

There are thus a number of worthwhile lines of research in molecular, cellular and evolutionary biology which could be based upon the results obtained above.

REFERENCES.

Aarts,H.J.M., Dendunnen,J.T., Lubsen,N.H. and Schoenmakers,J.G.G. Gene 59: 127-135 (1987).

Aguet,M., Grobke, M. and Dreiding, P. Virology 132: 211-216 (1984).

Allen,G., Fantes,K.H., Burke,D.C. and Morser,J. J. Gen.Viol. 63: 207-212 (1982).

Anderson,S., and Kingston,B.I. Proc. Natl. Acad. Sci. USA. 80: 6838-6842 (1983).

Amersham International plc. M13 Cloning and Sequencing Handbook., (1984).

Balkwill,F. Cytokines in Cancer Therapy. I.R.L. Press, Oxford. (1988).

Bird,A.,Taggart., M. Frommer,M., Miller,O.J., and Macleod,D. Cell 40:91-99. (1985).

Birnboim,H.C. and Doly,J. Nucl. Acids. Res. 7:1513-1523. (1979)

Blomstrom,D.C., Fahey,D., Kutny,R., Korant,B.D. and Knight,E. J. Biol. Chem. 261: 8811-8816 (1986).

Boissy and Astell. Gene 35: 179-185. (1985).

Branca,A.A. and Baglioni,C. J. Biol. Chem. 257: 13197-13200. (1982).

Branca,A.A., Faaltynek,C.R., D'alessandro,S.B and Baglioni,C. J.Biol. Chem. 257: 13291-13296. (1982).

Burke,D.C. Brit. Med. Bull. 41: 333-338. (1985).

Burke,D.C. and Shuttleworth,J. in Interferon 3, pp. 85-111. Eds.N.Finter and R.M. Friedman. Elsevier Science Pubs.Ltd BV. (1985).

Burrone,O.R. and Milstein,C. EMBO J 1, 345-349. (1982).

Cheng,Y.S.E., Colonno,R., and Yin,F.H. J. Biol. Chem. 258: 7746-7750. (1983).

Chirgwin,J.M., Przbyla,A.E., MacDonald,R.J. and Rutter,W.J. Biochemistry 18:5294-5299. (1979).

Clemens,M.J. and McNurlan,M.A. Biochem. J. 226: 345-360. (1985).

Cohen,B., Peretz,D., Vaiman,D., Benech,P. and Chebath,J. EMBO.J. 7: 1411-1419. (1988).

Collins,T., Korman,A.J., Wake,C.T. et al Proc. Natl. Acad.Sci.USA. 81 4917-4921. (1984).

Creasy,A.A., Eppstein,D.A., Marsh,Y.V., Khan,Z. and Merigan,T.C. Mol. Cell. Biol. 3: 780-786. (1983).

Cross,S.H. and Little,P.F.R. Gene 49: 9-22. (1986).

Denhardt,D. Biochem. Biophys. Res. Commun. 23:641-643. (1966)

Dorn,A., Bollekens,J., Staub,A., Benoist,C. and Mathis,D. Cell 50: 863-872. (1987).

Dumont,F.J.J. Interferon. Res. 8: 347-356. (1988).

Dumont,F.J., Palfree,R.G. and Coker,L.Z. J. Immunol. 137: 201-210. (1986).

Dynam,W.S. and Tjian,R. Nature. 316: 774-778. (1985).

Falmner,F.G. and Zachan,H.G. Nature. 310: 71-74. (1984).

Feinberg,A.P. and Vogelstein,B. Anal. Biochem. 132: 6-13. (1983).

Feinberg,A.P. and Vogelstein,B. Anal. Biochem. 137: 266-267. (1984).

Fergusson,R.J., Stewart,M.E., Cornbleet,M.A., Gregor,A., and Smyth,J.F. ICRF. Annual Report. (1987).

Foon,K.A., Sherwin,S.A. and Abrams,P.G.N. Engl. J. Med. 311: 1148-1152. (1984).

Forti,R.L., Mitchell,W.M., Hubbard,W.C., Workman,R.J. and Forbes,J.T. Proc. Natl. Acad. Sci. USA 81: 170-174. (1984).

Friedman,R.L., Manley,S.P., McMahon,M., Kerr,I.M. and Stark,G.R. Cell. 38: 745-755. (1984).

Friedman,R.L. and Stark,G.R. Nature. 314: 637-639. (1985).

Friedman-Einat,E., Revel,M. and Kimchi,A. Mol. Cell. Biol. 2: 1472-1480. (1982).

Gait,M.J., Matthes,H.W.D., Singh,M., Sproat,B.S. and Titmas,R.C. Nuc. Acid. Res. 10: 6243-6245. (1982).

Gastl,G., Aulitzky,W., Margreitor,R., Flener,R., Vancamp,B., Bernman,Z., Peetermans,M. and Huber,C. Brit. J. Haem. 61: 581-582. (1985).

Gewert,D.R., Coulomse,B., Castelino,M., Skup,D. and Williams,B.R.G. EMBO. J. 6: 651-657. (1987).

Giglioni,B., Casini,C., Mautovani,R., Merli,S. and Comi,P. EMBO.J. 3: 2641-2645. (1985).

Giulotto,E., Saito,I. and Stark,G.R. EMBO.J. 5: 2115-2121. (1986).

Goldstein,D. and Laszlo,J. Cancer Research 46:4315-4329. (1986).

Gorman,C., Moffat,L.F. and Howard,B.H. Mol. Cell. Biol. 2: 1044-1051. (1982a).

Gorman,C.M., Merlino,G.T., Willingham,M.C., Pastan,R. and Howard,B.H. Proc.Natl. Acad. Sci. USA. 79: 6777-6781. (1982b).

Gorman,C. in DNA Cloning, a Practical Approach ed D.M.Glover, I.R.L. Oxford. Vol.II. pp. 143-190. (1985).

Graham,F.L. and Van.Der.Eb, A.J. Virology 52: 456-467. (1973).

Gray,P.W., Leung,D.W., Pennica,D., Yelverton,E., Najarian,R., Simonsen,C., Derynck,R., Sherwood,P.J., Wallale,D.M., Berger,S.L., Levinson,A.D. and Goddel,D.V. Nature. 311: 433-438. (1982).

Haller,O. Curr. Top. Microb. Immunol. 92: 25-52. (1981).

Hardy,K. Bacterial Plasmids. Pub. Van. Nostrand. Reinhold (UK). (1981).

Hattori,M. and Sakaki,Y. Anal. Biochem. 152: 232-238. (1986).

Healey,G.B., Gelber,R.D., Trowbridge,R.N., Grundfast,K.M. Ruben,R.J. and Price,K.N. New. Engl. J. Med. 319: 401-407. (1988).

Hidaka,K., Gynheung,A., Ip,P., Kuwana,M. and Siminovitch,L. Som. Cell. Mol. Gen. 11: 109-115. (1985).

Hood,L., Kronenberg,M., Hunkapiller,T. Cell. 40: 225-229. (1985).

Horecker,B.L. and Morgan,J. Lymphokines. 9: 15-35. Academic Press, New York. (1985).

Isaacs,A. and Lindenmann,J. Proc. Roy. Soc. Brit. 147:258-273. (1957)

Israel,A., Kimura,A., Fournier,A., Felllous,M. and Kourilsky,P. Nature. 322: 743-746. (1986).

Jaye,M., de la Salle,H., Schamber,F., Balland,A., Kohli,V., Findeli,A., Tolstoshev,P. and Lecocq,J.P. Nucleic. Acid. Res. 11: 2325-2335. (1983).

Jeffreys,A.J., Wilson,V. and Thein,S.L. Nature. 314: 67-73. (1985).

Jonak,G.J. and Knight,E. Proc. Natl. Acid. Sci. USA. 81: 1747-1750. (1984).

Karin,M., Andersen,R.D., Slater,E., Smith,K. and Herschman,H.R. Nature. 286: 295-297. (1980).

Karin,M., Andersen,R.D. and Herschman,H.R. Eur. J. Biochem. 118:527-531. (1981).

Karin,M., Haslinger,A., Hotgreve,H., Richards,R.I., Krauter,P., Westphal,H.M. and Beato,M. Nature. 308: 513-519. (1984a).

Karin,M., Haslinger,A., Holtgreve,H., Cathala,G., Slater,E. and Baxter,J.D. Cell 36: 371-379. (1984b).

Karin,M. and Hotgreve,H. DNA. 3: 319-326. (1984c).

Karlsson,S. and Nienhuis,A.W. Ann. Rev. Biochem. 54: 1071-1108. (1985).

Kawai,S. and Nishizawa,M. Mol. Cell. Biol. 4: 1172-1174. (1984).

Kelly,J.M., Gilbert,C.S., Stark,G.R. and Kerr,I.M. Eur.J. Biochem. 153: 367-371. (1985).

Kelly,J.M., Porter,A.C.G., Chernajovsky,Y., Gilbert,C., Stark,G.R. and Kerr,I.M. EMBO J. 5:1601-1606. (1986)

Korant,B.D., Blomstrom,D.C., Jonak,G.J. and Knight,E. J.Biol. Chem. 259: 14835-14839. (1984).

Korber,B., Hood,L. and Stroyniwski,I. Proc. Natl. Acad. Sci.USA. 84: 3380-3384. (1987).

Kyte,J. and Dcolittle,R.F. J.Mol.Biol 157:105-132 (1982)

Laimins,L.A., Khoury,G., Gorman,C., Howard,B. and Gruss,P. Proc.Natl. Acad. Sci. USA. 79: 6453-6457. (1982).

Larner,A.C., Jonak,G., Cheng,Y-SE, Korant,B., Knight,E. and Darnell,J.E. Proc. Natl. Acad. Sci. USA. 81: 6733-6737. (1984).

Lathe,R.J. Mol. Biol. 183: 1-12. (1985).

Lengyel,P. Ann. Rev. Biochem 51: 251-282. (1982).

Leone,A., Pavlakis,G.N. and Hamer,D.H. Cell 40: 301-309. (1985).

Lindenmaier,W., Hauser,H., de Wilke,I.G. and Schutz,G. Nucl. Acids. Res. 10: 1243-1256. (1982).

Little,P.F.R. and Cross,S.H. Proc. Natl. Acad. Sci. USA 82:3195-3163. (1985)

Little,P.F.R. and Cross,S.H. Gene. 53: 275-281. (1987).

Lopata,M., Havercroft,J.C., Chow,L.T. and Cleveland,D.W. Cell. 32: 713-724. (1983).

Luster,A.D., Unkeless,J.C. and Ravetch,J.V. Nature. 315: 672-676. (1985).

Luster,A.D. and Ravetch,J.V. Mol. Cell Biol.7:3723-3731.(1987)

Maniatis,T., Fritsch,E.F., Sambrook,J. Molecular Cloning - A Laboratory Manual. Cold Spring Harbour. (1982).

Meinkoth,J. and Wahl,G. Anal. Biochem. 132: 267-284. (1984).

Miwa,K., Doyle,C., and Strominger,J.L. Proc. Natl. Acad. Sci. USA. 84: 4939-4943. (1987).

Moormann,R.J.M., Jongbloed,R. and Schoenmakers,J.G.G. Gene. 29: 1-9. (1984).

McKnight,S.L. Nucl. Acids. Res 8: 5949-5964. (1980).

McMahon,M. Thesis. University of London. (1985).

New England Biolabs. Catalogue. (1986/1987).

Newman,H.F.V., Bleehen,N.W., Galazka,A. and Scott,E. Cancer. 60: 2938-2940. (1987).

Orachansky,P., Goren,T and Rubinstein,M. in The Biology of the Interferon System. eds. E.de.Mayer. and H.Schellekens. Elsevier (1983).

Padgett,R.A., Wahl,G.M. and Stark,G.R. Mol. Cell.Biol. 2: 293-301. (1982a).

Padgett,R.A., Wahl,G.M. and Stark,G.R. Mol. Cell. Biol. 2: 302-307. (1982b).

Parker,B.A. and Stark,G.R. J. Virol. 31:360-369. (1979).

Porter,A.C.G., Chernajovsky,Y., Dale,T.C., Gilbert,C.S., Stark,G.R. and Kerr,I.M. EMBO.J. 7: 85-92. (1988).

Quesada,J.R., Reuben,J.. Manning,J.T., Hersh,E.M. and Gutterman,J.U. New. Engl. J.Med. 310: 15-18. (1984).

Rehberg,E., Kelder,B., Hoal,E.G. and Pestka,S. J. Biol. Chem. 257: 11497-11502. (1982).

Reich,N., Evans,B., Levy,D., Fahey,D., Knight,E. and Darnell,J.E. Proc. Natl. Acad. Sci. USA. 84: 6394-6398. (1987).

Resnitzky,D., Yarden,A., Zipori,D. and Kimchi,A. Cell. 46: 31-40. (1986).

Revel,M. and Chebath,J. T.I.B.S. 11: 166-170. (1986).

Rosa,F. and Fellous,M. Immunol. Today. 5: 261-262. (1984).

Ruddon,R.W. Cancer Biology. O.U.P. (1987).

de Saint Vincent,B., Delbruck,S., Eckhalt,W., Meinkoth,J., Vitto,L. and Wahl,G.M. Cell. 27: 267-277. (1981).

Saito,I. and Stark,G.R. Proc. Natl. Acad. Sci. USA. 83: 8664-8668. (1986).

Sanger,F., Nicklen,S., and Coulson,A.R. Proc. Natl. Acad.Sci. USA. 74: 5463-5467. (1977).

Sleigh,M.J. Anal. Biochem. 156: 251-256. (1986).

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

Stuart,G.W., Searle,P.F., Chen,H.Y., Brinster,R.L. and Palmiter,R.D. Proc. Natl. Acad. Sci. USA. 81: 7318-7322. (1984).

Staeheli,P., Horisberger,M.A. and Haller,O. Virology. 132: 456-461. (1984).

Staeheli,P, Haller,O., Boll,W., Lindenmann,J. and Weissmann,C.M. Cell.44: 147-158. (1986a).

Staeheli,P., Danielson,P., Haller,O. and Sutcliffe,J.G. Mol. Cell. Biol. 6: 4770-4774. (1986b).

Strander,H. Advances in Cancer Research. 46 (1986).

Thomas,H.C. and Scully,L.J. Brit. Med. Bull. 43: 374-380. (1985).

Thomas,P.S. Proc. Natl. Acad. Sci. USA. 77:5201-5205. (1980).

Tunacliff,A. Thesis. University of London. (1983).

Ullrich,A., Coussens,L., Hayflick,J.S. et al Nature. 309: 418-425. (1984a).

Ullrich,A., Berman,C.H., Dull,T.J., Gray,A. and Lee,J.M. EMBO.J. 3: 361-364. (1984b).

Umeda,T. and Niijima,T. Cancer 58:1231-1235. (1986)

Vasak,M. and Kagi,J.H.R. in Biomineralization and Biological Metal Accumulation. Biomin. Biol. Met. Acc. pp. 429-437. pub.D.Reidel.Publishing.Co. (1983).

Wallace,R.B., Shaffer,R., Murphy,R.F., Bonner,J., Hirose,T. and Itakura,K. Nucl. Acid. Res. 6: 3543-3557. (1979).

Wigler,M., Pellicer,A., Siverstein,S., Axel,R., Urlaub,G. and Chasin,L. Proc. Natl. Acad. Sci. USA. 76: 1373-1376. (1979).

Wood,W.I., Capon,D.J., Simonsen,C.C. et al. Nature. 312: 330-336. (1984).

Wood,W.I., Gitschier,J., Lasky,L.A. and Lawn,R.M. Proc. Natl. Acad. Sci. USA 82: 1585-1588. (1985).

Xanthondakis,S. and Hiscott,J. J.Biol. Chem. 262: 8298-8302. (1987).

Yarden,A., Shure-Gottlieb,H., Chebath,J., Revel,M. and Kimchi,A. EMBO.J. 3: 969-973. (1984).

Zoon,K.C., Arnheiter,H., Zur Nedden,D., Fitzgerald,D.S.P. and Willingham,M.C. Virology. 130: 195-203. (1983).