A TRANSCRIPTIONAL ANALYSIS OF THE HUMAN c-Ha-ras1 ONCOGENE

BY

NOEL F. LOWNDES B.A.

A Thesis Submitted for the Degree of Doctor of Philosophy in the Faculty of Medicine of Glasgow University

The Beatson Institute for Cancer Research,
Glasgow.

DECEMBER 1987

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

DECLARATION

This thesis has not been submitted as an exercise for a degree at any other university. Except where otherwise stated, the work herein has been carried out by the author. This thesis may be borrowed or copied upon request with permission of the librarian.

Noel F. Lowndes

To Mum and Dad

ACKNOWLEDGMENTS

I wish to express my gratitude to my supervisor, Dr. Maggi Allan, for her direction and support throughout this work. Her patience and encouragement are greatly appreciated. I thank the Cancer Research Campaign, of Britain, for their stipend which enabled me to work at the Beatson Institute for Cancer Research, Glasgow, and also the Faculty of Medicine, Columbia University, New York for further support. Additionally I would like to thank the following people: Dr. Arthur Banks, who provided laboratory work space when none was available, Ruth Gutman who kindly provided access to word processor facilities, Phil Oh for help with many of the diagrams, my family whose support is always invaluable and finally, Angela for her friendship, understanding, advice and the typing of this thesis.

SUMMARY

The pattern of transcription initiation for the human c-Ha-*ras*1 gene has been investigated using several experimental methods. These analyses demonstrated multiple clusters of transcription initiation sites distributed over an approximately 200 bp non-coding, upstream exon (termed exon -1), which is separated from the ATG codon by an 1040 bp intron. Mutational analysis of the promoter region identified a short positive regulatory element, located between positions -243 to -196, relative to the donor splice site of exon -1. This element contains known regulatory sequence motifs. Furthermore, a putative negative regulatory element, with an unusual DNA sequence, was identified between positions -103 to -34, relative to the same donor splice site. It has also been demonstrated that the human c-Ha-*ras*1 promoter region can function bidirectionally. The sequence directing the "reverse-orientation" promoter activity was located to between positions, -392 to -196, relative to the exon -1 donor splice site of the c-Ha-*ras*1 gene. Transcription of the c-Ha-*ras*1 gene was shown to be increased approximately 20 fold when covalently linked to the SV40 enhancer element. This result is the first direct demonstration that the SV40 enhancer can increase transcription of "housekeeping-type" genes and this result also has important implications for the possible methods of oncogenic activation of this gene. The 1040 bp intron located between exon -1 and the first coding exon (exon 1) was found to contain sequences within its 5' end, which were moderately repetitive within the human and mouse genomes, and homologous to abundantly transcribed, non-polyadenylated RNAs from various human cell lines. However, the functional significance, if any, of this result is unclear.

TABLE OF CONTENTS

INTRODUCTION

The human cellular Harvey ras oncogene (c-Ha-ras1) is a gene believed to be important in the process of oncogenesis (reviewed in Cooper, 1982; Weinberg, 1985; Bishop, 1985 and Levinson, 1986). An important question is whether the function of the c-Ha-ras1 encoded protein is so vital to the cell, that quantitative differences in its level of expression could lead to carcinogenesis? Therefore, this thesis will be introduced in two main sections, (I) the involvement of the ras genes in oncogenesis and (II) some aspects of gene regulation relevant to the expression of ras genes. In the first of these sections the ras gene family will be described, with particular emphasis on the so called "classical" ras genes, Harvey (Ha)-, Kirsten (Ki)- and N- ras. Some details of the expression pattern of the ras genes during malignancy, development and the cell cycle will then be discussed and finally, some recent speculations on the possible functions of the ras genes will also be discussed. In section II, some aspects of transcriptional regulation will be considered. To conclude this section distinct features of promoter regions for "housekeeping" genes, which parallel those of the ras genes, will be briefly discussed.

I. THE RAS GENES AND ONCOGENESIS

The first real progress towards identifying some of the genes which, under certain circumstances, may be implicated in the development of malignancy, came from the study of tumour inducing viruses, particularly the retroviruses. As early as 1910, Peyton Rous showed that a cell-free filtrate from chicken sarcomas could induce new sarcomas in chickens, but it was not until decades later that a retrovirus, now called Rous sarcoma virus (RSV), was identified as the sarcoma inducing agent. Later, a RSV-encoded gene was identified as the tumour inducing

agent within the virus and was called the src (for sarcoma) oncogene. This oncogene was the first of many to be discovered and to date at least 40 distinct oncogenes have been identified (reviewed by Bishop, 1985; Weinberg, 1985). However, it is now clear that although most of these oncogenes were first identified in tumour viruses, most are not endogenous to these viruses but are also present in normal cells, where they apparently function in normal cellular processes. To distinguish between the oncogenic versions of these genes and their normal cellular counterparts, the latter are termed proto-oncogenes. It is now known that the proto-oncogenes can be converted into oncogenes by various mechanisms which can be broadly grouped into two categories, quantitative and qualitative. Quantitative conversion results from abnormal levels of expression of the "normal" gene product, whereas qualitative conversion results from mutation of the proto-oncogene to produce an aberrantly acting, qualitatively different gene product. Since elevated transcription of the ras proto-oncogenes has recently been shown to be a mechanism whereby they are converted into oncogenes (see section I.c), a detailed understanding of the transcriptional regulation of these genes is now especially important.

(a) The ras genes: a family of highly conserved genes

In the genomes of all eukaryotes examined, at least one member of the ras multigene family has been detected. Organisms as diverse as the yeasts, Saccharomcyes cerevisiae and Schizosaccharomyces pombe (DeFeo-Jones et al., 1983; Powers et al., 1984; Fukui and Kaziro, 1985, Nadine-Davis et al., 1986), the slime mould, Dictyostelium discoideum (Reymond et al., 1984), the sea slug, Aplysia (Madaule and Axel, 1985) and the fruit fly, Drosophila melanogaster, (Neuman-Silberburg et al., 1984; Moser et al., 1985; Schejter and Shilo, 1985), all have genes

related to the more intensively studied ras genes of mammalian genomes. A detailed description of these intensively studied mammalian ras genes, sometimes referred to as the "classical" ras genes and which include the Harvey (Ha), Kirsten (Ki) and N-ras genes, will be presented in the next section.

Recently, other more distantly related members of the ras multigene family have been identified in mammalian genomes. Lowe et al. (1987) have reported a novel ras gene in human and murine cells which they have termed R-ras. The amino-acid sequence of the human R-ras gene product was found to be 55% homologous with that of the classical Ha-ras encoded protein. However, the R-ras sequence contains an amino-terminal extension of 26 residues, has an entirely different exon-intron structure and resides on a different chromosome to the "classical" ras genes. The Aplysia ras gene, which Madaule and Axel (1985) have called rho, shares approximately 35% homology in its protein product with the "classical" ras gene products. However, two putative human rho genes have also been identified, which are at least 85% homologous with the Aplysia rho gene. Thus the human genome contains at least three classical ras genes, one R-ras gene and possibly two ras-related rho genes. Furthermore, another ras-related gene, termed YPT1, has been isolated from S. cerevisiae (Gallwitz et al., 1983), whose protein product shows 38% homology with the "classical" ras proteins. Since both the rho and the YPT1 genes were identified purely by chance, in the course of the analyzing cloned DNA, it is likely that several other members of the ras multigene family may exist within the genomes of eukaryotes. Whether most, or only a few members of this emerging multigene family can act as oncogenes under certain conditions is not yet known. Interestingly, Padua et al. (1984) have identified a novel transforming gene from a human melanoma cell line, by DNA-mediated transfection into NIH 3T3 mouse fibroblasts, which shows only weak homology to the classical ras genes.

As yet the identity of this tentative member of the ras multigene family awaits further characterization.

The extraordinary evolutionary conservation of this multigene family, from unicellular eukaryotes to mammals, together with the oncogenic potential of specific mutants of at least some members, suggests that the ras multigene family must play a fundamental role in aspects of growth or development of eukaryotic cells.

(b) Detailed structure of the "classical" ras genes

The so called "classical" ras genes, Harvey (Ha), Kirsten (Ki) and N-ras, all encode immunologically related proteins (referred to as $p21^{ras}$) which have a molecular weight of 21,000 daltons (Levinson, 1986). The Harvey and Kirsten ras genes were originally discovered as the oncogenes of the Harvey and Kirsten strains of murine sarcoma viruses, Ha-MuSV and Ki-MuSV (Ellis et al., 1981), while the N-ras gene was identified by low stringency hybridization to a viral Harvey ras probe (Shimizu et al., 1983b). From sequence comparisons of a viral Ki-ras gene, a Kirsten pseudogene and Kirsten cDNA clones, with genomic clones (Shimizu et al., 1983a; McGrath et al., 1983; Capon et al., 1983b), it is known that the human c-Ki-ras2 locus (c-Ki-ras1 is a processed pseudogene) spans more than 38 Kb and is divided up into five coding exons separated by as much as 17 Kb (Figure 1). This large gene locus has a rather unusual organization of introns and exons. It has an additional non-coding upstream exon, called exon 0, and only four of the five available coding exons are ever represented in the mature transcripts of 5.5 Kb and 3.8 Kb. The vast majority of mature transcripts (approximately 99%) were spliced in such a manner that the 16 Kb of RNA sequence corresponding to the fourth exon and its flanking introns, were eliminated by the fusion of the third exon to the downstream fifth exon (Capon et al., 1983b). Hence, the

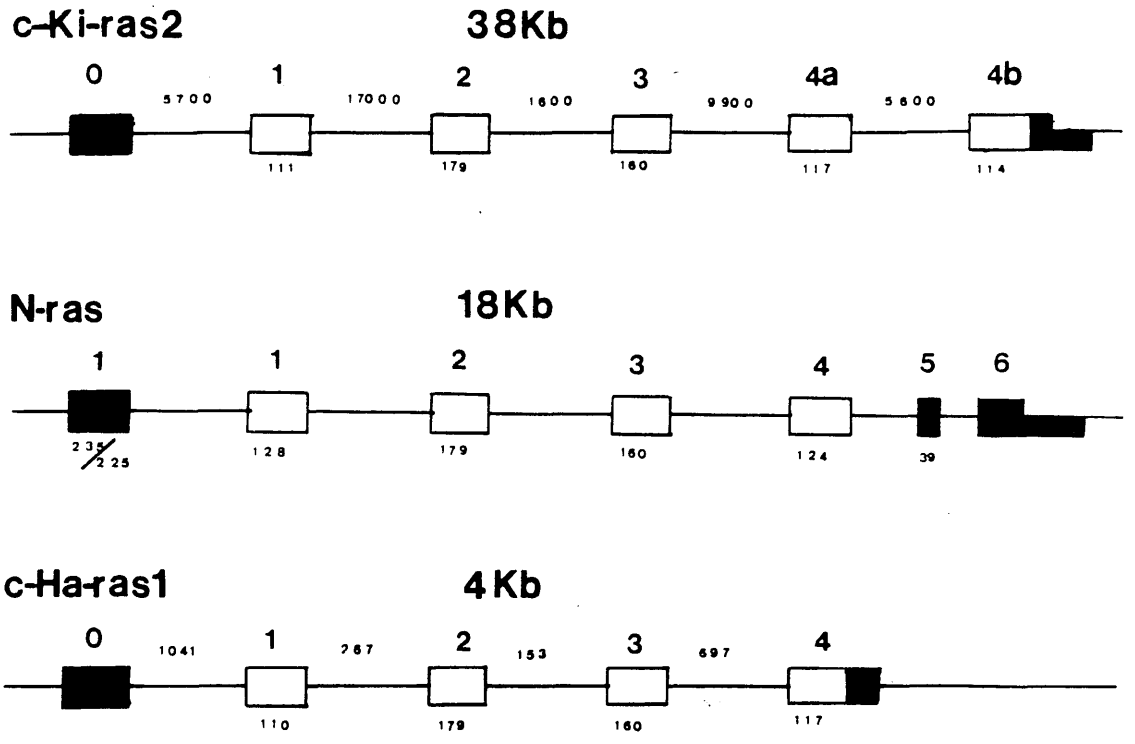

Figure 1 The "classical" human ras genes.

The human c-Ki-ras2, N-ras and c-Ha-ras1 genes are shown in schematic form. Open boxes indicate known coding exons and shaded boxes indicate non-coding exons. The overall size of each locus is given in kilo base pairs (Kb) and where known, the sizes of exons and introns are given in base pairs (bp). The diagram is not to scale.

fourth and fifth exons are termed exon 4A and 4B, respectively. This differential splicing of the precursor RNA to give an abundant species containing exon 4B and a rarer species containing exon 4A, does not explain the two discretely sized mature transcripts observed by northern blot analysis. Probes for exon 4A or for exon 4B detect similar proportions of each of these transcripts, which shows that each species contains sequences from both coding exons. Therefore, the two transcripts must differ at their 5' or 3' ends. This implies either the existence of two promoters, the upstream promoter being more efficient or, two polyadenylation sites, the more downstream one being used preferentially. An analogous situation has been observed in mouse adrenal tumour cells (George et al., 1985), where two c-Ki-ras transcripts of 5.2 Kb and 2.0 Kb are detected. Finally, it is interesting that exon 4A and exon 4B-encoded amino acid sequences are related for the first 14 residues and the last 5 residues at the carboxy terminus, but that the intervening amino acids are highly divergent (Shimizu et al., 1983a). Exon 4B-encoded amino acids include a polylysine stretch within the divergent region, thereby making the carboxy-terminal end significantly more basic. Thus, it is possible that the two alternative forms of c-Ki-ras may have different properties.

Hall and Brown (1985) used sequence comparisons of cDNA and genomic clones to define the structure of the human N-ras gene. A rather complex but familiar pattern emerged. Like c-Ki-ras, the human N-ras gene has a 5' untranslated exon (which is termed exon -1, rather than exon 0), four coding exons, but no alternative fourth coding exon. Instead, it has a small (39 bp) 3' untranslated exon 5, located between the fourth coding exon and exon 6, which contains the remainder of the 3' untranslated sequences (Figure 1). All seven exons are contained within three EcoR1 fragments spanning 18 Kb; that is, less than half the size of the c-Ki-ras locus. In

addition to determining the arrangement of introns and exons, Hall and Brown (1985) were able to show that the two N-ras specific transcripts of 4.3 Kb and 2.0 Kb were due to differential ultilization of polyadenylation sites. The 2.0 Kb transcript is less abundant since it uses a less efficient polyadenylation processing signal, AAUAUA, rather than the normal AAUAAA found just upstream of the polyadenylation site of the longer transcript. The 5' boundary of the N-ras transcription unit was also determined and shown to consist of two initiation or cap sites, separated by 10 bp at positions 225 ±2 bp and 235 ±2 bp upstream from the donor splice site of exon -1 (both appear to be used with equal frequency). The 5' untranslated region contains two AUG translation initiation codons, followed by termination codons 23 and 32 triplets further downstream, the last termination codon overlapping the AUG codon used for protein synthesis. Upstream of the two cap sites there are no TATA or CAAT boxes close to positions -30 and -80 respectively. Instead, the 100 bp preceding the cap sites contains approximately 65% A + T residues, in contrast to the overall approximately 75% G + C richness of the 5' end of the human N-ras gene. The only recognizable sequence motif in the promoter region of this gene is four GGCGGG hexanucleotide repeats, which potentially bind the transcription factor Sp1 (see section II.b). Unusually, only one of these repeats is present upstream of the cap sites, the other three being in the upstream non-coding exon -1. Thus, the human N-ras gene can be seen to have a very unusual genomic organization with one upstream non-coding exon and two downstream non-coding exons, one of which, exon 5, is only 39 bp long and may be a redundant vestigial exon, analogous to exon 4B of the Kirsten ras gene (see Figure 1).

The cloning of the human cellular Harvey ras oncogene, c-Ha-ras1 (c-Ha-ras2 is a processed pseudogene, O'Brien et al., 1983) was made possible by an initial series of experiments designed to identify and study DNA segments in

tumour cells which were responsible for the oncogenic conversion of those cells. High molecular weight DNA from tumour cells was transfected into normal NIH 3T3 mouse fibroblasts and foci of transformed cells were counted in the recipient NIH 3T3 monolayer culture. The observation that transformed foci were obtained with high molecular weight DNA from some tumour cell lines, demonstrated not only the existence of oncogenic DNA sequences in tumour cell DNA, but also facilitated the isolation of the DNA sequences involved. Thus, the transforming gene of the human bladder carcinoma cell lines T24 and EJ, now known to be identical cell lines (O'Toole et al., 1983), was cloned independently by three groups (Goldfarb et al., 1982; Pulciani et al., 1982; Shih and Weinburg, 1982) and more recently another allele of the same gene was cloned from a human squamous cell lung carcinoma (Kagimoto et al., 1985). A subsequent study of the bladder carcinoma oncogene and its related proto-oncogene present in normal cells, was facilitated by the later discovery that this oncogene was homologous to the transforming gene of the rat-derived Harvey murine sarcoma virus (Parada et al., 1982; Santos et al., 1982). This rat sarcoma virus gene, termed v-Ha-ras, had been acquired from the rat genome by retroviral transduction, and during the course of the study of this gene, both the rat and the human cellular homologues of v-Ha-ras were isolated (DeFeo et al., 1981; Chang et al., 1982). The human cellular homologue of the v-Ha-ras was found to correspond precisely to the normal version of the EJ/T24 bladder oncogene (Parada et al., 1982) and therefore, comparisons between oncogenic and proto-oncogenic versions of the human Harvey ras gene could be made. From this and later comparisons it was shown that oncogenic ras proteins differ from their normal homologues by a single amino acid substitution, usually at position 12 or 61 (Tabin et al; 1982; Taparowsky et al., 1983; Shimizu et al., 1983a; Yuasa et al, 1983; Capon et al, 1983b; see also the next section). The data from these studies are

strongly in favour of the qualitative hypothesis for oncogenic activation (see earlier), as it was only a qualitative change in the protein product, not a change in its level of expression, which appeared to be responsible for oncogenic activation. However, it is also clear that elevated expression of the normal human c-Ha-ras1 proto-oncogene may induce tumourogenic transformation. Chang et al. (1982) have shown that when this gene is ligated to strong transcriptional control elements (the long terminal repeats from a murine or feline retrovirus) the corresponding high $p21^{ras}$ levels produced, could induce tumourogenic transformation of NIH 3T3 mouse fibroblasts. Thus it appears that transformation by this gene can be accomplished by either quantitative or qualitative events (other examples of "quantitative transformation" are also discussed in section I.c). It is possible that quantitative activation could occur due to defects in the transcriptional regulation of c-Ha-ras1.

Sequence analyses of the c-Ha-ras1 gene and comparison with viral Harvey ras sequences have shown that there are four coding exons spread over less than 3 Kb (Reddy 1983; Capon et al., 1983a). However, the presence of an additional downstream coding exon, analogous to the situation in c-Ki-ras2 (see Figure 1), cannot be excluded. Interesting features of the c-Ha-ras1 sequence include the presence of potential donor and acceptor splice sites in the region upstream from the four coding exons. This implied the presence of an upstream non-coding exon analogous to exon 0 of c-Ki-ras, or exon -1 of N-ras (Figure 1). The presence of a putative upstream exon is also supported by comparison with the sequence reported for viral Harvey ras (Dhar et al., 1982). The viral sequence is homologous up to the upstream acceptor splice site in the genomic sequence and is again homologous to genomic sequences just after the upstream donor site (Figure 2). Alternatively, it has been suggested (Capon et al., 1983a) that the 1 Kb of DNA corresponding to the putative intron

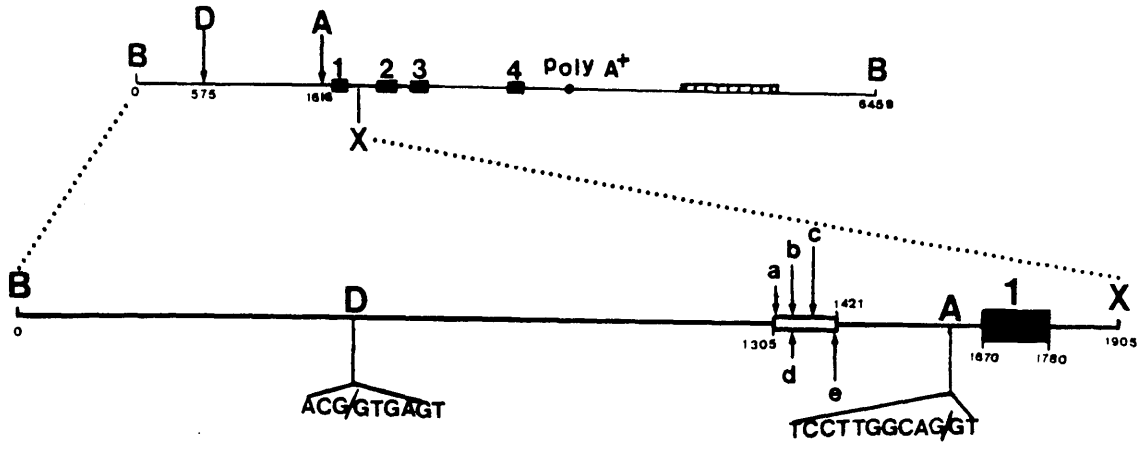

Figure 2 The human c-Ha-ras1 locus.

The approximately 6.4Kb BamHI fragment containing the human c-Ha-ras1 gene is shown in schematic form and numbered according to Reddy (1983). The four known coding exons are shown as shaded boxes, the large closed dot indicates the position of the polyadenylation site (poly A^+). The positions of two concensus splicing signals in the 5' region are indicated by vertical arrows, where D and A refer to the donor and acceptor splice sites respectively. Also shown is the position of a tandemly repeated 28 bp sequence present in the 3' flanking sequence (hatched open rectangle). An expanded view of the 1905 bp BamHI/XbaI fragment is shown below the line drawing of the c-Ha-ras1 locus. In addition, the sequence of the concensus splice sites are shown and the open rectangle indicates the position of the proposed TAAT/CAAT promoter region (Capon et al., 1983a and Reddy, 1983). The letters a to e indicate the positions of weak homologies to the concensus TATA or CAAT motifs. 1 = CCTTT; 2 = CTAAT; 3 = TAATTT; 4 = CCTTT and 5 = ATTAAG.

could have been deleted during the viral transduction of the cellular Harvey ras gene. This possibility is supported by the presence of 15 bp direct repeats flanking the 1 Kb of DNA in the genomic sequence (Capon et al., 1983a). It is interesting that the TATA and CAAT boxes proposed for this gene (Capon et al., 1983a; Reddy 1983 and Figure 2) occur within the 1 Kb putative intron and hence would not be used if there was an upstream exon. However, it should be pointed out that these proposed transcription signals are very poor fits to their concensus sequence motifs.

Finally, a particularly unusual feature of the human c-Ha-ras1 DNA sequence is the presence of a stretch of tandemly repeated DNA, nearly 1 Kb in length, located approximately 1 Kb downstream from the polyadenylation site (see Figure 2). It is composed entirely of tandem repeats of a 28 bp concensus sequence (Capon et al., 1983a) and the number of these repeats accounts for the size polymorphisms reported for the human c-Ha-ras1 locus (Krontiris et al., 1985). Interestingly, it has been suggested that some rare allelic variants of this polymorphism may be linked to genetic susceptibility to cancer in individuals carrying them (Krontiris et al., 1985). However, a recent report using a larger sample size of sporadic and familial melanoma cases has provided strong evidence against this hypothesis (Gerhard et al., 1987). Jeffreys et al. (1985) have reported similar tandem-repetitive sequences in human DNA which they have termed hypervariable "minisatellite" regions. As yet the tandemly repeated sequence closely linked to the human c-Ha-ras1 gene has no know function. However, Seeburg et al. (1984) speculate that it may have a role in transcriptional regulation as its presence restores a more strongly transforming phenotype to cells transfected with codon 12 mutants.

(c) Expression of ras genes during malignancy

The development of cancer is believed to be a

multistage process (Foulds, 1969; Cairns, 1975) and epidemiological studies have shown that tumours are the end product of a series of independent events, the exact number of which may vary depending on the particular target cell (Peto, 1977). Tumour development, at least in part, is believed to be a consequence of the activation of cellular proto-oncogenes by qualitative or quantitative mechanisms. The classical ras genes have dominant transforming activity when transfected into NIH 3T3 mouse fibroblasts (Cooper, 1982). This is believed to be due to various mutations clustered around amino acid codons 12 and 61 and at present there are at least 50 different mutations in five different ras codons known to confer varying degrees of transformation to NIH 3T3 cells (reviewed by Levinson, 1986). From an analysis of a large number of activated ras genes derived from tumours, it was found that changes at amino acid codons 12, 13 or 61 can activate the transforming potential of $p21^{ras}$. In addition, analysis of randomly mutated ras genes demonstrated that amino acid substitutions at residues 59 and 63 can also activate $p21^{ras}$ (Fasano et al., 1984). However, the best studied activating, or transforming, mutations are those affecting the codon which encodes the amino acid at position 12 of $p21^{ras}$. Seeburg et al (1984) have constructed twenty Ha-ras genes in vitro each encoding a different amino acid at position 12 and it was found that other than the normally occuring glycine, only proline had no transforming activity. Glycine and proline are the two amino acids predicted to disrupt protein α-helices and therefore, the transforming mutations are likely to exert their effect by allowing a longer α-helical region within this crucially important domain. Recently, a similar study has introduced mutant Ha-ras genes encoding 17 different amino acids at codon 61, into NIH 3T3 mouse fibroblast (Der et al., 1986). In this case, glutamine is the normally occuring amino acid at position 61 and of the 17 different mutants created, all but proline and glutamic acid, produced transforming

alleles. However, the transforming activity of the codon 61 Ha-ras mutants varied over a 1000-fold range in biological activity and increased levels of expression of the weakly transforming mutants, were required for morphological transformation of the NIH 3T3 cells.

It is possible that the NIH 3T3 transfection assay, routinely used to detect activated ras genes, may be biased in some way, such that, any involvement of ras genes in tumour development due to its overexpression, i.e. quantitative activation, may not be detectable. Thus, for the remainder of this section evidence in support of the quantitative hypothesis will be discussed. As mentioned earlier, the first direct experimental evidence for this hypothesis was the observation by Chang et al. (1982) that elevated expression of the normal c-Ha-ras1 gene, by placing it under the control of retroviral long terminal repeat control elements, could also induce tumourigenic transformation. Transfection of normal and mutant human c-Ha-ras1 genes linked to transcriptional enhancers, into primary cells rather than immortalized NIH 3T3 cells has also been reported (Spandidos and Wilkie, 1984; Yoakum et al., 1985). Transfection of the enhanced and mutated oncogene led to morphological transformation of primary cells, previously only possible with two cooperating oncogenes (Land et al., 1983; Ruley, 1983; Newbold and Overall, 1983). The enhanced "normal" c-Ha-ras1 caused immortalization of the primary cells, a property not normally associated with ras genes, but did not induce morphological transformation. Thus, this ras gene can have a range of biological effects depending on the progressive addition of quantitative and qualitative alterations and it may be that elevated ras expression is a more common event in tumour progression than previously anticipated. This suggestion is supported by the observation that only 10 to 15% of human tumours examined have been found to have a transforming mutation detectable in the NIH 3T3 transfection assay (reviewed by Varmus, 1984). It is

possible that elevated ras expression may have had an important role in the etiology of some of the remaining tumours examined.

Elevated ras expression has indeed been implicated with malignancy in a number of different studies. Slamon et al. (1984) have examined twenty different fresh human tumour types, from fifty-four patients, for the expression of fifteen oncogenes. They observed three different patterns of expression: no detectable transcription, transcription in some types of tumours only, and transcription in all tumours examined. It is to this third category that the two ras genes examined (Kirsten and Harvey) belong, along with the myc and fos genes. Furthermore, in sixteen of the fifty-four patients a comparison between apparently normal, adjacent tissue and the corresponding malignant tissue, could be made. This consistently showed elevation of ras specific transcripts in the malignant tissue. However, in three cases, the levels of all the expressed oncogenes examined were similar in the tumour and the histologically normal adjacent tissue. In these cases qualitative, rather than quantitative activation of the oncogenes responsible may have been more important. Alternatively, an oncogene not examined, or one yet to be identified, may have been responsible. Caution should be used in interpreting observations such as these, as in some cases, the differential expression between apparently normal and malignant tissue may be simply explained by cell-type differences between the two tissues. While this consideration is obviously important, the similar levels of expression of Harvey and Kirsten ras genes reported in all tissues thus far examined (see next section), may exclude these genes from this criticism.

Other studies have also correlated elevated ras expression with malignancy. Spandidos and Kerr (1984) have observed that malignant tumours of the colorectum have 4 to 14 times higher Harvey and 3 to 19 times higher Kirsten

specific transcripts than the corresponding normal colorectal mucosa. Interestingly, even higher levels of expression of these genes were detected in most of the premalignant lesions examined, i.e. 8 to 13 times higher Harvey and 9 to 20 times higher Kirsten specific transcripts. In an independent study (Gallick et al., 1985) the levels of ras protein ($p21^{ras}$) were shown to mirror almost exactly the levels of ras specific RNA observed by Spandidos and Kerr (1984). Similarly, enhanced ras protein levels have been detected in the majority of human colon and mammary carcinomas examined by Horan-Hand et al. (1984). Another pathway towards elevated ras expression which has been well documented, is amplification of ras genes. For example, transformation of NIH 3T3 murine fibroblast cells with the normal human c-Ha-ras1 gene, has been shown to result from elevated transcription from multiple copies of this proto-oncogene, rather than from spontaneous mutation of any of the transfected c-Ha-ras1 genes (Pulciani et al., 1985). Furthermore, Pulciani et al. (1985) were also able to detect 50 to 60 fold amplification of the human c-Ki-ras2 locus in one lung carcinoma out of 75 tumour biopsies. Schwab et al. (1983) have also detected an amplified (30 to 60 fold) c-Ki-ras gene in Y1 mouse adrenocortical tumour cells. In this latter study, c-Ki-ras specific mRNA and protein were shown to be correspondingly elevated. More recently, a human gastric carcinoma has been described (Bos et al., 1986), containing not only an amplified normal allele of the c-Ki-ras2 gene, but also a single mutated c-Ki-ras2 oncogene. Using a transformation assay, the mutated allele was shown to have replaced the usual glycine of codon 12 with a serine residue, a moderately transforming allele as defined by Seeburg et al. (1984). In this case it is possible that neither the serine mutation, nor the amplification itself, were sufficient for tumour development and that the second potential activating step occurred during tumour progression. However, the possible

involvement of other oncogenes was not examined by Bos et al. (1986). While the data described above can be interpreted as evidence for an involvement of expression of ras proto-oncogenesis in the development, it is also possible, given the likely involvement of elevated ras expression in cellular proliferation (see next section), that such increases may be a consequence of cell proliferation in tumour cells. Therefore, until it is more precisely known what levels of ras expression are required to trigger malignant growth in different cells, increased ras expression should be viewed cautiously as a model for such growth.

Finally, the animal model systems for carcinogenesis have been developed and have the great advantage that tumour progression can be studied from initial treatment with carcinogens, through benign to malignant lesions. One such experimental system has been well characterized (Burns et al., 1978) and is based on the chemical induction of tumours in mouse skin. Tumours are induced on the skin of sensitive mice by a single treatment with a mutagen, termed the tumour initiator, followed by multiple treatments with a tumour promoter. The tumour promoter is believed to be necessary for the rapid proliferation of appropriately mutated cells into benign papillomas. Most of these benign papillomas are tumour promoter-dependent since they regress once tumour promoter treatment is stopped. This suggests that the initial mutational event(s) cannot be sufficient for full malignant transformation. However, some of the benign papillomas progress to form autonomous benign lesions and malignant carcinomas, presumably determined by other mutational events. The early dominant mutational event has been shown to involve mutational activation of the mouse c-Ha-ras1 oncogene, which is also coupled to elevated expression of this gene in the papillomas relative to the normal epidermis (Balmain and Pragnell, 1983; Balmain et al., 1984). Thus, in this case, both activation and elevated expression of the mouse c-Ha-ras1 oncogene

occurred early in tumour progression. Since papillomas proliferate rapidly and reach a substantial size, ras gene activation and elevated expression are also directly correlated with proliferating cells. Recently, introduction of the Harvey murine sarcoma virus (HaMSV) directly into mouse epidermal cells has been shown to substitute for treatment with the initial carcinogen (Brown et al., 1986). In this case, the latency period before the appearance of the papillomas, and also for conversion of papillomas to carcinomas, was shorter after initiation with HaMSV than it was with initiating carcinogen. This is probably due to the viral promoter elements which induce higher levels of expression than are observed in chemically induced tumours.

In summary, it has been shown that oncogenic activation of the classical ras genes may not solely be a function of activating mutations clustered around codons 12 and 61. There is now evidence that oncogenic activation of these genes may also be caused by elevated expression and it is entirely possible that both these mechanisms may act together in the genesis of at least some tumours. Thus it is clear that the regulation of these important genes must be tightly controlled to prevent cells acquiring some neoplastic characteristics. No known regulatory defects in any ras genes have been identified and characterized in cells from naturally occuring tumours. The question of how ras expression is genetically controlled is unanswered.

(d) Expression of ras genes during development, cell proliferation and the cell cycle

A systematic investigation of proto-oncogene transcription during pre- and post-natal development of mice has been undertaken by Muller and colleagues (Muller et al., 1982, 1983). Most of the proto-oncogenes they examined were expressed transcriptionally in a tissue- and developmental stage-specific manner. In contrast, the

cellular Harvey and Kirsten ras genes were found to be expressed ubiquitously in mouse embryos, fetuses, placentas and extraembryonal membranes, as well as in various post-natal tissues. Furthermore, the level of expression of both of these ras genes represented a constant level of between 0.01% and 0.05% of the total cellular polyadenylated RNA. Hence, expression of these ras genes is unlikely to be involved with the differentiation of particular cell types. In Drosophila melanogaster, which has three genes homologous to Harvey ras probes (Shilo and Weinberg, 1981), ras expression from each of these genes was also found in all developmental stages examined. Interestingly, each gene encoded at least two distinct transcripts, the shorter transcript from each being more abundant in embryos, whereas the longer transcript was expressed ubiquitously at a constant level (Lov. et al., 1985).

Evidence that ras gene expression may be involved in cellular proliferation comes from studies on regenerating liver cells. Liver regeneration is one of the most striking examples of compensatory growth in mammals, and rat hepatocytes are known to respond to liver damage induced by partial hepatectomy or chemical injury, by undergoing a burst of DNA synthesis and replication. This burst of cell division proceeds synchronously for at least two cell cycles and it was observed that the levels of both Kirsten and Harvey ras transcripts increases 2 to 4-fold concommitant with the onset of DNA synthesis (Goyette et al., 1983; Goyette et al., 1984). To control for the general increase in transcriptional activity observed in this system, the levels of albumin transcripts were shown not to increase during liver compensatory growth. Furthermore, the peak in the amount of ras transcripts was shown to be both greater and to occur earlier, than that of α-fetoprotein transcripts, which are known to increase in regenerating liver. Goyette and colleagues interpreted their observations as indicating that there is regulated

transcription of the Harvey and Kirsten ras genes during the physiological growth process of liver regeneration. However, they did not eliminate the possibility that these ras transcripts are differentially stabilized during the regeneration process. Furthermore, the correlation of ras expression with DNA synthesis does not imply a causal involvement with cell growth. Interestingly, the levels of most other proto-oncogenes examined in this system remained unchanged, the exception being the elevation of the c-myc gene, which preceded that of the Harvey and Kirsten genes (Goyette et al., 1983, 1984).

The liver regeneration system suggests ras gene expression may be regulated within the cell cycle. In animal cells, the cell cycle has been divided into four distinct phases (reviewed by Pardee et al., 1978). They are; 1) the gap period between mitosis and the initiation of DNA synthesis, called G1 phase; 2) the synthesis of DNA, called S phase; 3) the gap period between DNA synthesis and mitosis, called G2 phase and 4) mitosis, called M phase. In addition, it is known that the crucial events for the regulation of growth occur mainly in the G1 phase and a "commitment" point in mid to late G1 has been defined (Pardue et al., 1978). At the commitment point, the cell may either enter the S phase, or cease proliferation and enter a prolonged G1 phase (sometimes called G0) referred to as quiescense. Campisi et al. (1984) have shown, using Balb/c 3T3 (A31) murine fibroblasts, that the relative abundance of c-Ha-ras1 specific transcripts was increased five to seven fold in S phase cultures. This cell cycle dependent elevation of ras transcripts was not affected in chemically transformed sub-lines of these cells. Furthermore, the elevation of the c-Ki-ras2 specific transcript was first detected in mid to late G1. This contrasts with the c-myc proto-oncogene whose expression is now known to be continuous throughout the cell cycle (Thompson et al., 1985). This analysis prompted Campisi et al. (1984) to speculate that ras gene expression is

involved in G1 regulatory events which prepare the cell for DNA synthesis and subsequent cell division. This speculation is supported by the observation that NIH 3T3 murine fibroblasts, induced to divide by the addition of serum to the culture medium, are unable to enter S phase following microinjection of anti-ras antibodies (Mulcahy et al., 1985). This observation provides strong evidence that the protein product of the ras proto-oncogene is required for the initiation of S phase in these cells. In the inverse experimental approach, Ferramisco et al., (1984) microinjected the T24/EJ bladder carcinoma cell line c-Ha-ras1 protein product into quiescent cells. After microinjection, DNA synthesis was induced, but cell division in at least one cell line examined required re-feeding with serum. These data also suggest that the c-Ha-ras1 function at least, is required for the initiation of S phase, but that cell division requires other regulatory events.

The remarkable evolutionary conservation of the ras multigene family (see section I.a) suggests that these genes encode functions that are essential to all eukaryotic cells. Therefore, the study of the function of ras genes during the development of lower eukaryotic organisms, which have simple life cycles and are more amenable to genetic analysis, should provide insight into the function of ras genes in the higher eukaryotes. One such suitable organism is the slime-mould, Dictyostelium discoideum, which is unusual in that the processes of cell division and differentiation are largely separate. That is, there is a proliferative amoeboid stage which when starved, ceases cell division and aggregates to differentiate into spore and stalk cells (Loomis, 1975). Expression of the single ras-related gene in D. discoideum appears to be linked to the mitotically active proliferative cells only (Reymond et al., 1984). Furthermore, the single ras-related protein, which is specifically precipitated by monoclonal antibodies raised against mammalian $p21^{ras}$, is detected only in

mitotically active cells (Pawson et al., 1985). The most intensively studied lower eukaryote is the yeast, Saccharomcyes cerevisiae. In this organism ras gene expression is also required early in the cell division cycle (CDC) as double mutants of the RAS1 and RAS2 alleles fail to germinate or to form buds (Tatchell et al., 1984; Kataoka et al., 1984), suggesting a block quite early in the yeast cell division cycle. However, comparisons between yeast and mammalian ras genes should be interpreted cautiously because of recently determined differences in their biochemical properties (Beckner et al., 1985), as discussed in the next section. Despite these differences, it is of interest that the human c-Ha-ras1 gene can be used to complement the inviability of yeast ras1$^-$ ras2$^-$ double mutants and conversely that a modified RAS 1 yeast gene can transform NIH 3T3 cells (Kataoka et al., 1985; DeFeo-Jones et al., 1985). Therefore, although aspects of the biochemistry of mammalian and yeast ras genes are known to have diverged, it is clear that functional homology must exist, as well as the obvious sequence homologies, between mammalian and yeast ras genes.

In this section, it has been observed that the expression of the Harvey and Kirsten ras genes appears to be ubiquitous in all cell types and developmental stages, accounting for constant levels of between 0.01% and 0.05% of the total polyadenylated RNA in non-malignant mammalian cells. Furthermore, evidence has been given which suggests that ras expression is cell cycle regulated, such that a peak of ras specific transcription is observed in mid to late G1, which is a period of the cell cycle known to be important for regulation of cell division. Furthermore, there is now direct evidence implicating ras gene function in the initiation of the S phase of the cell cycle.

(e) Function of the ras gene products

Recent biochemical and structural studies support

the hypothesis that ras proteins are involved in the transduction of extracellular signals in a manner analogous to the G proteins. The G proteins are known to transduce extracellular signals from their various receptor molecules to their appropriate effector enzymes (Stryer, 1983; Lefkowitz, 1983). These signals can be as diverse as a photon of light impinging on the retina, to a hormone molecule binding to a receptor molecule on the surface of a cell. Like G proteins, the mammalian ras gene products are associated with the inner leaflet of the cell membrane (Willingham et al., 1980), bind GTP (Scholnick et al., 1979) and they exhibit a slow GTPase activity (Sweet et al., 1984). Furthermore, the nucleotide sequence of the T_{α} subunit of bovine transducin (the G protein involved in light transduction), has been determined and found to have several domains closely homologous to the mammalian ras proteins and to various other GTP binding proteins (Lochrie et al., 1985; Tanabe et al., 1985; Medynski et al., 1985; Yatsunami and Khorana, 1985). Hurley et al. (1984) have also observed significant homology between the N-terminal sequence of the human $p21^{ras}$ proteins and the two mammalian G proteins, G_s and G_i, which are known to interact with adenylate cyclase. In the yeast, Saccharomcyes cerevisiae, genetic and biochemical analyses have suggested that the main physiological function of the RAS gene products, is the control of adenylate cyclase activity (Toda et al., 1985). This suggests that the yeast ras proteins are functional equivalents of the mammalian G proteins controlling adenylate cyclase. In this yeast the level of cAMP is known to be involved in cell cycle control (Matsumoto et al., 1982). Conversely, it is now known that the mammalian Harvey, Kirsten and N-ras gene products are not controlling elements of mammalian adenylate cyclase (Beckner et al., 1985). This functional divergence between yeast and mammalian ras proteins suggests that the putative G protein function of the mammalian ras proteins must be coupled to effector enzymes other than adenylate cyclase.

A good candidate for the ras effector enzyme is phospholipase C (Berridge and Irvine, 1984; Berridge, 1984), which is responsible for the generation of two specific second messengers from inositol lipids, the so-called phosphatidylinositol (PI) turnover cycle (see review by Berriddge and Irvine, 1984). Of the two second messengers generated, diacylglycerol (DG) acts in the plane of the cell membrane to activate protein kinase C, generating a cascade of events ultimately leading to an increase in cellular pH. The other second messenger generated, inositol trisphosphate ($InsP_3$), is released into the cytoplasm where it leads to a mobilization of intracellular calcium stores. Increased cellular pH and calcium concentration are characteristic features of the action of growth factors and are the two major ionic events implicated in cell proliferation. Evidence for the hypothesis that the mammalian ras proteins are candidates for the unknown G protein(s) that modulate the PI turnover cycle comes from studies on ras-transformed cells. Fibroblast cell lines have been transformed with Harvey, Kirsten and N-ras genes and the levels of PI turnover cycle catabolites, including DG and $InsP_3$, were shown to be elevated with respect to the untransfromed cells (Fleischman et al., 1986; Wolfman and Macara, 1987). Also, Wakelam et al. (1986) have constructed an NIH 3T3 murine fibroblast cell line containing the unmutated "normal" N-ras gene under the control of a glucocorticoid inducible promoter. In uninduced cells the level of $p21^{N\text{-}ras}$ protein is negligible; however, induced cells contain large amounts of this protein. This is correlated with a dramatic increase in the levels of inositol phosphates and an increase in DNA synthesis when the cells are treated with certain growth factors. The fact that several growth factors elicited this response to varying degrees, suggests that $p21^{N\text{-}ras}$ interacts with a spectrum of growth factor receptors. Furthermore, since the expression of $p21^{N\text{-}ras}$ caused such a dramatic response, this suggests that in

these cells the concentration of the usual complimentary G protein is limiting. These observations led Wakelam et al. (1986) to propose that normal $p21^{N\text{-}ras}$ is a complimentary protein, akin, if not identical to, the putative G protein which allows growth factor receptors to stimulate inositol lipid metabolism by activating phospholipase C.

Given the large number of ras genes now known to exist (see section I.a), it is possible that most may serve only a few common effector enzymes, like phospholipase C, but couple to many distinct signal receptor molecules in a manner analogous to $p21^{N\text{-}ras}$. Thus, the proteins of the ras multigene family could respond to many different signals in an integrated way. Previously discussed observations on the pattern of ras expression during development and the cell cycle (section I.d), suggested that ras gene expression is involved in the control of cell growth, possibly by having a role in the initiation of DNA synthesis. However, it is also possible that other members of the ras multigene family, which are more divergent in sequence from the classical ras genes, may have a function in different cellular processes. For example, the S. cerevisiae ras-related YPT1 gene product, which is nearly 40% homologous to the mammalian $p21^{ras}$ gene products (Gallwitz et al., 1983), seems to be involved in microtubule organization and function (Schmitt et al., 1986). Furthermore, the Schizosaccharomyces pombe RAS1 protein, which is highly homologous to the S. cerevisiae RAS1 and RAS2 proteins, is not involved in vegetative growth nor regulation of cAMP levels, but is required for mating (Fukui and Kaziro, 1985; Fukui et al., 1986).

As yet no known functions for other ras-related genes, such as the rho and R-ras genes found in the genomes of higher eukaryotes (see section I.a), have been reported. However, some studies have even suggested that the mammalian ras genes may also function in the differentiation of certain cell types. Treatment of rat pheochromocytoma cells (PC12) with nerve growth factor

(NGF) results in the induction of a number of phenotypic characteristics of sympathetic neurones, including cessation of cell division and outgrowth of neuronal processes (neurites). Microinjection of the oncogenic (T24/EJ) form of the human Ha-ras protein, but not the proto-oncogenic form, mimics the effect of NGF treatment on these cells (Bar-Sagi and Feramisco, 1985). Furthermore, the NGF-induced differentiation was inhibited when monoclonal antibodies raised against viral Harvey ras protein, which recognize all rat $p21^{ras}$ species, were microinjected into PC12 cells (Magag et al., 1986). Thus, in this system, $p21^{Ha\text{-}ras}$ appears to be involved in the induction of neurite formation in PC12 cells, but only an oncogenic form of this protein provides a sufficiently powerful signal for this differentiation process. It is not known whether the classical ras proteins are typically involved in other differentiation processes. However, it may be that $p21^{ras}$ involvement in neuronal development is a pecularity of PC12 cells.

Finally, Birchmeier et al. (1985) have shown $p21^{ras}$ involvement in another highly specialized cellular function. They have shown by microinjection, that both normal and oncogenic Harvey ras proteins can induce maturation (meiosis) of Xenopus oocytes. This maturation is independent of progesterone treatment, which is the normal physiological inducer of oocyte maturation. The $p21^{Ha\text{-}ras}$ induced oocyte maturation is nearly 100-fold more potent for the oncogenic version of the protein than it is for the normal wild type version. As oocyte maturation involves the stimulation of oocytes arrested at prophase of meiosis to complete meiosis, this phenomenon may be related to the ability of ras proteins to advance somatic cells through the mitotic cell division cycle.

In summary, there is strong evidence that ras proteins are of vital importance to all cells. This provides a basis for the hypothesis, that elevated expression of ras genes can directly lead to oncogenesis. Moreover, as

already discussed, elevated ras expression has been correlated with oncogenesis. Many observations suggest that the ras gene products act at a key regulated step in the transduction of extracellular information into the cell, in a manner analogous to the various known G proteins. Similarities have been found between these G proteins and the ras proteins and it has been hypothesised that the mammalian classical ras proteins regulate the production of the second messengers, diacylglycerol and inositol trisphosphate. The possibility has been raised that members of the ras multigene family of closely related proteins may differ in functional properties, such as the reception of different extracellular signals or coupling to different effector enzymes. The intriguing possibility that the mammalian ras gene products may function in the differentiation or maturation of certain cell types, has also been suggested.

II. ASPECTS OF GENE REGULATION

Multiple levels of control are exerted on the genetic information contained within the cell nucleus. Nuclear architecture, chromatin structure, transcription, RNA transport, RNA stability, translation, and processing of RNA and protein products, are all levels at which genetic information may be regulated. However, as this thesis is concerned with the transcriptional regulation of the human c-Ha-ras1 gene, this discussion will be limited to aspects of gene regulation which are likely to have relevance to the transcription of this gene. Later discussion will focus on the hypothesis that transcription of the ras genes is more like that of the "housekeeping" genes, than that of the more often studied "luxury-protein" genes. "Housekeeping genes" are those whose products are involved in various metabolic processes common to all cells. Examples of such processes are energy production, energy

utilization and DNA synthesis. The so called "luxury-protein" genes are those which encode and produce proteins, usually in bulk, in specific types of terminally differentiated cells. Such genes are exemplified by the globin genes (reviewed by Karlsson and Nienhuis, 1985).

(a) Transcriptional control signals

Genes are transcribed by one of three different RNA polymerases in higher eukaryotes. The class of RNA polymerase used for transcription depends on recognisable structures (DNA or DNA-protein binding complexes) surrounding the gene (Baker and Platt, 1986; Nevins, 1983; Enver, 1985). These structures constitute the "promoter" and for protein coding genes they allow the promoter to be recognized by RNA polymerase II (pol II) and its associated transcription factors (recently reviewed by Gidoni et al., 1984; Dynan and Tjian, 1985; Serfling et al., 1985; Ptashne, 1986; McKnight and Tjian, 1986). Unlike prokaryotic RNA polymerases, purified pol II requires protein factors for promoter recognition and correct initiation of transcription *in vitro* (Manley et al., 1980). Some of these factors seem to be required for initiation at all promoters, whereas others are required for initiation at specific promoters only (Dynan and Tjian, 1983a). The sequence-specific DNA binding proteins can activate transcription and probably allow the transcription machinery to discriminate between different promoters.

(i) The TATA box

Early experiments on "luxury-protein" genes involved functional testing of *in vitro* manipulated cloned DNA and demonstrated the importance of several concensus sequences, located upstream from these genes. One such concensus sequence, TATAAA, called the TATA box, is located 25-30 bp upstream from the transcription start site (Breathnach and

Chambon, 1981). Mutation of the TATA box in a sea urchin histone gene (Grosschedl and Birnsteil, 1981) and in the SV40 early gene region (Ghosh et al., 1981; Benoist and Chambon, 1981), generates 5' (upstream) heterogeneity in the transcripts produced, although the overall level of RNA synthesis was not significantly reduced. In the case of Drosophila heat shock genes, a protein, or proteins, have been shown to bind to the TATA box, independently of heat shock, both in in situ chromatin (Wu, 1984) and on free linear DNA (Wu, 1985). Davidson et al. (1983) have reported a HeLa cell factor which is necessary for in vitro transcription from conalbumin and adenovirus late gene promoters and which also binds to the TATA box region. Weintraub (1985) has used novel high resolution mapping techniques to show that the TATA box region of the major late promoter in adenovirus chromatin, is a dominant site of S1 and DNAse I hypersensitivity. This suggests that the DNA structure at the TATA box is in an altered conformation, perhaps caused by protein binding. Recently, the technique of photofootprinting in vitro, has been used to detect transcription-dependent changes in the photomodification of yeast TATA boxes (Selleck and Majors, 1987). It is believed that the altered sensitivity to photomodification results from a DNA structural change, mediated by the binding of the TATA-specific protein(s) in yeast. This structural change, induced by binding, could be a similar or related feature to that detected by S1 and DNAse I nucleases as a hypersensitive site.

(ii) The CAAT box

Further upstream from the TATA box, usually at position -70 to -80 relative to the transcriptional start site, an additional conserved sequence of concensus GG($^{C}/_{T}$)CAATCT, the so called CAAT box, has been found in several cellular and viral protein-coding genes (Benoist et al., 1980; Efstratiadis et al., 1980; Breathnach and

Chambon, 1981; Shenk, 1981). However, with the accumulation of DNA sequence information on the promoter regions of many more genes, it is now known, that the CAAT consensus sequence is not so widespread in promoters, as that of the TATA box. The CAAT box is believed to function as a transcriptional activator, as mutation of the CAAT boxes of the herpes simplex virus (HSV), thymidine kinase, the human and mouse β-globin and the Xenopus hsp 70 genes, reduced RNA synthesis by as much as 10-fold in vivo (McKnight et al., 1984; Graves et al., 1986; Charnay et al., 1985; Myers et al., 1986; Bienz and Pelham, 1986). Furthermore, deletion of the human α-globin CAAT sequence, reduces transcription in vivo, at least 25-fold (Mellon et al., 1981).

At least two different protein complexes are now known to bind CAAT sequence motifs, one isolated from HeLa cells, called CAAT-binding transcription factor, CTF (Jones et al., 1985) and the other from rat liver, called CAAT binding protein, CBP (Graves et al., 1986; Jones et al., 1985). Cohen et al. (1986) have also reported an activity which binds to mouse α- and β-globin promoter CAAT sequences, but it is not yet known if this factor(s) correspond to CTF or CBP, or even to a third distinct CAAT box binding factor. The factors, CTF and CBP, have been shown to have distinct biochemical properties and sequence requirements (Graves et al., 1986; Jones et al., 1987). For example, their chromatographic and heat stability properties are different, and they do not generate identical footprint patterns on the HSV1 tk promoter. Furthermore, mutation of the CCAAT sequence to GCAAT, which inhibits transcription in HeLa cells, eliminates binding of CTF but enhances binding of CBP. Jones et al. (1987) have also shown that CTF is identical to the previously identified adenovirus DNA replication stimulating factor, NF-1. This surprising result, indicates that CTF/NF-1 can have multiple functions, as it activates both viral DNA replication and transcription. Multiple functions for

CTF/NF-1 are also suggested by its complex multimeric nature. The purified CTF/NF-1 activity consists of a heterogeneous population of polypeptides of 52 to 66 kilodaltons (Jones et al., 1987). Transcriptional activation by CTF/NF-1 is likely to be determined by the context of the binding site, in relation to the location of other transcription factors. This suggestion comes from the observation that the high affinity CTF-binding site, located upstream of the human β-globin promoter, at about position -200, does not contribute to β-globin gene expression, whereas the low affinity binding site at position -80, is fully functional (Jones et al., 1987). A possible explanation for such differential behaviour, is that transcriptional activation requires interaction of CTF with other linked transcription factors. Such an interaction is suggested by the demonstration that the inverted CAAT box at position -55 of the Xenopus hsp 70 gene is required for long distance activation of the promoter by the upstream heat shock elements (HSEs). However, the CAAT box was not required if the HSEs were moved close to the TATA box (Bienz and Pelham, 1986). A similar effect was noticed for the SV40 enhancer region, as maximal long-range activation of β-globin genes by this enhancer, requires a CAAT box located within 80 bp of the TATA box (Dierks et al., 1983). Another role for the CAAT box binding proteins which bind the inverted CAAT box at position -55 of the hsp 70 gene, is suggested by the apparent cell-type-specific regulation of this gene during oogenesis. The hsp 70 gene is constitutively expressed in Xenopus oocytes (Bienz, 1984) and this level of expression is unaffected by heat shock, whereas hsp 70 transcripts are undetectable in all somatic cells prior to heat shock. The HSEs and the CAAT box were the only sequences which could be defined by mutational techniques as regulators of this cell-type specific response (Bienz, 1986). Thus, cell-type specific activation of the Xenopus hsp 70 gene does not require a cell-type specific promoter element. To explain

this observation, Bienz (1986) suggested that oocytes contain high levels of the heat shock transcription factor (HSTF) and/or the CAAT-binding activity, which leads to "permissive" expression of hsp 70, while somatic cells contain subcritical amounts of at least one of the two factors. Alternatively, enhanced interaction between HSTF and the CAAT-binding activity may occur in oocytes due to oocyte-specific forms of either one of the binding activities.

(iii) The GC box

Another upstream sequence motif which is found in many genes is the GC box. This motif has a concensus sequence of 5' $(^{G}/_{T})$GGGCGG$(^{G}/_{A})(^{G}/_{A})(^{C}/_{T})$ 3', although analysis of the human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter, suggests that other related sequences may also be functional GC boxes (Kadonaga et al., 1986). The GC box was first identified as the binding site of a factor isolated from human HeLa cells, called Sp1. This factor was found to activate transcription of the early transcription unit of the simian virus 40 (SV40) genome (Dynan and Tjian, 1983a). As the SV40 early promoter is central to any discussion of the Sp1 transcription factor, but also because the early promoter shares some characteristics with the c-Ha-ras1 gene promoter and the promoters of most housekeeping genes, it is worthwhile considering this promoter in some detail.

The 300 bp region that lies between the early and late SV40 transcription units, contains at least four specific types of regulatory elements and functions as a bidirectional promoter unit for early (anticlockwise) and late (clockwise) viral transcription (reviewed by McKnight and Tjian, 1986 and see Figure 3-1). Proceeding clockwise from the early to the late transcriptional units, the first regulatory elements encountered are three T antigen binding sites, the middle one of which overlaps the early RNA

initiation sites. These binding sites, when bound by T antigen, function to repress early gene transcription. Also, within this region is the origin of DNA replication, which is also bound by T antigen. Adjacent to and overlapping the replication origin, is a TATA-like sequence motif which specifies the locations of early gene transcriptional initiation sites and directly following this motif are three direct repeats of a G + C rich 21 bp sequence. Each 21 bp repeat contains two copies of the concensus GC box sequence and complete deletion of this regulatory element severely limits early promoter function. Finally, two more tandem repeats of 72 bp each are found to the right of the 21 bp repeats on the "late" side of the 300 bp SV40 control region. It is from the 72 bp repeat, nearest to the late transcription unit that late transcription is initiated. However, deletion of both 72 bp repeats also reduced early transcription, whereas deletion of just one has virtually no effect on early transcription and virus viability. Like the 21 bp repeats, the 72 bp elements stimulate transcription in either orientation but unlike the 21 bp repeats, the 72 bp elements can have their effect even when placed thousands of base pairs away from the 300 bp SV40 control region. This is the basis of the so called enhancer activity, which will be discussed more fully in the next section.

Dynan and Tjian (1983b) have shown that Sp1 selectively binds to the 21 bp repeats using the *in vitro* DNAse I footprinting technique. Further DNAse I footprinting analysis, along with dimethyl sulphate (DMS) methylation protection experiments, have shown that Sp1 binds in a sequence-specific manner to the GC boxes, that occur twice in each 21 bp repeat (Gidoni et al., 1984; Gidoni et al., 1985; Dynan and Tjian, 1985). However, Sp1 forms strong contacts with only three of the six GC boxes and weakly contacts two others. This is because steric constraints and differences in binding affinities between the six potential sites limits binding to only five sites

at any one time. It is believed, that when the 21 bp repeat region is fully occupied, Sp1 binds as five protamers, each of more than 100 kilodaltons (Kadonaga and Tjian, 1986). These promoters lie predominantly on the same face of the DNA helix and each is capable of protecting approximately 20 bp of DNA from DNAse I (Gidoni et al., 1984). Furthermore, contacts have been detected only in the major groove of the DNA helix and on one strand only (Gidoni et al., 1984; McKnight and Tjian, 1986), an observation which has not been observed with the more intensively studied, prokaryotic DNA binding proteins.

The transcription factor Sp1 is also known to interact with other viral promoters, such as the promoters of the herpes simplex virus (HSV) immediate early genes -4/5 and 3 (IE-4/5 and IE-3) and also the HSV tk gene. Like the SV40 300 bp control region, which directs the synthesis of both early and late transcripts in divergent directions, the IE-4/5 and IE-3 transcripts are also divergently transcribed from an approximately 700 bp control region. Unlike the SV40 control region, this 700 bp control region has 12 GC boxes, which have been shown to bind Sp1 in vitro and are spaced at greater intervals and in either orientation (Jones and Tjian, 1985). Promoter activity correlates with the Sp1 binding sites proximal to each IE gene. Interestingly, the more distal Sp1 binding sites also closely flank the TAATGARAT elements implicated in transactivation of IE transcription, by virion-associated proteins (Jones and Tjian, 1985). It is believed that the Sp1 protein, bound to the GC boxes, potentiates this induction of IE transcription in vivo (Jones and Tjian, 1985). Preston et al. (1984) have also shown that deletion of at least one of the distal Sp1 binding sites, decreases the level of transcriptional induction of both the IE-4/5 and IE-3 genes. Therefore, at least at this site there is overlap between the sequences required for viral activation and transcription.

The role of Sp1 in HSV transcription is not linked

solely to IE gene transcription, it is also implicated in transcription of the "delayed-early" tk gene (Jones et al., 1985). For this gene, mutagenesis experiments have established that only two GC boxes are required (McKnight et al., 1981; McKnight and Kingsbury, 1982). One is located in a region termed distal region I (dsI) and is adjacent to a TATA box, called the proximal signal (ps). The other is located further upstream in distal region II (dsII), along with an inverted CAAT box. The distal GC box is orientated in the direction of transcription, while the dsI GC box is orientated away from the direction of transcription. Sp1 binding is appreciably greater to dsII than to dsI (Jones et al., 1985). Interestingly, an inversion mutant, which exchanges the positions of dsI and dsII, thereby bringing the inverted CAAT box into close proximity with the TATA box of ps, can now function in the absence of dsI (McKnight et al., 1984). As disruption of either dsI or dsII, in their normal positions, is known to severely reduce promoter function in vivo (McKnight, 1982), this observation suggests, that Sp1 bound to ps may simply act by linking the factors bound at dsII to those found at the TATA box containing ps region. The effect of a weak Sp1 binding site in dsI, would be to reduce the effect of the strong binding site in dsII. This would allow transcription of the delayed-early tk gene, only when the levels of the Sp1 factor recruited to the vicinity of the viral genome, were high enough to favour binding to the weaker Sp1 binding site of dsI. The possibility that these proteins may interact in an "interdigitated" manner, possibly forming a framework to which other regulatory proteins may bind, is also suggested by their positioning along the same face of the DNA helix (McKnight and Tjian, 1986).

Recently, it has become apparent that Sp1 may also interact with several cellular genes, as a number of GC box-containing promoters have been described. Thus far these genes are mainly members of the "housekeeping" class

and the various patterns of GC boxes associated with some of them will be discussed later (see section II.c). As yet only one of these housekeeping genes, the mouse dihydrofolate reductase (dhfr) gene, has actually been tested for Sp1 binding *in vitro* (unpublished data reported in, Kadonaga et al., 1986; Dynan and Tjian, 1985). For other housekeeping genes, Sp1 binding has only been predicted from the DNA sequence. However, GC boxes have also been described for the human metallothionein I_A and II_A promoters (Richards et al., 1984) and these have been shown to bind Sp1 (unpublished data reported in Kadonaga et al., 1986). The metallothionein I_A and II_A Sp1 binding sites are interspersed with other important regulatory sequences, such as the metal regulatory elements (MREs) and basal level enhancer sequences (BLEs). Finally, a DNA segment cloned from the monkey genome by homology with the SV40 origin of replication, which functions *in vivo* as a bidirectional promoter element (Saffer and Singer, 1984), contains two Sp1-binding regions with closely spaced, tandemly arranged GC boxes (Dynan et al., 1985).

The TATA, CAAT and GC-box motifs and the trans-acting factors that have been identified which bind to them, have been described in detail. Given the large number of genes which contain these motifs, it is likely that they generally represent constitutive transcriptional elements for the genes containing them. The amount of factor available to these elements within different cell types and also their actual arrangement within promoter regions, could be used to modulate gene activity. However, to explain tissue-specific transcription during development and differentiation of different cell types, tissue-specific promoter elements are believed to be involved. These include the *Drosophila* heat shock element, HSE, which is recognized by the heat shock transcription factor, HSTF (Parker and Topol, 1984), the glucocorticoid response element, GRE (Miesfeld et al., 1986), the metal response element, MRE (Richards et al., 1984) and the

octameric element essential for the efficient use of immunoglobulin promoters in B cells (Singh et al., 1986). It is also possible that these site-specific DNA-binding proteins participate in different cellular processes. For example, the octamer element important in immunoglobin gene transcription is also found as a regulatory motif in histone H2B and U2 snRNA genes (Singh et al., 1986; Sive and Roeder, 1986). It can also be specifically bound by another cellular activity required for adenovirus replication (Jones et al., 1987). These observations support the emerging notion that some DNA binding activities may be redundant in nature, allowing them to participate in many different regulatory events. The specificity of such events, whether transcriptional or replicational, being determined, at least in part, by the context of the particular binding site.

(iv) Enhancer elements

In addition to the sequence modules discussed above, the enhancer elements are characterized by many individual motifs clustered together. These elements have unusual properties which were first observed for the SV40 enhancer (Banerji et al., 1981; Moreau et al., 1981; Fromm and Berg, 1983). They include: 1) strong transcriptional activation (up to 1000-fold) of a linked gene, 2) transcriptional activation is independent of enhancer orientation and 3) the enhancer can function over distances of greater than 1000 bp from an upstream or downstream position, relative to the transcriptional initiation sites. The SV40 enhancer is referred to as promiscuous because of its ubiquitous role in stimulating RNA polymerase II transcription in nearly all mammalian cell types and also in amphibian and even algal cells (reviewed by Serfling et al., 1985). The magnitude of the stimulation however, depends on the promoter examined. Nuclear run-on transcription assays suggest that this enhancer increases the density of RNA

polymerase II molecules within the linked gene (Treisman and Maniatis, 1985; Weber and Schaffner, 1985). Both *in vitro* and *in vivo* competition experiments implicate *trans* acting factors in this activity (Sassone-Corsi et al., 1985; Wang and Calame, 1986). *In vitro* DNA-protein binding techniques have identified specific factors which interact with the SV40 enhancer and also with various cellular regulatory sequences (Bohman et al., 1987; Lee et al., 1987).

Unlike the promiscuous SV40 enhancer many other viral enhancers have been described which show cell-type specificity, for example, the SV40-related mouse polyoma virus is about four times more active in mouse cells than in primate cells. An explanation for this observation is suggested by experiments showing that the polyoma enhancer competes only moderately for the SV40 enhancer factors (Sassone-Corsi et al., 1985). Viral enhancer elements exhibiting varying degrees of cell-type specificity have also been detected in the early region of adenoviruses, the immediate early regions of herpes viruses and in the long terminal repeats of retroviruses (Serfling et al., 1985).

Cellular enhancers are usually found in the 5' flanking sequences of genes (typically between position -100 and -300 bp relative to the cap site(s)), where they may physically and functionally overlap upstream promoter elements, such that the distinction between these two types of regulatory elements becomes unclear. In the metallothionein (MT) genes of mouse and man, the upstream regions contain constitutive sequence elements, interspersed with the regulatory ones, such as a TATA box, GC boxes and basal level enhancer (BLE) sequences (Stuart et al., 1986; Carter et al., 1984; Karin et al., 1984; Haslinger and Karin, 1985). Cellular enhancers are also found within, and 3', to their respective genes. In fact, the first cellular enhancer discovered, was found in the intron separating the rearranged V_HDJ_H and C_H segments of the immmunoglobulin heavy-chain gene (Gillies et al., 1983;

Banerji et al., 1983). This discovery was of particular importance as this immunoglobulin enhancer was the first regulatory element identified capable of stimulating transcription in a cell-type specific manner. Recently, a 3' tissue-specific enhancer has been found approximately 400 bp 3' to the polyadenylation signal of the chicken adult β-globin gene (Choi and Engel, 1986). Thus the unique ability of enhancers to stimulate transcription even when placed downstream of genes in vitro, has now been realized in vivo.

Various models for enhancer function have been suggested and perhaps the model most often suggested is that enhancers provide a bidirectional entry site for RNA polymerase molecules, or for the other factors associated with the transcription complex (Wasylyck et al., 1983). However, at least for the Kappa immunoglobulin enhancer, this model has been shown to be inapplicable and thus it is possible that different enhancers may have different mechanisms of action (Atchison and Perry, 1986). Other models include: 1) A general domain "opening" of the Kappa locus, caused by the enhancer altering the torsional stress in the DNA (see review by Luchnik, 1985); 2) Direct interactions of the enhancer with other components of the chromatin fibre may direct the linked gene to transcriptionally active regions of the nucleus. This speculation is supported by the presence of a nuclear matrix associated region, adjacent to the mouse Kappa immunoglobulin gene enhancer (Cockerill and Garrard, 1986). Finally, a fourth model involves interactions between proteins bound at enhancers and proteins bound at promoters. An important observation for models of enhancer action is that immunoglobulin producing myeloma cells have been found, in which endogenous heavy chain genes remain active, even though the enhancer has been completely deleted. This suggests that the heavy chain enhancer is not continually required for immunoglobulin transcription (Wabl and Burrows, 1984; Klein et al., 1984). In this

case, the enhancer function is believed to be required only for the establishment of transcriptional activity and not for the maintenance of this transcriptionally active state. In support of this speculation, an enhancer-less heavy chain gene, which is transcriptionally active in its endogenous location, becomes enhancer-dependent when introduced into plasmacytoma cells by transfection (Zaller and Eckhardt, 1985).

In summary, the above observations suggest that enhancers are likely to be a hetergeneous group of regulatory elements and therefore the models outlined here may be relevant to some, but not to others. Among the heterogeneous enhancer elements known, those of the SV40 virus and various retroviral LTR enhancers, are believed to enhance transcription from the human c-Ha-ras1 gene (Chang et al., 1982; Spandidos and Wilkie, 1984). However, no enhancer-like activity for any sequences near the cellular ras genes has yet been adequately demonstrated.

(b) Distinct features of promoters for 'housekeeping" genes

The "housekeeping" genes are generally ubiquitously expressed in all tissues and most, like the classical ras genes, may be expressed at nearly constant levels throughout development and differentiation. Therefore, it is likely that these genes have different mechanisms controlling their expression, than those controlling the highly regulated "luxury-protein" genes. As yet little information is available describing the details of either of these levels of control. It is important to remember that the "luxury-proteins" and "housekeeping" genes represent the extremes of gene control. There is likely to be a full spectrum of regulatory mechanisms between these two extremes, which in some cases may overlap. In general, a number of striking differences between the promoters for these two types of genes have been recently noticed. Of

the "housekeeping" genes categorized so far, all have an exceptionally high G + C content in their 5' flanking sequences. For example, the human adenosine deaminase (ADA) promoter has been shown to be contained within a 135 bp fragment, which is 82% G + C (Valero et al., 1985) and the 5' flanking sequences of the adenine phosphoribosyl transferase (APRT) gene is 66% G + C (Dush et al., 1985). Another particularly noteworthy feature, is the absence of TATA and CAAT boxes from nearly all of these housekeeping genes, an observation first noted for the mouse hypoxanthine phosphoribosyl transferase (HPRT) gene (Melton et al., 1984). Rare exceptions to this general rule are known, such as the superoxide dismutase 1 (SOD-1) gene which is known to have a TATA box (Levanon et al., 1985). The absence of a TATA box may help to explain another property common to many of these genes. That is, the use of multiple transcription initiation sites, as the TATA sequence motif is believed be to involved in determining the specific site of transcriptional intitation (see section II.a.i). The hamster 3-hydroxyl-3-methylglutaryl coenzyme A (HMG CoA) reductase gene has five major clusters of multiple initiation sites, spanning approximately 90 bp (Reynolds et al., 1984). The mouse dihydrofolate reductase (DHFR) gene also has multiple initiation sites, clustered at positions 500-560 bp, 270-310 bp and a major initiation site 115 bp, upstream from the ATG translational initiation codon (McGrogan et al., 1985; Crouse et al., 1985). Interestingly, the mouse DHFR gene has been reported in two conflicting reports, to have a bidirectional promoter. Both reports describe divergent transcription from the promoter region of this gene. However, Farnham et al. (1985) describe the opposite-strand RNAs as small (180-240 nucleotides), abundant in nuclear RNA, non-polyadenylated and with heterogenous 5' ends. This contrasts with the opposite-strand RNA of Crouse et al. (1985) which is at least 14,000 nucleotides long and appears to code for a

protein. There is no good explanation for these conflicting data, but it should be noted that the discrete 3' end of the small nuclear RNA, described by Farnham and colleagues, corresponds to a good concensus donor splice site, suggesting a processing event.

The only conspicuous concensus sequences in the 5' flanking region of most of these "housekeeping" genes, are multiple copies of the previously discussed GC box (see section II.a.iii). These motifs may be arranged symmetrically, as in the human ADA promoter (Valerio et al., 1985) and the mouse APRT promoter (Dush et al., 1985), which have five and four copies of the repeat respectively. They may also be asymmetrically arranged, as are the five repeats of the hamster HMG Co A reductase promoter (Reynolds et al., 1984). As previously discussed, functional binding of the GC box binding factor, Sp1, is independent of the orientation of these repeats and herein may lie a clue to the possible bidirectional transcription from "housekeeping" promoters, such as the mouse DHFR gene. Bidirectional transcription from a promoter region is not without precedent, as the 300 bp SV40 control region containing six GC boxes, is essential for both early and late viral transcription (section II.a.iii). Furthermore, the SV40 late transcription unit uses no obvious TATA box and has multiple transcriptional initiation sites (Fromm and Berg, 1982; Brady et al., 1984).

The G + C rich 5' flanking sequences of "housekeeping" genes have been suggested by Bird (1985) to be related to the previously reported HpaII tiny fragments, HTFs (Cooper et al., 1983). These HTFs represent about 1% of the genomes of a wide range of vertebrates and this fraction is derived from islands of non-methylated, G + C rich DNA, containing HpaII restriction endonuclease recognition sites, at approximately 15 times their frequency in bulk DNA. Cloning of three, non-ribosomal, mouse HTF islands, along with Cot analysis of the entire HTF subfraction of the mouse genome, has shown that HTF islands are mostly low

copy number sequences (Bird et al., 1985). Indeed, the three clones hybridized to discrete transcripts on northern blots of mouse polyadenylated RNA. These results suggest that HTF islands are associated with genes and by using the criteria of HpaII site frequency and lack of methylation, putative HTF islands have been identified around some genes. The mouse HPRT gene has a highly G + C rich 5' flanking region (Melton, et al., 1984), which is hypomethylated and has clustered HpaII sites, but the rest of the gene is extensively methylated (Wolf, et al., 1984; Yen, et al., 1984). The hamster APRT, the mouse DHFR and the chicken α2(I) collagen genes are also hypomethylated (Stein, et al., 1983; McKean, et al., 1982). Furthermore, this pattern of hypomethylation appears to be invariant in all cells examined, including sperm, which agrees well with the constitutive patterns of expression of these genes. In vitro methylation of the hamster APRT gene also supports this pattern, as CpG methylation in the 5' region inhibited expression of this gene, whereas 3' methylation had no effect (Keshet, et al., 1985). Further circumstantial support for the hypothesis that hypomethylated HTF islands may be involved in the expression of "housekeeping" genes, comes from the observation that these sites are nuclease hypersensitive in the human glucose-6-phosphate dehydrogenase (G6PD) and HPRT genes (Wolf and Migeon, 1985).

The hypothesis that the c-Ha-ras1 proto-oncogene performs some sort of "housekeeping" function essential to all dividing cells. (see section I.d), suggests that some aspects of the regulation of this gene could resemble those of other "housekeeping" genes. From the above discussion of "housekeeping-type" promoters, it is noticeable that these promoters have an exceptionally high G + C content, which may be hypomethylated. They generally lack the usual concensus sequence motifs often associated with "luxury-protein" genes, like the TATA and CAAT boxes and also the various tissue-specific motifs known. Furthermore,

they often specify multiple transcriptional initiation sites and usually have multiple copies of the GC box.

The DNA sequence of the human c-Ha-ras1 proto-oncogene has been determined for the 1669 bp upstream from the ATG translational initiation codon (Reddy, 1983; Capon, et al., 1983a). This analysis revealed a number of interesting features (see section I.b and Figure 2) including: 1) the general G + C richness of the 5' flanking sequences, 2) a cluster of TATA-like and CAAT-like sequence motif, 3) a potential acceptor splice site 53 bp upstream from the ATG codon, 4) a potential donor splice site 1093 bp upstream from the ATG codon and 5) a pair of 15 bp direct repeats located adjacent to each of the upstream splice site. However, at the time the work presented in this thesis was begun, there were no RNA mapping data available to support any of these features. Furthermore, the information recently accumulated on the promoter structures of the various "housekeeping" genes was not yet available. The initial results presented here attempted to provide S1 mapping data in support of the strongly proposed promoter region corresponding to the TATA- and CAAT-like sequence motif (Reddy, 1983; Capon, et al., 1983a). However, this analysis provided data in support of an upstream non-coding exon, 5' to the donor splice site 1093 bp upstream from the ATG codon. This upstream region has no TATA or CAAT boxes, is 80% G + C rich and has eleven randomly repeated (in either orientation) hexanucleotide core sequences corresponding to GC boxes. These GC boxes include three which have been characterized as strong binding sites for the Sp1 transcription factor (Kadonaga, et al., 1986). Furthermore, multiple transcriptional initiation sites have been identified in this region, which has been shown to function as a bidirectional promoter. Finally, in this thesis a mutational analysis of the promoter region identified was used to identify sequences important for the regulation of this gene. The effect of a known element (the SV40 enhancer) on c-Ha-ras1 transcription was also examined

and shown to increase c-Ha-*ras*1 transcription by 20 fold. The purpose of this thesis was to investigate the transcriptional regulational of the human c-Ha-*ras*1 proto-oncogene. This investigation allowed some interesting speculations to be made concerning the question of whether transcriptional regulation of this gene, may in certain circumstances, be implicated in carcinogenesis.

MATERIALS AND METHODS

I. GROWTH OF BACTERIA AND GENERAL MANIPULATIONS OF DNA

(a) Growth and maintainance of bacteria

The bacterial strain HB101 (Boyer and Roulland Eussoix, 1969) was grown in LB (Luria-Bertani) medium (1% bactotryptone, 5% bacto-yeast extract, 1% NaCl, pH 7.5) or occasionally in N medium (1% glucose, 1% bactotryptone, 0.1% bacto-yeast extract, 0.8% NaCl, 0.03% $CaCl_2$) and maintained on LB or N medium plates (LB or N medium containing 1.5% Difco minimal agar). Long term storage of HB101, as well as HB101 containing plasmids was by making aliquots of stationary phase culture 20% with respect to glycerol and storing at -20°C.

(b) Preparation of calcium competent cells

A 10 ml overnight cluture of HB101 cells was grown from a single bacterial colony. This 10 ml overnight culture was used to inoculate 1L of LB medium and this culture was then shaken at 37°C until the cell density, as measured by absorbance at A_{650} was 0.6 to 0.8. The cells were chilled on ice for 30 minutes and pelleted at 5000 rpm for 5 minutes at 4°C. The supernatant was decanted, the bacterial pellet resuspended in 1L of ice cold 10mM $MgSO_4$ and incubated on ice for 30 minutes. The cells were pelleted as before, resuspended in 500 ml of ice cold $CaCl_2$ and again incubated on ice for 30 minutes. The supernatant was decanted and the cells resuspended in 50 mls of ice cold 50mM $CaCl_2$ and 20% glycerol. The cell suspension was distributed as 1 ml aliquots in Eppendorf tubes, snap-frozen in liquid nitrogen and stored in liquid nitrogen or in a -135°C freezer.

(c) Transformation of calcium competent cells

A tube of competent HB101 cells was removed from storage in liquid nitrogen and thawed on ice. A 10 ul aliquot of the 20 ul ligation reaction (see earlier) was removed and placed on ice, to which was added, 40 ul of ice cold TNE buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1mM EDTA), 20 ul of ice cold 10X TCM buffer (100 mM Tris pH 7.0, 100mM $MgCl_2$, 100 mM $CaCl_2$) and 125 ul of competent HB101 cells. This suspension was incubated on ice for 20 minutes, heat shocked at 42°C for 1.5 minutes and allowed to sit at room temperature for 10 minutes. One ml of LB or N medium was added to each tube and the tubes were shaken at 37°C for 60 minutes to allow the antibiotic resistance gene encoded by the plasmids to be expressed. The tubes were briefly (15 seconds) spun in a microcentrifuge and the cells resuspended in 200 ul or 400 ul of LB or N medium and 200 ul aliquots were plated onto LB or N medium plates containing 100 ug/ml ampicillin (Sigma). The plates were incubated at 37°C overnight.

(d) Rapid screening of bacterial colonies (Birnboim and Doly, 1979)

Bacterial colonies were picked from plates and used to inoculate 1.5 ml of LB medium in Eppendorf tubes, which were then left shaking at 37°C overnight. Cells were spun down (15 seconds in a microcentrifuge) and resuspended in 100 ul of ice cold lysis solution (25 mM Tris pH 8.0, 50 mM glucose, 10 mM EDTA, 2mg/ml lysozyme). The resuspended cells were incubated on ice for 15 minutes, 200 ul alkaline- SDS (0.2 M NaOH, 1% sodium dodecylsuphate) was added and the suspension incubated on ice for a further 5 minutes. Finally, 150 ul of ice cold High Salt solution (3M sodium acetate pH 4.8) was added, the contents of the tube were gently mixed by inversion and incubated on ice for a

further 15 minutes. After being spun in a microcentrifuge for 3 minutes the supernatants were removed into fresh tubes and 0.6 volumes of ice cold propan-2-ol (Fisher) were added to precipitate the plasmid DNA. The precipitated plasmid DNA was washed with 80% ethanol, dried, redissolved in 50 ul of 1X T.E. buffer plus 50 ug/ml RNase A (T.E. buffer is 10 mM Tris pH 8.0, 1 mM EDTA) and incubated at 70°C for one hour. This incubation helps inactivate bacterial DNases which may contaminate the plasmid DNA isolated in this way and the RNase A digests contaminating RNA. Diagnostic restriction endonuclease digestions can now be performed on 10 ul aliquots of the crudely purified plasmid DNAs.

(e) Growth and isolation of plasmid DNA

A 5 ml aliquot of an overnight culture of bacteria, containing the plasmid of interest, was inoculated into 500 ml of fresh LB medium containing 100 ug/ml ampicillin (Sigma). The culture was grown at 37°C for 6 - 8 hours when 85 mg of chloramphenicol (Sigma) was added to give a final concentration of 170 ug/ml. The culture was then shaken at 37°C overnight. For some plasmids the chloramphenicol amplification step was not necessary so cultures containing these plasmids were simply allowed to grow overnight in LB medium containing ampicillin.

The isolation of plasmid from overnight cultures was essentially a scaled up version of the method described for the rapid screening of bacterial colonies (Birnboim and Daly, 1979). Cells were harvested by centrifugation at 6000 rpm at 4°C for 10 minutes in 500 ml bottles, washed by resusending in 50 ml of ice cold 50mM Tris pH 8.0 and pelleted as before. The cells were then resuspended in 50 ml of freshly prepared lysis solution (25 mM Tris(pH 8.0), 50 mM glucose, 10 mM EDTA, 5 ug/ml lysozyme) and incubated on ice for 30 minutes. Then 80 mls of fresh alkaline-SDS (see earlier) was added, mixed gently and incubated on ice for 5

minutes. Lastly, 40 ml of ice cold High Salt solution (3M potassium acetate, pH 4.8) was added, the mixture was gently mixed and incubated on ice for 15 minutes, before centrifugation at 8000 rpm for 5 minutes at room temperature. The supernatant was decanted into fresh 250 ml bottles and 0.6 volumes of propan-2-ol was added and mixed and spun immediately at 8000 rpm for 5 minutes at room temperature. The pellet was allowed to drain dry, resuspended in 5 ml of T.E. buffer (10 mM Tris pH 8.0, 1mM EDTA) and loaded onto cesium chloride (BRL)-ethidium bromide equilibrium gradients. Cesium chloride-ethidium bromide equilibrium gradients were 0.95 g/ml cesium chloride, 0.5 mg/ml ethidium bromide and made up to 10 ml with the crudely purified plasmid DNA and T.E. buffer. These gradients were ultracentrifuged at 40,000 rpm for 36 hours at 21°C in a Beckmann Type-65 rotor or Type -70 rotor. The banded plasmid DNA was extracted, the ethidium bromide removed with propan-2-ol and the DNA dialysed against dH_2O at 4°C overnight.

(f) Agarose gel electrophoresis

Agarose gels were generally made up to 1%, 1.5% or 2% agarose in 1X TEA (40 mM Tris pH 8.0, 1 mM EDTA, 4 mM sodium acetate) or in 1X TBE (89 mM Tris-borate, 89 mM boric acid, 2mM EDTA, pH 8.0) by boiling. Ethidium bromide was added to a final concentration of 0.5 ug/ml and the gel was poured into an appropriately sized horizontal gel apparatus and allowed to solidify with an appropriate comb in place. Running buffer (1X TEA or 1X TBE containing 0.5 ug/ml ethidium bromide) was added and the samples were loaded into wells. Gels were generally run at 25 V overnight or at 100-150 V for 1 to 6 hours, and were photographed immediately under ultra violet (u.v.) light. Occasionally low melting point agarose gels were used for DNA fragment isolation (see below). These gels were prepared in a similar fashion using BRL low melting point

agarose and were always poured and electrophoresed in the cold room ($4^{o}C$).

(g) Polyacrylamide gel electrophoresis

Polyacrylamide gels (5 - 12%) were made up from a 50% stock solution (crosslinked 29:1, acrylamide: bis-acrylamide). TBE was added to a final concentration of 1X and dH_2O added to give the final volume of 50 mls. Ammonium persulphate was added to 0.03% and N,N,N,N-tetramethylethlenediamine (temed) was added to 0.08% to catalyse the polymerization reaction. The solution was poured immediately between taped, siliconized glass plates to give a gel, upon polymerization, with dimensions 150 mm X 180 mm X 1 mm. The running buffer was 1X TBE and the gel was run at 150 Volts for 3-6 hours or at 25-50 Volts overnight. Following electrophoresis these gels were stained in 0.5 ug/ml ethidium bromide and photographed under u.v. light.

(h) Denaturing polyacrylamide gel electrophoresis

To visualize single stranded nucleic acid fragments, an acrylamide gel containing urea, which is a strong denaturant, was used. These gels were used to visualize S1 nuclease protection products, primer extension products and Maxam Gilbert sequencing products (see analysis of DNA and RNA). Depending on the size of the fragments to be visualized these gels were from 6 - 10% acrylamide (crosslinked 19:1 acrylamide: bisacrylamide), 7M urea (Pharmacia) and 1X TBE. Ammonium persulphate was added to 0.03% and N,N,N,N-tetramethylethylenediamine (temed) was added to 0.08%. Gel dimensions were 180mm X 380mm X 0.5mm. These gels were run at a constant 45 Watts for 1.5 - 2.0 hours in 1X TBE with a precisely fitting aluminium plate clipped onto the larger glass plate to prevent uneven migration of bands. It is important to prerun these gels

for at least one half hour prior to loading the sample, so that, the gel can heat up to its running temperature. Following electrophoresis the gels were autoradiographed at -70°C with an intensifying screen (Cronex) overnight or longer until the bands were visualized.

(i) Isolation of DNA fragments from low melting agarose gels (Smith, 1980)

After running the gel (see earlier section) the desired band was cut out under u.v. illumination and a 4 - 5 fold excess of sterile dH_2O was added. The agarose was melted at 70°C for 10 minutes and placed at 37°C for 5 minutes to lower its temperature. One volume of cold phenol was added, the mixture vortexed immediately for one minute and cooled on ice for 5 minutes. The phases were separated by centrifugation in an Eppendorf microcentrifuge for 2 minutes. The aqueous phase was removed into a fresh tube and the phenol phase was re-extracted with a small volume of dH_2O. The pooled aqueous phases were phenol extracted once more to eliminate any remaining agarose. Following centrifugation, the aqueous phase was removed and either butan-2-ol extracted to reduce its volume and ethanol precipitated, or ethanol precipitated directly.

(j) Isolation of DNA fragments from acrylamide gels

This procedure routinely gave yields of at least 90%, therefore it was the method generally employed for isolating both radioactively labelled and non-radioactively labelled DNA fragments. Following autoradiography, or ethidium bromide staining and u.v. visulization, the appropriate band was sliced from the gel and placed into a heat sealed 1ml pipette tip that contained a small amount of siliconized glass wool packed into the tip. The acrylamide was finely minced and Elution Buffer (0.2M sodium acetate pH 5.3, 0.1% SDS) was added to cover the

acrylamide. The top of the pipette tip was sealed and this mixture was incubated at 37°C overnight. The liquid was drained into an Eppendorf tube and the acrylamide left behind was rinsed twice with a small amount of elution buffer. The collected solution was then ethanol precipitated on dry ice. Following precipitation, the pellet was washed carefully twice with 80% ethanol to remove any SDS and the nucleic acid was resuspended in sterile dH_2O.

(k) Dephosphorylation of DNA

The 5' terminal phosphates were removed from linear DNA molecules prior to 5' end labelling with Kinase (see later) and from linearized vectors to reduce background due to self ligation while cloning. Two different methods were employed. For 5' end-labelling, bacterial alkaline phosphatase (BAP, Worthington) was used to remove the terminal 5' phosphates, and in order to prevent vectors from self-ligating during cloning, calf intestinal alkaline phosphatase (CIAP, Boehringer) was used. For the BAP reaction, 1 to 20 picomol of 5' ends were brought to a volume of 50 ul with 50mM Tris pH 8.8, 0.5 units Worthington BAP was added (this BAP was prepared and its activity assayed as directed by the manufacturer) and the reaction incubated at 60°C for 30 minutes. The reaction was stopped with 0.5 ul of 0.1M EDTA and extracted twice with phenol chloroform, once with ether, and then ethanol precipitated. The CIAP reaction was simply carried out by adding 1 ul of CIAP (Boehringer) to the restriction digest after ensuring that digestion was complete, and further incubating at 37°C for 1 hour. The reaction was stopped with 10 ul of 0.1M EDTA and all the protein was removed as for the BAP reaction.

(l) Creating blunt ended fragments

Overhanging 5' ends were filled in, using the DNA polymerizing activity of the Klenow fragment of E. Coli DNA polymerase I, exactly as described in the Molecular Cloning Manual (Maniatis et al., 1982). However, after the end filling reaction, the fragments were extracted with phenol/chloroform and then precipitated with ethanol, prior to ligation.

Overhanging 3' ends were made blunt-ended by using the 3' exonuclease activity of bacteriophage T4 DNA polymerase. Again, this procedure is descibed in detail in the Molecular Cloning Manual (Maniatis et al., 1982).

(m) Ligations

Ligations were performed in 20 ul reaction volumes, as the concentration of ends and the concentration of ligase are the critical factors. Optimum efficiencies were obtained when 1 - 10 ng of vector DNA was ligated to a 3 - 5 fold molar excess of insert. Appropriate dilutions of vector and insert DNA were made, to which 4 ul of 5X Ligation Salts (150 mM Tris pH 7.5 at $4^{o}C$, 150 mM NaCl, 38 mM $MgCl_2$), 2 ul of 10X Ligation Mix (2.5 mM ATP (P-L Biochemicals), 20 mM dithiothreitol (Sigma), 10mM spermidine (Sigma), 2mM EDTA, 1mg/ml BSA) and 2 units of T4 DNA ligase (Boehringer), were added and the volume made up to 20 ul with dH_2O. The reactions were incubated at room temperature for 1-2 hours (for ligation of overhaning ends) or at $4^{o}C$ overnight followed by 1 hour at room temperature (blunt end ligations).

Linker ligations were done by resuspending the ethanol precipitated blunt ended fragments (see earlier) in 9 ul of dH_2O and adding 4 ul of 5X Ligation Salts, 2 ul of 5X Ligation Mix, 4 ul of phosphorylated linkers and 2 units of T4 DNA ligase. In this study two different octameric linkers were used; BamH1 (d(pC-G-G-A-T-C-C-G)) NEB lot 6-59; one A_{260} unit/50 ul) and HindIII (d(pC-G-G-A-T-C-C-G)) NEB lot 8-100; one A_{260} unit/50ul).

The linker ligation reactions were left incubating at 4°C overnight and for 1 hour at room temperature the next day. An aliquot of 49 ul dH_2O ws then added, 10 ul of restriction buffer (10X BamH1 buffer or 10X HindIII buffer), 10 ul of 1mg/ml BSA (BRL) and 10 ul of restriction endonuclease (25u/ul BamH1 (NEB) or 20u/ul HindII (NEB)). The vast excess of restriction endonuclease was required because of the molar excess of linkers. The digestion was left all day and overnight at 37°C to ensure complete digestion, extracted with phenol/chloroform, precipitated with ethanol and run on an 8% polyacrylamide gel (cross-linked 29:1, acrylamide:bis-acrylamide) to separate the digested linkers from the linkered fragment.

II. GROWTH AND TRANSFECTION OF HUMAN CELLS IN CULTURE

(a) Growth and maintainence of human cells

The cell lines used in this study are, EJ human bladder carcinoma cells, which are now known to be identical to the T24 human bladder carcinoma cell line (O'Toole et al., 1983), A431 human epidermal cells, derived from a human vulval carcinoma, HeLa human fibroblasts cells and Cos7 monkey Kidney cells. All of these cells were obtained from the Beatson Institute for Cancer Research, Glasgow. A431, HeLa and Cos7 cells were grown in Special Liquid Medium (SLM) containing penicillin, supplied by Gibco, supplemented with 10 - 15% fetal calf serum (Gibco) and 200 mM L-glutamine (Gibco) to give a final concentration of 2 mM. EJ cells were grown in SLM plus L-glutamine but only 5% fetal calf serum. HeLa cells were grown in Dulbecco's Modified Eagles (DME) medium supplemented with 10% fetal calf serum. All cells were grown at 37°C in chambers containing 5% CO_2 or pre-gassed with an air mixture containing 5% CO_2 and grown in tightly sealed flasks (Corning). As these cell lines are all

adherent, confluent monolayers were passaged by aspirating away the medium, washing the cells with PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.0) and adding 1 -2 mls of trypsin- EDTA (0.25% trypsin (Gibco), 0.1% EDTA in PBS). The trypsin -EDTA was left in contact with the cells until the cells rounded up and dissociated from the plate. 5 - 10 mls of fresh medium was added to each disk or flask, the cells were pipetted gently several times to break up clumps of cells and then added to fresh dishes or flasks in the ratio of 1ml cells suspension to 10 ml of fresh medium. To protect all cell strains against genetic divergence and microbiological contaminations, cell cultures growning in log phase were harvested and frozen slowly at $1^{o}C$ per minute in the presence of 10% glycerol as a protective agent in 1 ml ampoules, each containing 2.5 - 10 X 10^6 cells. The ampoules were then permanently stored in liquid nitrogen until required, when a single ampoule was rapidly thawed by immersion in tepid water and the cells injected directly into culture medium for further propagation. In general any cells grown continuously in culture for long periods of time were replaced every 3 months with a fresh seed taken from the liquid nitrogen freezer.

(b) Transfection of cultured cells by calcium phosphate precipitation (Wigler et al., 1979)

Twenty-four hours prior to transfection, cells were passaged to give on the day of transfection either, 50% confluent cultures for HeLa cells or 70% confluent cultures for Cos7 cells. Usually two 80 cm^2 dishes or flasks were transfected for each plasmid of DNA of interest. Calcium phosphate precipitates were formed for each 80 cm^2 dish or flask by bringing 40 ug of the plasmid DNA of interest up to a volume of 900 ul with sterile 0.1X TE buffer (1mM Tris at pH 8.0, 0.1mM EDTA), adding 100 ul of 2.5 M $CaCl_2$ (Mallinkrodt, filter sterilized) and adding all of the

above to a separate tube, while vortexing, containing 1 ml of 2X Hepes Buffered Saline (HBS) (2X HBS is 50 mM Hepes, 280 mM NaCl, 1.5 mM Na_2HPO_4 accurately brought to pH 7.1 at room temperature). The DNA - $CaPO_4$ precipitate was allowed to form by letting the vigorously vortexed mixture sit at room temperature for exactly 30 minutes, after which all of the above was added to one 80 cm^2 flask or dish and left for 8-16 hours at 37°C. Following this treatment, the medium containing the finely grained DNA -$CaPO_4$ precipitate was aspirated off and fresh medium added to allow cells to recover. The transfected cells were harvested 48 hours following addition of the frsh medium.

(c) Isolation of total RNA from cultured cells (Chirgwin et al., 1979)

Both transfected and non-transfected cells were washed with PBS and trypsinized as above, followed by centrifugation (1000 rpm at room temperature) and further washing of the cell pellet with PBS. The volume of the cell pellet was estimated and for an approximately 100 ul cell pellet, 800 ul of Lysis buffer (5M guanidineisothiocyanate (Fluka), 50 mM Tris pH 7.5, 10 mM EDTA) and 50 ul of neat 2- mercaptoethanol (Sigma) were added to lyse the cells. The cell suspension was vortexed vigorously and passed through a fine hypodermic needle which sheared genomic DNA and hence reduced the viscosity. The lysed cell suspension was then left at room temperature for 20 minutes and then 5 ml of filter sterilized 4M LiCl (Sigma) was added. After mixing, the lysed cell suspension was left at 4°C for at least 20 hours to specifically precipitate the RNA, followed by centrifugation at 6500 rpm for 1.5 hours at 4°C. The supernatant was removed and discarded, the RNA pellet washed by resuspension in 3 ml of 3M LiCl, centrifuged at 6500 rpm for 30 minutes at 4°C and the RNA pellet dissolved in 200 ul of TE buffer. The dissolved RNA was extracted twice with phenol/chloroform, first with 0.5

volumes of phenol/chloroform and second with 1 volume of phenol/chloroform, then precipitated with ethanol and redissolved in 30 - 40 ul of sterile dH_2O. An aliquot of 2 ul was removed and used to determine the yield of RNA by u.v. spectrophotometry.

This procedure was also used in a scaled up version to isolate total RNA from large cultures (up to twenty 175 cm^2 flasks) of EJ and A431 cells.

(d) Purification of polyadenylated RNA

Polyadenylated RNA was isolated from the total RNA prepared as above by oligo-dT selection (Favaloro et al., 1980). Total RNA was brought to 1 ug/ul. To this, one volume of 2X dT Binding buffer (20 mM Tris pH 7.5, 1M NaCl, 0.1 % (w/v) Sarcosyl (Sigma)) was added to bring the mixed solution to 1X dT Binding buffer (10 mM Tris pH 7.5, 500 mM NaCl, 0.1% (w/v) sarcosyl). This solution was then applied to an oligo-dT cellulose (BRL) column prepared in 5 ml hypodermic syringe by blocking off the opening to the syringe with a disc of Whatman 2.5 cm GF/C glass microfibre filter, adding oligo-dT celluolse up to the 0.5 ml mark on the syringe and washing the column carefully with 1X dT Binding buffer. Up to 10 mg of total RNA can be processed per 1 ml of oligo-dT cellulose. After the prepared RNA sample was added to the column, the initial flow through was reapplied to the column to ensure complete binding of the polyadenylated RNA. The column was then washed 3 times with 2 ml of 1X dT Binding buffer, per ml of column volume and washed 3 times more with 2 ml of dT Midwash buffer (10 mM Tris pH 7.5, 100 mM NaCl, 0.1% sarcosyl), per ml of column volume. Finally, the bound polyadenylated RNA was eluted off the column into four separate fractions by washing the column 4 times with 2 ml of dT Elution buffer (5 mM Tris pH 7.5, 1 mM EDTA, 0.1% (w/v) sodium dodecylsulphate (Sigma)), per ml of column volume. Each fraction was assayed for RNA content by u.v.

spectrophotometry.

III. Analysis of RNA

(a) Preparation of end-labelled single-stranded DNA fragments

5' end-label DNA restriction fragments were first dephosphorylated, by treatment with bacterial alkaline phosphatase (see previous section), then precipitated with ethanol and redissolved in 2.5 ul of 10 mM Tris pH 8.0, 17.7 ul of Kinase Buffer 1 (20 mM Tris 9.5, 1 mM spermidine (Sigma), 0.1 mM EDTA). The mixture was incubated at 90°C for 2 minutes, followed by quenching on ice. This incubation facilitates the Kinase reaction by opening up the ends of the molecule. Next, 2.5 ul of Kinase buffer 2 (500 mM Tris pH 9.5, 100 mM $MgCl_2$, 50 mM dithiothreitol (DTT, Sigma), 50% (v/v) glycerol), and 20 units of polynucleotide Kinase (P-L Biochemicals) were added. The mixture was then transferred to another tube containing 125 uCi of γ-^{32}P-ATP (> 5000 Ci/mmol, Amersham) which had been freeze dried to a volume of approximately 2.5 ul. The reaction was incubated at 37°C for 30 minutes, stopped with 0.1 M EDTA and co-precipitated with 40 ug of yeast tRNA in preparation for strand separation (see later).

Fragments to be 3' end labelled were designed so that they had recessed 3' ends which could therefore be filled in using the enzyme T4 DNA polymerase I (BRL). A disadvantage of this technique is that DNA fragments with blunt-ends or 3' overhanging ends cannot easily be labelled in this way. The amount of DNA equivalent to 5 to 20 picomol of 3' ends was calculated and after phenol/chloroform extraction, the dry ethanol precipitated pellet was resuspended in 3 ul of dH_2O. To this, 15 ul of 2X T4 buffer (130 mM Tris pH 8.0, 13 mM $MgCl_2$, 10 mM dithiothreitol (Sigma)), 1 ul of a 5 mM solution of each of

dATP, dGTP and dTTP (obtained as 100 mM solutions from P-L Biochemicals), 8 ul --^{32}P-dCTP (> 7000 Ci/mmol, Amersham) and 10 units of T4 DNA polymerase I was added. The reaction was incubated at 11^{o}C for 3 hours, stopped with 10 ul of 0.1 M EDTA and prepared for strand separation as for 5' end labelled probes. It is important to note that the choice of labelled deoxyribonucleotide triphosphate used depends on the presence in the 5' overhang of its complementary deoxyribonucleotide triphosphate.

Both 5' and 3' end-labelled DNA fragments were strand separated as follows. Strand separation gels, of the appropriate concentration were prepared from a 50% acrylamide stock solution (crosslinked 29:1, acrylamide: bisacrylamide), using 0.6X TBE running buffer. The gel dimensions were 180 mm X 380 mm X 1 mm and these gels were run at 100 - 300 Volts in 1X TBE running buffer, for at least 15 hours. The samples were prepared for strand separation by co-precipitating 40 ug of 30% DMSO mix (30% dimethylsulphoxide (Sigma), 1X TBE, 0.1% bromophenol blue, 0.1% xylene cyanol) or 80% formamide mix (80% formamide (Fluka), 1X TBE, 0.1% bromophenol blue, 0.1% xylene cyanol). When fully resuspended, the samples were placed into a boiling water bath for 4 - 5 minutes, quenched immediately on ice and loaded onto the gel. Usually the labelled fragments to be strand separated in this way were of such high specific activity that autoradiography for 2 minutes at room temperature with an intensifying screen (Cronex) was sufficient to allow detection of the bands.

(b) S1 nuclease analysis

End-labelled strand separated DNA probes were used for S1 nuclease analysis performed by the method of Berk and Sharp (1977) as modified by Weaver and Weissman (1979). The appropriate RNA sample was freeze dried with the end labelled DNA probe, together with sufficient yeast tRNA (P-L Biochemicals) to give a specified total of nucleic

acid, usually 20 ug. The freeze dried nucleic acid was resuspended in 10 ul of Formamide Hybridization Buffer (FHB is: 80% formamide (Fluka), 400 mM NaCl, 40 mM Pipes pH 6.4 (Sigma) 1mm EDTA) pipetted into a siliconized glass capillary tube and the ends heat sealed. The capillary tubes were placed into glass test tubes containing dH_2O, immersed in a 90°C water bath for 10 minutes to denature the nucleic acids and immediately transferred to another water bath, set at the appropriate hybridization temperature, and incubated overnight. Whenever a new probe was used, control hybridizations were done to determine empirically, the optimum hybrization temperature and also to ensure that the probe excess was reached.

Overnight hybridizations were stopped by quenching the capillaries in -20°C ethanol and flushing out the capillaries with ice cold S1 nuclease assay buffer (250 mM NaCl, 30 mM sodium acetate pH 4.6, 1 mM $ZnSO_4$ (Sigma), 200 ug/ml calf thymus DNA (Sigma)). 300 - 1000 units of S1 nuclease (Boehringer) were added and the reaction incubated at 37°C for 90 minutes unless S1 nuclease digestion temperature controls were required when the temperature was varied between 21°C and 45°C. After S1 nuclease digestion, the reactions were phenol/chloroform extracted once, ethanol precipitated in the presence of 10 ug of yeast tRNA, resuspended in formamide loading buffer (80% formamide, 1X TBE, 0.1% bromophenol blue, 0.1% xylene cyanol), denatured for 2 minutes at 100°C, chilled on ice and the reaction products analysed on denaturing polyacrylamide gels, (see earlier) followed by autoradigraphy.

(c) Exonuclease VII analysis

Exonuclease VII analysis (Berk and Sharp, 1978) can be used to independently confirm data already obtained with S1 nuclease. Exonuclease VII is a processive exonuclease that, unlike S1 nuclease, releases small oligonucleotides

from the 3' and 5' ends of single stranded DNA (Chase and Richardson, 1964). Analyses using this enzyme were performed almost identically to S1 nuclease analysis (Proudfoot, personal communication). However, the capillaries were flushed out with ice cold Exonuclease VII buffer (10 mM Tris pH 7.8, 13 mM KCl, 10 mM EDTA), 4 units of exonuclease VII (BRL) were added to the reaction mixture and the reaction incubated at 37°C for 2 hours. The rest of the procedure was as above for S1 nuclease analysis.

(d) Primer extension analysis

Primer extension analysis was performed with short, 5' end-labelled, strand-separated DNA probes. These probes were prepared and hybridized to RNAs, as has previously been described for S1 nuclease and exonuclease VII analysis (see above). The hybridizations were performed at 45-52°C and the hybridized nucleic acids were flushed from their capillaries with 200 ul of ice cold 0.2M sodium acetate pH 4.8, phenol extracted once, ethanol precipitated and resuspended in 29 ul of sterile dH_2O, which had been treated with diethyl pyrocarbonate (Sigma). To the resuspended nucleic acid, 10 ul of 5X Reverse Transcriptase buffer (250 mM Tris pH 8.2, 250 mM KCl, 10 mM $MgCl_2$, 5 mM dithiothreitol), 10 ul of a solution containing all four deoxynucleotide triphosphates, each at 5 mM (Pharmacia), and 1 ul AMV reverse transcriptase at 25 units/ul (Boehringer), was added. The reaction was incubated at 42°C for 3 hours, with the addition of a further 10 units of AMV reverse transcriptase and 2 ul of the 5 mM deoxyribonucleotide triphosphate mix, after 1 hour and 2 hours of incubation. The reaction was stopped with 5 ul of 0.1 M EDTA and the RNA was alkali hydrolyzed by the addition of 10 ul of 1 M NaOH and incubated at 42°C for 1 hour. Phenol red indicator was added and the reaction titrated against 1M HCl until the solution was neutralized.

The neutralized reaction was phenol/chloroform extracted twice, ethanol precipitated and the reaction products analysed on denaturing polyacrylamide gels, followed by autoradiography (see earlier).

(e) Northern blot analysis

RNA was electrophoresed on 1% denaturing agarose gels as follows. 2g of agarose (BRL) was autoclaved in 147 mls of dH_2O for 20 minutes. After cooling to approximately 50^oC, 20 mls of 10X MOPS buffer (0.2M morpholinopropanesulphonic acid (Sigma), 50 mM sodium acetate (Sigma), 5 mM EDTA, pH 7.0) and 33 mls formaldehyde (Fluka) were added and the molten gel was poured onto appropriate gel moulds. The freeze dried RNA samples (usually 20 ug total or poly A^- RNA, or 5 to 10 ug poly A^+ RNA) were resuspended in 25 ul of freshly prepared sample buffer (50% formamide (Fluka), 2.2M formaldehyde, 1X MOPS buffer), denatured at 65^oC for 5 minutes, chilled on ice and electrophoresed at 30 Volts overnight or 100 to 140 Volts for 3 to 4 hours with buffer circulation. After photography, the gel was equilibrated to 20X SSC (3M NaCl, 0.3 M sodium citrate) by soaking in 20X SSC for 15 minutes. Excess gel was trimmed away and the dimensions of the gel measured. One piece of nitrocellulose (Sartorius), four pieces of filter paper (Whatman 3mm) and a 10-15 cm stack of paper towels were cut to the size of the gel. The nitrocellulose filter was briefly wet in dH_2O followed by soaking for 30 minutes in 20X SSC. Meanwhile, a platform was fitted into a rectangular pyrex dish containing 500 ml of 20X SSC. Two appropriately sized pieces of wet filter paper were placed over the platform so that the ends formed wicks. The gel was placed on top and all air bubbles removed before the pre-wet nitrocellulose filter was placed on top of the gel, again making sure all air bubbles were removed. Next the four pieces of Whatman 3mm filter paper were placed on top of the nitrocellulose filter, the first

two were pre-wet in 20X SSC, and finally the stack of paper towels was placed on top along with a small weight. To ensure that blotting only occured through the gel, saran wrap was placed alongside the edges of the gel, this also prevented excessive evaporation. Blotting was allowed to continue at least overnight. After blotting, the nitrocellulose filter was removed, rinsed briefly in 5X SSC, placed between two sheets of Whatman 3mm filter paper and baked at 80°C for 2 hours.

The nitrocellulose filter was placed in a polystyrene bag and heat sealed as close to the filter as possible, carefully removing air spaces. Prehybridization solution (50% formamide, 5X SSC, 5X Denhardts, 50 mM Na_2HPO_4/NaH_2PO_4 pH 6.5, 0.1% SDS, 500 ug/ml salmon sperm DNA) was added to a final volume of at least 200 ul/cm^2 per filter (note, 100X Denhardts is 2% ficoll (Sigma), 2% polyvinyl pyrrolidone-40 (Sigma), 2% w/v Bovine Serum Albumin (Sigma)). The bag was carefully sealed to remove all air bubbles and the filter was incubated at 42°C for 6-16 hours with gentle shaking. The prehybridization solution was removed and hybridization solution (50% formamide, 5XSSC, 1X Denhardts, 20 mM Na_2HPO_4/NaH_2PO_4 pH 6.5, 0.1% SDS, 10% dextran sulphate (Sigma), 100ug/ml salmon sperm DNA, which contained labelled probe at a concentration of not more than 10 ug/ml was added to give a volume of 100 ul/cm^2 per filter. The bag was sealed as before and the filter incubated at 42°C for 16-48 hours with gentle shaking.

Filters were washed in solutions of increasing stringency to remove any nonspecific hybridization. The first wash was with 500 ml of 2X SSC, 0.1% SDS at room temperature for 20 minutes. All other washes were performed at 65°C for 20 minutes and consisted of a 2X SSC; 0.1% SDS wash; two 0.5X SSC; 0.1% SDS washes, and one or two 0.1X SSC; 0.1% SDS washes. To remove traces of SDS the filter was washed in 0.1X SSC at room temperature for 10 minutes. Following the washes the filter was briefly blot dried and autoradiographed at -70°C with an intensifying screen

(DuPont).

(f) Nick translation of DNA

Nick translations (Rigby et al., 1977) were performed exactly as directed by the Amersham Nick Translation Kit, using 50 - 100 ul of α-^{32}P-dCTP (> 3000 Ci/mmol). Separation of incorporated label from unincorporated label was by chromatography through G-50 sephadex (Pharmacia) columns, equilibrated with 0.3M NaCl, 0.1% SDS, 20 mM Tris pH 7.5 and 1 mM EDTA. Fractions collected off the first radioactive peak were pooled and used directly.

(g) Oligo labelling (or Second Strand Synthesis) of DNA

A recent technique has been developed (Feinberg and Vogelstein, 1983) for radiolabelling DNA restriction endonuclease fragments to high specific activity. 50ng of DNA resticition fragment was brought to a volume of 32 ul with dH_2O and denatured in boiling water for 10 minutes. Without quenching on ice, 10 ul of OLB buffer, 2 ul of 10 mg/ml BSA (BRL), 5 ul α-^{32}P-dCTP (> 3000 Ci/mmol, Amersham) and 0.5 ul of Klenow fragment (Boehringer Mannheim) were added. The reaction was mixed well and incubated at room temperature for 2.5 hours (OLB buffer is 248 mM Tris pH 7.4, 24mM $MgCl_2$, 56 mM 2-mercaptoethanol (BDH), 0.12mM in each of dATP, dGTP, dTTP (Pharmacia), 1M hepes pH 6.6 (Sigma), 30 OD units/ml calf thymus hexadeoxyribonucleotides (Pharmacia)). To further increase the specific activity of certain probes, α-^{32}P-dGTP (> 3000 Ci/mol) was used as well as α-^{32}P-dCTP. In this case OLB buffer without the dGTP component was used. The reaction was stopped by the addition of 100 ul of 0.5 M EDTA and since the incorporation of labelled nucleotide triphosphate is usually greater than 50% it was usually not necessary to purify the precursor nucleotide triphosphate from the labelled product prior to hybridization. In this case the

specific activity was determined by spotting a 1 ul sample onto a DEAE cellulose filter (Whatman DE81) and measuring Cherenkof radiation. The filter was then washed with a phosphate buffer such as 20X GeneScreen buffer (0.5M Na_2HPO_4/NaH_2PO_4, pH 6.5), dried and Cherenkof radiation remeasured. The latter reading gives the incorporated counts per minute and can be subtracted from the previous reading to determine the percentage incorporation.

A modification of this technique was also used, in which DNA restiction fragments were labelled without purification from low melting point agarose (Feinberg and Vogelstein, 1983a).

IV. ANALYSIS OF DNA

(a) Southern blotting and hybridization

Cellular DNA was analysed by the technique of Southern blot analysis (Southern, 1977). The DNA was digested to completion with restriction endonucleases and run on 1% agarose gels at 40 Volts overnight in a 1X TEA buffer, for a 20 cm long gel. After staining in 0.5 ug/ml of ethidium bromide, the gel was photographed and the DNA transferred to GeneScreen hybridization transfer membranes (NEN) by the capillary blot procedure, exactly as recommended by the manufacturer. Hybridization to the DNA was performed using the dextran sulphate plus formamide method (i.e. GeneScreen hybridization method III).

(b) DNA sequencing

Single stranded end-labelled DNA molecules were sequenced using a modification of the technique developed by Maxam and Gilbert (1980). Freeze dried end-labelled DNA was resuspended in 11 ul of dH_2O and 2.5 ul aliquots were dispensed into four siliconized Eppendorf tubes. The

guanosine (G) specific cleavage reaction was performed as follows. To the aliquot of end-labelled DNA, 98 ul of DMS buffer (50 mM sodium cacodylate (Sigma), 1 mM EDTA , pH 8.0) was added and the mixture was placed on ice. Upon addition of 0.5 ul of dimethylsulphate (DMS, Sigma) the mixture was incubated at 20°C for four minutes, after which 24 ul of DMS "stop buffer" (1.5 M sodium acetate, 1.0 M mercaptoethanol, 100 ug/ml yeast tRNA (P-L Biochemicals)) and 40 ul of absolute ethanol were added. The mixture was then stored at -70°C for 15 minutes. The adenosine and guanosine (A + G) specific cleavage reaction was performed as follows. To the aliquot of end-labelled DNA, 11 ul of dH_2O was added and the mixture was placed on ice. Upon addition of 2.5 ul of piperidine formate (4% v/v formic acid, adjusted to pH 2.0 with piperidine (Fisher)) the mixture was incubated at 30°C for 70 minutes. The mixture was then freeze-dried, resuspended in 10 ul of dH_2O and freeze-dried once more. The cytosine (C) specific cleavage reaction was performed as follows. To the aliquot of end-labelled DNA, 5 ul of 5 M sodium chloride was added and the mixture placed on ice. Upon addition of 15 ul of hydrazine (Kodak), the mixture was incubated at 20°C for 8 minutes. Meanwhile, the thymine and cytosine (T + C) specific cleavage reaction was set up as follows. To the aliquot of end-labelled DNA, 6 ul of dH_2O was added and the mixture placed on ice. Upon addition of 15 ul of hydrazine, the mixture was incubated at 20°C for 6 minutes. Both the C and the T + C specific cleavage reaction mixtures were then placed on ice and 60 ul of hydrazine "stop buffer" (0.3 M sodium acetate, 0.1 mM EDTA, 50 ug/ml yeast tRNA) and 250 ul of absolute ethanol was added to each. These reaction mixtures were then stored at -70°C for 15 minutes. The G, T + C and C specific reaction mixtures were then centrifuged at 10,000 rpm for 5 minutes. The resulting pellets were resuspended in 60 ul of 0.3 M sodium acetate, 200 ul of absolute ethanol was added and the mixtures were stored at -70°C for 15 minutes. Following

centrifugation as before, the supernatant was removed and the pellets were washed twice with 200 ul of 70% ethanol. The dried pellets from all four reaction mixtures were then resuspended in 100 ul of 10% v/v piperidine (Sigma), incubated at 90°C for 30 minutes and lyophilized overnight. The dried reaction mixtures were twice resuspended in 20 ul of dH_2O and re-lyophilized prior to gel electrophoresis on 6% denaturing polyacrylamide gels.

V. CAT ASSAYS

Chloramphenicol acetyl transferase (CAT) assays were used as an indirect means of determining the promoter activity, if any, of DNA fragments inserted immediately upstream of the bacterial CAT gene, contained in the plasmid pLW2 (Gaffney et al., 1985). HeLa cells were transiently transfected as described above and harvested as follows.

The tissue culture medium was carefully removed from the cells, the cells were washed with 5 mls of PBS buffer and then soaked in 1 ml of CAT-TEN buffer (40 mM Tris, pH 7.4; 150 mM NaCl; 1mM EDTA) for 5-15 minutes. The cells were scraped off the dishes, pipetted into Eppendorf tubes on ice, centrifuged at 10,000 rpm for 30 seconds at 4°C and the supernatant removed. The cell pellet was then resuspended in 100 ul of 0.25 M Tris (pH 7.8), dispersed by pipetting and immediately transferred to an ethanol/ dry-ice bath. The cells were lysed by three successive cycles of freezing and thawing (37°C) and centrifuged at 10,000 rpm for 8 minutes at 4°C. The supernatant was removed and used as a crude cell extract for the CAT assay.

The levels of CAT activity contained within each cell extract were measured by incubating (37°C for 30 minutes) 20 ul aliquots of each cell extract in a 50 ul mixture containing, 250 mM Tris (pH 7.8), 3 mg/ml acetyl coenzyme A (Sigma) and 0.5 uCi of ^{14}C-chloramphenicol (NEN, 100

uCi/ml;60 mCi/mmol). Each reaction was stopped by the addition of 300 ul of ethyl acetate (Fisher) and vortexed vigorously for exactly 15 seconds. The timing is important because the amount of ^{14}C chloramphenicol extracted into the organic phase varies with time. After vortexing, the reaction mixture was centrifuged at 10,000 rpm for 8 minutes at room temperature and the organic phase was removed and dried down in a dessicator. The dried ^{14}C chloramphenicol was redissolved in 20 ul of ethyl acetate and spotted onto SIL-G silica gel thin layer chromatography plates (Kodak). Ascending chromatography was performed in 95% chloroform (Fisher): 5% methanol (Fisher), after which the chromatograms were autoradiographed overnight at room temperature without an intensifying screen.

RESULTS

I. MAPPING THE PROMOTER SEQUENCES FOR THE HUMAN c-Ha-ras1 GENE

Promoter sequences of genes are genetic regulatory elements which direct correct levels of transcription of linked genes from specific initiation sites. To understand transcriptional regulation of the human c-Ha-ras1 gene, it is therefore necessary to identify the transcriptional initiation site(s) for this gene. Subsequent analysis of the linked promoter sequences controlling the levels of transcription from these sites can then be undertaken. Sequence data from Capon et al. (1983a) and Reddy (1983), as discussed in the Introduction, suggested that the promoter for the c-Ha-ras1 gene might be located about 250 bp upstream from the translational initiation site. This suggestion was based largely on weak homologies to the TATA and CAAT boxes, which are known to regulate transcription initiation in certain genes (Introduction, II.b.i). Thus initial attempts at mapping the transcription initiation sites were focused on the region containing the putative TATA/CAAT promoter.

(a) S1 mapping analysis of the putative TATA/CAAT promoter region

To identify transcription initiation sites in the region identified as the putative promoter region, the initial approach of S1 mapping analysis was used, as described by Berk and Sharp (1977) and modified by Weaver and Weissman (1979). A series of DNA restriction fragments spanning the putative promoter region were used as S1 hybridization probes. These fragments were isolated from pT24 (Figure 1-1), a plasmid containing the c-Ha-ras1

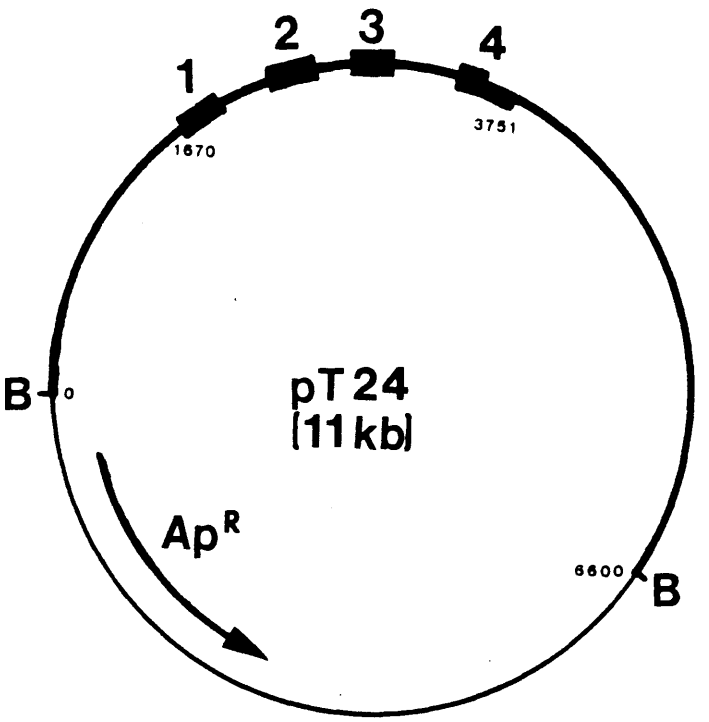

Figure 1-1 The plasmid pT24.

Plasmid pT24 comprises the 6.4 Kb BamHI fragment (heavy line) containing the human c-Ha-*ras*1 oncogene from the EJ/T24 bladder carcinoma cell line, inserted into the vector pBR322. Numbering of the BamHI insert is as in Reddy (1983), where 1670 gives the position of the ATG codon and 3751 gives the position of the polyadenylation site. The positions of the four coding exons are indicated by shaded rectangles. The position of the selectable ampicillin resistance (Ap^R) gene within the pBR322 sequence is shown by the arrow. B = BamHI.

oncogene derived from the EJ/T24 human bladder carcinoma cell line (Introduction, I.b), cloned as a 6.4 Kb BamHI fragment, into the unique BamHI site of pBR322. The pT24 plasmid was obtained from Dr. Alan Balmain (Beatson Institute, Glasgow). Strand separated DNA fragments labelled with γ-^{32}P-ATP at the 5' termini were hybridized to polyadenylated (poly A^+) RNA from the EJ/T24 human bladder carcinoma cell line. RNA from this cell line was used, as transcription of the c-Ha-ras1 gene was believed to be ten fold elevated with respect to other cell lines available (Dr. Alan Balmain, personal communication). To control for probe artifacts each probe was also hybridized to yeast transfer RNA (tRNA), as this RNA contains no sequences homologous to ras genes.

Initially a 1146 bp Sau3A fragment, extending from the middle of exon 1 to approximately 1 Kb upstream of exon 1, was isolated from pT24 (Figure 1-2). This fragment was 5' end-labelled with γ-^{32}P-ATP, strand-separated and the strand labelled at the Sau3A site within exon 1, was hybridized at 55^{o}C for 16 hours in 80% formamide hybridization buffer (FHB, see Materials and Methods) to 10 ug of poly A^+ EJ/T24 RNA plus 10 ug of yeast tRNA, or to 20 ug of yeast tRNA. Following hybridization, the hybrids were digested with 300 units of S1 nuclease at 37^{o}C for 90 minutes and the products denatured and sized on a 6% denaturing polyacrylamide gel. Data representative of six independent S1 mapping experiments, using this fragment, are given in Figure 1-2. A number of DNA fragments were protected from S1 digestion. The major bands form a cluster approximately mapping to positions -40 to -50 relative to the ATG translational initiation codon, which is taken as position +1. Three minor bands were frequently observed, mapping just upstream of the main cluster, at positions -70 $\pm$2, -75 $\pm$3 and -82 $\pm$3. However, no bands were detected in the yeast control lane, indicating that these protected species are unlikely to be probe artifacts. Furthermore, no protected bands were observed mapping to the region

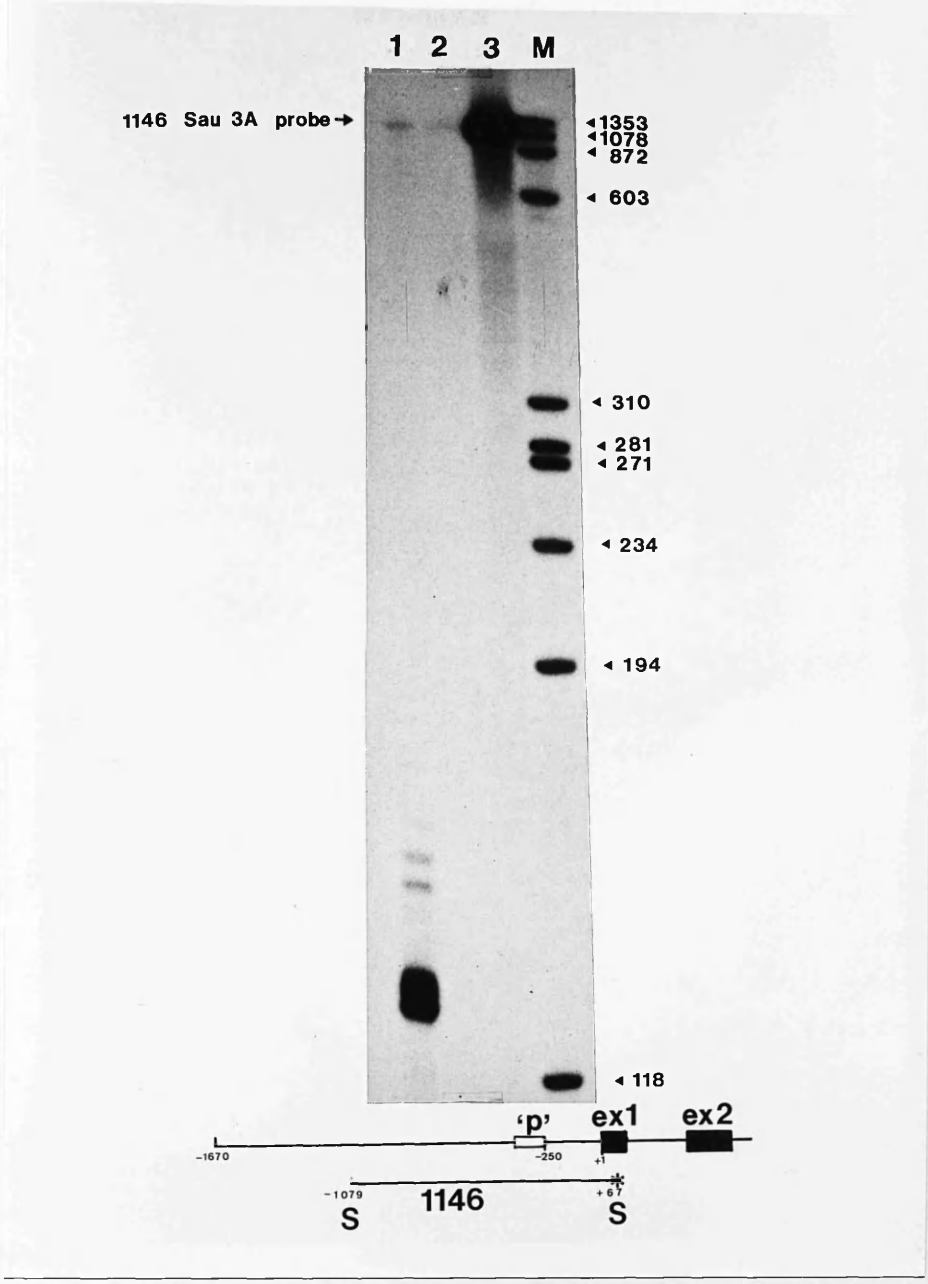

Figure 1-2 S1 mapping analysis of EJ poly A^+ RNA using the 1146 nucleotide Sau3A-Sau3A probe.

RNA was hybridized for 16 hours at 55°C to γ-^{32}P-ATP 5' end-labelled anti-sense 1146 Sau3A-Sau3A probe DNA. The probe was labelled in the first coding exon as indicated by the asterisk in the line drawing and all hybridizations were carried out in probe excess. Hybrids were treated with S1 nuclease, as described in the main text, and digestion products were denatured and separated on a 6% denaturing polyacrylamide gel. Autoradiography was for 10 days at -70°C with an intensifying screen. Lanes: (1) 7ug of EJ poly A^+ and 13 ug of yeast tRNA; (2) 20 ug of yeast tRNA; (3) probe alone. Markers (M) are provided by HaeIII fragments of ØX174, whose sizes are indicated by closed arrow heads. The line drawing shows the positions of the first two coding exons (ex1 and ex2) and the position of the putative TATA/CAAT promoter (P) as an open box. Numbering is relative to the ATG translational initiation codon. S = Sau3A.

containing the proposed TATA and CAAT boxes, as would be expected if these motifs were elements of the c-Ha-ras1 promoter. To confirm this result, the independent 213 nucleotide PstI-HaeIII probe, mapping from position -182 to +31, and the 415 nucleotide XmaI-XmaI probe, mapping from position -441 to -26, were labelled as shown in Figures 1-3 and 1-4, and hybridized to EJ poly A^+ RNA at 55°C and 45°C respectively. The lower hybridization temperature was used in the case of the 415 nucleotide XmaI-XmaI probe as the sizes of the hybrids formed were expected to be much smaller than those detected by the 213 nucleotide PstI-HaeIII probe. The protected hybrids were digested with 300 units S1 nuclease at 37°C and the products denatured and sized on 8% or 10% denaturing polyacrylamide gels. As shown in Figures 1-3 and 1-4, these independent probes confirm the presence of c-Ha-ras1 specific messenger RNA (mRNA) with putative 5' termini clustered between positions -40 to -50 and approximately at positions -70, -75 and -80 relative to the ATG codon. This confirms the putative 5' termini observed with the 1146 nucleotide Sau3A-Sau3A probe. The weakness of the banding pattern observed with the 213 nucleotide PstI-HaeIII probe (Figure 1-3) is believed to be caused by a combination of the reduced efficiency of 5' end-labelling at a blunt end and because half the amount of EJ poly A^+ RNA was used in the experiment. The result obtained with the 415 nucleotide XmaI-XmaI probe was particularly useful for estimating the sizes of the protected fragments observed, as the position of the 5' end-label was only 20 nucleotides from the main cluster of bands observed. Additionally, this 415 nucleotide XmaI-XmaI probe did not detect RNA termini which would correspond to termini produced by the proposed TATA/CAAT promotor. In this experiment fainter bands were also observed in the negative control (Figure 1-4, lane 2). These fainter bands were occasionally observed in experiments using the 415 nucleotide XmaI-XmaI probe and are now known to correspond to secondary structure within

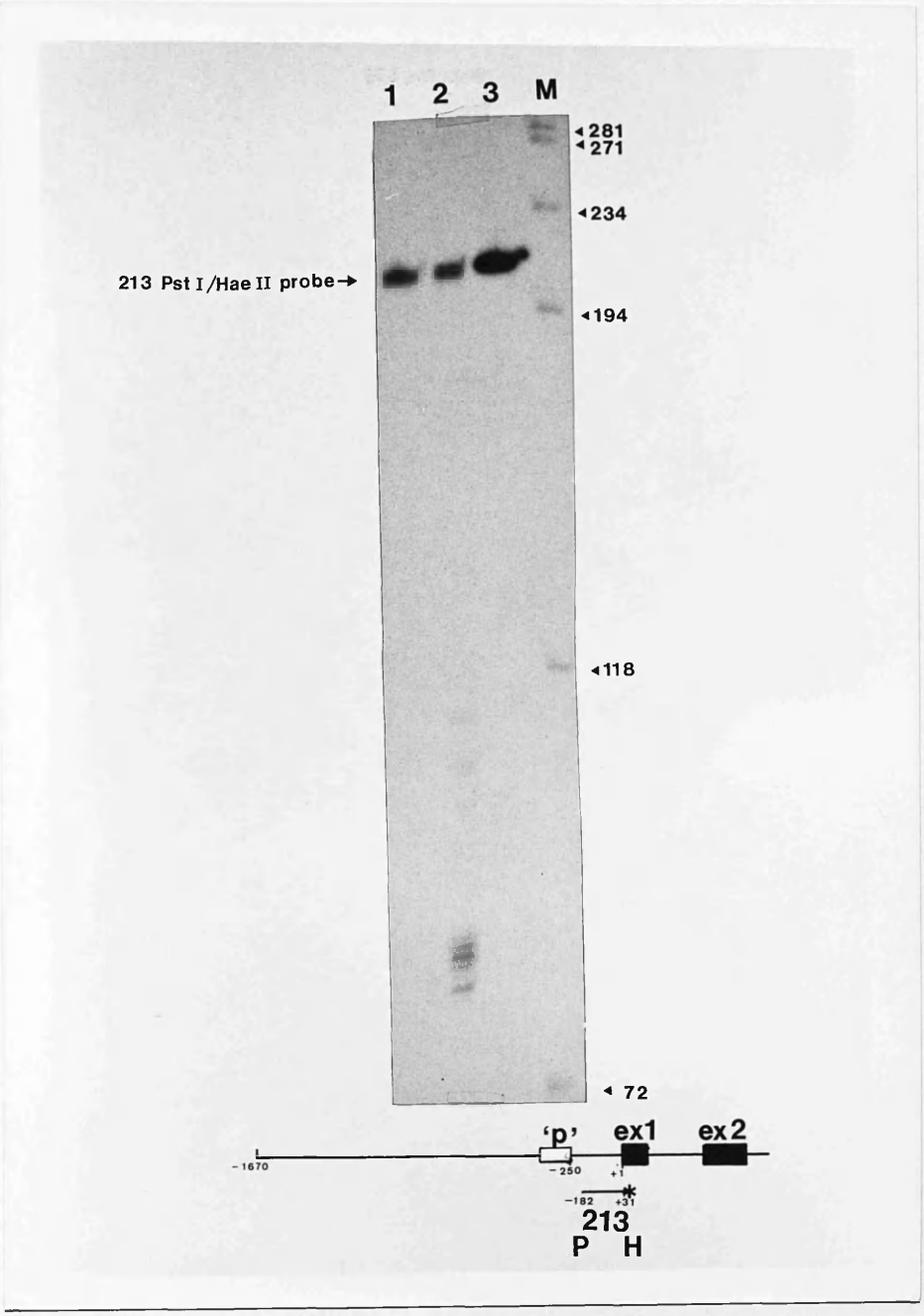

Figure 1-3 S1 mapping analysis of EJ poly A+ RNA using the 213 nucleotide Pstl-HaeIII probe.

RNA was hybridized for 16 hours at 55°C to γ-^{32}P-ATP 5' end-labelled anti-sense 213 nucleotide Pst-HaeIII probe DNA. The probe was labelled in the first coding exon as indicated by the asterisk in the line drawing and all hybridizations were carried out in probe excess. Hybrids were treated with S1 nuclease, as described in the main text, and digestion products were denatured and separated on an 8% denaturing polyacrylamide gel. Autoradiography was for 10 days at -70°C with an intensifying screen. Lanes: (1) 4 ug of EJ poly A$^+$ RNA and 16 ug of yeast tRNA; (2) 20 ug of yeast tRNA; (3) probe alone. Markers (M) are provided by HaeIII fragments of 0X174, whose sizes are indicated by closed arrow heads. The line drawing is as in Figure 1-2. P = PstI; H = HaeIII.

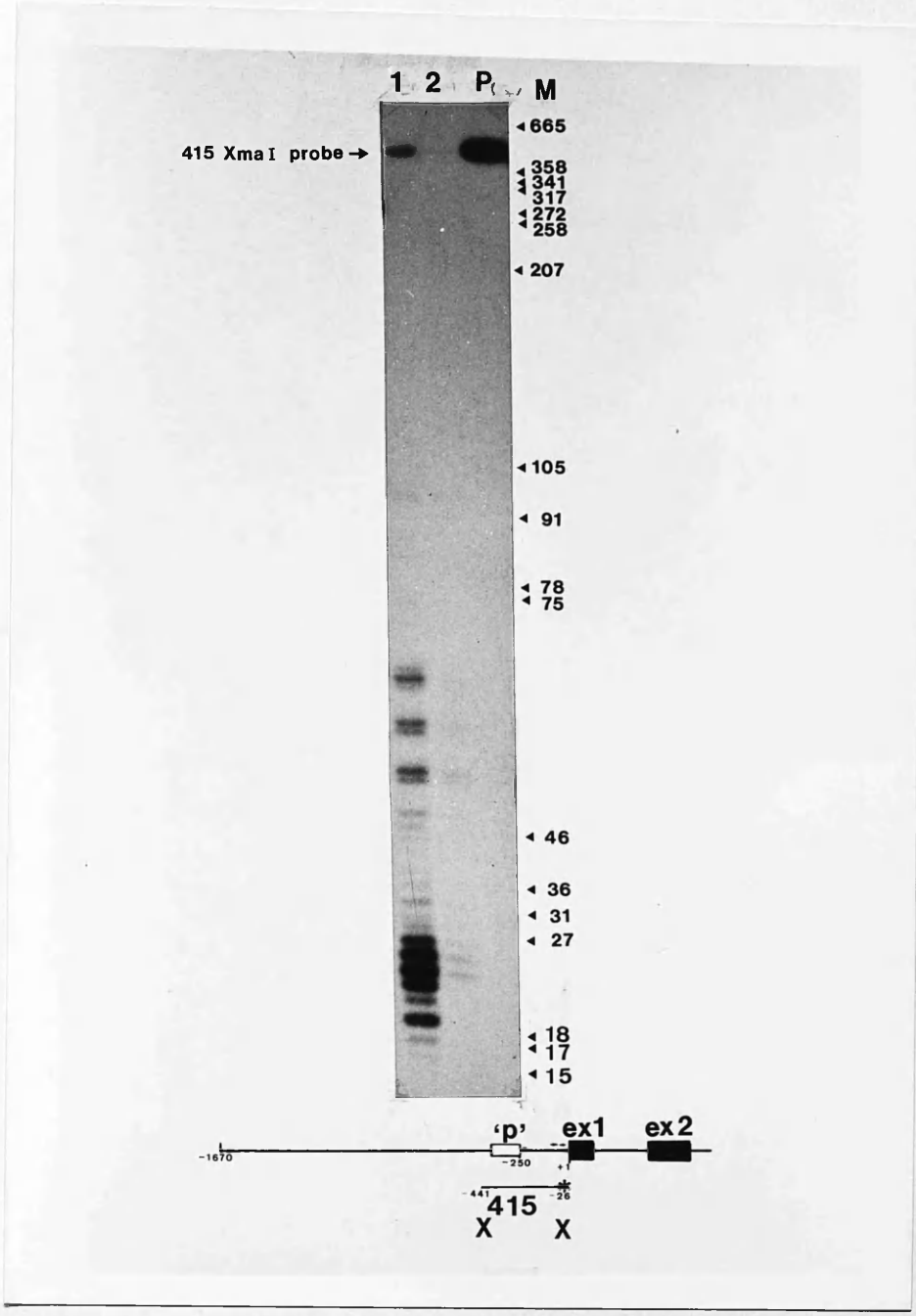

Figure 1-4 S1 mapping analysis of EJ poly A^+ using the 415 nucleotide XmaI-XmaI probe.

RNA was hybridized for 16 hours at 45°C to γ-^{32}P-ATP 5' end-labelled anti-sense 415 nucleotide XmaI-XmaI probe DNA. The probe was labelled in the 5' untranslated sequence of exon 1, as indicated by the asterisk in the line drawing and all hybridizations were carried out in probe excess. Hybrids were treated with S1 nuclease, as described in the main text, and digestion products were denatured and separated on a 10% denaturing polyacrylamide gel. Autoradiography was for 5 days at -70°C with an intensifying screen. Lanes: (1) 7 ug of EJ polyA^+ and 13 ug of yeast tRNA; (2) 20 ug of yeast tRNA; (3) probe alone. Markers (M) are provided by Sau3A fragments of pBR322 whose sizes are indicated by closed arrow heads. The line drawing is as in Figure 1-2 except that the small inverted arrows represent a potential stem-loop structure centered at position -54. X = XmaI.

the probe.

The technique of S1 mapping analysis using 5' end-labelled probes allows the detection of either genuine RNA initiation sites, or, 5' RNA termini generated by digestion at a splice junction. Inspection of the DNA sequence immediately upstream from the region where the major cluster of S1 protected fragments map, does indeed reveal an acceptor splice site (Capon et al., 1983a). Furthermore, closer examination of the DNA sequence in this region revealed a potential stem-loop structure. Figure 1-5 shows this secondary structure, which consists of a stem of 20 base-paired nucleotides and a 26 nucleotide loop. By assigning the rules of Tinoco et al. (1973), for estimating the most stable secondary structure for the RNA equivalent of this structure, a free energy of approximately -30 kCal at 25^{o}C in neutral buffer of moderate or high ionic strength, was calculated. Therefore, it is possible that this stem-loop structure could be stable under the conditions of S1 nuclease digestion. Taken together with the S1 mapping analysis, there are three possible explanations for these data: 1) that the clusters of S1 protected fragments map to genuine RNA transcription initiation sites, 2) that they are artifacts due to the strong stem-loop structure occuring in the probe molecule and that S1 cleavage at this structure prevents detection of RNA species initiating from the putative TATA/CAAT promoter, 3) that they correspond to 5' termini generated by digestion at a splice junction present in a longer message initiating transcription further upstream than the corresponding intervening sequence.The possibility that the stem-loop structure may interfere with the detection of longer transcripts is supported by the S1 mapping analysis using the 415 nucleotide XmaI-XmaI probe (Figure 1-4) since the bands observed with this probe accurately fit the model of the stem-loop structure shown in Figure 1-5. This does not, however, rule out the other two hypotheses.

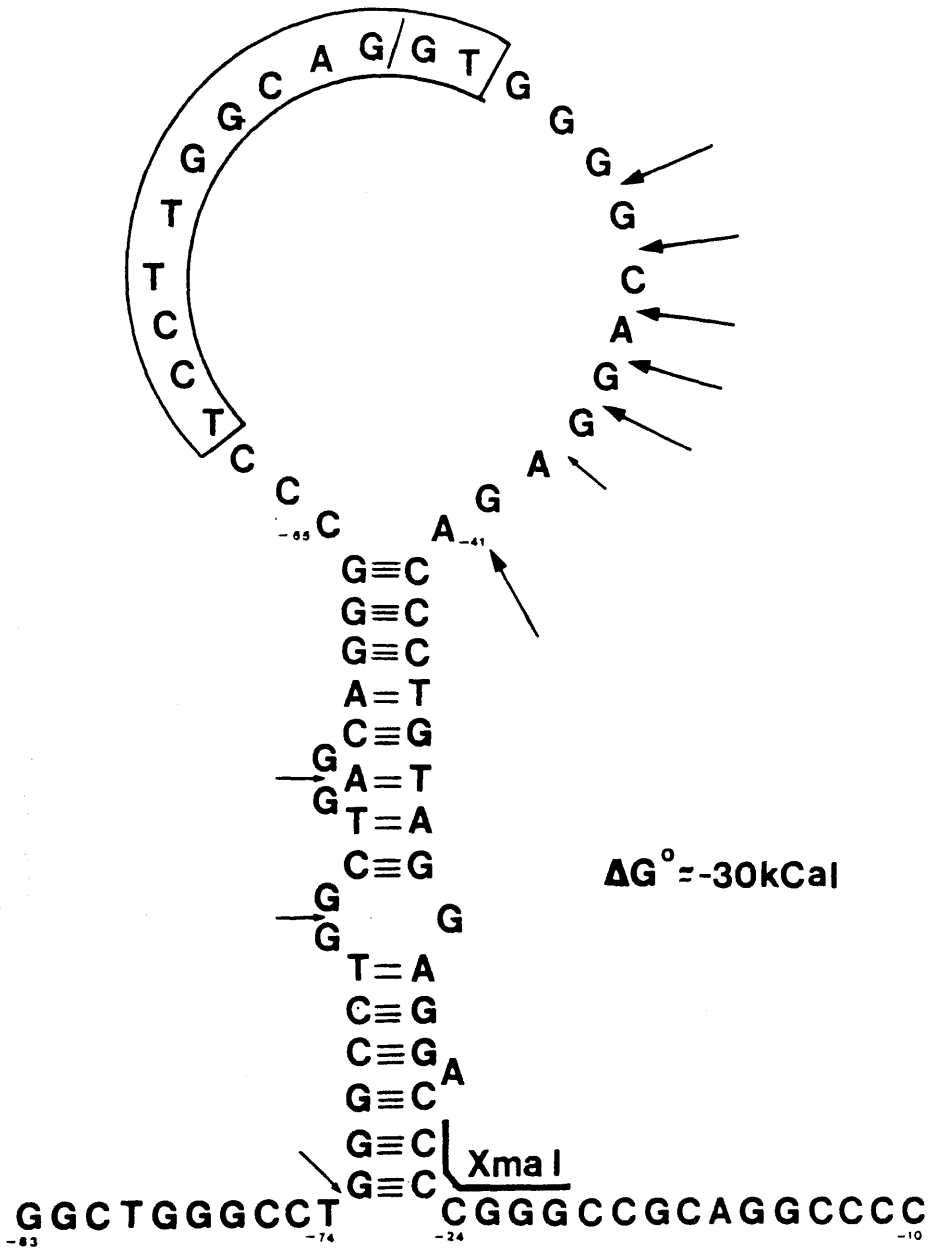

Figure 1-5 The stem-loop structure 24 bp upstream from the ATG translational initiation codon.

This secondary structure consists of a stem of at least 20 bp and a 26 nucleotide loop. It is shown for the sense strand of the double stranded DNA molecule. The free energy value (-30 kCal at 25°C) for the RNA equivalent of this structure was calculated from the rules of Tinoco et al. (1973). Numbering is relative to the ATG codon and the position of the XmaI site used for 5' end-labelling in Figure 1-4 is shown. The arrows mark the approximate position of the S1 protected fragments shown in Figure 1-4, where the size of the arrow is an indication of the intensity of the signal observed. The position of the concensus acceptor splice sequence within the loop region is shown by the enclosed curves and the splice site is at position -53. The horizontal dashes indicate hydrogen bonds.

(b) Exonuclease VII mapping analysis

To further investigate which of the above hypotheses is correct, an alternative mapping analysis, using exonuclease VII rather than S1 nuclease, was carried out. Exonuclease VII is also a single-stranded specific nuclease, but unlike S1 nuclease, it cannot cut single-stranded regions within hybrid molecules. Its activity is limited to exonucleolytic attack at single-stranded 5' and 3' ends, from which it sequentially releases small oligonucleotides (Chase and Richardson, 1964). Thus, by using exonuclease VII mapping it should be possible to distinguish whether the RNA termini mapping between positions -40 to -82 are genuine transcription initiation sites or whether the stem-loop structure in the probe allows S1 cleavage at this single stranded region thus preventing the detection of RNA(s) initiating further upstream, from the putative TATA/CAAT promoter. Furthermore, if the probe used for exonuclease VII mapping analysis completely spans the intervening sequence expected on the basis of the splice site hypothesis, then a very large protected fragment would be expected.

The 1146 nucleotide Sau3A-Sau3A single stranded probe, 5' end-labelled with γ-^{32}P-ATP at position +71 within the first coding exon and extending to position -1079, was prepared as before. This probe was hybridized at 55^{o}C for 16 hours in 80% FHB to either 5 ug of EJ/T24 poly A^{+} and 15 ug of yeast tRNA, or to 20 ug of yeast tRNA. Following hybridization, the hybrids were incubated at 37^{o}C for 90 minutes with varying amounts of S1 nuclease or exonuclease VII and the products were denatured and sized on a 6% denaturing polyacrylamide gel (Figure 1-6). The S1 mapping data was as before, with the major cluster of bands mapping to positions -40 to -50 and the three minor bands mapping to positions -70 $\pm$2, -75 $\pm$3 and -82 $\pm$3 (lanes 2 and 3). To determine the sensitivity of these bands to a higher

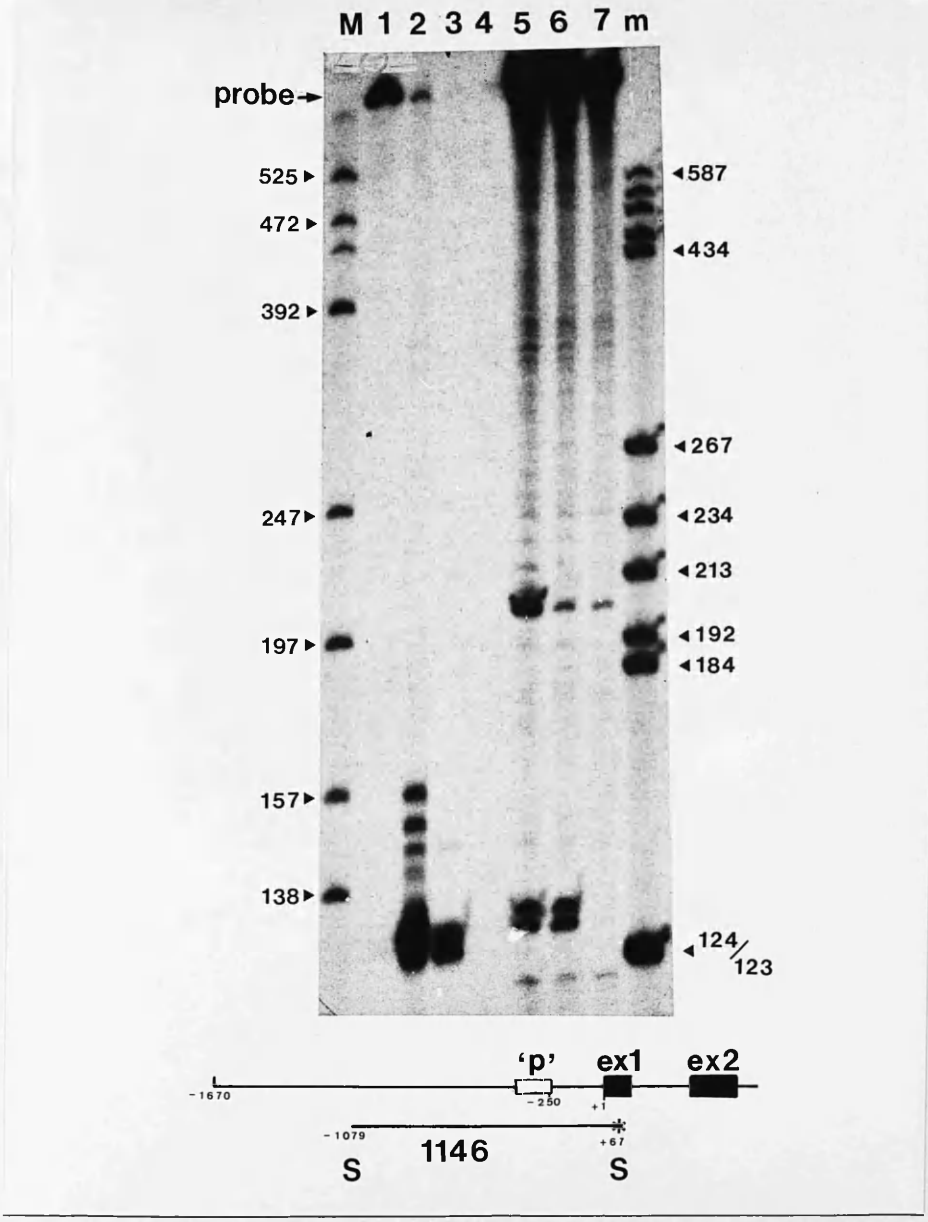

Figure 1-6 S1 and Exonuclease VII mapping analyses of EJ poly A^+ RNA using the 1146 nucleotide Sau3A-Sau3A probe.

RNA was hybridized to the γ-^{32}P-ATP 5' end-labelled 1146 nucleotide Sau3A-Sau3A probe exactly, as indicated in Figure 1-2. Hybrids were treated with 300 units or 900 units of S1 nuclease (Boehringer) for 1.5 hours at 37°C or with 5 units or 15 units of exonuclease VII (BRL) for 2 hours at 37°C. Digestion products were denatured and separated on a 6% denaturing polyacrylamide gel. Autoradiography was for 12 days at -70°C with an intensifying screen. Lanes: (1) probe alone; (2) 5 ug of EJ poly A^+ RNA and 15 ug of yeast tRNA, 300 units of S1; (3) 5 ug of EJ poly A^+ RNA and 15 ug of yeast tRNA, 900 units of S1; (4) 20 ug of yeast tRNA, 300 units of S1; (5) 5 ug of EJ poly A^+ and 15 ug of yeast tRNA, 5 units of exonuclease VII; (6) 5 ug of EJ poly A^+ and 15 units of yeast tRNA, 15 units of exonuclease VII; (7) 20 ug of yeast tRNA, 5 units of exonuclease VII. Markers are provided by RsaI (M) and HaeIII (m) fragments of 0X174 whose sizes are indicated by the closed arrow-heads. The line drawing is as in Figure 1-2. S = Sau3A.

concentration of S1 nuclease, the hybrids in lane 3 were digested with a 3 fold excess of S1 nuclease. This results in the loss of signal expected for such a large excess of enzyme but interestingly, the two higher molecular weight bands in lane 2, mapping to positions -75 $\pm$3 and -82 $\pm$3, disappear completely in lane 3. This suggests that these bands may be probe artifacts due to the stem-loop structure in the probe molecule, as increasing the concentration of S1 nuclease would be expected to increase the frequency of a second cut by S1 nuclease at a position within the loop region nearer to the 5' end-label. Thus S1 nuclease cuts, at sites within the secondary structure further away from the 5' end-label, would be observed less often. Following digestion with exonuclease VII, the only protected fragments not observed in the negative control (yeast tRNA; lane 7), map approximately to positions -55 and -60, relative to the ATG codon, which corresponds to a position within the loop region of the putative stem-loop structure in Figure 1-5. Since exonuclease VII digestion results in protected fragments approximately mapping to the same region as the major cluster of S1 nuclease digestion products, this result strongly suggests that genuine 5' RNA termini (generated either by transcription initiation or by digestion at the splice junction) exist in this region. Thus this result effectively rules out the possibility that an RNA species is initated from the region of the proposed TATA/CAAT promoter, mapping approximately 190 bp upstream from the S1/exonuclease VII protected fragments. Furthermore, these bands map to a position close to the consensus acceptor splice site at position -53, of sequence 5' TCCTTGGCAG/GT 3', where the underlined G residue is the only point of divergence from the acceptor splice site consensus, 5' PyPyPyPyPyPyX($^{C}/_{T}$)AG/G($^{G}/_{T}$) 3', of Breathnach and Chambon (1981), in which Py is pyrimidine and X is any base. The observation of two protected fragments mapping close to this consensus acceptor splice site, even after digestion with a 3-fold excess of exonuclease VII (lane 6),

may be explained by the mechanism of action of this enzyme. Exonuclease VII rarely digests to a flush end because it releases oligonucleotides rather than mononucleotides. Therefore, the protected fragments observed are often slightly longer than those observed by S1 mapping analysis (Dr. Nick Proudfoot, personnal communication). Thus there are two possible explanations for these S1 nuclease and exonuclease VII mapping data: 1) the bands observed with both techniques reflect the putative splice junction but the complex banding pattern observed with S1, but not with exonuclease VII, is caused by the stem-loop structure occuring in the probe molecule, or 2) some of the bands observed are due to genuine transcriptional initiation sites mapping to a region approximately 40 to 50 bp upstream of the ATG codon (see Figure 1-5). Therefore, primer extension analysis was performed in order to distinguish between these two possibilities.

(c) Primer extension analysis

The preceding S1 nuclease mapping analysis suggested the existence of 5' RNA termini mapping to a major cluster between positions -40 to -50, with minor termini approximately at positions -70, -75 and -82. With exonuclease VII mapping 5' termini mapping between positions -55 to -60, were observed. These analyses do not distinguish whether the protected fragments observed with these two techniques reflect 5' RNA termini fgenerated by digestion at the splice junction or genuine transcriptional initiation sites. Both possible explanations of the data were further confused by the potential stem-loop structure occuring within each of the probes used. If the 5' termini detected by S1 and exonuclease VII analysis represent genuine initiation sites, they will also be detected by primer extension. However, if a splice junction exists at this position (-54, relative to the ATG codon), then

primers from within the first coding exon should extend through the acceptor splice site at position -54, up to the actual 5' terminus or termini, located 5' to the corresponding donor splice site. Furthermore, the actual 5' termini are very likely to be located further upstream than the Sau3A site at position -1079, that is, the upstream Sau3A site of the 1146 nucleotide Sau3A-Sau3A probe used in Figure 1-6, since as previously mentioned, a much longer exonuclease VII protected fragment should have been observed if the corresponding donor splice site occurred within this probe. A candidate for the donor splice site has been reported at position -1094 (Capon et al., 1983a). Its sequence, 5' ACG/GTAGT 3', is a good match for the donor splice site concensus sequence, 5' (A/C)AG/GT(A/G)AGT 3' (Breathnach and Chambon, 1981), as the second cytosine residue in the c-Ha-ras1 sequence is the only point of divergence between the two. Alternatively, if the bands mapping to the region between positions -40 to -82, represent genuine transcriptional initiation sites, as inferred from the S1 nuclease and Exonuclease VII mapping analyses, then these sites should be confirmed by primer extension analyses.

Several short DNA restriction fragments were isolated and used for primer extension analysis with EJ or A431 cell line RNAs (A431 cells are a human vulval carcinoma-derived cell line, which were available at the Beatson Institute, Glasgow). However, only one of these fragments, the 51 bp AluI/PvuII fragment, mapping from position +13 to +65 within exon 1, could be successfully extended with reverse transcriptase. This fragment was 5' end-labelled with γ-^{32}P-ATP, strand-separated and the strand labelled at position +65 (see Figure 1-7) was hybridized at 50^{o}C for 16 hours in 80% FHB to 10 ug of EJ poly A^{+} RNA, 10 ug of A431 poly A^{+} RNA or to 10 ug of yeast tRNA. The hybrids were extended by treatment with reverse transcriptase (Materials and Methods) and the resulting extension products were denatured and sized on an 8% denaturing polyacrylamide gel.

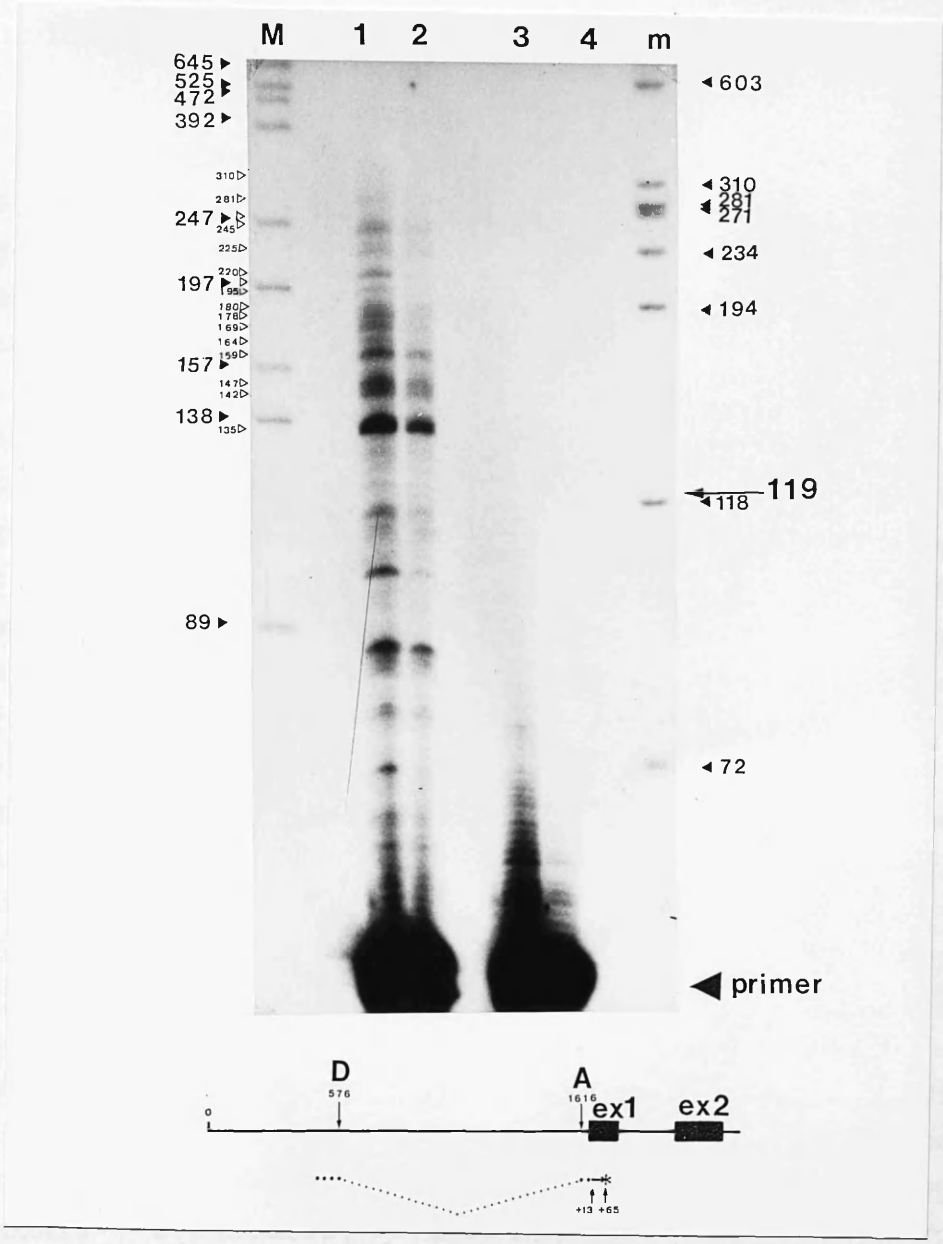

Figure 1-7 Primer extension analysis of EJ and A431 RNAs using the 51 nucleotide AluI-PvuII primer.

RNA was hybridized at 52°C to γ-^{32}P-ATP 5'. end-labelled anti-sense 51 nucleotide AluI-PvuII primer DNA. The primer was labelled within the first coding exon at position +65 relative to ATG, as indicated in the line drawing by the asterisk. All hybridizations were carried out using approximately 200 cps of the primer. Hybrids were treated with 25 units of AMV reverse transcriptase (Boehringer) for 3 hours at 42°C and extended products were denatured and separated on an 8% denaturing polyacrylamide gel. Autoradiography was for 36 hours at -70°C with an intensifying screen. Lanes: (1) 10 ug of EJ poly A$^+$ RNA; (2) 10 ug of A431 poly A$^+$ RNA; (3) and (4) 10 ug of yeast tRNA. Markers are RsaI (M) and HaeIII (m) fragment of ØX174, whose sizes are indicated by the closed arrow heads. The line drawing shows the position of the primer which extends from positions, +13 to +65, relative to the ATG codon. The region upstream of D to which the extended products above position 119 nucleotides map, is indicated by the large dots. The small dots represent the region of RNA removed by the splicing event. The arrow at position 119 nucleotides shows the position of the splice site. The large arrow-head marks the position of the primer and the open arrow-heads indicate the positions and sizes of major extension products above position 119 nucleotides. The donor (D) and acceptor (A) splice sites are numbered according to Reddy (1983).

Data representative of primer extension analyses using this fragment are given in Figure 1-7. Multiple extension products of varying intensity were observed, which do not reflect the patterns observed with the preceding S1 nuclease and exonuclease VII analyses. Furthermore, extension products are observed which would map to a region upstream of the S1 nuclease and exonuclease VII protected fragments. Thus, this primer extension analysis suggests that the 51 nucleotide AluI-PvuII primer is extended through the acceptor splice site at position -54, up to 5' termini located 5' to the corresponding donor splice site. There are at least 16 extension products greater in size than 119 nucleotides (that is, the position of the acceptor splice site). These are approximately sized in Figure 1-7 and correspond to putative transcriptional initiation sites mapping approximately to positions, 16, 23, 28, 40, 45, 50, 59, 61, 76, 86, 101, 106, 126, 135, 162 and 191 bp upstream of the corresponding donor splice site. Those extension products smaller than 119 nucleotides, the position which marks the acceptor splice site (see Figure 1-7), are likely to be artifacts due to premature termination of reverse transcriptase, perhaps caused by RNA secondary structures. Examination of overexposed autoradiographs revealed extension products at least as large as 310 nucleotides which would indicate transcripts initiating at least 191 bp upstream from the donor splice site. From the complexity of the banding pattern observed, it is possible that not all of the extended products correspond to genuine 5' RNA termini. Another possibility is suggested by the fact that the primer was isolated from the first coding exon of the c-Ha-_ras_1 gene. This exon is known to be highly homologous to other _ras_ genes (Introduction, section I.b), therefore it is possible that some of the extension products could be derived from hybridization of the primer to other "_ras_-like" species (Introduction, section Ia,b). While this work was in progress, Ishii et al. (1985) carried out primer extension

analysis of A431 cell RNA, using the same 51 nucleotide AluI-PvuII primer described here. They suggested that extension products observed above the band numbered 159 in Figure 1-7 are artifactual, but that the extension products corresponding to those numbered 135, 142, 147 and 159 in Figure 1-7 are genuine transcriptional initiation sites. This claim is based on their reported sequencing of the extended primer, which showed that the sequences of the extended products above the 135, 142, 147 and 159 products were different from the sequence of the human c-Ha-ras1 genomic clone. Unfortunately, their sequence data was not presented to support their hypothesis that the larger extended products were derived from cross hybridization to transcripts from other gene(s). In an attempt to resolve this question, the cell line EJ focus 8.3.5, was obtained from Dr. Alan Balmain (Beatson Institute, Glasgow) and RNA was isolated for primer extension analysis. EJ focus 8.3.5 is a primary focus cell line, obtained by tranfection of high molecular weight EJ cell line DNA, into NIH 3T3 mouse fibroblasts. Using the techniques of Southern and northern blot analysis, EJ focus 8.3.5 has been shown to be positive for the EJ c-Ha-ras1 oncogene (it is contained on a 23 Kb EcoR1 restriction fragment present in 3 to 4 copies per cell) and that it expresses this gene at high levels (Dr. Alan Balmain, personal communication). Figure 1-8 shows the result obtained when 50 ug of EJ focus 8.3.5 total RNA and 50 ug of mouse Friend cell total RNA, were used for hybridization against the 5' end-labelled 51 nucleotide AluI-PvuII primer (Friend cells are a murine erythroid leukaemia cell line and were used as a mouse RNA negative control). The RNAs were hybridized at 52^{o}C for 16 hours and the resulting hybrids were extended with reverse transcriptase as described before and in the legend for Figure 1-8. In this experiment 10 ug of EJ poly A^{+} RNA was used as a positive control and 40 ug of yeast tRNA as a negative control. As can be seen from comparison of lanes 1 and 3 in Figure 1-8, essentially the same pattern of

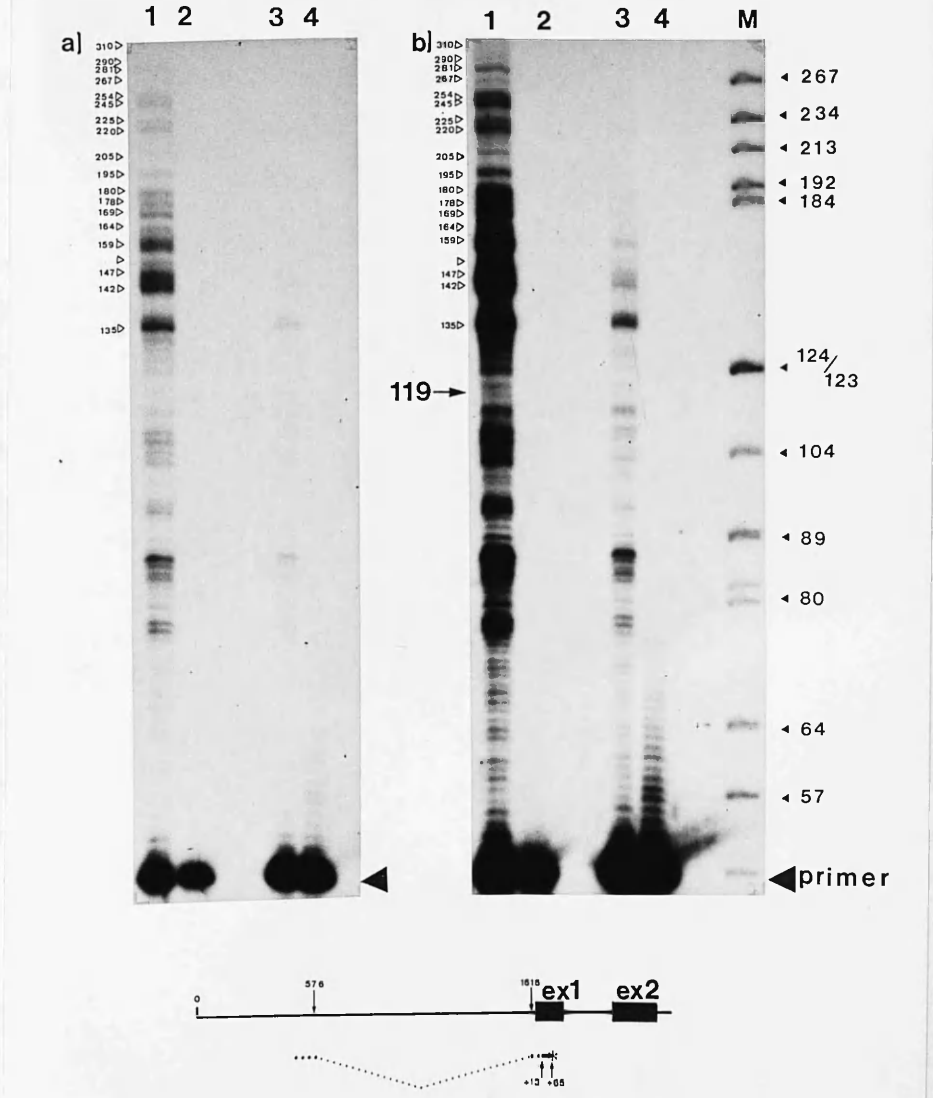

Figure 1-8 Primer extension analysis of EJ and EJ focus 8.3.5 RNAs using the 51 nucleotide AluI-PvuII primer.

RNA was hybridized at 52°C to γ-^{32}P-ATP 5' end-labelled anti-sense 51 nucleotide AluI-PvuII primer DNA. The primer was labelled within the first codon exon as indicated by the asterisk in the line drawing. All hybridizations were carried out using approximately 100 cps of primer. The hybrids were incubated with reverse transcriptase as described in the legend to Figure 1-7, except that RNA was alkali hydrolysed following reverse transcription. Extended products were denatured and separated on an 8% polyacrylamide gel. Autoradiography was for, (a) 3 hours or (b) 16 hours, at -70°C with an intensifying screen. Lanes: (1) 50 ug of EJ focus 8.3.5 total RNA; (2) 50 ug of Friend cell total RNA; (3) 10 ug of EJ poly A$^+$ RNA; (4) 40 ug of yeast tRNA. Markers are HaeIII fragments of pBR322 (M) and HaeIII fragments of 0X174 (m) whose sizes are indicated by the closed arrow-heads. The position and sizes of the major extension products are indicated by open arrow-heads.

extended products can be seen with both the EJ focus 8.3.5 and EJ RNAs. These extended products map approximately to positions 16, 23, 28, 40, 45, 50, 59, 61, 76, 86, 101, 106, 126, 135, 148, 162, 171 and 191 bp upstream of the corresponding donor splice site (in this experiment two extra extension products of size 267 and 290 nucleotides were detected, mapping to 148 and 171 bp upstream of the donor splice site). However, the intensity of the signal observed for the EJ focus 8.3.5 total RNA is at least 100 times greater than the EJ poly A^+ RNA, assuming that 50 ug of total RNA is equivalent to about 1 ug of poly A^+ RNA (Figure 1-8, lanes 1 and 3). No signal can be detected even at long exposures (up to 1 week), in lane 2, in which 50 ug of Friend cell total RNA was used. This result confirms that under the conditions used, all the extension products observed in Figure 1-8 are derived from extension on human c-Ha-ras1 RNA.

In summary, the data in this section provide evidence for human c-Ha-ras1 specific transcripts, mapping at least 191 bp upstream from the putative donor splice site corresponding to the acceptor splice site at position -54 relative to the ATG codon. These data imply that the c-Ha-ras1 gene has an upstream exon, of at least 191 bp, which is analogous to the human N-ras and c-Ki-ras1 genes (Introduction, Figure 1). Furthermore, sequence comparisons support the location of this upstream exon as being immediately upstream from the reported donor splice site at position -1094 (Capon et al., 1983a). Chichutek and Duesberg (1986) have compared the sequences of the human c-Ha-ras1 gene with that of various sarcoma viruses containing the Harvey ras gene and also with that of the rat c-Ha-ras2 pseudogene. These analyses demonstrated that certain sarcoma viruses and the rat c-Ha-ras2 pseudogene are homologous with the human c-Ha-ras1 genomic sequence up to the acceptor splice site at position -54, with homology resuming upstream of the donor splice site at position -1094. In addition, Ishii et al. (1985) report that the

sequence of their extended 51 nucleotide AluI/PvuII primer reads through the -53 acceptor splice site, to be continued for 40 nucleotides upstream of the -1094 donor splice site. Thus the c-Ha-ras1 upstream exon, which will now be termed exon -1 by analogy with the human N-ras upstream exon, is separated from the first coding exon by a 1040 bp intron. Since the pattern of primer extension products is complex (Figures 1-7 and 1-8), it is possible that not all of these products correspond to genuine 5' termini in the c-Ha-ras1 mRNA. An attempt to eliminate any artifactual extension products was made by using actinomycin D and also by replacing dGTP with 7-deaza-dGTP (7-deaza-2'-deoxyguanosine triphosphate), during the extension reaction. These reagents are believed to inhibit RNA secondary structure, or to relieve premature termination of the reverse transcriptase (Mizusawa et al., 1986). However, no significant difference was observed in the extension products obtained with or without these reagents (data not shown).

(d) S1 nuclease mapping analyses of the c-Ha-ras1 upstream exon

The preceding primer extension analyses and the sequence comparisons of Chichutek and Duesberg (1986), have demonstrated that the 5' boundary of exon -1, as defined by the initiation site of the most upstream transcript detected, is likely to be located at least 191 bp upstream from the -1094 donor splice site. To independently confirm this finding, and to determine more precisely which of the complex pattern of primer extension products represent genuine initiation sites, single-stranded 5' end-labelled probes were designed for S1 nuclease mapping analyses of this region. However, S1 nuclease mapping analyses in this region of the genome is hampered by its unusual G + C richness (the 576 bp upstream of the donor splice site at position -1094 is 80% G + C). S1 hybridization probes

which are G + C rich are more difficult to strand separate and require more stringent DNA:RNA hybridization conditions. In fact, Hess et al. (1985) have reported their inability to perform S1 analysis on a region of approximately 90% G + C richness upstream of the human a1 globin gene, although the primer extension technique was successful. A similar observation has been made by Treisman et al. (1982).

A 389 bp XcyI restriction fragment, mapping from positions 411 bp to 21 bp upstream from the donor splice site of exon -1, was isolated from plasmid pT24. This fragment was 5' end-labelled using γ-^{32}P-ATP and strand separated using 40% DMSO or 80% formamide buffers (see Materials and Methods). Additionally, some of the 5' end-labelled 389 bp XcyI fragment was treated with the restriction endonuclease, XhoI, prior to strand separation. XhoI removes 32 bp from the 5' end of the 389 bp XcyI fragment converting it into a 357 bp XhoI/XcyI fragment. Figure 1-9 is an example of the species observed upon strand separation of these 5' end-labelled restriction fragments. Unusually the 5' end-labelled 389 bp XcyI fragment is strand-separated into four major single-strand species following denaturation with DMSO (Figure 1-9a, lane 2) and three major single-stranded species with formamide (Figure 1-9a, lane 3, labelled I, II and III). Interestingly, form X (Figure 1-9a, lane 2) is specific only to denaturation by 80% DMSO as opposed to 80% formamide. These alternate forms of the two strands are an indication of the secondary structure forming potential of this region of the c-Ha-ras1 gene. No other fragments further upstream or downstream of this region, gave multiple single stranded forms upon strand separation. Figure 1-9b is a strand separation gel of the XhoI digested 5' end-labelled 389 bp XcyI fragment, the resultant 357 bp XhoI/XcyI fragment is labelled only at the XcyI site 21 bp upstream from the donor site of exon -1. Therefore, only one strand can be visualized by autoradiography. In this

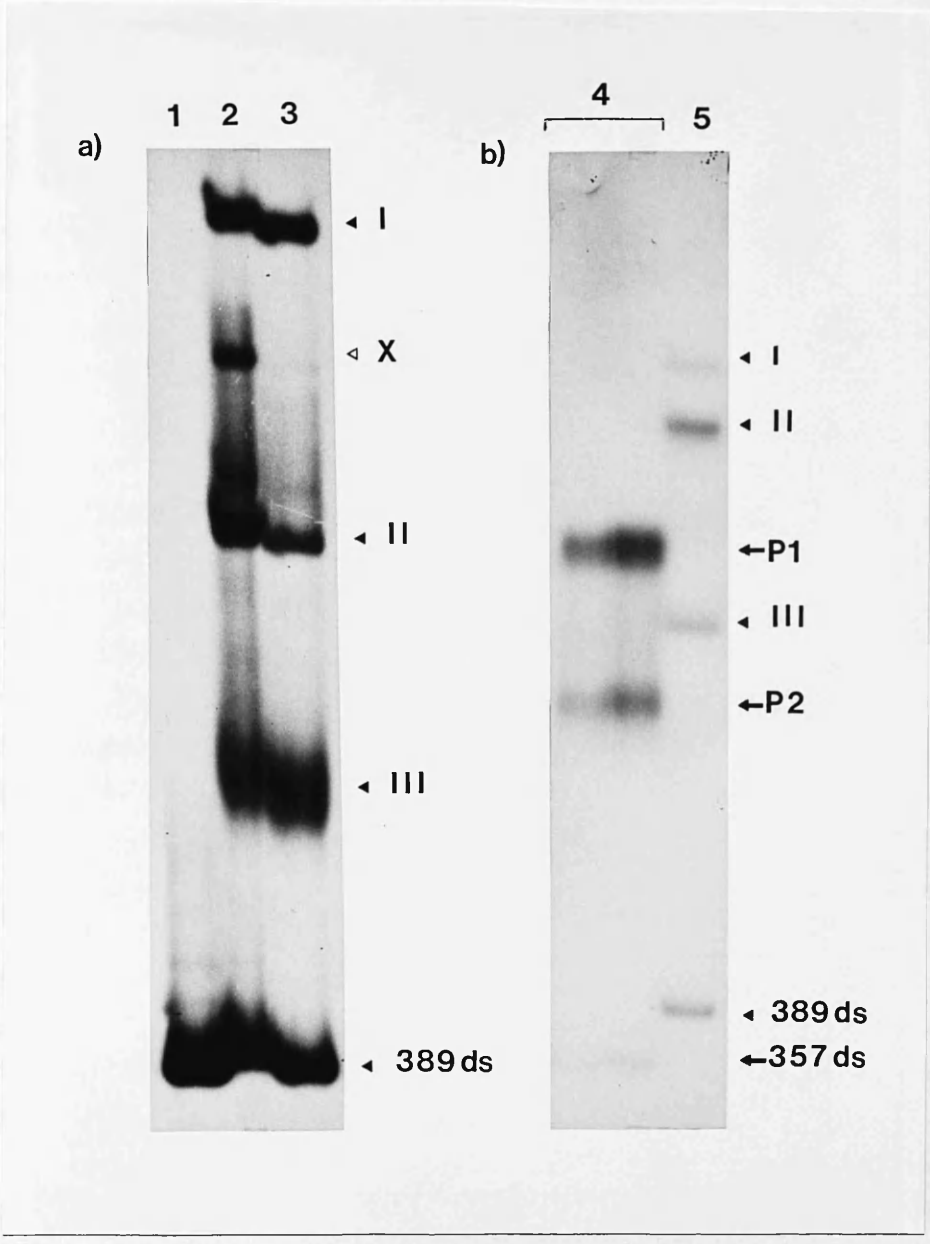

Figure 1-9 Strand separation gels of the 389 bp XcyI and the 357 bp XhoI/XcyI DNA fragments.

Double stranded DNAs were γ-32P-ATP 5' end-labelled, denatured with either dimethylsulphoxide (DMSO) or formamide as described in the Materials and Methods. The denatured DNAs were separated on strand separation gels (Materials and Methods) and autoradiography was for 2 minutes at room temperature with an intensifying screen. (a) a 5% strand separation gel electrophoresed at 300 V for 16 hours; (b) a 6% strand separation gel electrophoresed at 200 V for 16 hours. Lanes: (1) undenatured 389 bp XcyI DNA; (2) 389 bp XcyI DNA denatured with 80% DMSO; (3) 389 bp XcyI DNA denatured with 80% formamide; (4) 357 bp XhoI/XcyI DNA denatured with 80% DMSO (This DNA was obtained by isolating the 389 bp XcyI double stranded DNA from gel (a) and digesting it with XhoI); (5) 389 bp XcyI double stranded DNA denatured with 80% DMSO (used as a marker). Single-stranded forms I, II and III used in Figure 1-10 are indicated by closed arrow-heads, whereas single stranded forms P1 and P2 used in Figure 1-14 are indicated by the arrows. Form X (lane 2) was not used in any S1 nuclease mapping experiments. ds = double stranded form.

case two alternative forms of this anti-sense strand were observed, presumably of different secondary structure (Figure 1-9b, forms P1 and P2).

To identify potential transcription initiation sites upstream of the donor splice site of exon -1, single stranded probe forms I, II and III (see Figure 1-9a, lane 3), derived from the 5' end-labelled 387 bp XcyI fragment, were hybridized at 55°C in 80% FHB for 16 hours, to 4 ug of EJ poly A^+ RNA and 16 ug yeast tRNA, or to 20 ug of yeast tRNA. Following hybridization, the hybrids were digested with 300 units of S1 nuclease at 37°C for 90 minutes. The products were denatured and sized on a 6% denaturing polyacrylamide gel. As shown in Figure 1-10, only form II of the probe gave protected fragments which did not appear in the negative control lane (compare lanes 5 and 6). This signal consisted of multiple protected fragments ranging in size from approximately 180 to 98 nucleotides (mapping to positions 201 to 119 bp upstream of the donor splice site). The major clusters of protected fragments are bracketed and denoted by the letters A, B, C and D in Figure 1-10, where A corresponds to fragments between positions -210 to -191, B to fragments between positions -175 to -161, C to fragments between positions -151 to -141 and D to fragments between positions -128 to -119, all relative to the donor splice site of exon -1. This S1 nuclease analysis independently confirms the primer extension analysis of the preceding section (Figures 1-7 and 1-8) as both techniques identify corresponding transcriptional initiation sites in the region mapping between approximately 191 bp to 119 bp upstream of the donor splice site. However, it was not possible in this experiment to confirm the sites identified by primer extension analyses below position -119 relative to the donor splice site. It is clear that under these conditions of hybridization and S1 digestion, all three forms of the probe can form considerable secondary structures, as detected by the number of probe artifacts. Therefore, to confirm that probe form II is an anti-sense

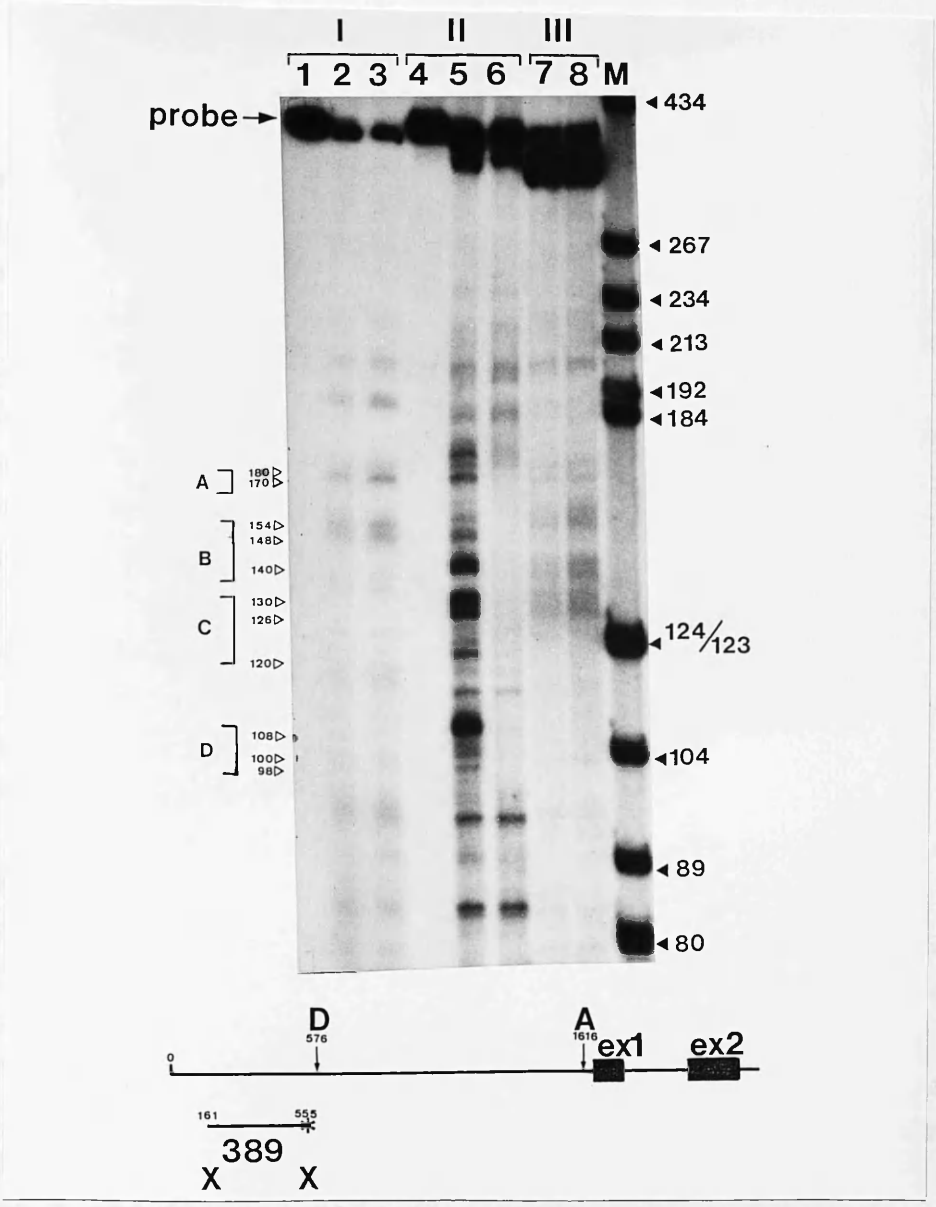

Figure 1-10 S1 mapping analysis of EJ poly A^+ using the 389 nucleotide XcyI-XcyI probe.

RNA was hybridized for 16 hours at 55°C to single-stranded forms I, II and III (Figure 1-9). The probe was labelled at a position 21 nucleotides upstream from the donor splice site (D) as indicated by the asterisk in the line drawing and all hybridizations were carried out in probe excess. Hybrids were treated with 300 units of S1 nuclease (Boehringer) for 1.5 hours at 37°C and digestion products were denatured and separated on a 6% denaturing polyacrylamide gel. Autoradiography was for 10 days at -70°C with an intensifying screen. Lanes: (1) an aliquot of unhybridized form I probe; (2) 5 ug of EJ poly A^+ RNA and 15 ug of yeast tRNA; (3) 20 ug of yeast tRNA; (4) an aliquot of form II probe; (5) 5 ug of EJ poly A^+ RNA and 15 ug of yeast tRNA; (6) 20 ug of yeast tRNA; (7) 5 ug of EJ poly A^+ RNA and 15 ug of yeast tRNA; (8) 20 ug of yeast tRNA. Lanes (1), (2) and (3) are with form I; lanes (4), (5) and (6) are with form II and lanes (7) and (8) are with form III probes respectively. Markers (M) are HaeIII fragments of pBR322 whose sizes are denoted by the closed arrow-heads. The line drawing is as in Figure 1-7. Protected fragments are indicated by open arrow-heads and these are arbitrarily grouped into clusters A to D as shown. X= XcyI.

form of the 389 nucleotide XcyI-XcyI probe, it was subjected to the Maxam-Gilbert sequencing protocol and the reaction products analysed on a 6% denaturing polyacrylamide gel. Figure 1-11 confirms that probe form II of the γ-^{32}P-ATP 5' end labelled 389 nucleotide XcyI-XcyI probe is indeed anti-sense, as it is complementary to the sequence of the sense-strand published by Capon et al. (1983a) and Reddy (1983). An interesting feature of the sequence that facilitated its identification are the runs of 4 or 5 guanosine residues, followed by a single cytosine residue (see Figure 1-11).

To determine if different hybridization conditions would improve the weak signal observed in Figure 1-10, a range of different hybridization temperatures were used with probe forms P1 and P2 of the 357 nucleotide XhoI-XcyI probe (Figure 1-9b). The 5' end-labelled 389 bp XcyI fragment was prepared, digested with XhoI to remove the 5' end-label from the sense strand and strand separated as in Figure 1-9. In Figure 1-12 both forms P1 and P2 of the resulting 357 nucleotide XhoI-XcyI probe were hybrizized for 16 hours to 5 ug EJ poly A$^+$ RNA and 15 ug yeast tRNA in 80% FHB, at temperatures of; 59^{o}C (lanes 1 and 5), 65^{o}C (lanes 2 and 6) and 72^{o}C (lanes 3 and 7). The yeast tRNA (20 ug) negative control was hybridized at 65^{o}C (lanes 4 and 8). These hybridization controls revealed essentially the same pattern of protected fragments as in Figure 1-10 and that the optimum signal is observed at 65^{o}C. Furthermore, there are fewer probe artifacts at these higher temperatures, than was observed at 55^{o}C for forms I, II and III of the 389 nucleotide XcyI-XcyI probe (Figure 1-10). It is not known why form P2 of the 357 nucleotide XhoI-XcyI probe, does not form hybrids at the lowest temperature used (lane 5), but does at the two higher temperatures (lane 6 and 7).

The 357 nucleotide XhoI-XcyI probe (form P1) was also used for S1 nuclease digestion controls. Increasing the amount of S1 nuclease used to digest the hybrids formed,

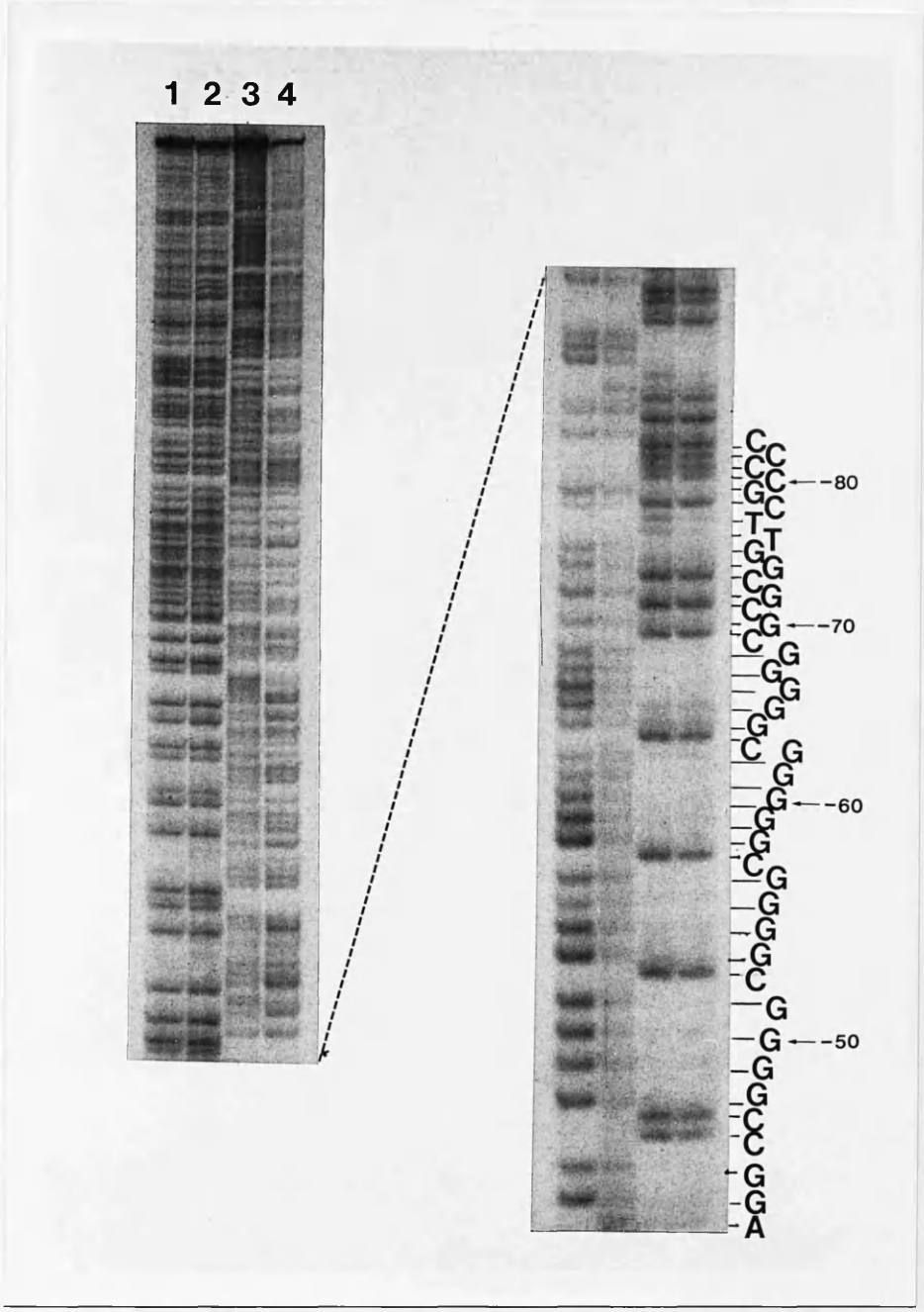

Figure 1-11 Maxam-Gilbert sequencing of the 5′ end-labelled 389 nucleotide XcyI-XcyI form II probe.

The γ-^{32}P-ATP 5′ end-labelled 389 nucleotide XcyI-XcyI form II probe (Figures 1-9a and 1-10) was subjected to the Maxam-Gilbert sequencing protocol as described in the Materials and Methods. Reaction products were denatured and separated on a 6% denaturing polyacrylamide gel. Autoradiography was for 10 days at -70°C without an intensifying screen. Lanes: (1) G reaction; (2) G + A reaction; (3) T + C reaction and (4) C reaction. The numbers indicate distances relative to the donor splice site.

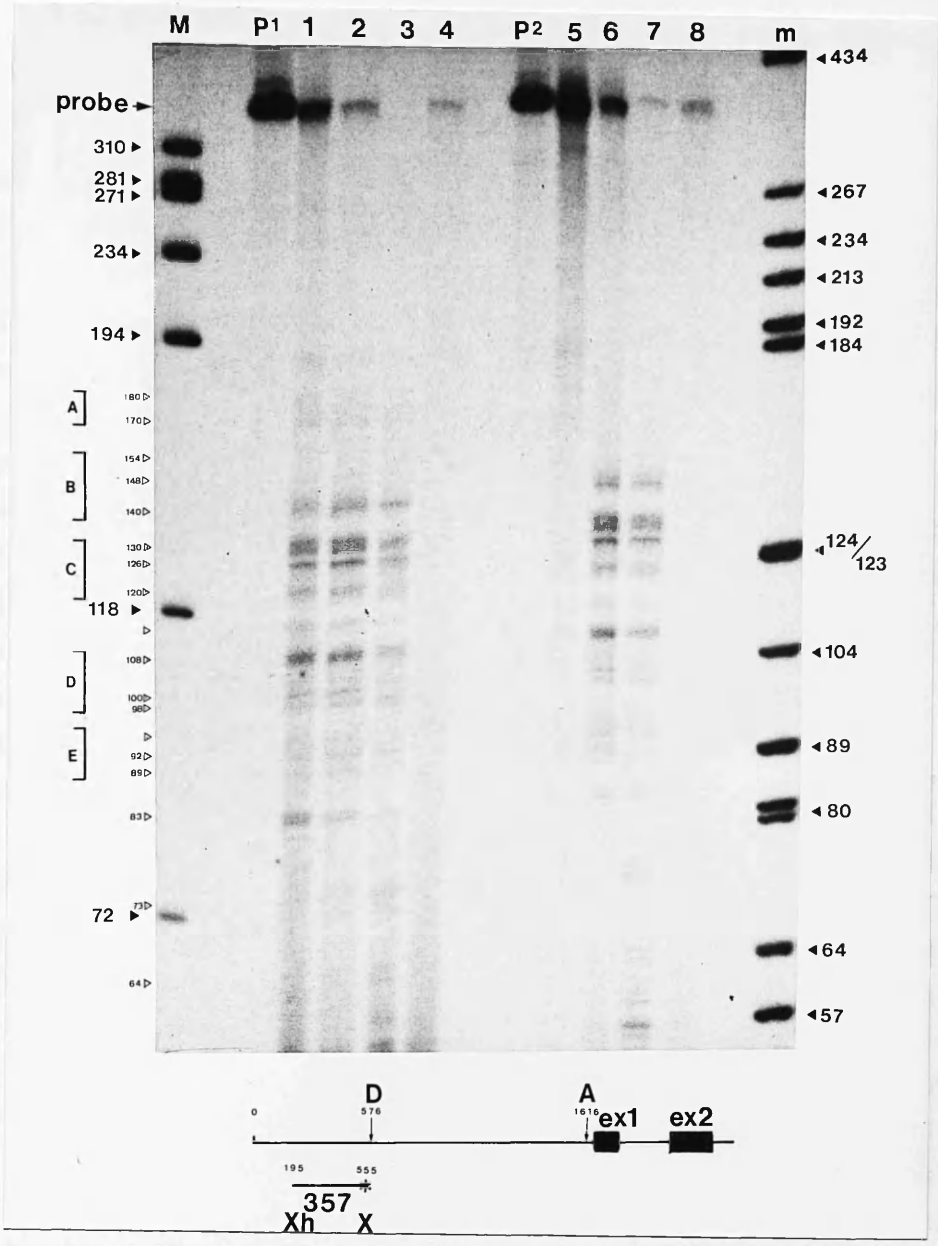

Figure 1-12 S1 mapping analysis of EJ poly A^+ RNA using the 357 nucleotide XhoI-XcyI probe under different hybridization conditions.

RNA was hybridized for 16 hours at 59°C, 65°C or 72°C to either form P1 or form P2 of the γ-^{32}P-ATP 5' end-labelled anti-sense 357 nucleotide XhoI-XcyI probe DNA. The probe was labelled at a position 21 nucleotides upstream from the donor splice site (D) at position 576, as indicated by the asterisk in the line drawing and all hybridizations were carried out in probe excess. Hybrids were treated with 300 units of S1 nuclease (Boehringer) for 1.5 hours at 37°C and digestion products were denatured and separated on a 6% denaturing polyacrylamide gel. Autoradiography was for 20 days at -70°C with an intensifying screen. Lanes: (P1) an aliquot of unhybridized probe form P1; (1), (2) and (3) 5ug of EJ poly A^+ RNA and 15 ug of yeast tRNA hybridized to P1 at 59°C, 65°C and 72°C respectively; (4) 20 ug of yeast tRNA hybridized to P1 at 65°C; (P2) an aliquot of unhybridized probe form P2; (5), (6) and (7) 5 ug of EJ poly A+ RNA and 15 ug of yeast tRNA hybridized to P2 at 59°C, 65°C and 72°C respectively; (8) 20 ug of yeast tRNA hybridized to P2 at 65° C. Markers are HaeIII fragments of 0X174 (M) and HaeIII fragments of pBR322 (m) as indicated by the closed arrow-heads. The line drawing is as in Figure 1-7. Protected fragments are indicated by the open arrow-heads and are arbitrarily grouped into clusters A to E as indicated. Xh= XhoI; X= XcyI.

should preferentially digest probe artifacts, whereas the signal due to the detection of genuine DNA:RNA hybrids should only decrease linearly upon increasing the S1 concentration. In Figure 1-13, the 357 nucleotide XhoI-XcyI probe was hybridized at 65°C in 80% FHB, for 16 hours, to 10 ug EJ poly A^+ RNA and 10 ug yeast tRNA and the hybrids formed were digested with 300 units (lane 1), 600 units (lane 2) or 1000 units (lane 3) S1 nuclease. The yeast tRNA negative control (20 ug) was digested with 300 units of S1 nuclease (lane 4). All digestions were at 37°C for 90 minutes, the products were denatured and sized on a 6% denaturing polyacrylamide gel. As Figure 1-13 shows, none of the major protected species observed appear to be probe artifacts. Only the weak band of approximately 186 nucleotides and the bands below 57 nucleotides appear to be artifactual. In this experiment another cluster of protected fragments could be identified (denoted as E, corresponding to 5' RNA termini mapping between positions -115 and -110, relative to the donor splice site) and weaker fragments of approximately 84, 73 and 64 nucleotides were also detected (mapping to positions -103, -94 and -85). Thus, S1 nuclease mapping independently confirms the primer extension analysis by localizing transcriptional initiation sites in the region mapping between positions -191 to -84 bp relative to the donor splice site (Figures 1-7, 1-8, 1-10, 1-12 and 1-13).

To confirm this result with an independent S1 probe, the 195 bp MspI restriction fragment, mapping from 241 bp to 46 bp upstream from the donor splice site of exon -1, was isolated from plasmid pT24 (Figure 1-1). The 195 nucleotide MspI-MspI anti-sense strand, γ-^{32}P-ATP 5' end-labelled 46 nucleotides upstream from the donor splice site, was prepared and hybridized at 60°C or 65°C, in 80 % FHB, to 4 ug EJ poly A^+ RNA and 60 ug yeast tRNA. Hybridization to the yeast tRNA negative control (20 ug) was at 60°C. Figure 1-14 shows that S1 digestion of the hybrids formed, confirms the clusters of bands previously

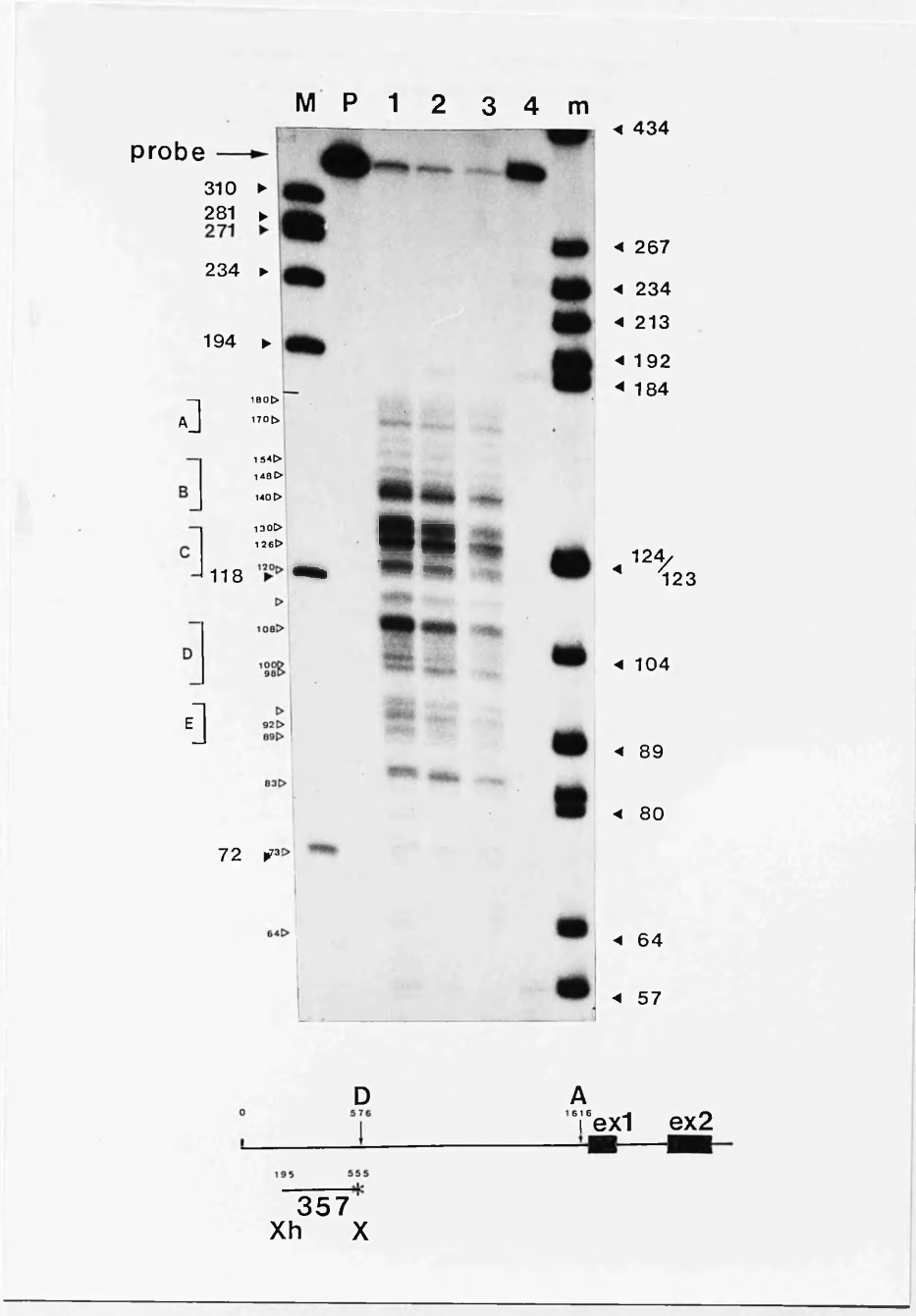

Figure 1-13 S1 mapping analysis of EJ poly A^+ RNA using the 357 nucleotide XhoI-XcyI probe with increasing amounts of S1 nuclease.

RNA was hybridized for 16 hours at 65°C to form P1 of the γ-^{32}P-ATP 5' end-labelled anti-sense 357 nucleotide XhoI-XcyI probe DNA. The probe was labelled as before (Figure 1-12) and all hybridizations were carried out in probe excess. Hybrids were treated with 300, 600 or 1000 units of S1 nuclease (Boehringer) for 1.5 hours at 37°C and digestion products were denatured and separated on a 6% denaturing polyacrylamide gel. Autoradiography was for 10 days at -70°C with an intensifying screen. Lanes: (P) an aliquot of unhybridized probe form P1; (1), (2) and (3) 10 ug of EJ poly A^+ RNA and 10 ug of yeast tRNA digested with 300, 600 and 1000 units of S1 nuclease respectively; (4) 20 ug of yeast tRNA digested with 300 units of S1 nuclease. Markers are HaeIII fragments of 0X174 (M) and HaeIII fragments of pBR322 (m) as indicated by the closed arrow-heads. The line drawing is as in Figure 1-7. Protected fragments are indicated by open arrow-heads and are arbitrarily grouped into clusters A to E as indicated. Xh= XhoI; X= XcyI.

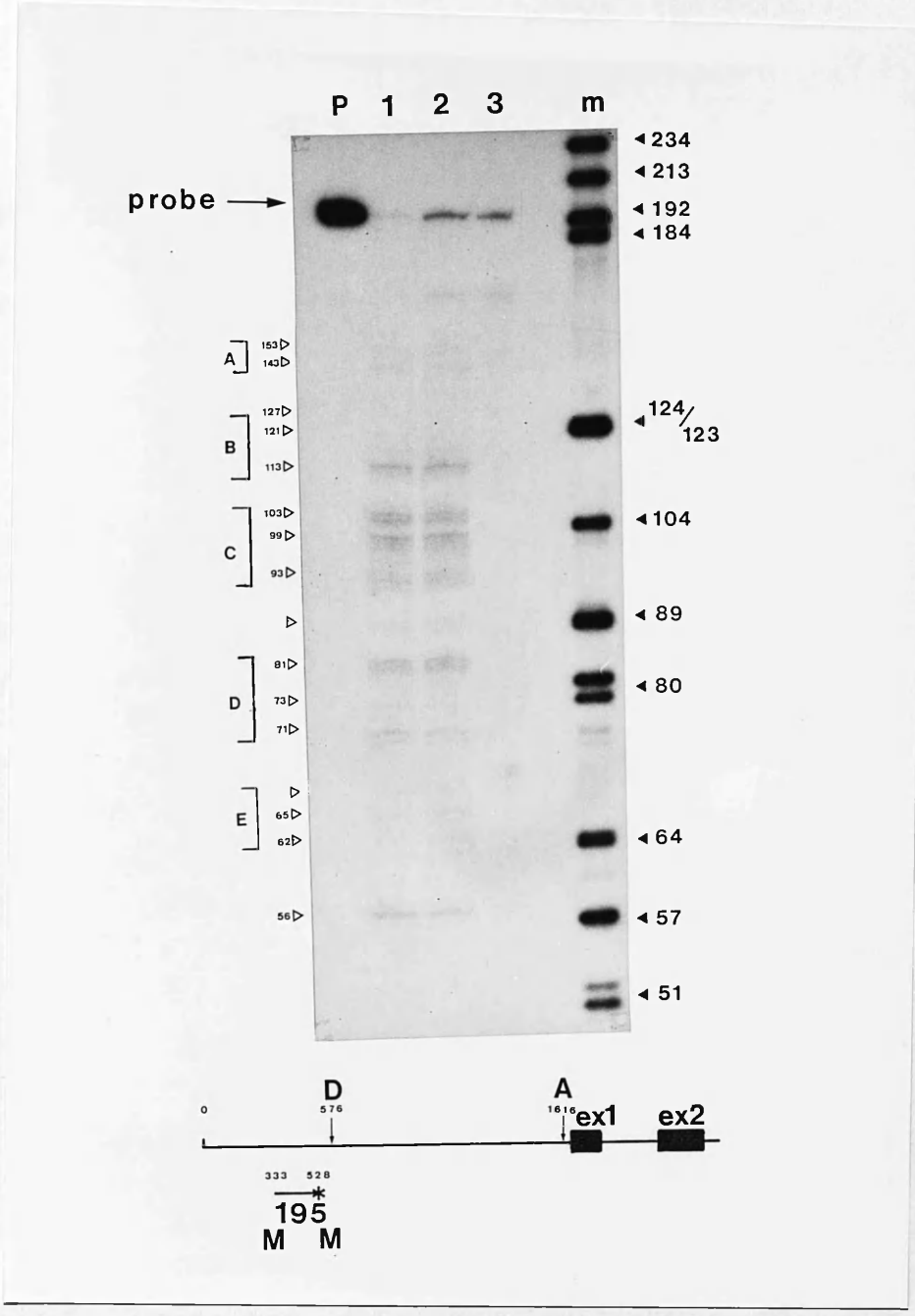

Figure 1-14 S1 mapping of EJ poly A^+ RNA using the 195 nucleotide MspI-MspI probe.

RNA was hybridized for 16 hours at 60°C or 65°C to γ-^{32}P-ATP 5' end-labelled anti-sense 195 nucleotide MspI-MspI probe DNA. The probe was labelled as indicated in the line drawing at a position 48 nucleotides upstream from the donor splice site (D) at position 576, and all hybridizations were carried out in probe excess. Hybrids were treated with 300 units of S1 nuclease (Boehringer) for 1.5 hours at 37°C and digestion products were denatured and separated on a 6% denaturing polyacrylamide gel. Autoradiography was for 12 days at -70°C with an intensifying screen. Lanes: (P) an aliquot of unhybridized probe; (1) and (2) 4 ug of EJ poly A^+ RNA and 16 ug of yeast tRNA hybridized to the probe at 60°C and 65°C respectively; (3) 20 ug of yeast tRNA hybridized to the probe at 60°C. Markers are HaeIII fragments of pBR322 (m) as indicated by the closed arrow-heads. The line drawing is as in Figure 1-7. protected fragments are indicated by the open arrow-heads and are arbitrarily grouped into clusters A to E as indicated. M= MspI.

observed with the 389 nucleotide XcyI-XcyI probe (Figure 1-10) and with the 357 nucleotide XhoI-XcyI probe (Figures 1-12 and 1-13). These five clusters are labelled A to E as before. However, protected fragments corresponding to the weak protected fragments of approximately 73 to 64 nucleotides in Figure 1-13 (mapping to positions -94 and -84), were not detected in Figure 1-14 with the 195 nucleotide MspI-MspI probe. This is presumably because of the weak signal observed in this experiment (only 4 ug of EJ poly A^+ RNA was used) and because transcripts initiating transcription at positions -94 and -84 would produce very small protected fragments with the 195 nucleotide MspI-MspI probe. In Figure 1-14 the 5' end-label is located 25 nucleotides further upstream from the exon -1 splice site compared with the 357 nucleotide XhoI-XcyI probe. Thus, the gel banding pattern is lowered by 25 nucleotides.

In summary, the S1 mapping analyses described in this section have provided evidence for multiple c-Ha-ras1 mRNA 5' termini, clustered into a broad domain located between approximately 200 to 84 bp upstream of the -1094 donor splice site (see Figure 1-13). The multiple 5' termini have been broadly grouped into five main clusters and three minor bands mapping to positions -103, -94 and -84 bp relative to the -1094 donor splice site (Figure 1-13). These S1 protected fragments correspond to extension products of between approximately 320 to 203 nucleotides in the primer extension analyses presented earlier (Figures 1-7 and 1-8). Furthermore, the S1 protected fragments in cluster A correspond closely to the extension products mapping to position -191, relative to the donor splice site, those in cluster B correspond closely to extension products mapping between positions -179 to -162, those in cluster C correspond closely to extension products mapping between positions -148 to -135 and those in clusters D and E correspond to extension products mapping between positions -126 and -106. The three S1 protected fragments mapping between positions -103 and -84 correspond to

extension products mapping between positions -101 and -86. Thus the multiple 5' termini mapping between 191 to 84 bp upstream from the donor splice site of exon -1, appear to correspond to genuine transcriptional initiation sites since they have been confirmed by two independent techniques and using multiple probes. Furthermore, it has been shown by using total RNA isolated from murine NIH 3T3 cells transfromed with EJ high molecular weight DNA and which express the human c-Ha-ras1 gene at high levels, that none of the bands observed are derived from cross-hybridization to another "ras-like" gene (Figure 1-8). The RNA mapping data is summarized schematically in Figure 1-15. Unfortunately, the 85 bp region immediately upstream of the -1094 donor splice site could not be mapped by S1 nuclease analysis with 5' end-labelled probes. Therefore, it was not possible to independently confirm the primer extension products mapping to positions -76, -61, -59, -50, -45, -40, -28, -23 and -16 relative to the donor splice site of exon -1 (Figures 1-7 and 1-8).

(e) Northern blot analysis

To further confirm the presence of an upstream exon, the independent technique of northern blot analysis was used. EJ poly A^+ RNA (5 ug) and EJ poly A^- RNA (25 ug) were electrophoresed on 1% formaldehyde gels and transferred onto nitrocellulose paper (Materials and Methods). The 591 bp Sau3A restriction fragment, mapping from positions -1079 to -1670, was nick-translated and used as the probe. As a positive control, an identical blot was probed with the nick-translated 571 bp XmaI restriction fragment, which contains sequences homologous to the whole coding region of the human c-Ha-ras1 fourth coding exon (117 bp) and 232 bp of the 3' untranslated region. This fragment was chosen because the fourth coding exon is the region of greatest divergence between the different human ras genes (Introduction, section I.b). Figure 1-16 shows

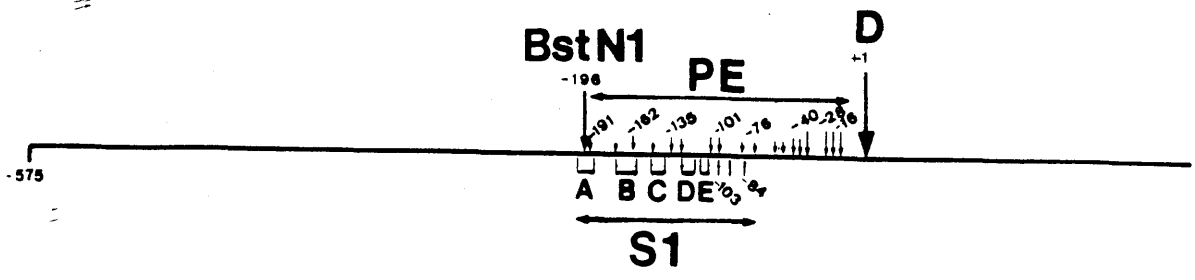

Figure 1-15 Summary of RNA mapping analysis.

A line drawing of the region upstream of the donor splice site (D), numbered as +1, which is 1094 bp 5′ to the ATG codon (not shown). The position of the BstNI site which approximately defines the 5′ limit of exon -1 is shown. Also shown are the map positions, relative to D, of the clusters of bands (A to E) detected by S1 nuclease mapping (S1). The primer extension products (PE) are indicated by the arrows. For a complete list of the S1 nuclease protected fragments, see Figure 1-13 and for the primer extension products, see Figure 1-8. The 5′ termini of the primer extension products also identified by Ishii et al. (1985), are located between positions -40 and -16.

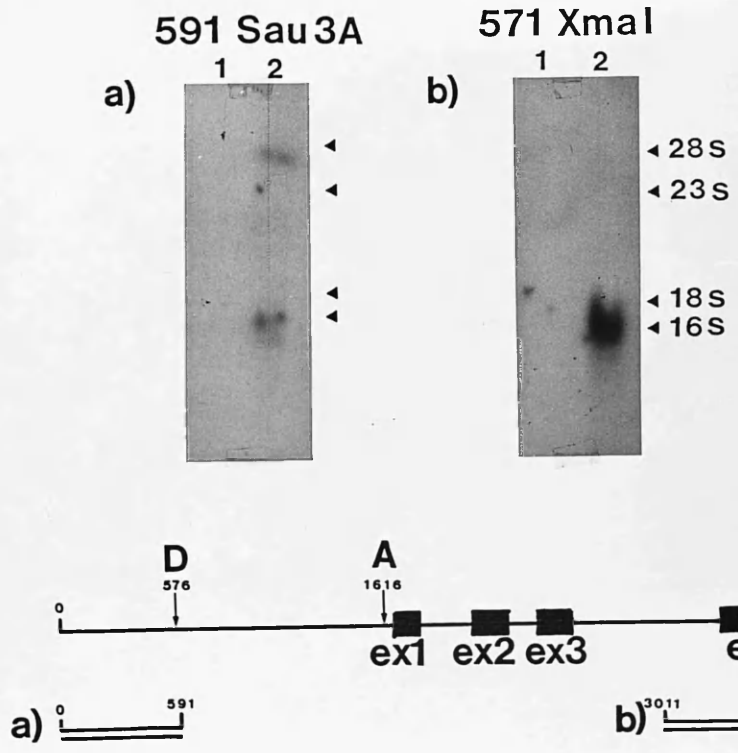

Figure 1-16 Northern blot analysis of EJ RNA using probes homologous to exon -1 and exon 4.

RNAs were electrophoresed on 1% formaldehyde agarose gels, blotted onto nitrocellulose membranes and hybridized at 42°C overnight to nick-translated probes, a) 591 bp Sau3A DNA (9.2 X 10^8 cpm/ug) and b) 571 bp XmaI DNA (9.2 X 10^8 cpm/ug). Unhybridized probe was removed by serial washing to a final stringency of 0.1X SSC; 0.1% SDS, at 65°C (Materials and Methods). Autoradiography was for 2 days at -70°C with an intensifying screen. Lanes: (1) 25 ug of EJ poly A^- RNA; (2) 5 ug of EJ poly A^+ RNA. Marker RNAs are human (28 S and 18 S) and bacterial (23 S and 16 S) ribosomal RNAs as indicated by the closed arrow-heads. The line drawing shows the positions of the probe fragments relative to the four coding exons. Numbering is as in Reddy (1983) and the consensus splice sites at positions 576 and 1616 are shown. D = donor splice, A = acceptor splice site; A^+ = polyadenylation site.

autoradiographs of both blots exposed for two days at -70° C with an intensifying screen. It is clear that both the 591 bp Sau3A and 571 bp XmaI probes, hybridize to identically sized RNA molecules of approximately 1.4 Kb. This confirms the data obtained by primer extension since it shows that in c-Ha-ras1 mRNA, sequences homologous to the upstream probe are contiguous with sequences homologous to the fourth coding exon probe. The most likely explanation for this result, is that the 591 bp Sau3A probe hybridizes to sequences in the c-Ha-ras1 mRNA corresponding to an upstream, non-coding exon for this gene. The large difference in signal detected with these two probes can only be partly explained by the two times greater specific activity of the 571 bp XmaI probe. Another explanation could be that the 571 bp XmaI probe is likely to form a larger, more stable, hybrid with c-Ha-ras1 mRNA, than does the 591 bp Sau3A probe. This is particularly true if, as is suggested by the primer extension analyses (Figure 1-7 and 1-8), the majority of c-Ha-ras1 transcripts are initiated close to the donor splice site of exon -1. In addition, it is possible that hybridization of the 591 bp Sau3A probe to c-Ha-ras1 mRNA sequences corresponding to the upstream, non-coding exon may be inhibited by the demonstrated secondary structure forming potential of this region (Figures 1-9 and 1-10). Another interesting feature of the result shown in Figure 1-16 is that the 591 bp Sau3A probe appears to hybridize to another larger (approximately 6 Kb), less abundant RNA species, whereas the 571 bp XmaI, fourth exon probe, does not. An interpretation of this feature will be discussed later (Discussion, section I).

The presence of an upstream, probably non-coding exon for the human c-Ha-ras1 gene is analogous to that of the human N-ras gene, which has recently been shown to have an upstream, non-coding exon of 235 bp (Hall and Brown, 1985). Primer extension and S1 nuclease analyses have demonstrated that the human c-Ha-ras1 upstream exon is at least 191 bp in size. To further investigate the size of the c-Ha-ras1

exon -1, the 196 bp BstN1, the 195bp MspI and the 178bp PvuII DNA restriction fragments were isolated from pT24 for northern blotting analysis (Figure 1-17). The 196 bp BstN1 fragment maps from 392 bp to 196 bp upstream from the donor splice site of exon -1 and the 195bp MspI fragment maps from 244 bp to 48 bp upstream from this splice site. The 178 bp PvuII fragment was used as a positive control, as it overlaps exon 1 by 119 bp. These three restriction fragments were designed to be of approximately equal size (the largest is only 18bp larger than the smallest) and due to their small size and G + C richness, both α--^{32}P-dGTP and α-^{32}P-dCTP were used to label each fragment, by the technique of oligo-labelling (Materials and Methods). This resulted in probes of nearly equal size, with approximately equal specific activities. These probes were then hybridized to three identical northern blots containing, 0.04 ng ~~0.4 ug~~ of the 1905 bp BamHI/XbaI DNA restriction fragment (Figure 1-17, lanes 1), 10 ug A431 poly A$^+$ RNA (lanes 2) and 10 ug EJ poly A$^+$ RNA (lanes 3). The 1905 bp BamHI/XbaI DNA restriction fragment was included for more accurate quantitation of the amount of hybridization of each of the probes to c-Ha-ras1 mRNA. Figures 1-17b and 1-17c clearly show that the 195 bp MspI probe hybridizes to the 1.4 Kb RNA species, as does the 178 bp PvuII probe, whereas the 196 bp BstNI probe does not (Figure 1-17a). This result suggests that the extreme 5' end of the human c-Ha-ras1 mRNA in both these cell lines, is likely to be located somewhere upstream of the MspI site at position 528 (-48, relative to the donor splice site of exon -1) and somewhere downstream of the BstNI site at position 380 (-196, relative to the donor splice site). Furthermore, as the 3' end of the 195 bp MspI probe is 48 bp upstream from the donor splice site (see Figure 1-17) it cannot hybridize to any of the initiation sites described by Ishii et al. (1985). Thus this result independently confirms that the region upstream of these initiation sites is transcribed. However, the signal detected with the 195bp MspI probe is

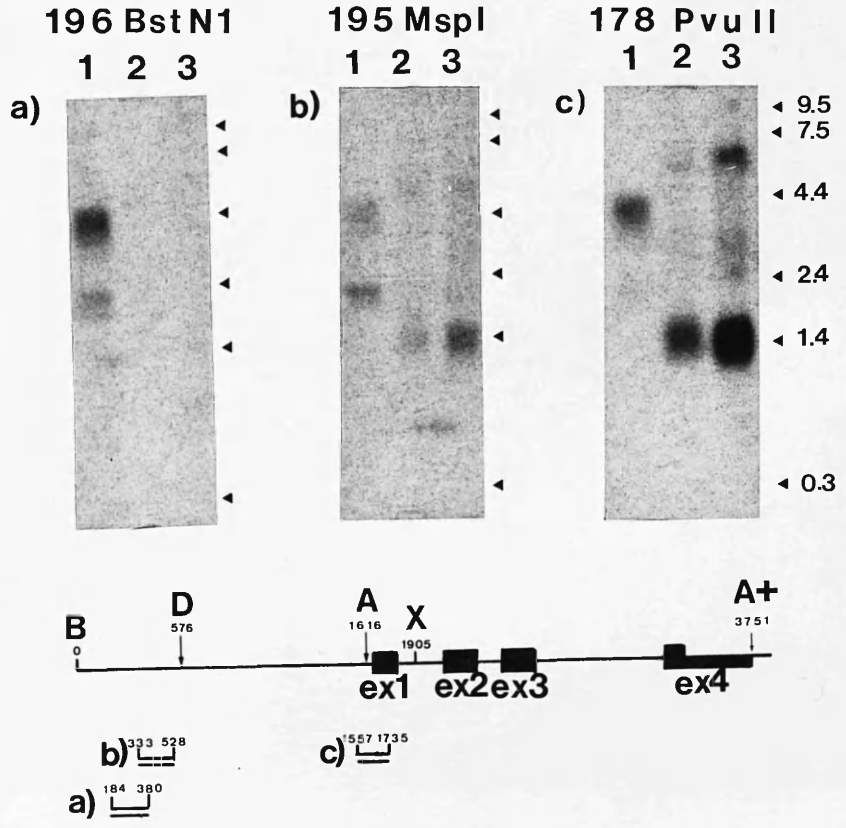

Figure 1-17 Northern blot analysis of A431 and EJ RNA using probes from the upstream exon -1 region and exon 1.

RNA and DNA controls were electrohoresed on 1% formaldehyde agarose gels, blotted onto nitrocellulose membranes, and hybridized to oligo-labelled probes, (a) 196 bp BstNI DNA (1.3 x 10^9 cpm/ug), (b) 195 bp MspI DNA (1.4 X 10^9 cpm/ug) and (c) 178 bp PvuII DNA (1.7 x 10^9 cpm/ug). Hybridization was at 42°C overnight and unhybridized probe was removed by serial washing to a final stringency of 0.1X SSC; 0.1% SDS, at 65°C (Materials and Methods). Autoradiography was for 3 days at -70°C with an intensifying screen. Lanes: (1) 0.04 ng of 1.9 Kb BamHI/XbaI DNA fragment; (2) 10 ug of A431 poly A^+ RNA; (3) 10 ug of EJ poly A^+ RNA. The line drawing is as in Figure 1-16. B = BamHI, X = XbaI.

approximately 10 fold weaker than that detected with the 178 bp Pvu II probe, which can form a hybrid of 119 bp with the c-Ha-ras1 mRNA. The low signal observed with the 195 bp MspI probe may suggest that this probe is not available for hybridization to the major initiation sites. Thus, the 5' termini of the majority of the c-Ha-ras1 transcripts may be located either completely 3' to, or quite close to, the 3' end of the 195 bp MspI probe. Alternatively, sequences in this region may not be as available for hybridization due to the extra potential of this region for forming secondary structures (Figures 1-9 and 1-10). This suggestion is supported by the decreased hybridization of the 195 bp MspI probe, relative to the 178 bp PvuII probe, to the 1905 bp BamHI/XbaI DNA control fragment, even though these probes have nearly identical specific activities (Figure 1-17). Finally, both the 195 bp MspI and the 178 bp PvuII probes also appear to detect the same less abundant RNA species of approximately 6.0 Kb (Figure 1-17), as did the 591 bp Sau3A probe in Figure 1-16.

In summary, the northern blotting analyses described above confirm the existence of an upstream, possibly non-coding exon (exon -1), 1040 bp upstream from the ATG codon of the first coding exon. Furthermore, this analysis has defined the 5' boundary of this c-Ha-ras1 upstream exon, as being close to the BstN1 site 196 bp upstream from the donor splice site (see Figure 1-17). This is in good agreement with the primer extension and S1 nuclease analyses presented earlier. Additionally, evidence was obtained for a less abundant, approximately 6.0 Kb RNA species, with sequences homologous to at least exon -1 and exon 1, but not exon 4 (Figures 1-16 and 1-17).

II. CAT ANALYSIS: DISSECTION AND FUNCTIONAL ANALYSES OF THE HUMAN c-Ha-ras1 PROMOTER REGION

Experiments described in the preceding section, using

S1 nuclease, primer extension, northern blot and sequence analyses, have located the promoter region of the human c-Ha-*ras*1 gene to a region immediately upstream of the donor splice site of exon -1, which is 1094 bp upstream of the ATG codon. It was decided to functionally confirm this promoter region by inserting the 591 bp Sau3A restriction fragment, mapping between positions 575 bp upstream and 15 bp downstream of the donor splice site of exon -1, in front of a "promoter-less" bacterial chloramphenicol acetyltransferase (CAT) gene. Promoter activity could then be measured by determining the levels of CAT activity in cells transfected with this construct. In addition, to search this promoter region for putative control regions, a series of constructs were made in which various deletion fragments covering this region were placed in front of the "promoter-less" CAT gene.

(a) Construction of CAT vectors

The human c-Ha-*ras*1 promoter region (Figure 2-1a), contained within the 591 bp Sau3A fragment was digested to yield the series of restriction fragments shown in Figure 2-1b. The 591 bp Sau3A fragment, designated fragment 1, was cloned in both orientations into the unique BamH1 site of the CAT vector, pCO (Figure 2-2). This vector is a "promoter-less" form of the vector pLW2 (Gaffney et al. 1985), obtained from Dr. Barklie Clements (Institute of Virology, Glasgow).The resulting constructs were designated pRasCAT1D and pRasCAT1R, where D and R refer respectively to the direct and reverse orientations of the c-Ha-*ras*1 fragment relative to the CAT gene. In Figure 2-2 a diagram of these constructs is given, along with the expected sizes of diagnostic digestion products, following digestion with BamHI or PstI/XhoI. The 2% agarose gel (Figure 2-2b) confirms that the clones used for CAT analysis are indeed pRasCAT1D and pRasCAT1R.

The c-Ha-*ras*1 promoter fragments 2 (390 bp SmaI) and 3

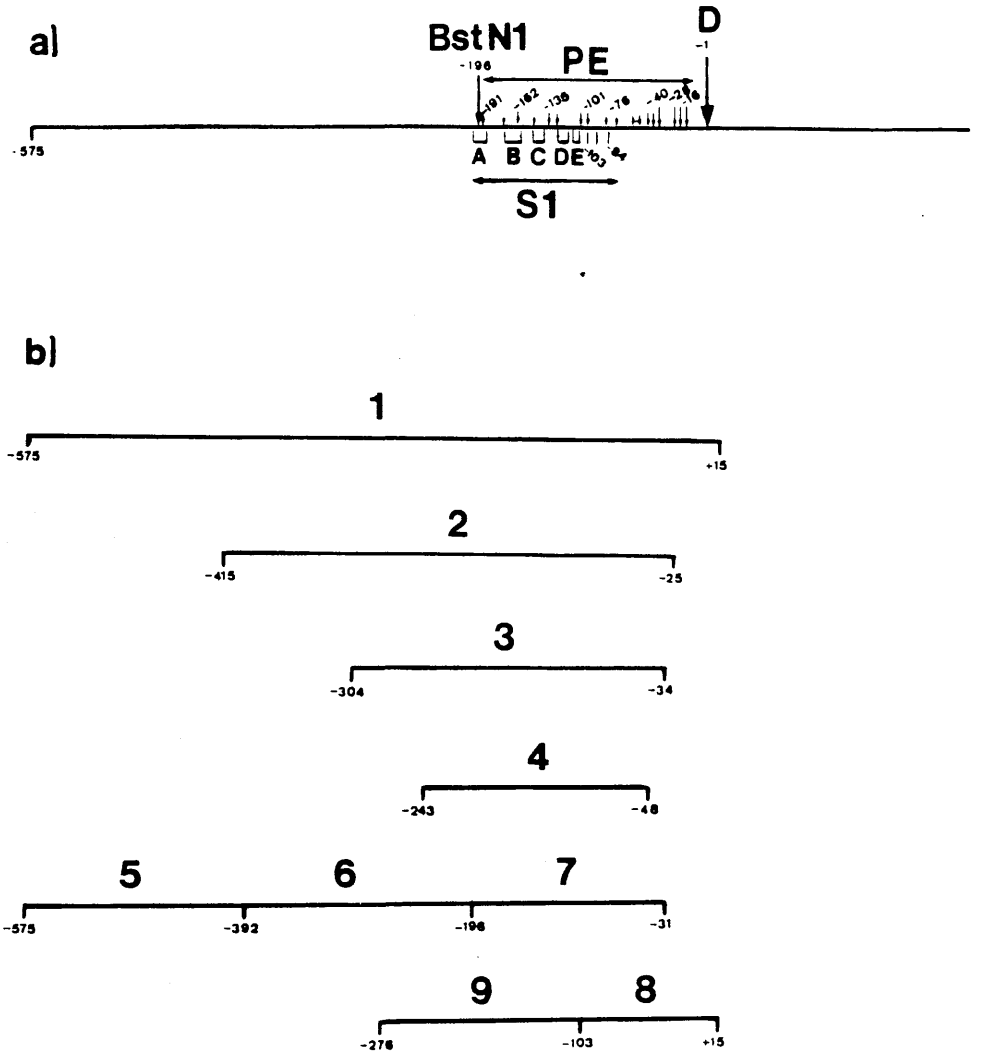

Figure 2-1 DNA fragments used in CAT constructs.

a) A line drawing of the region upstream of the exon -1 donor splice site (D), which summarizes the RNA mapping data (see Figure 1-15). b) The nine fragments used in the CAT constructs numbered relative to D. 1= 591 bp Sau3A; 2= 390 bp SmaI; 3= 270 bp MnlI; 4= 195 bp MspI; 5= 184 bp Sau3A/BstNI; 6= 196 bp BstNI; 7= 165 bp BstNI; 8= 118 bp SstII/Sau3A and 9 = 173 bp SstII.

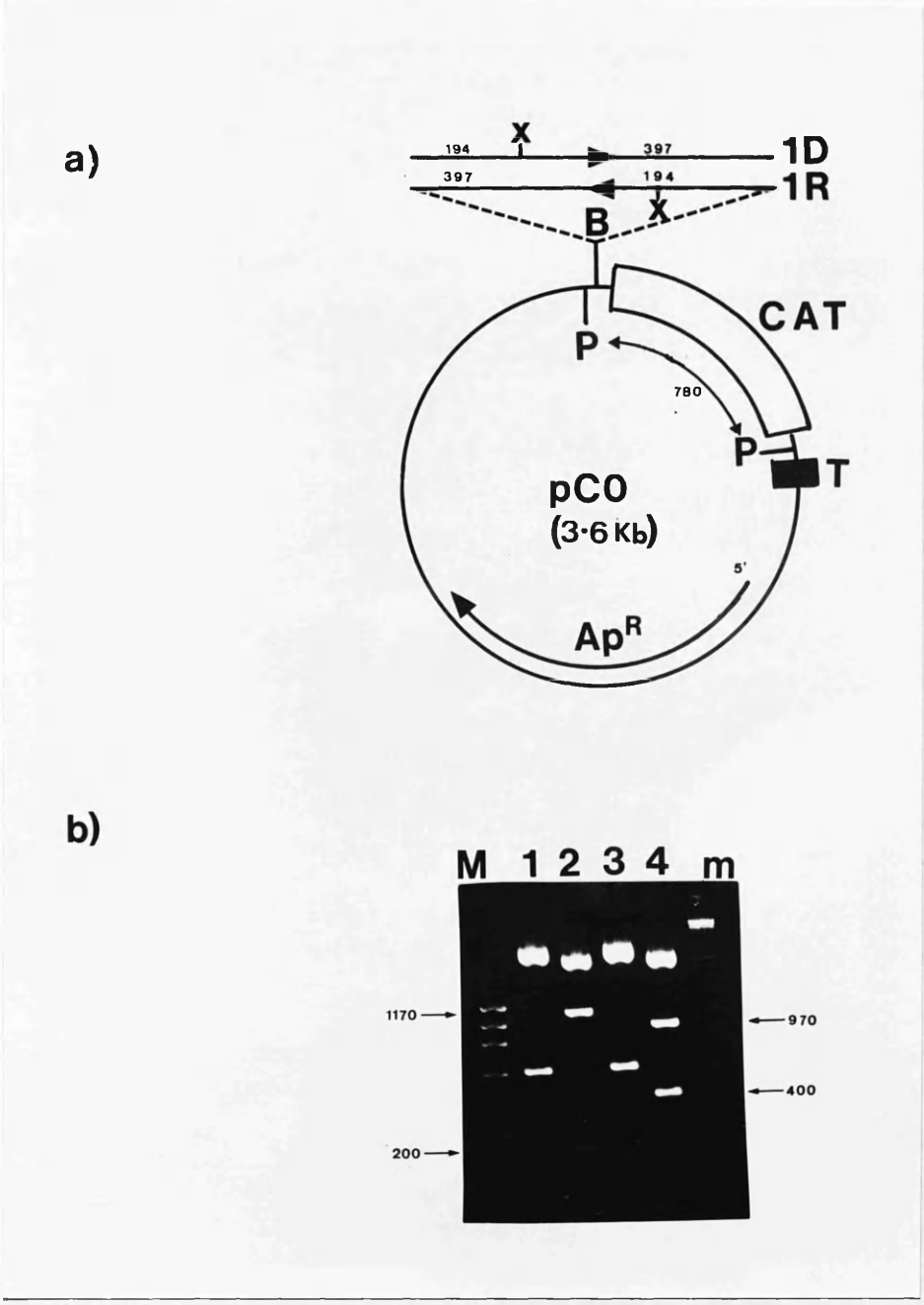

Figure 2-2 The pRasCAT1D and pRasCAT1R constructs.

a) A schematic drawing showing the sizes of fragments expected from diagnostic BamHI and PstI/XhoI digestions of pCO, into which fragment 1 (Figure 2-1) has been inserted. pCO is a derivative construct of pLW2 (Gaffney et al., 1985) in which the 200 bp BamHI fragment containing the HSV-I IE gene-4/-5 promoter has been removed. The position of the bacterial CAT gene (open box) and ampicillin (Ap^R) resistance gene are shown, as is the 100 bp PstI/XbaI terminator sequence (T, closed box), from the HSV-I IE gene-4/-5. B= BamHI, P= PstI. The short vertical lines, marked X on the insert fragments, are XhoI sites and the numbers on either side are distances from XhoI to BamHI in bp. There are no XhoI sites in pCO. b)a 2% agarose gel in which pRasCAT1D was digested with BamHI (lane 1) and PstI/XhoI (lane 2), pRasCAT1R was digested with BamHI (lane 3) and PstI/XhoI (lane 4). Marker DNAs were ØX174/HaeIII (M) and lambda/HindIII (m) fragments. Arrows indicate the expected sizes for the diagnostic digestion products.

(270 bp MnlI) were prepared for cloning into the BamHI site of pCO, by addition of BamHI linkers (see Materials and Methods). The 5' overhanging ends of fragments 4 (195 bp MspI), 5 (184 bp Sau3A/BstNI), 6 (196 bp BstNI) and 7 (165 bp BstNI) were filled in using the Klenow fragment of E.coli DNA polymerase, subsequently ligated to BamHI linkers and inserted into the BamHI site of pCO. Fragment 8 (118 bp SstII/Sau3A) had a 3' overhang at the SstII site and a 5' overhang at the Sau3A site. The 3' overhang was removed using the 3'-5' exonuclease activity of T4 DNA polymerase and the 5' overhang simultaneously filled-in using the 5'-3' polymerase activity of T4 DNA polymerase. The resulting blunt-ended fragment was ligated to BamHI linkers and inserted into the BamHI site of pCO. The 2% agarose gels in Figure 2-3b show pRasCAT2D and 2R, pRasCAT6D and 6R and pRasCAT8D and 8R, diagnostically digested with BamHI and also Pst/BglI double digested to determine orientations. As can be seen, all digests gave the correctly sized diagnostic restriction fragments. Similarly, the 2% agarose gels in Figure 2-4b show pRasCAT3R, pRasCAT4D, pRasCAT5D and 5R, and pRasCAT7D and 7R, diagnostically digested with BamHI and also PstI/SstII double digested to determine orientations. Digestion of pRasCAT5D and 5R and pRasCAT7D and 7R gave the correctly sized diagnostic restriction fragments. However, construct pRasCAT3D was obtained in a later round of cloning experiments and diagnostically digested for the direct orientation using a PstI/SstII double digestion, which released the correctly sized 40 bp fragment (Figure 2-4a). The diagnostic gel is not shown in this case. Unfortunately, subsequent diagnostic digestions of the construct shown in lanes 7 and 8 of Figure 2-4b reveal that this construct was not, in fact, construct pRasCAT4R. It was not possible in the time remaining to complete this thesis to obtain pRasCAT4R.

Finally, another series of constructs was prepared by cloning fragment 9 (the 173 bp SstII fragment, shown in

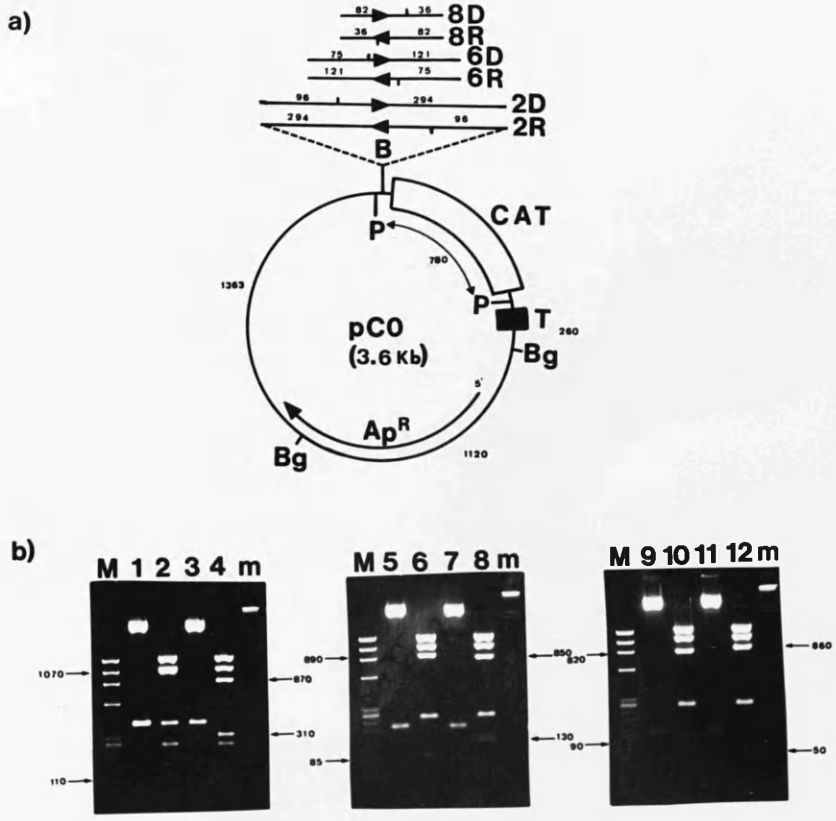

Figure 2-3 The pRasCAT2D and 2R, pRasCAT6D and 6R, and pRasCAT8D and 8R constructs.

a) A schematic drawing showing the sizes of fragments expected from diagnostic BamHI and PstI/BglI double digestions of pCO, into which fragments 2,6 and 8 (Figure 2-1) have been inserted. B= BamHI, Bg= BglI, P= PstI. The short vertical lines on the insert fragments indicate BglI sites and the numbers on either side are distances from BglI to BamHI in bp. The distances in bp between the diagnostic restriction sites are indicated on the vector. b) 2% agarose gels showing diagnostic digestions. All the odd numbered lanes are BamHI digestions and all even numbered lanes are PstI/BglI double digestions. Lanes 1 and 2 = pRasCAT2D; lanes 3 and 4 = pRasCAT2R; lanes 5 and 6 = pRasCAT6D; lanes 7 and 8 = pRasCAT6R; lanes 9 and 10 = pRasCAT8D; lanes 11 and 12 = pRasCAT8R. Marker DNAs are as in Figure 2-2. Arrows indicate the expected sizes for the diagnostic digestion products.

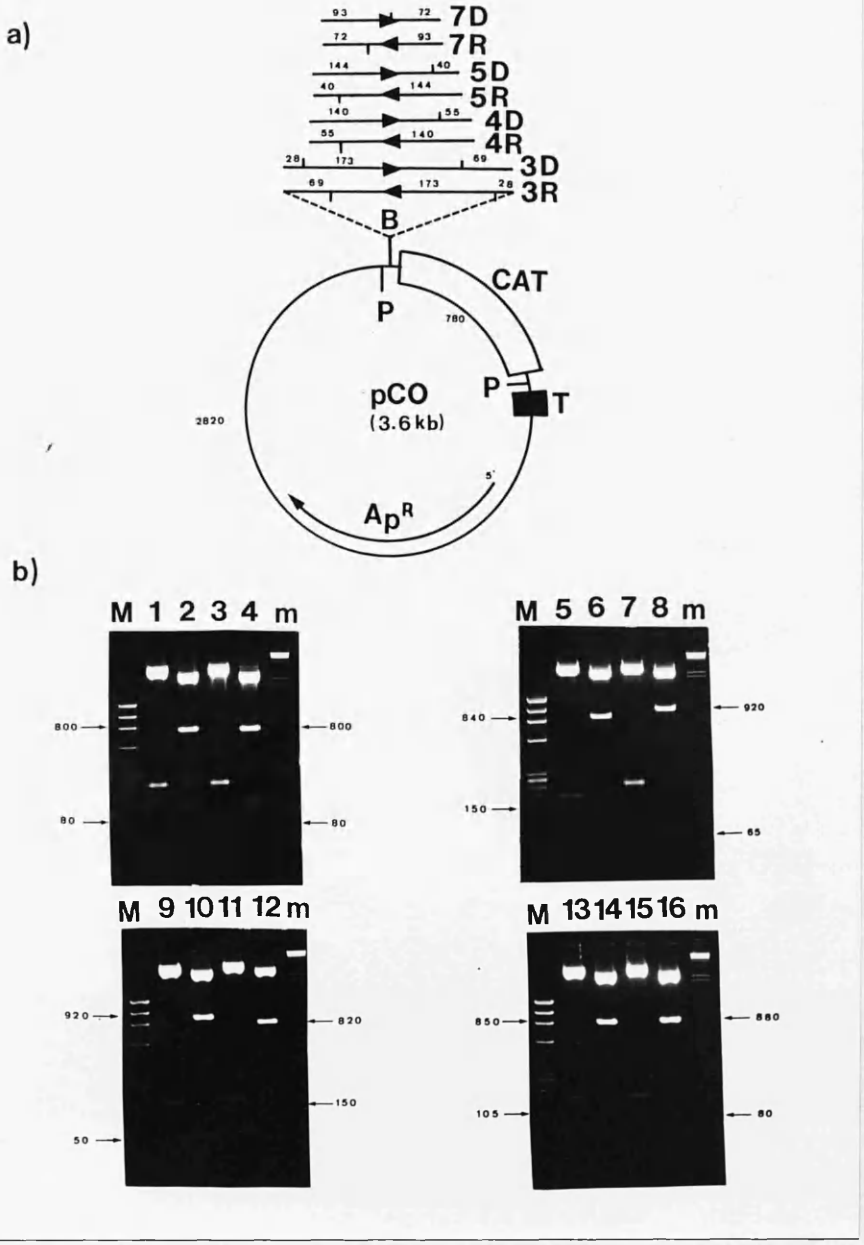

Figure 2-4 The pRasCAT3R, pRasCAT4D, pRasCAT5D and 5R, and pRasCAT7D and 7R constructs.

a) A schematic drawing showing the sizes of fragments expected from diagnostic BamHI and PstI/SstII digestions of pCO, into which fragments 3,4 5 and 7 (Figure 2-1) have been inserted. B= BamHI, P= PstI. The short vertical lines on the insert fragments indicate SstII sites and the numbers on either side are distances from SstII to BamHI in bp. There are no SstII sites in pCO. b) 2% agarose gels showing diagnostic digestions. All the odd numbered lanes are BamHI digestions and all the even numbered lanes are PstI/SstII double digestions. Lanes 1 to 4 - pRasCAT3R (in this cloning experiment, both clones isolated were pRasCAT3R); lanes 5 and 6 - pRasCAT4D; lanes 9 and 10 - pRasCAT5R ; lanes 11 and 12 - pRasCAT5D ; lanes 13 and 14 - pRasCAT7D; lanes 15 and 16 - pRasCAT7R. The construct shown in lanes 7 and 8 was not used. Marker DNAs are as in Figure 2-2. Arrows indicate the expected sizes for the diagnostic digestion products.

Figure 2-1b) into the unique HindIII site, located in the polylinker region of the plasmid pC0, immediately 5' to fragment 8 in pRasCAT8D. Fragment 9 was treated with T4 DNA polymerase to create blunt ends at the SstII site, prior to the addition of HindIII linkers (Materials and Methods).Following addition of HindIII linkers, the 173 bp SstII fragment was then ligated into the HindIII site of pRasCAT8D. Figure 2-5a shows the resulting constructs, where the 2% agarose gels in Figure 2-5b show diagnostic restriction digestions of each of the clones isolated. In lanes 1 and 5 of Figure 2-5b, BamHI digestion of these constructs confirms that fragment 8 is indeed present, and in lanes 2 and 6 the orientation of fragment 8 has been determined by PstI/BglI double digestion, as before (Figure 2-3). This confirms that fragment 8 is in the direct orientation in both constructs. In lanes 3 and 7, HindIII digestion of these constructs confirms that fragment 9 is present and in lanes 4 and 8 the orientation of fragment 9 was determined by HinfI digestion. The HinfI digestion shown in lane 4 reveals that fragment 9 is present as a dimer, arranged in a head-to-tail fashion, as a fragment of exactly the same size as the 173 bp (SstII) HindIII insert was obtained (marked with an asterisk in Figure 2-5b, lane 4). Thus, this construct was designated pRasCAT8D.9D.9D. The HinfI digestion shown in lane 8 confirms that fragment 9 is present as a monomer in the reverse orientation, therefore this construct was designated pRasCAT8D.9R. During a later round of cloning experiments, fragment 9 was successfully cloned as a monomer in the direct orientation, to give the construct pRasCAT8D.9D (data not shown). Similarily fragment 9 was also cloned into the unique HindIII site in the polyliker of pC0 during subsequent cloning experiments. This generated the constructs pRasCAT9D and pRasCAT9R, which were diagnostically digested with HindIII and orientated by HinfI digestion (data not shown).

(b) Analysis of CAT vectors

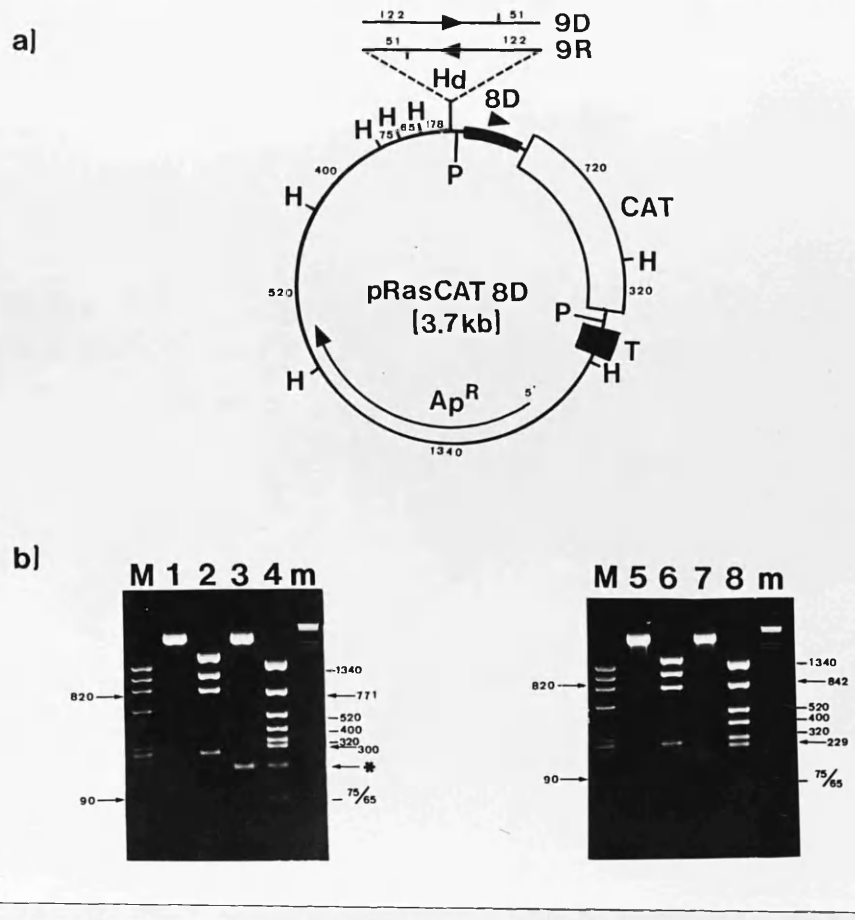

Figure 2-5 The pRasCAT8D.9D.9D and pRasCAT8D.9R constructs.

a) A schematic drawing showing the sizes of fragments expected from diagnostic digestions of pRasCAT8D (Figure 2-3), into which fragment 9 (Figure 2-1) has been inserted. H= HinfI; Hd= HindIII; P= PstI. The position of the BamHI linkered fragment 8 in the direct orientation is shown by the closed box and arrow-head. The short vertical lines on the insert fragments indicate HinfI sites and the numbers on either side are distances from HinfI to HindIII in bp. The differences in bp between the diagnostic restriction sites are indicated on the vector. b) 2% agarose gels showing diagnostic digestions. Lanes 1 and 5 are BamHI digests specific for fragment 8. Lanes 3 and 7 are HindIII digests specific for fragment 9. Lanes 2 and 6 are PstI/BglI double digests to confirm the orientation of fragment 8 (Figure 2-3). Lanes 4 and 8 are HinfI digests to determine the orientation of fragment 9. Marker DNAs are as in Figure 2-2. Arrows to the left of each gel indicate the positions of the diagnostic fragments expected for fragment 8 in the direct orientation, after PstI/BglI double digestion. Arrows to the right of each gel give the positions of the diagnostic fragments expected for both orientations of fragment 9, after HinfI digestion. The asterisk indicates the fragment arising due to tandem duplication of fragment 9. Horizontal lines indicate non-diagnostic fragments following HinfI digestion.

To functionally confirm the presence of promoter sequences in the region upstream of the donor splice site of exon -1, the constructs pRasCAT1D and pRasCAT1R (see Figure 2-2), along with pC0, the "promoter-less" version of these constructs, were introduced into exponentially growing human HeLa cells by the calcium phosphate precipitation technique (Materials and Methods). Crude cellular extracts were prepared from the transfected cells 48 hours after transfection by scraping the cells off the surface of the petri dishes and lyzing the cells by repeated freeze-thawing. After centrifugation, the supernatant obtained was used as a crude cellular extract and the CAT activity in each was assayed (Materials and Methods). An aliquot of each crude extract was incubated for 30 minutes at 37°C with acetyl coenzyme A and ^{14}C labelled chloramphenicol. The acetylated products were separated from unacetylated ^{14}C-chloramphenicol by ascending chromatography on SIL-G, silica gel thin layer chromatography (TLC) plates in 95% chloroform: 5% methanol. The TLC plates were air-dried and autoradiographed at room temperature overnight. An example of such an analysis is shown in Figure 2-6. Cells transfected with pC0 (HeLa/pC0) contained undetectable CAT activity, whereas HeLa/pRasCAT1D cells contained strong CAT activity. This CAT activity was equivalent to the activity observed in HeLa cells following transfection with a construct containing the strong promoter from the Herpes Simplex Virus type I (HSV-I) immediate early gene-4/-5 (IE gene-4/-5), cloned into the BamHI site in the polylinker of pC0 (Dr. Maggi Allan, personal communication). Unexpectedly, HeLa cells transfected with pRasCAT1R also produced similar levels of CAT activity. Thus, the human c-Ha-ras1 promoter region, contained in the 591 bp Sau3A restriction fragment, mapping from positions -575 to +15 relative to the exon -1 donor splice site, can function as a strong promoter directing CAT gene expression in HeLa cells regardless of its

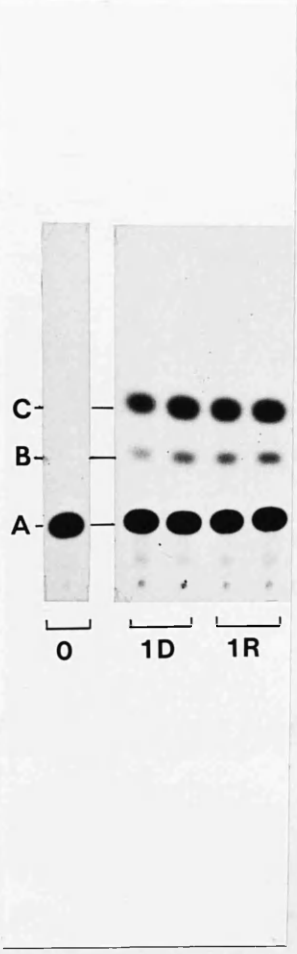

Figure 2-6 A representative autoradiograph of CAT assays performed after transient expression of pC0, pRasCAT1D and pRasCAT1R in HeLa cells.

HeLa cells at mid-log stage were transfected with 40ug of plasmid DNA per 80 cm^2 tissue culture dish. 48 hours after transfection, cell lysates were prepared and assayed for CAT activity (Materials and Methods). The letters A, B and C indicate the positions of the unacetylated ^{14}C-chloramphenicol, the 1-acetylchloramphenicol and the 3-chloramphenicol, respectively. 0 = pC0; 1D = pRasCAT1D and 1R = pRasCAT1R.

orientation. These data suggest that the human c-Ha-ras1 promoter may function as a bi-directional promoter in vivo. Bi-directional promoter regions have been described in other systems, such as the SV40 300bp control region (see Introduction, section II) and similarities between these and the c-Ha-ras1 promoter region will be discussed in a later section. Furthermore, this finding also provides strong functional confirmation of the RNA mapping analyses presented in the preceding section which suggested that human c-Ha-ras1 promoter sequences are located at least 200 bp upstream of the exon -1 donor splice site. To further localize the sequences within the 591 bp Sau3A fragment possessing promoter activity, and to investigate the regulatory sequences required for full promoter activity, the series of CAT deletion constructs shown in Figures 2-3 and 2-4 was used in transient expression assays in HeLa cells.

(1) Mutational analysis of the "direct-orientation" c-Ha-ras1 promoter

The 591 bp Sau3A c-Ha-ras1 promoter fragment was digested to provide nine fragments, numbered 1 to 9 as shown in Figure 2-1, and these fragments were inserted, in both orientations, into the unique BamHI site in the polylinker of the "promoter-less" CAT construct (pCO) as described earlier (Figures 2-3 and 2-4). This series of constructs was transiently introduced into exponentially growing HeLa cells by the calcium phosphate precipitation technique and CAT assays performed as previously described. Figure 2-7 shows representative autoradiographs of some assays after approximately 16 hours exposure at room temperature. Figure 2-8 summarizes data derived from all the CAT constructs and includes a line drawing which shows the positions of the initiation sites detected by S1 mapping and primer extension analyses (sections, I.c and

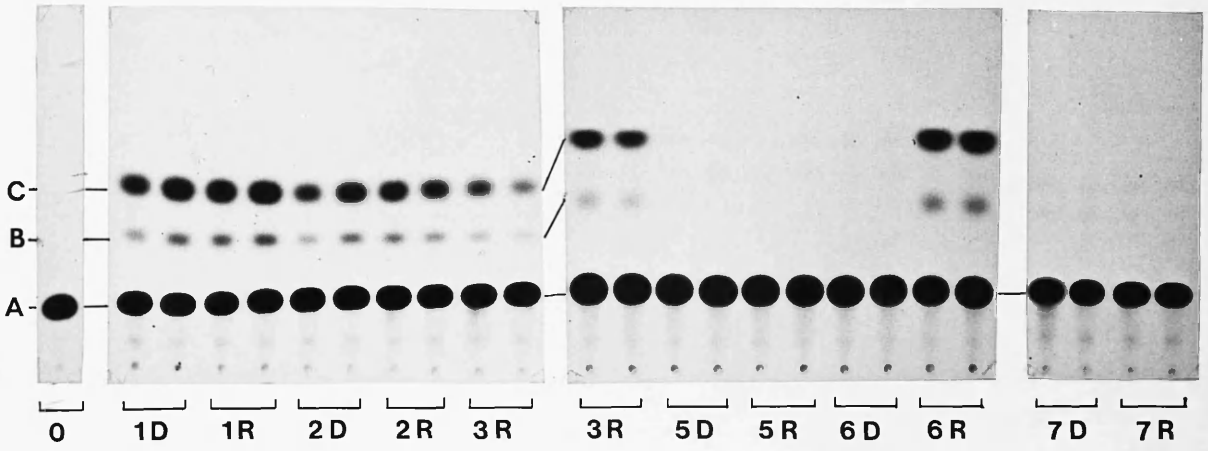

Figure 2-7 A representative autoradiograph of CAT assays performed after transient expression of pRasCAT constructs in HeLa cells.

HeLa cells at mid-log stage were transfected with 40 ug of plasmid DNA per 80 cm^2 tissue culture dish. 48 hours after transfection cell lysates were prepared and assayed for CAT activity (Materials and Methods). The letters A, B and C indicate the positions of the unacetylated ^{14}C-chloramphenicol, the 1-acetylchloramphenicol and the 3-acetylchloramphenicol, respectively. 0 = pC0 and the numbers 1, 2, 3, 5, 6 and 7 refer to fragments used in the construction of the pRasCAT series of constructs, where D or R refers to the direct or reverse orientations respectively.

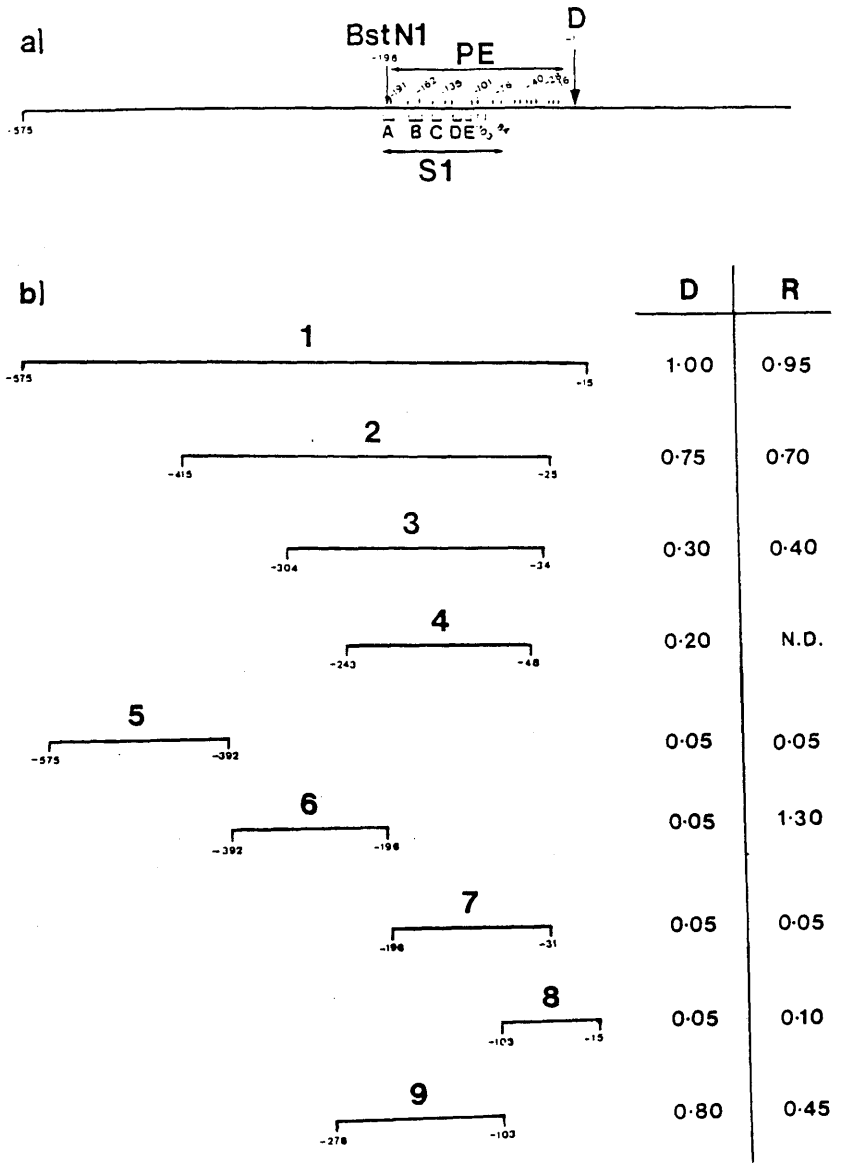

Figure 2-8 Summary of CAT data.

a) A line drawing of the region upstream of the exon-1 donor splice site (D), which summarizes the RNA mapping data (see Figure 1-15). b) The nine fragments used in the CAT constructs listed alongside a table which gives numbers representing conversions of ^{14}C-chloramphenicol, into its acetylated products. These conversions have been normalized against the conversion observed for pRasCAT1D, which has been given a value of 1.00. These results were obtained from several experiments in each case.

I.d). In all cases, several repetitions of each assay, using two independent plasmid DNA preparations of each construct, were used. The numbers represent average conversions of ^{14}C chloramphenicol into its acetylated products and have been normalized against the conversion observed for pRasCAT1D, which is given a value of 1.00. As can be seen in Figure 2-8, pRasCAT2D, which contains all the transcriptional initiation sites identified in the preceding section, except the two nearest the donor splice site at positions -16 and -23 relative to the donor splice site of exon -1 (Figure 2-8b), produced approximately 75% of the CAT activity observed with pRasCAT1D. Although this slight reduction in CAT activity is likely to be within the bounds of experimental error, an alternative interpretation might be the loss of the first two initiation sites which correspond to strong bands in the primer extension analyses presented earlier (Figures 1-7 and 1-8). Another explanation might be the absence of positive regulatory elements upstream of position -415. The construct pRasCAT3D produces approximately 30% of the CAT activity observed with pRasCAT1D (Figure 2-8b), but does not contain the first three transcriptional initiation sites at -16, -23 and -28, which correspond to strong bands in the primer extension analyses. However, it is still not clear whether this lower CAT activity is caused by the absence of 5' flanking sequences upstream of position -304 or whether this reduction is caused by the loss of three strong transcriptional initiation sites. The construct pRasCAT4D, which does not contain the first five initiation sites detected by primer extension analyses at positions -16 to -45 (Figures 1-7 and 1-8), produces approximately 20% of the CAT activity of pRasCAT1D (Figure 2-8b). Once again this result could be explained either by loss of initiation sites or by the loss of regulatory sequences. However, it is clear that the initiation sites upstream of position -48 (that is, upstream of those previously suggested by Ishii et al., (1985) to be the only initiation

sites), must be functionally active and account for at least 20% of the c-Ha-ras1 transcripts. Conversely, as pRasCAT4D does not contain the first four initiation sites detected as strong primer extension products in Figures 1-7 and 1-8, these data could suggest that these four initiation sites may indeed account for the majority of the c-Ha-ras1 transcripts. However, as will be shown, preliminary data obtained with construct, pRasCAT9D, suggest that it is more likely that the low signal obtained with pRasCAT4D results from the loss of regulatory, rather than transcriptional initiation sites.

Constructs pRasCAT5D and pRasCAT6D, contain sequences entirely upstream of the initiation sites identified in the preceding section, and both of these constructs produced similar CAT activities to the "promoter-less" pC0 (Figure 2-6 and 2-9). This result strongly suggests that there are no transcriptional initiation sites upstream of position -196, in agreement with the RNA mapping data. Interestingly, pRasCAT7D which contains all the transcriptional initiation sites upstream of position -31, including the strong site at position -40 (that is, it lacks only the three initiation sites nearest the donor splice site and is therefore, equivalent in this respect to pRasCAT3D) produced no more CAT activity than construct pC0. Thus, important promoter regulatory sequences located between position -196 and -304, appear to have been deleted in this construct. Furthermore, pRasCAT4D does not contain any of the strong initiation sites between positions -16 and -40, but nevertheless produced approximately 20% of the CAT activity observed with pRasCAT1D, which is similar to that observed with pRasCAT3D. This observation strongly suggests that the positively acting regulatory sequences located upstream of position -196 mainly reside within the 47 bp of sequence between positions -196 and -243 (Figure 2-8b). Examination of the sequence of this 47 bp region (Capon et al., 1983a, Reddy 1983) revealed that it contains a CAAT box and a Sp1 binding site (see Discussion, section

II and Figure 5-1). It is possible that these sequence motifs, known to have regulatory function (see Introduction, section II), may be the positively acting regulatory elements responsible for the CAT activity of pRasCAT3D and pRasCAT4D. However, additional sequence elements, other than the transcriptional initiation sites these constructs lack, may be required to provide CAT activity comparable to pRasCAT1D. A similar result to that obtained with pRasCAT7D was observed with pRasCAT8D; that is, pRasCAT8D produced similar CAT activity to pCO. This construct contains all of the initiation sites up to position -103, therefore it appears that important promoter regulatory sequences have also been deleted in this construct.

Finally, the construct pRasCAT9D contains only the transcriptional initiation sites detected upstream of position -103, that is, all of the clusters of initiation sites detected by S1 mapping analyses and the weak primer extension products greater than 220 nucleotides observed in Figures 1-7 and 1-8 (see also Figure 2-8a). This construct produced approximately 80% of the CAT activity observed for pRasCAT1D. However, the data obtained with prasCAT9D is preliminary, in that it was obtained with only one DNA preparation, and must be interpreted cautiously until it has been confirmed. Nevertheless, since pRasCAT7D has the same transcriptional initiation sites as pRasCAT3D, but produces no more CAT activity than the "promoter-less" construct pC0, this result suggests that positively acting regulatory element(s) are localized between positions -196 and -276 (Figure 2-8b). Furthermore, as pRasCAT9D produced greater CAT activity than pRasCAT4D, it is likely that the 47 bp sequence, between positions -196 and -243 within pRasCAT4D, and which has been identified as a positive regulatory element, requires additional sequence information between positions -243 to -276 for increased CAT activity. In addition, as pRasCAT9D produces significantly greater CAT activity than pRasCAT3D (although

pRasCAT3D contains 28 bp more 5' and 79 bp more 3' c-Ha-ras1 promoter sequence information). This suggests that pRasCAT3D (and pRasCAT4D) may also contain negatively acting regulatory elements. This/these negatively acting regulatory element(s) is/are likely to be located between positions -34 and -103, as constructs pRasCAT3D and pRasCAT7D differ in sequence information only at their 5' ends, a region shown to have positively acting regulatory activity. However, confirmation of this putative negative regulatory element and also the putative positive regulatory element located between positions -276 to -243, awaits confirmation of the result obtained with pRasCAT9D.

(2) Reconstruction of the "direct-orientation" c-Ha-ras1 promoter

As pRasCAT8D was known from the previous analysis to produce negligible CAT activity (Figure 2-8b), it was decided to attempt to restore promoter activity to this construct by addition of fragment 9. This fragment was inserted in both orientations into the HindIII site immediately upstream of fragment 8 in pRasCAT8D and the resulting constructs (Figure 2-5, data for pRasCAT8D.9D not shown) were then transfected into HeLa cells and CAT assays performed as before. Figure 2-9 summarises the data obtained in these experiments using two independent plasmid DNA preparations of each construct, except in the case of pRasCAT8D.9D in which only one plasmid DNA preparation was used. The construct pRasCAT8D.9D produced approximately 40% of the CAT activity observed with pRasCAT1D and approximately 40 fold more activity than pRasCAT8D alone. Thus, fragment 9 has to a large extent restored promoter function to pRasCAT8D, confirming that there are positively acting regulatory elements within fragment 9. However, as fragment 9 on its own in construct pRasCAT9D produced approximately 80% of the CAT activity observed with pRasCAT1D, this result also supports the earlier suggestion

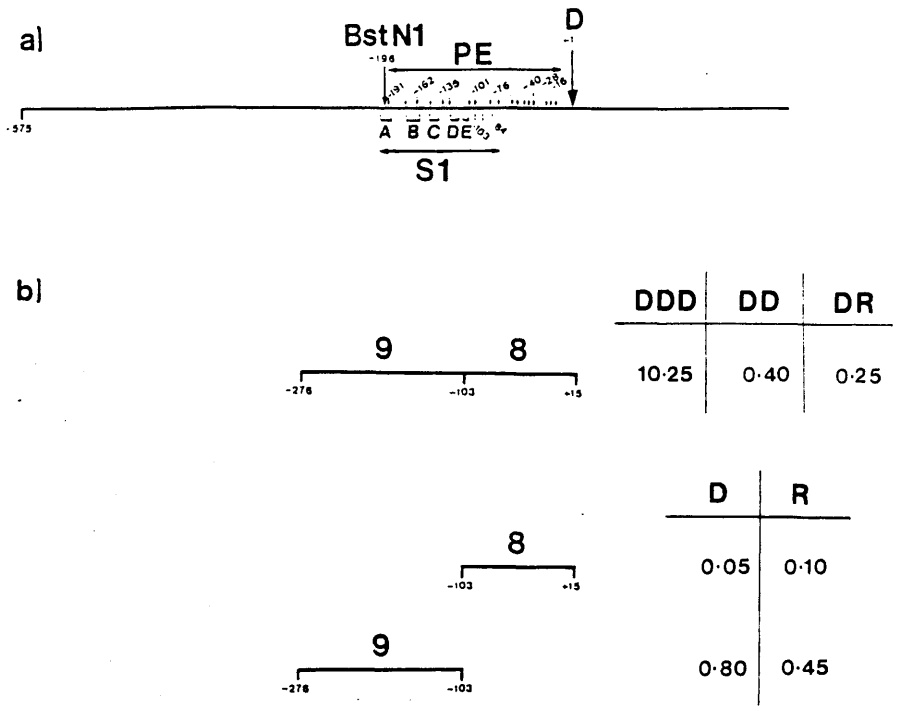

Figure 2-9 Summary of CAT data.

a) A line drawing of the region upstream of the exon-1 donor splice site (D), which summarizes the RNA mapping data (see Figure 1-15). b) Fragments 8 and 9 used in the CAT constructs either on their own (pRasCAT8D, pRasCAT8R, pRasCAT9D and pRasCAT9R), or covalently linked (pRasCAT8D.9D.9D, pRasCAT8D.9D and pRasCAT8D.9R), are shown alongside a table which gives numbers representing conversions of ^{14}C-chloramphenicol into its acetylated products, normalized against the conversion observed for pRasCAT1D, which has been given a value of 1.00.

that sequences between -34 and -103 act in this assay as a negative regulatory element. Interestingly, this 69 bp sequence has a 90% G + C content and contains 5 repeats of the sequence 5'$G(C)_n$, where n= 4 or 5, four of which are tandemly repeated (Figure 1-11 shows the sequence of the anti-sense strand).

Fortuitously, a construct was obtained in which fragment 9 was present as a tandem duplication in the direct orientaion, pRasCAT8D.9D.9D (Figure 2-5). This construct produced approximately 10 fold more CAT activity than pRasCAT1D, approximately 25 fold more CAT activity than pRasCAT8D.9D and approximately 1000 fold more CAT activity than pRasCAT8D. In fact, pRasCAT8D.9D.9D produced 10 fold greater CAT activity than the strong promoter from the HSV-I IE gene-4/-5, cloned into pC0 (Dr. Maggi Allan, personal communication). As pRasCAT8D.9D.9D has duplicated the positively regulatory elements between positions -196 and -243, and also all the transcriptional initiation sites located upstream of position -103, it is likely that duplication of these important sequences is responsible for the elevated CAT activity of this construct. This result also suggests that fortuitous duplications in this region of the human c-Ha-ras1 promoter, in vivo, would have the potential to increase c-Ha-ras1 promoter function to ten times its normal level. Interestingly, the construct pRasCAT8D.9R produced approximately 25% of the CAT activity observed with pRasCAT1D, which is just over half the CAT activity observed for pRasCAT8D.9D. This result suggests that the regulatory elements present within fragment 9 can operate in either orientation. The bi-directional nature of the c-Ha-ras1 promoter region will be more fully discussed in the next section.

(3) Mutational analyses of the "reverse-orientation" c-Ha-ras1 promoter

Promoter fragments 2 to 9 (Figure 2-1) were also

obtained in the reverse orientation relative to the CAT gene. Thus the sequences within pRasCAT1R responsible for the reverse orientation CAT gene expression (Figure 2-6) could be more accurately localized. Using pRasCAT2R, approximately 70% of the CAT activity obtained with pRasCAT1R was observed (Figure 2-6 and 2-8), suggesting that most of the sequences required for the "reverse-promoter" activity are located within fragment 2. The construct pRasCAT3R produced approximately 40% of the CAT activity obtained with pRasCAT1R (Figure 2-6 and 2-9). While a less than 2 fold difference in CAT activity may be within the bounds of experimental error, this result could suggest that sequences located between positions -415 and -304 may either positively regulate the "reverse-orientation" promoter activity, or may provide "reverse-orientation" transcriptional initiation sites. Unfortunately construct pRasCAT4R was not obtained, since this construct would have allowed a more accurate location of the "reverse-promoter" activity. The almost undetectable CAT activity obtained with with pRasCAT5R, pRasCAT7R and pRasCAT8R (Figures 2-6 and 2-8), suggest that the "reverse-promoter" activity must be located between positions -392 and -196. Indeed, pRasCAT6R, containing sequences between positions -392 and -196 directs levels of CAT activity comparable with pRasCAT1R. Furthermore pRasCAT9R produces approximately 45% of the CAT activity observed for pRasCAT1R (which is similar to the level of CAT activity produced by pRasCAT3R). This result suggests that sequences located between positions -196 and -276 also provide positively acting sequence information for the "reverse-orientation" promoter, as well as for the "direct-orientation" promoter. However, as pRasCAT3R and pRasCAT9R provide less than half the CAT activity observed for pRasCAT1R and pRasCAT6R, it is likely that additional sequences upstream of position -304 are required for full "reverse-orientation" promoter activity. The analyses presented above using a range of mutants have located the

region required for full "reverse- orientation" promoter activity to the region between -196 to -392 (Figure 2-8b). At least some of this activity has been further located at the region between positions -196 and -276. Furthermore, as pRasCAT6D produces only negligible CAT activity, this result suggests that the "reverse-promoter" activity can be physically uncoupled from the "direct-promoter" activity. This is presumably because fragment 6 does not provide any transcriptional initiation sites for the "direct-orienation" promoter, which further supports the RNA mapping analyses of section I. Since no RNA analyses which detect opposite strand transcripts have been obtained, it is not possible at this point to dissect the relative importance of transcriptional initiation sites, as opposed to regulatory elements, in these constructs. The relevance of the data presented in this thesis, which identifies regulatory sequences important for regulation of the activity of the human c-Ha-ras1 gene, will be discussed in terms of known regulatory elements. These data are summarized in Figure 5-1.

III. THE EFFECT OF SV40 REGULATORY SEQUENCES ON TRANSCRIPTION OF THE c-Ha-ras1 GENE

It has been demonstrated that for many genes encoding "luxury-function" proteins, the presence of a covalently linked active enhancer increases transcription of the linked gene (de Villiers et al., 1982; Wasylyk et al., 1983; Allan et al., 1984). However, it is not known if enhancers can increase transcription of genes like the human c-Ha-ras1 gene which does not contain a functional TATA box and, as has been shown in this thesis, contains a large number of transcriptional initiation sites, rather than a single, discrete site. Furthermore, if transcription of the human c-Ha-ras1 gene is increased by a linked

enhancer, it is not known if the enhancer affects transcription from all of the transcriptional initiation sites equally, or if the enhancer effect is specific to a subset of the transcriptional initiation sites. Such a differential enhancer effect on transcriptional initiation sites, within the same gene, has been observed with the human ε-globin gene when it is linked to the SV40 enhancer (Allan et al., 1984). Transcription from the major initiation site of the ε-globin gene was increased, but transcription from the initiation site at position -200 relative to the ε-globin ATG codon, was unaffected. Interestingly, it is possible that increased transcription of the c-Ha-ras1 proto-oncogene, by activation of a nearby enhancer or by fortuitous insertion of an enhancer, is a potential event involved in oncogenic activation of this gene in vivo (see Introduction, section I.c and Discussion).

To investigate the effect of the SV40 enhancer sequence on the expression of the human c-Ha-ras1 gene, the 6.4 Kb BamHI restriction fragment of construct pT24 (Figure 1-1), was cloned into the unique BamHI sites of the SV40 vectors, pSVOD (Mellon et al., 1981) and pSVED (Proudfoot et al., 1984). The vector pSVOD contains pBR322 sequences, from which the sequences inhibitory to replication in mammalian cells have been removed (Luskey and Botchan, 1981), and the 311 bp EcoRII G fragment, of the SV40 genome (Figure 3-1). This 311 bp EcoRII fragment contains 55 bp of one of the SV40 72 bp repeats, all three 21 bp repeats, the TATA-like sequence, the early RNA initiation sites and approximately 120 bp of sequence homologous to early RNAs (Introduction, II.a and Figure 3-1). Therefore, pSVOD contains the SV40 origin of replication, but not a functional SV40 enhancer (McKnight and Tjian, 1986). The vector pSVED contains the 342 bp PvuII/HindIII fragment, spanning the complete SV40 300 bp control region and essentially differs from pSVOD only in the possession of two full copies of the 72 bp enhancer repeats (Figure 3-1).

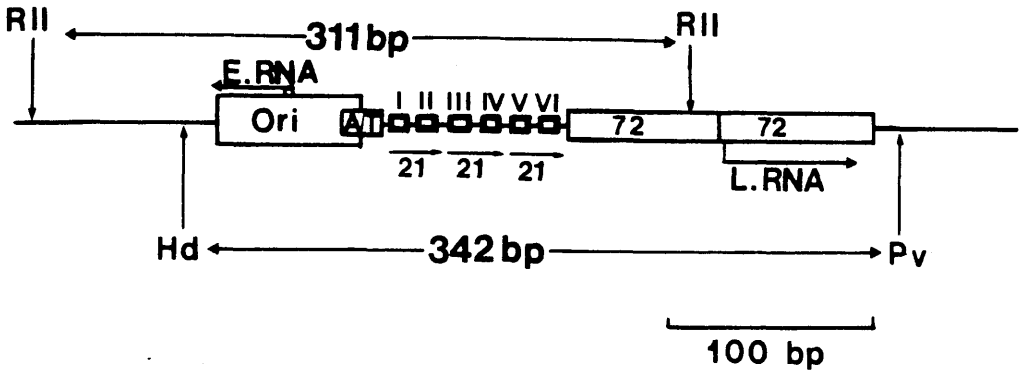

Figure 3-1 Fragments from the SV40 control region used in the pSVOD and pSVED plasmids.

The various *cis* acting control sequences of the SV40 300 bp control region are shown schematically. The origin (ori), "TATA-like" sequence (A/T) and the 72 bp repeats (72) are indicated by open rectangles. The GC boxes of the 21 bp repeats are numbered I to VI, where the horizontal arrows indicate the three 21 bp repeats. The directions and the 5′ ends of both the early (E.RNA) and late (L.RNA) transcripts are also indicated, as are the 311 bp EcoRII and 342 bp HindIII/PvuII fragments used in the plasmids pSVOD and pSVED, respectively. RII = EcoRII; Hd = HindIII; Pv = PvuII (Adapted from McKnight and Tjian, 1986).

These constructs, pSVODras1D and pSVEDras1D, containing the c-Ha-ras1 6.4 Kb BamHI fragment at their unique BamHI sites, are shown in Figure 3-2. Since these constructs contain the SV40 origin of replication, but not T antigen, replication can occur only in Cos7 cells, in which T antigen is endogeneously supplied (Gluzman, 1981). Therefore these constructs were introduced into exponentially growing Cos7 cells, by the calcium phosphate precipitation technique. Total RNA was prepared 48 hours after transfection and c-Ha-ras1 specific mRNA molecules were primer extended, using the 51 AluI-PvuII primer as before (Results, I.c) This single-stranded γ-^{32}P-ATP 5' end-labelled primer, mapping from position +68 to +119, relative to the ATG codon, was hybridized at 52^{o}C for 16 hours, in 80% FHB, to 50 ug of total RNAs isolated from pSV0Dras1D/Cos7 and pSVEDras1D/Cos7 transfected cells and untransfected Cos7 cells. The resulting hybrids were treated with reverse transcriptase as before (Materials and Methods) and the extended products denatured and sized on an 8% denaturing polyacrylamide gel. As can be seen from Figure 3-3, the pattern of extension products is essentially identical to the pattern previously observed with EJ, EJ focus 8.3.5 and A431 cell RNAs (Figures 1-7 and 1-8). A comparison of lanes 1 and 2 shows that the presence of the SV40 enhancer element, present in construct pSVEDras1D (lane 2), enhances c-Ha-ras1 transcription approximately 20 fold more than when the functional enhancer is absent, as in construct pSVODras1D (lane 1). Furthermore, the enhancer effect appears to increase transcription from each initiation site proportionately and cannot therefore be specific to a subset of the initiation sites used. It is also noteworthy that the SV40 enhancer has only an approximately 20 fold increase in transcription with the c-Ha-ras1 gene, whereas with globin genes the corresponding increases are 50 - 200 fold (Allan et al., 1984). Unfortunately, the reverse orientation of the c-Ha-ras1 gene relative to the SV40 regulatory sequences

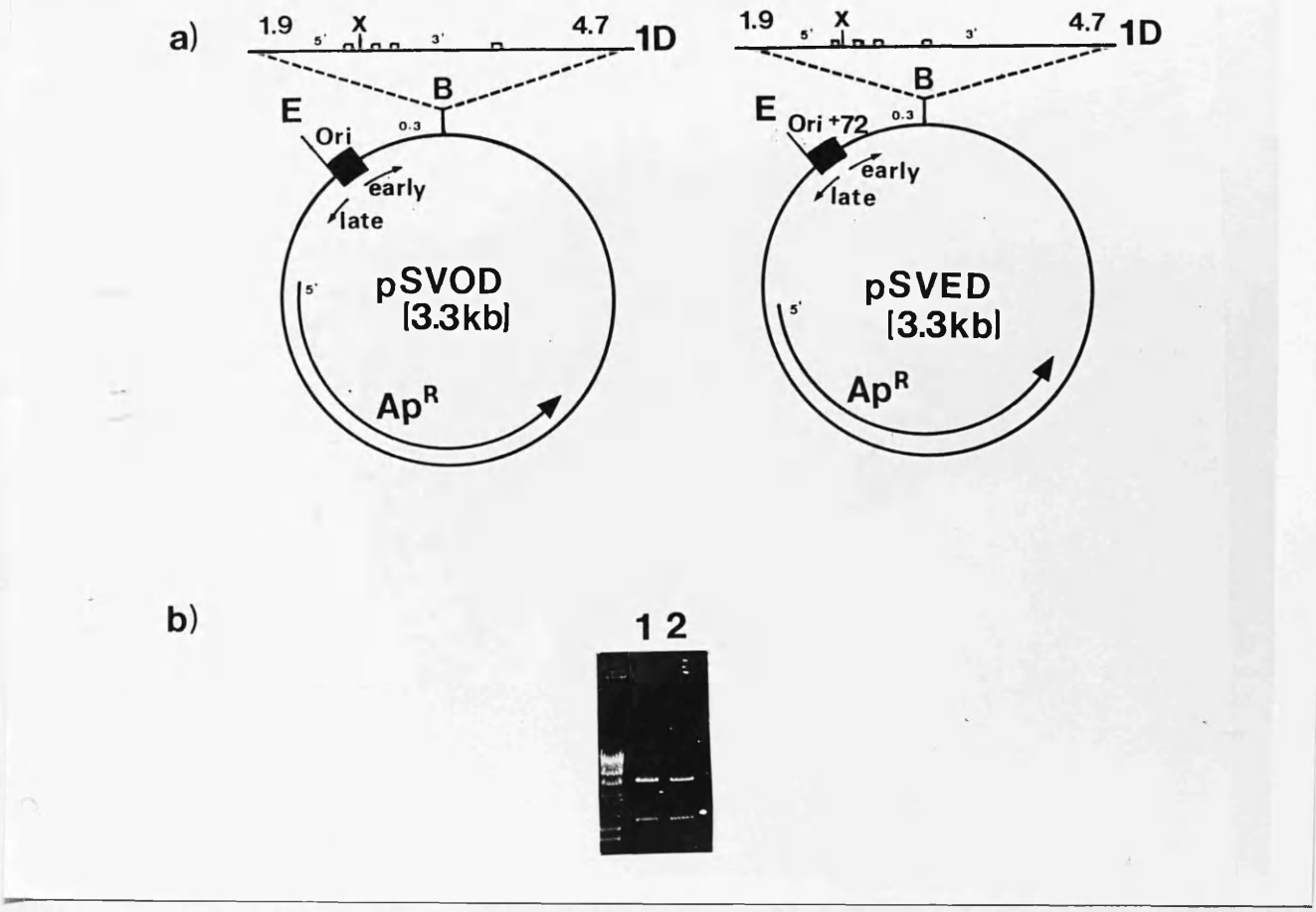

Figure 3-2 The pSVODras1D and pSVEDras1D constructs.

a) A schematic drawing showing the sizes of fragments expected from diagnostic EcoRI/XbaI double digestions of pSVOD and pSVED, into which the 6.4 KB BamHI fragment, containing the human c-Ha-ras1 gene (Figure 1-1), has been inserted in the direct orientation. The open boxes indicate the relative positions of the c-Ha-ras1 coding sequences. pSVOD and pSVED are essentially pBR322, from which the sequences inhibitory to repliction in mammalian cells have been removed (Mellon et al., 1981), into which the 311 bp EcoRII (ori) or 342 bp HindIII/PvuII fragments (ori + 72) have been inserted (Mellon et al., 1981; Proudfoot et al., 1984). E = EcoRI; B = BamHI and X = XbaI, b) a 0.8% agarose gel showing diagnostic EcoRI/XbaI digestions of either pSVODras1D (lane 1) or pSVEDras1D (lane 2). The gel shows the diagnostic fragments of 7.4 Kb and 2.4 Kb expected for the direct (5′ to 3′) orientation of the c-Ha-ras1 6.4Kb BamHI insert.

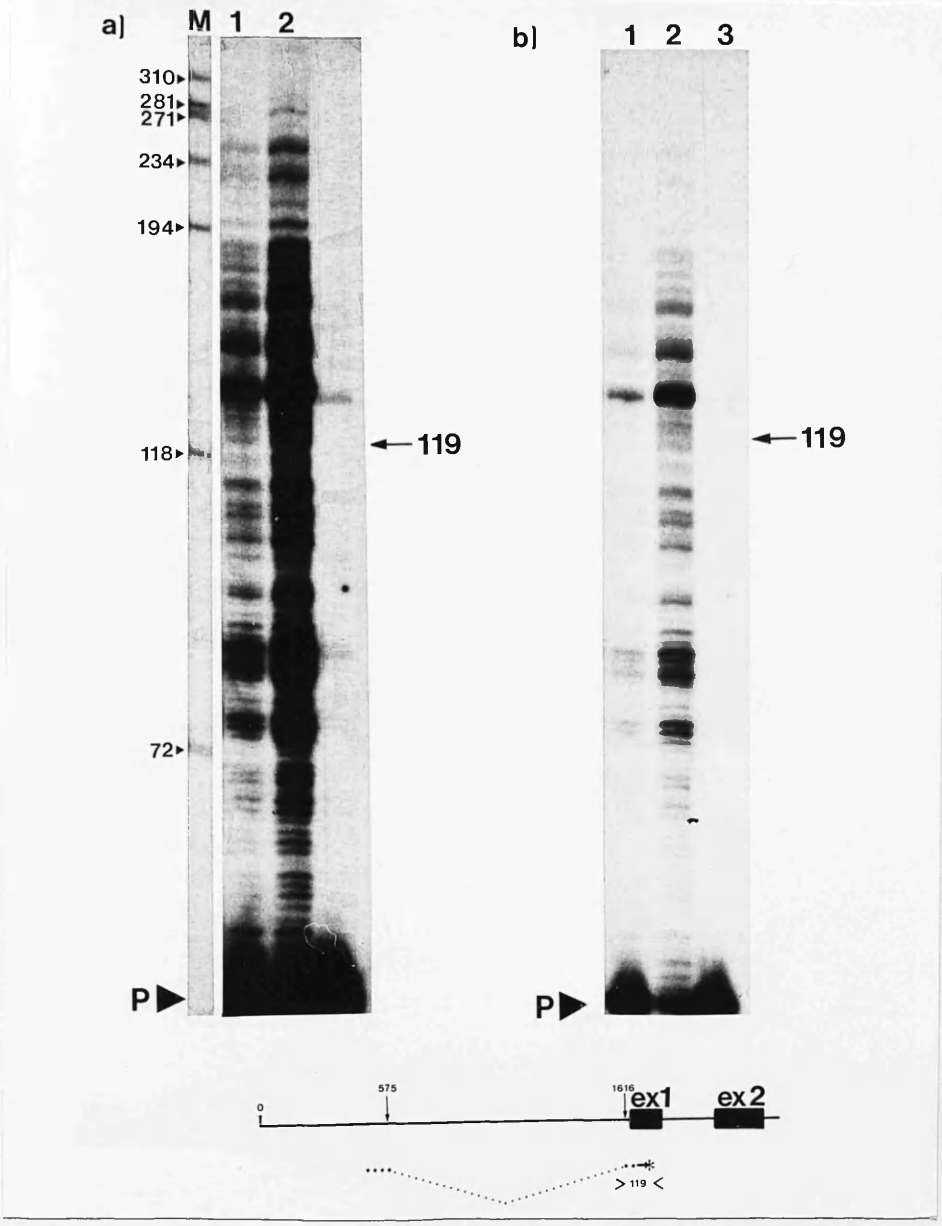

Figure 3-3 Primer extension analysis of total RNA isolated from Cos7 cells transiently transfected with pSVODras1D and pSVEDras1D.

Cos7 cells at late-log phase were transfected with 20ug/75 cm^2 of either pSVODras1D or pSVEDras1D. Total RNA was prepared 48 hours later and hybridized at 52°C to γ-^{32}P-ATP 5' end-labelled anti-sense 51 nucleotide AluI-PvuII primer DNA. The primer was labelled within the first coding exon as indicated by the asterisk in the line drawing. All hybridizations were carried out using approximately 200 cps of primer. The hybrids were incubated with 25 units of AMV reverse transcriptase (Boehringer) as described in the legend of Figure 1-7, except that RNA was alkali hydrolysed following reverse transcription. Extended products were denatured and separated on an 8% denaturing polyacrylamide gel. Autoradiography was for (a) 16 hours or, (b) 3 hours, at -70°C with an intensifying screen. Lanes: (1) 50ug of Cos7/pSVODras1D total RNA; (2) 50ug of Cos7/pSVEDras1D total RNA and (3) 50ug of Cos7 total RNA. Markers are HaeIII fragments of ØX174 (M), whose sizes are indicated by the closed arrow-heads. The line drawing is as in Figure 1-7, except that the distance from the acceptor splice site at position 1616 to the 5' end-label is indicated by >119<. The arrow at position 119 nucleotides shows the position of the splice junction and the large closed arrow-heads (P) mark the position of the primer.

used here was not obtained. Therefore it was not possible to assay any enhancer activity in the oppposite orientation.

IV. THE UPSTREAM INTRON OF THE HUMAN c-Ha-ras1 GENE CONTAINS REPETITIVE SEQUENCES

(a) Northern blot analysis indicates the presence of abundant RNAs homologous to sequences within the upstream intron

During the analysis of the transcription pattern of the c-Ha-ras1 gene, using the northern blotting technique, it was noticed that, when probes homologous to the intron between exon -1 and exon 1 were used , that RNAs from various human sources gave an unusual banding pattern when hybridized to these probes (data not shown). This banding pattern consisted of up to six bands which were superimposed onto a relatively uniform smear. Thus it was decided to pursue this further using a more detailed analysis. The RNAs used were: total and poly A^- RNA from EJ cells, total RNA from EJ focus 8.3.5 cells (a cell line derived from a focus induced in NIH 3T3 murine fibroblasts by transfection with high molecular weight EJ DNA, see Results I.c), total and poly A^- RNA from K562 cells (a human erythroleukemia cell line) and total RNA from human white blood cells. These RNAs were denatured, electrophoresed on 1% formaldehyde agarose gels and transferred to nitrocellulose paper (Materials and Methods). Identical blots were probed with, a) the nick-translated 1905 bp BamHI/XbaI restriction fragment, mapping from 1669 bp upstream to 235 bp downstream from the ATG translational initiation codon of the human c-Ha-ras1 gene and b) nick-translated BS-9 plasmid DNA (Ellis et al., 1980). This plasmid contains viral Ha-ras sequences from the Harvey murine sarcoma virus (HaMSV) genome in a 460 bp

EcoRI fragment cloned into pBR322. Therefore, it contains a spliced form of the murine c-Ha-ras1 gene and should only hybridize to RNA sequences homologous to the exons of the human c-Ha-ras1 gene. Hybridizations were performed exactly as described in the Materials and Methods. In both cases excess, unhybridized probe was removed using high stringency washing conditions, (the final two washes were for 20 minutes each at 65°C in 0.1X SSC; 0.1% SDS). As shown in Figure 4-1a, when hybridized against the 1905 bp BamHI/XbaI c-Ha-ras1 probe, total and poly A^- RNAs from human cells, produced a strong signal composed of up to six bands, superimposed on to a relatively uniform smear, as had been observed previously in earlier experiments. The most intense signal corresponded to a band slightly larger than the 28 S ribosomal RNA (rRNA) marker and other prominent bands were observed at approximately 26 S, 20 S, 17 S and 14 S. Interestingly, EJ focus 8.3.5 total RNA (Figure 4-1a, lane 3), which is essentially N1H 3T3 murine fibroblast RNA, containing transcripts from the transfected human EJ c-Ha-ras1 gene, did not produce the intense signal observed with the human RNAs. In this case only the approximately 1.4 Kb c-Ha-ras1 mRNA was observed. Southern blot analysis has shown that EJ focus 8.3.5 cells are positive for the EJ c-Ha-ras1 oncogene, which is contained on a 23 Kb EcoRI restriction fragment present in 3 to 4 copies per cell (Dr. Alan Balmain, personal communication). An identical blot to that shown in Figure 4-1a was probed with the nick-translated BS-9 plasmid (Figure 4-1b). However, this probe did not produce the abundantly transcribed human specific RNA. With this probe the 1.4 Kb c-Ha-ras1 transcript was clearly observed with EJ and with EJ focus 8.3.5 total RNAs (lanes 2 and 3), as both of these cell lines are known to have elevated expression of the human c-Ha-ras1 gene. However, in the exposure shown, the c-Ha-ras1 transcript was only weakly detected with K562 total RNA (lane 5) and not at all with human white cell total RNA (lane 6). The high molecular weight signal

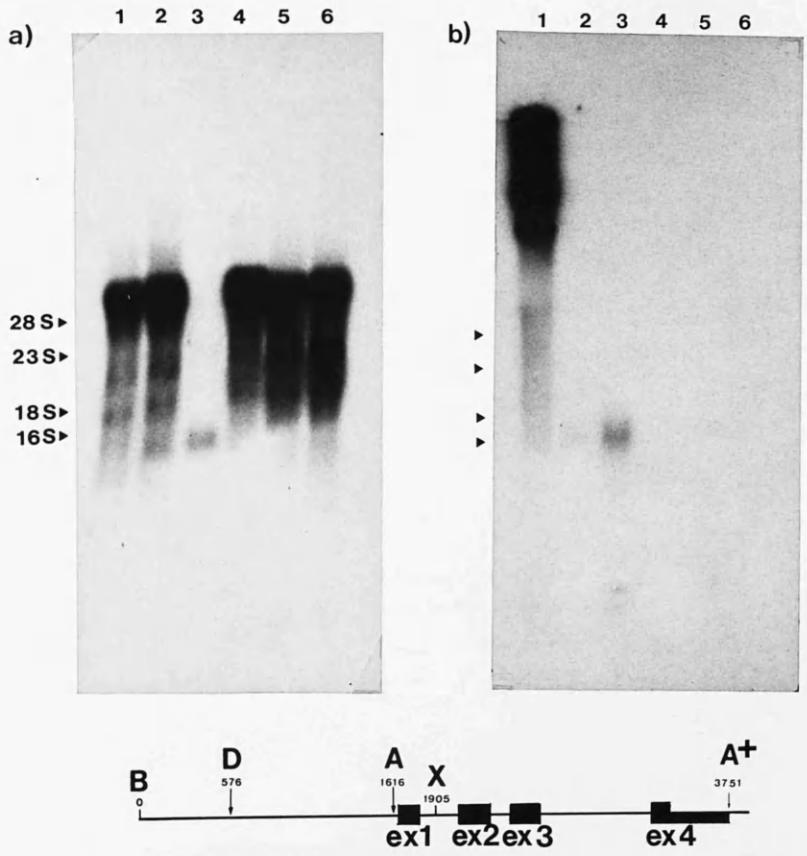

Figure 4-1 Northern blot analysis of human RNAs and EJ focus 8.3.5 total RNA.

RNAs were electrophoresed on 1% denaturing agarose gels, blotted onto nitrocellulose membranes and hybridized at 42°C overnight to nick-translated probes: (a) 1905 bp BamHI/XbaI DNA (1.5×10^8 cpm/ug) or (b) BS-9 plasmid DNA (1.6×10^8 cpm/ug). The 1905 bp BamHI/XbaI restriction fragment is shown in the line drawing and the BS-9 plasmid contains sequences homologous to the Harvey murine sarcoma virus contained in pBR322 (Ellis et al., 1980). Unhybridized probe was removed by serial washing to a final stringency of 0.1X SSC; 0.1% SDS, at 65°C (Materials and Methods). Autoradiography was for (a) 50 hours or (b) 80 hours, at -70°C with an intensifying screen. Lanes: (1) 20 ug of EJ poly A- RNA; (2) 20 ug of EJ total RNA; (3) 20 ug of EJ focus 8.3.5 total RNA; (4) 20 ug of K562 poly A^- RNA; (5) 20 ug of K562 total RNA and (6) 20 ug of human white cell total RNA. Marker RNAs are human (28S and 18S) and bacterial (23S and 16S) ribosomal RNAs, as indicated by the closed arrow-heads. The line drawing is as in Figure 1-16, where B = BamHI and X = XbaI.

detected in Figure 4-1b, lane 1 is believed to be contaminating plasmid DNA.

The data presented above suggest that sequences contained within the 1905 bp BamHI/XbaI probe, but not within the BS-9 plasmid (which does not contain sequences homologous to the murine Harvey ras introns), are homologous to the abundantly transcribed human specific RNAs. As earlier northern data (Figure 1-16) show that the 591 bp Sau3A probe, which is homologous to the first 591 bp at the 5' end of the 1905 bp BamHI/XbaI probe, did not hybridize to the abundant RNAs described here, whereas the 1905 bp BamHI/XbaI probe did. These data suggest that the region of the 1905 bp BamHI/XbaI probe which hybridizes to the abundant RNAs, must correspond to sequences within the upstream intron, or the 125 bp of the first intron 5' to the XbaI site (Figure 4-1).

To further test the possibility that sequences within the human c-Ha-ras1 upstream intron are homologous to the abundant RNAs, EJ poly A^+ and poly A^- RNAs, were denatured, electrophoresed on 1% denaturing gels, transferred to nitrocellulose paper and probed with the 377 bp Sau3A/DdeI fragment, mapping from positions -1079 to -702 and the 453 bp BstNI fragment, mapping from position -623 to -170, relative to the ATG codon. These fragments were labelled by the technique of second strand synthesis (or oligo labelling) using α-^{32}P-dCTP, hybridized to identical blots containing EJ poly A^+ and poly A^- RNAs. Excess, unhybridized probe was removed using high stringency washing conditions as before. Figure 4-2 shows that only the 377 bp Sau3A/DdeI probe detected the abundant RNAs observed with the 1905 bp BamHI/XbaI probe (Figure 4-1). From these data, it is clear that the region containing sequences homologous to the abundant RNAs, present in various human cells, lies approximately between the Sau3A site at position -1079 and the BstNI site at position -596, relative to the ATG codon.

In summary, the data presented in this section suggest

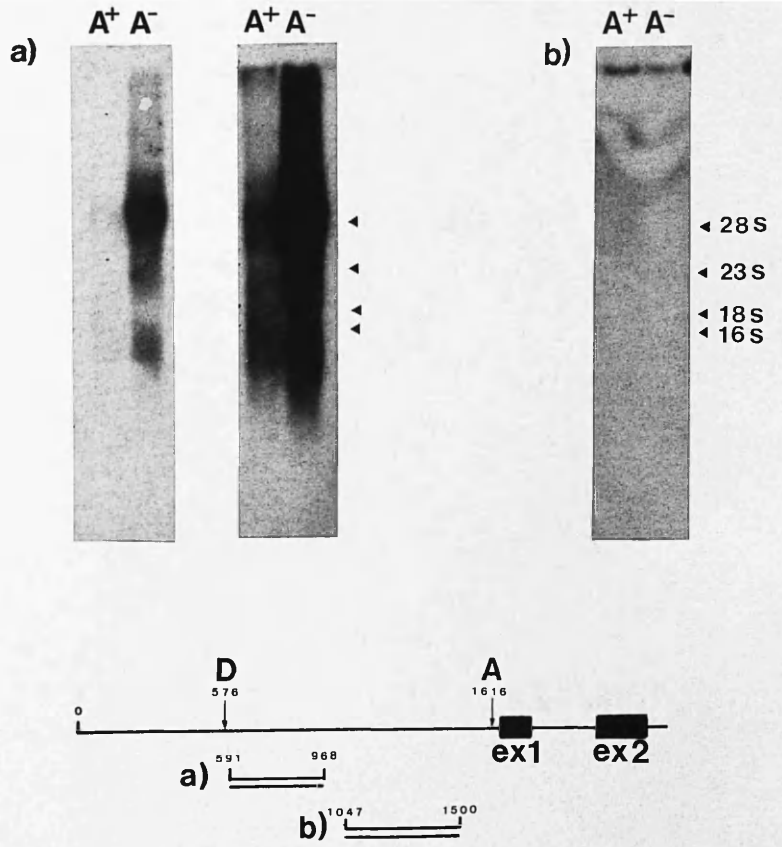

Figure 4-2 Northern blot analysis of EJ RNA using probes homologous to the upstream intron.

RNAS were elecrophoresed on 1% denaturing gels blotted onto nitrocellulose membranes and hybridized at 42°C overnight, to oligo-labelled probes: (a) 377 bp Sau3A/DdeI DNA (1.0 x 10^9 cpm/ug) or (b) 453 bp BstNI DNA (1.4 x 10^9 cpm/ug). Unhybridized probe was removed by serial washing to a final stringency of 0.1X SSC; 0.1% SDS, at 65°C (Materials and Methods). An autoradiograph of the same blot (probed with the 377 bp Sau3A/DdeI probe) after either 3 hours or 16 hours exposure is shown in (a) and an autoradiograph of an identical blot (hybridized to the 453 bp BstNI probe) after a 16 hour exposure is shown in (b). Autoradiography was performed as described in Figure 4-1. Lanes were either, A^+ = 3 ug of EJ poly A^+ RNA or A^- = 20 ug of EJ poly A^- RNA. Marker RNAs are human (28S and 18S) and bacterial (23S and 16S) ribosomal RNAs,as indicated by the closed arrow-heads. The line drawing is as in Figure 1-16, where the relative positions of probes (a) and (b) are indicated.

that a region of approximately 456 bp, mapping to the 5' half of the upstream intron of the human c-Ha-ras1 gene, contains sequences homologous to some highly transcribed, heterogeneously sized, non-polyadenylated RNAs. These RNAs range in size from approximately 30 S (~5Kb) to, at least as small as, 14 S (~1Kb). However, these data do not provide information as to whether the repetitive region of the c-Ha-ras1 upstream intron is itself, transcribed.

(b) Southern blot analysis of the region containing sequences homologous to the abundant RNAs

The abundant, heterogeneously sized RNA molecules, homologous to the 5' half of the human c-Ha-ras 1 intron between exon -1 and exon 1, may correspond to abundant transcripts derived from a few DNA sequences, or alternatively from repetitive DNA sequences within the human genome. To resolve this question, the abundance of DNA sequences homologous to the c-Ha-ras1 upstream intron was determined. EJ cell line DNA, normal human colon DNA and mouse liver DNA were digested with either BamHI, or EcoRI and electrophoresed on a 0.8% agarose gel. The size fractionated genomic DNA was denatured, Southern blotted onto a GeneScreen filter and hybridized to oligo-labelled probes exactly as described in the GeneScreen manual. The probes used were, a) the 591 bp Sau3A fragment, which contains exon -1 mapping from positions -1669 to -1079 relative to the ATG codon, and b) the 1146 bp Sau3A fragment, mapping from position -1079 to +64, which includes the upstream intron and 120 bp of exon 1. As shown in Figure 4-3, the 591 bp Sau3A probe hybridized to the appropriately sized genomic fragments expected for BamHI or EcoRI digested human DNA. The human c-Ha-ras1 gene is found within an approximately 6.6 Kb BamHI fragment, or within an approximately 23 Kb EcoR1 fragment (Pulciani et al., 1982; Shih and Weinburg, 1982; Goldfarb et al., 1982). This probe also detects a BamHI genomic fragment of approximately 3.5

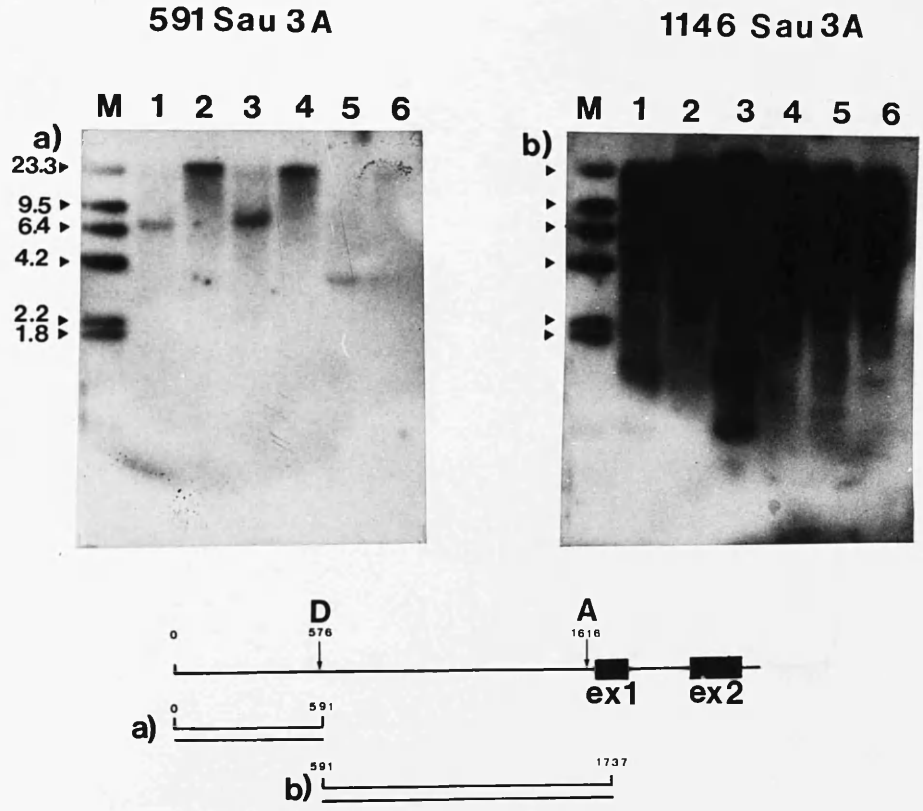

Figure 4-3 Southern blot analysis of human and mouse DNAs.

DNAs were digested with either BamHI or EcoRI, electrophoresed on 1% agarose gels, blotted onto GeneScreen membranes and hybridized at 42°C overnight to oligo-labelled probes, (a) 591 bp Sau3A DNA (6.6 x 10^8 cpm/ug or (b) 1146 bp Sau3A DNA (4.0 x 10^8 cpm/ug). Unhybridized probe was removed by serial washing to a final stringency of 0.1x SSPE; 0.1% SDS, at 65°C (Materials and Methods). Autoradiography was for 36 hours in both cases, at -70°C with an intensifying screen. Lanes: (1) and (2) 10 ug of EJ DNA; (3) and (4) 10 ug of human colon DNA; (5) and (6) 10 ug of mouse liver DNA. The DNA in the odd numbered lanes was BamHI digested and the DNA in even numbered lanes was EcoRI digested. Marker DNAs (M) are HindIII fragments of lamba phage DNA, whose sizes are indicated by the closed arrow-heads. The line drawing is as in Figure 1-16, where the relative positions of the probes (a) and (b) are indicated.

Kb in mouse liver DNA, but barely detects the approximately 23Kb EcoRI genomic fragment, expected for mouse genomic DNA. When the 1146 bp Sau3A probe is hybridized to an identical blot, an intensely hybridizing smear is observed in all six lanes, even after the high stringency washes used to remove excess probe (the final two washes were for one hour each at 60^{o}C, in 0.1X SSPE; 0.1% SDS). However, in shorter exposures (not shown) slightly less hybridization is seen with the mouse DNA. Moreover, each lane has more intense bands of discrete sizes, superimposed onto the background smear, although no two lanes have an identical pattern. Interestingly, these superimposed bands are slightly polymorphic not only between human DNA and mouse DNA, but also between the two different human DNAs.

This Southern blot analysis suggests that sequences within the upstream intron of the human c-Ha-ras1 gene are moderately repetitive throughout the human genome and that these sequences have some homology with similar sequences within the mouse genome. The northern blot analysis of the preceding section suggested that sequences homologous to the 5' half of the upstream intron, were homologous to abundantly transcribed, heterogeneous RNAs in human cells, but not to any RNAs present in mouse (NIH 3T3) cells. Therefore, the moderately repetitive sequences within the human c-Ha-ras1 upstream intron, are homologous to both human and mouse genomic DNAs and also to some transcribed human sequences, but not to any detectable transcribed mouse sequences.

DISCUSSION

I. IDENTIFICATION OF THE TRANSCRIPTIONAL INITIATION SITES FOR THE HUMAN c-Ha-ras1 GENE

Until recently, little information was known about the sequence motifs involved in the transcription of the "housekeeping" genes. For this reason early attempts described in this thesis to determine the 5' transcriptional boundary of the human c-Ha-ras1 gene, focused on the region immediately upstream from the first coding exon which contained a putative TATA/CAAT promoter region (Capon et al., 1983a; Reddy, 1983 and see Figure 2). S1 mapping analysis of this region identified a series of protected fragments consisting of a major cluster of bands between positions -40 and -50, relative to the ATG translational initiation codon and three minor bands mapping to positions -70 ±2, -77 ±3 and -82 ±3 (Figures 1-2, 1-3, 1-4 and 1-6). However, a strong, potentially stable, stem loop structure was identified in the DNA sequence from this region. This structure could be recognised by S1 nuclease, if it occurred in the probe molecule during S1 nuclease digestion, and would explain the pattern of bands observed in the S1 mapping analyses (Figure 1-5). This secondary structure and also the occurrence of a consensus acceptor splice site at position -53 (Capon et al., 1983a; Reddy, 1983), suggested that the S1 protected fragments did not map to genuine transcriptional initiation sites. There were three possible explanations for this data, 1) that the protected fragments observed represented genuine transcriptional initiation sites, 2) that they were artifacts due to the stem-loop structure and that the true 5' termini of the c-Ha-ras1 RNA are located further upstream of the stem-loop structure, (perhaps initiated by the putative TATA/CAAT promoter region), or 3) that, since there is an acceptor splice site at position -53 the protected fragments reflect a processing event in c-Ha-ras1

RNA and that the fortuitous occurrence of the stem-loop structure at this site accounts for the multiple protected bands observed. Exonuclease VII mapping analysis (Figure 1-6) did not eliminate the first possibility, as two major protected fragments mapping within approximately 10 bp of the major cluster of S1 protected fragments were also obtained with exonuclease VII. The possibility that transcripts which originated upstream of the stem-loop structure are not detected because of internal cleavage by S1 nuclease at the stem-loop structure, could be eliminated as no protected fragments mapping upstream of the stem-loop structure to a region expected for the putative TATA/CAAT promoter were detected. In addition, S1 mapping analyses with probes 5' end-labelled at positions upstream of the stem-loop structure, did not produce protected fragments when hybridized to EJ RNA (data not shown).

Taken together with the occurrence of an acceptor splice site at position -53, these analyses provide support for the hypothesis that the protected fragments observed reflect the processing of RNA at the accceptor splice site at position -53. The ability of the probe molecules to form the stem-loop structure is likely to explain the confusing occurrence of the multiple protected fragments. However, it is unlikely that this stem-loop structure has any functional significance, as inspection of the recently determined sequence of the mouse c-Ha-ras1 gene, did not reveal any analogous structure at this region (Dr. Alan Balmain, personal communication). To test the hypothesis that a processing event does occur at the putative acceptor splice site, the independent technique of primer extension analysis was used (Figures 1-7 and 1-8). A primer from within the first coding exon of the c-Ha-ras1 gene was extended through the position of the acceptor splice site to produce multiple extension products. These extension products formed a series of at least eighteen resolvable bands of varying intensities, which presumably map between 191 bp to 16 bp upstream of the corresponding donor splice

site, reported at position -1095 (Capon et al., 1983a). To confirm this, S1 nuclease mapping probes, 5' end-labelled at positions upstream of the donor splice site, were used in an attempt to detect c-Ha-ras1 mRNA 5' termini.

When probes labelled at positions 21 and 46 nucleotides upstream from the donor splice site were used for S1 nuclease mapping analysis, protected fragments were detected mapping as far upstream from this site as approximately 210 bp (e.g. Figure 1-13). The multiple protected fragments observed in these experiments were arbitrarily grouped into five clusters (A through E) and three weaker fragments (Results, section I.d). It is possible that all the bands detected in these S1 nuclease mapping experiments correspond to genuine transcriptional initiation sites. This suggestion is strengthened by the observation that the bands in cluster A correspond to extension products mapping close to the extension product at positon -191 (relative to the donor splice site of exon -1), the bands in cluster B correspond to the extension products between positions -179 and -162, those in cluster C correspond to the extension products between positions -148 and -135 and those in clusters D and E correspond to the extension products between positions -126 and -106. Furthermore, the three S1 protected fragments mapping between positons -103 and -84 correspond to extension products between positions -101 and -86 (Results, sections I.c and I.d, also see Figure 5-1). Hence, the upper limit to the size of the upstream exon (called exon -1), as estimated by these two independent techniques, must be at least 191 bp. It is difficult to be certain of the precise location of each initiation site, in this rather complex situation, since at least some of the S1 protected fragments may correspond to secondary structures in the probe molecule. In this respect, it has been demonstrated that this region has considerable secondary structure forming potential (Figures 1-9 and 1-10). Furthermore, some of the primer extension products detected, may correspond

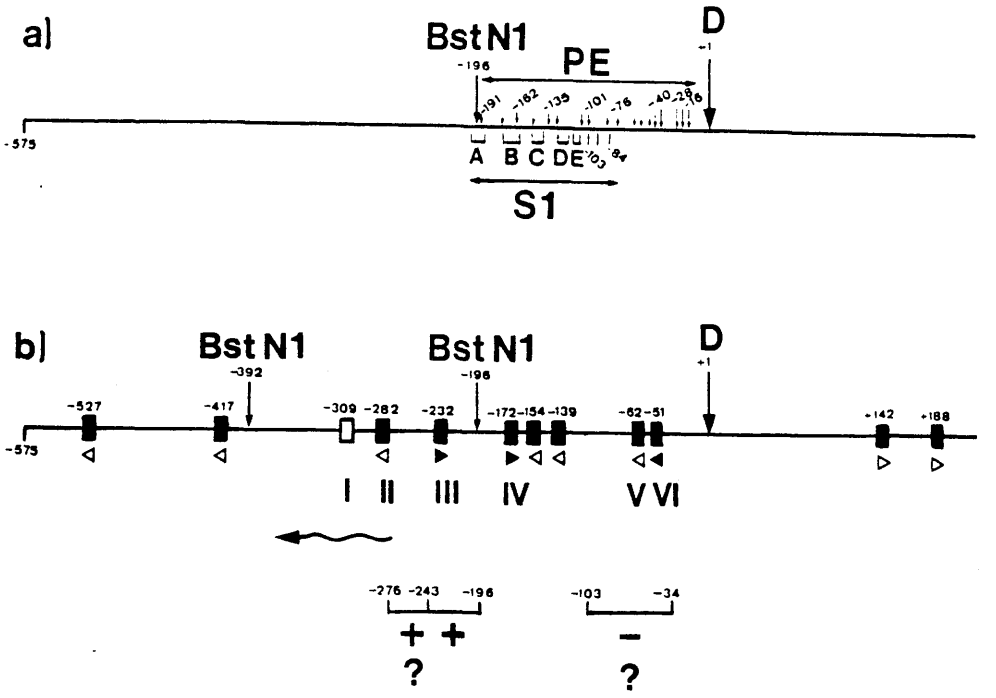

Figure 5-1 Summary of the RNA mapping analysis and the mutational analysis of the c-Ha-ras1 promoter region.

a) A line drawing of the region upstream of the donor splice site (D), numbered as +1. The position of the BstNI site, which approximately defines the 5′ limit of exon -1 is shown. Also shown are the map positions, relative to D, of the clusters of bands (A to E) detected by S1 nuclease mapping (S1). The primer extension products (PE) are indicated by the arrows. The 5′ termini of the extension products also identified by Ishii et al., (1985), are located between positions -40 and -16. b) A line drawing of the same region as that shown in (a). The positions of the two BstNI sites which define the ends of fragment 6, used in constructs pRasCAT6D and pRasCAT6R, are shown. Also shown are the positions of GC boxes (closed boxes), where the arrow-heads give their orientations. Solid arrow-heads indicate putative high affinity binding sites and open arrow-heads indicate low or medium affinity binding sites, for the Sp1 transcription factor (Kadonaga et al., 1986). The roman numerals I to VI are demonstrated Sp1 binding sites, as determined by in vitro DNAse I footprinting studies (Ishii et al. 1986). Sp1 binding site number I is shown as an open box since it does not contain a core GGGCGG hexanucleotide repeat. Also shown below this line drawing are the positons of the positive (+) and negative (-) regulatory elements defined in this thesis, where the question marks indicate the putative elements. The wavy arrow indictes the direction of transcription and the possible position of the putative opposite-strand transcript for the "reverse-orientation" promoter, defined in this thesis.

to premature termination of the reverse transcriptase. Nevertheless, these two independent techniques have independently confirmed that multiple human c-Ha-ras1 mRNA 5' termini exist approximately between 191 and 84 bp upstream of the donor splice site (that is, the primer extension products above position 203 nucleotides in Figures 1-7, 1-8 and 3-3). Thus the situation with the human c-Ha-ras1 gene is strikingly reminiscent of at least one other "housekeeping" gene. The HMG CoA reductase gene also has multiple transcription initiation sites, which in this case are grouped into five major clusters and several minor bands, spread over a region of approximately 90 bp (Reynolds et al., 1984).

The short distance between the c-Ha-ras1 exon -1 donor splice site and the primer extended product mapping approximately 84 bp upstream of it (i.e., the position nearest the donor splice site to which an S1 protected fragment mapped), prevented the independent confirmation of these putative initiation sites by S1 mapping analysis, with 5' end-labelled probes. As this region contains the four strong extension products described by Ishii et al. (1985), and also some of the extension products described in this study, it would be desirable to confirm these extension products with an independent technique.

Uniformly labelled single stranded probes, offer a possibility for independently mapping some of the putative initiation sites identified by primer extension analyses between positions -84 and -16. To this end, the 591 bp Sau3A fragment mapping from position -575 to +15 relative to the exon -1 donor splice site, has been cloned in both orientations into the vector pSP64. This vector will allow the production of uniformly labelled single stranded RNA probes complimentary to this insert, by transcription with the bacterial SP6 RNA polymerase. These probes could then be used for RNase protection experiments, to detect transcripts originating from either strand in the region covered by the 591 bp Sau3A fragment. However, it is still

unlikely that such probes could detect hybrids corresponding to those expected for the putative initiation sites of Ishii et al. (1985), as the longest hybrid (corresponding to site 4, see Figures 1-7 and 1-8) would only be 40 bp long. Therefore, in order to independently confirm the presence of these sites it would be necessary to construct a cDNA clone, which could then be used to design 5' end-labelled probes for S1 mapping analyses. Ultimate proof of the occurrence of a true transcriptional initiation site however, depends on the identification of a cap structure at the 5' termini of the transcripts themselves.

The report from Ishii et al. (1985), describes the use of the same 51 nucleotide AluI-PvuII primer from the first coding exon of the human c-Ha-ras1 gene, as used in this study. These authors suggests that the extension products observed above the bands numbered 135, 142, 147 and 159 in Figures 1-7 and 1-8, are derived from cross hybridization to transcripts from some other gene or genes. Of critical importance to this suggestion is their claim that when they sequenced their extended primer, the sequences of the extended products above the extension products numbered 135, 142, 147 and 159, were different from the sequence of the human c-Ha-ras1 genomic clone. Unfortunately, they did not present their sequence data. To test their suggestion, primer extension analyses were repeated using RNAs derived from c-Ha-ras1 transfected cells. When EJ focus 8.3.5 RNA was used (Figure 1-8), the same pattern of bands was observed as before for EJ and A431 cell RNAs. EJ focus 8.3.5 cells are mouse NIH 3T3 fibroblasts, containing the human c-Ha-ras1 gene from EJ/T24 bladder carcinoma cells, and they produce greatly elevated levels of human c-Ha-ras1 RNA. The pattern of banding observed with this RNA can only be derived from the human c-Ha-ras1 gene and not from any mouse genes, since no cross hybridization to mouse RNA was observed. Similarly, the primer extension analysis using RNA derived from monkey Cos7 cells transfected with the

SV40 based vectors, pSVED.ras1D and pSVOD.ras1D (Figure 3-3), must detect only RNA derived from the 6.6 Kb BamHI insert containing the human c-Ha-ras1 gene, since only negligible cross hybridization to monkey Cos7 control RNA occurs. Since this analysis also produced the same pattern of bands as was observed with human EJ and A431 cell RNA, although with nearly a 100-fold stronger signal, it is clear that the 5' transcriptional boundary of the human c-Ha-ras1 gene is further upstream, than that proposed by Ishii et al. (1985).

Further support for the suggestion that the 5' transcriptional boundary of the human c-Ha-ras1 gene is further upstream than position -40 comes from sequence comparisons. Chichutek and Duesberg (1986) have compared the sequences of the human c-Ha-ras1 gene with that of various Harvey ras-containing sarcoma viruses and the rat c-Ha-ras2 processed pseudogene. The Harvey sarcoma virus (HaSV) and the rat c-Ha-ras2 processed pseudogene, were found to be homologous with the human c-Ha-ras1 DNA sequence for 127 bp and 145 bp respectively, from the donor splice site of exon -1. Therefore, this sequence analysis strongly suggests that the 5' transcriptional boundary of the human c-Ha-ras1 gene is at least 145 bp upstream from the exon -1 donor splice site, which provides further support for the data presented in this thesis. However, this sequence analysis disagrees with the estimated size of 40 bp for exon -1, obtained by Ishii et al. (1985).

To further define the 5' boundary of the human c-Ha-ras1 transcriptional unit, northern blot analyses were also performed (Figures 1-16 and 1-17). These analyses revealed that probes containing sequences up to, but not 5' to, the BstNI site 196 bp upstream from the donor splice site, hybridized to the same sized RNA species as did probes homologous to the coding exons of the human c-Ha-ras1 gene. This suggests that the 5' boundary of exon -1, must be close to the BstNI site 196 bp upstream from the donor splice site. This is in close agreement with the

S1 mapping and primer extension data presented in this thesis. Furthermore, the 195 bp MspI probe, does not contain any of the transcriptional initiation sites defined by Ishii et al. (1985), but still hybridizes to the 1.4 Kb c-Ha-ras1 message. Thus, these data strongly suggest that sequences upstream of the MspI site, mapping to a position 48 bp upstream from the donor splice site, are indeed transcribed and contiguous with the 1.4 Kb c-Ha-ras1 message. In addition, this analysis was designed to allow a quantitative comparison of the signals detected by the upstream exon -1 and exon 1 probes (Figure 1-17). The 195 bp MspI probe, mapping from 243 bp to 48 bp upstream from the exon -1 splice site, hybridized approximately 10-fold less efficiently to c-Ha-ras1 RNA, than did the 178 bp PvuII probe, which had 129 bp of exon 1 sequences available for hybridization. This less efficient hybridization could suggest that the majority of c-Ha-ras1 transcripts may initiate near the donor splice site and therefore produce a less stable, smaller hybrid, than the 129 bp hybrid formed between the 178 bp PvuII probe and the c-Ha-ras1 RNA. Alternatively, the reduced signal with the 195 bp MspI probe, may be explained by the demonstrated strong secondary structure forming potential of this region (Figures 1-9 and 1-10), which may interfere with the hybridization between the probe and the c-Ha-ras1 RNA.

Interestingly, northern blot analysis using probes specific for, exon -1 and exon 1, revealed a less abundant RNA species of approximately 6.0 Kb, whereas only the expected 1.4 Kb c-Ha-ras1 transcript was observed with an exon 4 probe. This observation suggests that the human c-Ha-ras1 gene, like the c-Ki-ras2 and N-ras genes, has two differently sized mature transcripts (Shimizu et al., 1983a; McGrath et al., 1983; Capon et al., 1983b and Hall and Brown, 1985). In the case of the human N-ras gene, it is known that the 4.3 and 2.0 Kb transcripts, observed by northern blot analysis, are generated by the use of two polyadenylated processing signals. Furthermore, it is known

that the 2.0 Kb transcript is less abundant since it uses the less efficient polyadenylation processing signal (Introduction, I.b). The situation is even more analogous to the human c-Ki-ras2 gene, where the two mature transcripts are 5.5 and 3.8 Kb respectively, both of which contain predominantly exon 4B information but also a rarer species containing exon 4A information (Introduction, I.b). Thus it is possible that the human c-Ha-ras1 gene may also contain an alternative fourth coding exon located further downstream. Alternatively, the exon -1 and exon 1 probes may cross hybridise to a transcript from another ras-like gene. A similar higher molecular weight transcript has been noticed occasionally in mouse RNA probed with mouse c-Ha-ras1 probes (Dr. Alan Balmain, personal communication) and also in human RNAs probed with human Ha-ras specific probes (Parada et al., 1982; Ellis et al., 1982).

The 5' transcriptional boundary of the human N-ras gene has been accurately determined (Hall and Brown, 1985). This gene also has an upstream non-coding exon, called exon -1, and S1 mapping has been used to define two transcriptional initiation sites separated by 10 bp at positions 225 $\pm$2 bp and 235 $\pm$2 bp upstream from the exon -1 donor splice site (Introduction, I.b). These data contrast with that for the human c-Ha-ras1 gene, where the region containing the 5' transcriptional boundary has multiple transcription initiation sites spread over a region of approximately 190 bp. However, for the human N-ras gene, the 100 bp region containing the two transcriptional initiation sites is approximately 65% A + T rich, in contrast with the overall approximately 75% G + C richness of the 5' end of the c-Ha-ras1 gene. Interestingly, this notable difference between the structures of the promoter regions for these two human ras genes, may indicate significant differences in their mechanisms of transcriptional regulation.

A recent report from Honkawa et al., (1987) has described S1 nuclease mapping analyses using total and poly

A^- RNAs isolated from the EJ human bladder carcinoma cell line. These investigators used two uniformly labelled single-stranded probes mapping between positions -364 to -125 or -575 to -125, relative to the donor splice site of exon -1. These probes detected one small cluster of putative transcriptional initiation sites mapping to positions -275, -272 and -271, and two further sites at positions -205 and -204, relative to the donor splice site. However, no data from an independent technique was provided to support their S1 mapping analyses. Data from this thesis provides strong evidence against the possibility that RNAs, initiating at the cluster centered at position -272, are contiguous with the c-Ha-ras1 gene. Northern blot analysis using the 196 bp BstNI probe mapping between positions -392 and -196 (Figure 1-17) did not detect the 1.4 Kb human c-Ha-ras1 mRNA in poly A^+ RNAs isolated from either the EJ or the A431 cell lines. However, the 195 bp MspI probe which is homologous to the upstream non-coding exon -1, identified here, and the 178 bp PvuII first coding exon probe (Figure 1-17), did hybridize to 1.4 Kb message. This strongly suggests that transcripts originating upstream of position -196 and contiguous with the c-Ha-ras1 gene are not present in poly A^+, nor poly A^- RNAs, from the EJ and A431 cell lines, since they should be able to form at least an 82 bp hybrid with the 196 BstNI probe. Furthermore, neither the S1 mapping analyses (Figures 1-10, 1-12, 1-13 and 1-14) nor the primer extension analyses (Figures 1-7, 1-8 and 3-3), presented here, provide any evidence for such transcripts. Additionally, analyses of the CAT activities produced by construct pRasCAT6D, which contains c-Ha-ras1 sequence information between positions -392 to -196, revealed that although it contained all the regulatory sequences necessary for expression, it lacked initiation sites (Results, section II.b). The cluster at position -204, identified by Honkawa et al., (1987), is approximately 3 bp upstream from the most 5' transcription initiation site identified in this thesis. No bands mapping

below position -204 could be observed in the autoradiograph shown by Honkawa et al., (1987) as the autoradiograph was cut off immediately below the cluster at position -204 (that is, only 79 nucleotides upstream from the 5' end of their probe). A possible explanation for the mapping data obtained by Honkawa et al., (1987), given that the human c-Ha-ras1 promoter region has been identified as a bidirectional promoter region (Results, section II.b.3), is that their probes may be detecting transcription initiation sites for the putative opposite-strand RNA molecules (Figure 1-15). This is supported by the fact that the transcription initiation sites which they identified, correspond to the region shown by CAT analysis to have "reverse-orientation" promoter activity (Figure 2-8). Furthermore, fragment 6 which contains these apparent transcription initiation sites, but not the transcription initiation sites identified in this thesis, directs the synthesis of CAT activity only when it is in the reverse-orientation relative to the CAT gene (Figure 2-8).

II. FUNCTIONAL ANALYSIS OF THE HUMAN c-Ha-ras1 PROMOTER REGION

(a) CAT analysis defines functionally important sequences for the c-Ha-ras1 promoter

The 591 bp Sau3A fragment, mapping from 575 bp upstream to 15 bp downstream of the exon -1 donor splice site, was assayed for promoter activity. This fragment, when inserted in front of the bacterial chloramphenicol acetyl transferase (CAT) gene, was found to direct the synthesis of the CAT gene product at high levels in HeLa cells (Figure 2-6). Thus, this analysis confirms functionally that the region containing the transcriptional initiation sites identified by RNA mapping analysis possesses has promoter activity. To further delineate the

region containing transcriptional initiation sites, and to determine the sequences within the 591 bp Sau3A fragment required for promoter activity, the restriction fragments shown in Figure 2-1b were subcloned into the "promoter-less" pC0 vector. Subsequent transient expression assays performed with these constructs revealed that in order to produce CAT activity, the fragments used must contain at least some of the transcriptional initiation sites identified in the region between positions -196 to -16 and must also contain sufficient sequence information upstream of these sites. Fragments 5 and 6 contain none of the initiation sites identified by S1 nuclease mapping and primer extension analyses. Thus they produce no more CAT activity, in the direct orientation, than the "promoter-less" control construct, pC0. However, fragments 7 and 8 both contain most of the transcriptional initiation sites (Figure 2-1) but do not produce significant CAT activity in the direct orientation, suggesting that neither of these fragments contain appropriate regulatory sequences. Fragment 2 contains virtually all the initiation sites identified, lacking only the two nearest the donor splice site of exon -1, and correspondingly produced only slightly less CAT activity than fragment 1 when in the direct orientation. Fragment 4, in the direct orientation, produced approximately 20% of the CAT activity observed for fragment 1 (Figure 2-8). In this fragment the first four initiation sites identified as major extension products in Figures 1-7 and 1-8 are missing and furthermore, it contains 47 bp more upstream sequence than fragment 7. Thus these data provide further evidence for the functional significance of the initiation sites, detected in this study, which are upstream of those described by Ishii et al. (1985) and suggests that these upstream initiation sites account for at least 20% of the human c-Ha-ras1 transcripts. It would have been interesting to S1 nuclease map the initiation sites used by these constructs, using 5' single-stranded probes, end-labelled within the CAT gene.

Such an analysis would determine quantitatively the distribution of the initiation sites used by these constructs and could also confirm the initiation sites detected by primer extension analysis with EJ and A431 cell RNA. pRasCAT1D is the construct of choice for such an analysis, as it contains the most 5' information and such information may affect the specification of the initiation sites used. This consideration is particularly important as the effect of upstream sequences on a monkey "SV40-like" promoter, has been reported to influence the position and relative efficiency of the apparent RNA initiation sites used by this promoter (Saffer and Singer, 1984). Thus, analysis of the initiation sites used by pRasCAT1D may not eliminate the possibility that sequences upstream of fragment 1, or those within the CAT gene itself, may have an effect on transcription. It would be interesting for future studies to assay the effect on transcription, if any, of sequences upstream of fragment 1, using this system. To this end, the lambda Charon 9A phage designated T24-15A (Pulciani et al., 1982), which contains the c-Ha-ras1 gene from the T24 human baldder carcinoma cell line, was obtained from Dr. Mariano Barbacid. This lambda phage has approximately 5 Kb of extra 5' flanking sequence, relative to fragment 1.

Comparisons of the CAT activity produced by fragments 4 and 7 in the direct orientation, revealed that although fragment 4 contains fewer initiation sites, it produced 4 times greater CAT activity than fragment 7 (Figure 2-8). Thus, it is likely that the 47 bp of extra 5' flanking sequence of fragment 4, contains important positive regulatory elements (Figure 5-1). This sequence,

$$\overset{232}{\text{CGGGGTGGGGCGGGGCCCGATGGCGCGCAG}}\overset{212}{\text{CCAATGGTAGGCCGCGC}}$$

contains a GC hexanucleotide repeat and a CAAT box (underlined) beginning at positions -232 and -212

respectively, relative to the exon -1 donor splice site. The regulatory significance of these sequence motifs within other promoters has been extensively discussed in the introduction. Interestingly, both of these sequence motifs, within the human c-Ha-ras1 promoter region, have been shown in in vitro binding assays to be bound by their respective transcription factors, Sp1 (Ishii et al., 1986) and CTF (Jones et al., 1987). However, the data reported in the present study, provides the only data suggesting functional significance for these two sequence motifs in transcription of the c-Ha-ras1 gene, as they occur in the 47 bp region which had been identified as a positive regulatory element in CAT assays. Confirmation that this 47 bp region has a regulatory function also comes from the analysis of constructs containing both fragments 8 and 9 covalently linked to the CAT gene (Figure 2-9). Fragment 8 produced only negligible CAT activity on its own, although it contains all the transcription initiation sites up to position -103 (Figure 2-7). However, when fragment 9 in the direct orientation was inserted 5' to fragment 8 in the construct pRasCAT8D, promoter activity was increased by 16 fold, relative to fragment 8 in the direct orientation on its own. Fragment 9 contains the 47 bp regulatory region, plus an additional 33 bp of upstream sequence (Figure 2-9) and therefore provides further support for the functional significance of this region (Figure 5-1). When fragment 9 was placed in the reverse orientation 5' to fragment 8, resulting in the construct pRasCAT8D9R (Figure 2-5), the positive regulatory effect of fragment 9 was reduced by less than half, relative to when it was in the direct orientation (Figure 2-9). This slight reduction in the positive regulation by fragment 9, when it is in the reverse orientation, is likely to be within the bounds of experimental error. However, as fragment 9 must have transcription initiation site(s) for the "reverse-orientation" promoter (Results, section II.b), it is also possible that these initiation sites may be used

less efficiently than those used by the "direct-orientation" promoter. Interestingly, the construct pRasCAT8D9D9D, containing fragment 9 as a tandem duplication in the direct orientation (Figure 2-5), was also obtained. When pRasCAT8D9D9D was analysed for CAT activity in HeLa cells, it was found to produce over 10 fold greater activity than pRasCAT1D (Figure 2-9). Thus, duplication of fragment 9 accounts for the highly elevated CAT activity observed with construct pRasCAT8D9D9D, and it is possible that the 47 bp regulatory region within fragment 9 is essential for this elevated activity. However, it is not known whether it is the Sp1 binding site, the CAAT box, both of these sequences, or additional sequences within the 47 bp regulatory region, which are responsible for the regulatory function observed. It is also possible that sequences within the 33 bp of sequence upstream of the 47 bp regulatory region may have an effect. To determine exactly which sequences are responsible, it would be necessary to mutate this region at a finer scale than in the analyses presented here. Addition of the initiation sites located downstream of fragment 4 to pRasCAT4D, would also be useful in determining the minimum sequence information required for the restoration of comparable promoter activity to that observed for fragment 1. It is possible that fragment 4, plus the missing initiation sites, may contain sufficient regulatory sequences for "full" promoter activity. This seems unlikely since pRasCAT8D9D produced only 40% of the CAT activity associated with pRasCAT1D. Furthermore, fragment 9 contains fewer transcription initiation sites than fragment 4, but has 33 bp of additional 5' flanking sequence and produces approximately 4 times the CAT activity of fragment 4, when both are assayed in the direct orientation. This result suggests that the extra 33 bp of 5' flanking sequence within fragment 9, relative to fragment 4, may indeed contain additional sequence information for promoter activity. However, the possibility that this result

reflects the presence in fragment 4 of a negative regulatory element cannot, at present, be excluded. Furthermore, support for this interpretation comes from the observation that fragment 9 in the direct orientation, produces 3 fold more CAT activity than fragment 3 in the direct orientation. This is despite the fact that fragment 3 contains 28 bp more 5' flanking sequences and 69 bp more 3' flanking sequences. However, to confirm the presence of this putative 69 bp negative regulatory element, pRasCAT9D must be used in multiple experiments. Additionally, it would be of some interest, in this context, to add smaller subclones of fragment 4 to the 3' side of fragment 9 in pRasCAT9D, in an attempt to further delineate the position of this putative negative regulatory element. As noted earlier (Results, section II.b.2), the 69 bp sequence corresponding to the putative negative regulatory element has a G + C content of nearly 90% and contains 5 repeats of the sequence 5' $G(C)_n$ on the sense strand (where n = 4 or 5), four of which are tandemly repeated. It is not known how this unusual sequence could be exerting a negative effect on promoter activity. It is possible to imagine models whereby this G + C sequence may inhibit the passage of the transcription complex, either by inhibiting to some extent the denaturation of the DNA strands, or by allowing the formation of an unusual secondary structure.

Although the CAAT box located 212 bp upstream of the human c-Ha-ras1 exon -1 donor splice site can bind strongly to the CAAT-binding transcription factor, CTF (Jones et al., 1987), this fact itself does not necessarily implicate this motif in transcriptional regulation. This is particularly true for CTF, as it is known to be identical to the previously identified adenovirus DNA replication stimulating factor, NF-1 (Jones et al., 1987 and see Introduction section II). Thus CTF/NF-1 may be involved in both transcriptional and replicational processes. It is likely that the positioning of the CAAT box relative to other regulatory elements, e.g., the TATA box, may be the

important factor in determining whether the CAAT box has a transcriptional or replicational role. In this respect, it is interesting that the high affinity CTF binding site, located 200 bp upstream of the human α-globin ATG codon, does not contribute to α-globin gene expression, whereas the low affinity binding site at position -80 is fully functional (Jones et al., 1987).

Notable features of the entire c-Ha-ras1 promoter region are its high G + C content (the 576 bp upstream of the exon -1 donor splice site are 80% G + C rich) and the eleven randomly spaced GC hexanucleotide repeats (Figure 5-1b). In this study it has been shown that a 47 bp region which includes the GC hexanucleotide repeat centered at position -232 (Figure 5-1b) is important for promoter activity. Ishii et al. (1986) have established that the Sp1 transcription factor binds to five of these core hexanucleotide repeats, plus a sixth related sequence (Figure 5-1b). It is not known from this study if the two most upstream hexanucleotide repeats shown in Figure 5-1b, are bound by Sp1. Ishii et al. (1986) have also suggested that only binding sites IV, V and VI are required for promoter activity. This suggestion was made on the basis of a restriction fragment, mapping from position -149 to + 90 relative to the exon -1 donor splice site (that is, it contains 75 bp more 3' sequence than any of the constructs used here), which produced a low level of CAT activity when placed in front of the bacterial CAT gene and transiently expressed in monkey CV1 cells. These data suggest that sequences between positions +15 to +90 may contain further positive regulatory information, as fragments 7 (position -196 to -31) and 8 (position -103 to +15) produced only negligible CAT activity when transfected into HeLa cells. In support of this assumption, Reddy (1983) has reported that deletion of sequences between the XmaIII sites at positions +19 and +324, drastically reduced the transforming activity of the c-Ha-ras1 gene from T24 cells. However, preliminary data, obtained by comparing the levels

of c-Ha-ras1 mRNA produced by a construct containing this 305 bp XmaIII deletion, with an unmutated construct, did not reveal any differences in the levels of mRNA produced when these constructs were transiently expressed in monkey Cos7 cells (data not shown). An alternative explanation for the reported CAT activity of the fragment between positions -149 to +90 (Ishii et al., 1986), could be because this fragment contains sequences known to be homologous to abundant non-polyadenylated RNAs (Results, section IV). It is possible that sequences between positions +15 and +90, may in fact contain information for transcribing a member of these abundant, non-polyadenylated RNAs and that this information is responsible for the CAT activity observed for the construct described by Ishii et al. (1986).

The high G + C content and the high frequency of restriction endonuclease sites which contain the CpG dinucleotide in the promoter region of the c-Ha-ras1 gene, suggests that this region is a good candidate for a HTF island (Cooper et al., 1983; Bird et al., 1985; Dr. Adrian Bird, personal communication). HTF islands are hypomethylated regions of genomic DNA known to be associated with unique, transcribed sequences (Introduction, section II). Circumstantial evidence for a regulatory role for HTF islands comes not only from their usual location in the 5' flanking sequences of "housekeeping-type" genes, but also from the observation that these regions are hypersensitive in the human glucose-6-phosphate dehydrogenase (G6DP) and the hypoxanthine phosphoribosyl transferase (HPRT) genes (Wolf and Migeon, 1980). Furthermore, in vitro methylation of the hamster adenine phosphoribosyl transferase (APRT) 5' region inhibited expression, whereas 3' methylation had no effect (Keshet et al., 1985). However, support for the hypothesis that the promoter region of the human c-Ha-ras1 gene constitutes a HTF island, awaits thorough examination of the methylation status of this region of the genome.

(b) The human c-Ha-ras1 promoter region functions bidirectionally

An unexpected result was obtained when pRasCAT1R, containing fragment 1 in the reverse orientation (Figure 2-2), was transiently expressed in HeLa cells. This construct was found to produce similar CAT activity to pRasCAT1D, in which fragment 1 is in the direct orientation (Figures 2-6 and 2-8). Thus the human c-Ha-ras1 promoter region contained in fragment 1, mapping from position -575 to +15, can function as a strong promoter directing CAT gene expression in HeLa cells regardless of its orientation. Dissection of this "reverse-orientation" promoter activity by subcloning the fragments shown in Figure 2-1 in the reverse orientation, allowed the sequences directing this activity to be further localized. Fragment 6, mapping between positions -392 and -196, produced a "reverse-orientation" promoter activity slightly greater than that observed for fragment 1 in the reverse orientation. Furthermore, fragments 5, 7 and 8 produced negligible CAT activities, confirming that the "reverse-orientation" promoter activity maps between position -392 and -196 relative to the donor splice site of exon -1 (Figure 2-8 and Figure 5-1). It is possible that the positively acting regulatory sequences previously identified between positions -243 and -196 for the "direct-orientation" promoter, may also have a regulatory function for the "reverse-orientation" promoter. This interesting possibility could be tested by a further mutational analysis of this region. However, it is more difficult to interpret CAT data determined for "reverse promoter" regulation, since transcription initiation sites have not been mapped for the "reverse-orientation" promoter. Therefore, it is not known whether lack of CAT activity, for a given fragment in the reverse-orientation, is the result of lack of initiation sites, regulatory elements, or both. Interestingly, fragment 6 has

"reverse-orientation" promoter activity only, presumably because this fragment lacks transcription initiation sites in the direct orientation. This result suggests that the "reverse-orientation" promoter activity can be functionally uncoupled from the "direct-orientation" promoter activity. Fragment 3 in the reverse-orientation produced only a "reverse-orientation" promoter activity of approximately 30% of the activity observed with fragment 6 and approximately 40% of the activity observed with fragment 1 (Figure 2-8 and Figure 5-1). An almost identical result was obtained with fragment 9 in the reverse-orientation (Figure 2-8). Thus, sequences important for the "reverse-orientation" promoter activity have been deleted in constructs pRasCAT3R and pRasCAT9R. Hence, the minimum sequence information required for the "reverse-orientation" promoter activity must lie between positions -276 to -196. Finer-scale mutational techniques and RNA mapping analyses would be necessary to determine more accurately which sequences contain information for transcriptional initiation and which contain regulatory information, for the "reverse-orientation" promoter activity.

The bidirectional nature of the human c-Ha-ras1 promoter region has raised the question of whether or not transcripts are initiated on the anti-sense strand in vivo. Furthermore, as previously discussed, it is possible that Honkawa et al. (1987) have actually identified an opposite-strand RNA in this region (Discussion, section I). However, 5' end-labelled sense strand probes, derived from the restriction fragments designated 1, 2 and 4 in Figure 2-1, did not detect transcripts with EJ poly A^+ RNA (data not shown). These data suggest that if opposite strand, divergently transcribed RNA species exist, they were not detected in the present study because they are either too short, non-polyadenylated, of low abundance or very unstable in vivo. This situation may be analogous to the mouse DHFR gene promoter region (described in the Introduction, section II.b), from which divergent, opposite

strand transcription has been reported (Farnham et al., 1985; Crouse et al., 1985). However, in this case some confusion exists as the opposite strand RNAs have been variously described as small (180-240 nucleotides), abundant in nuclear non-polyadenylated RNA, or as a long (at least 14,000 nucleotides) protein-encoding divergent transcription unit. Bidirectionally active promoter regions, other than the mouse DHFR promoter, are not without precedent, as the previously discussed SV40 300 bp and the HSV-1 700 bp control regions (Introduction, section II) are also bidirectional promoter regions, promoting transcription of opposite-strand RNAs from the same regulatory region. Also, Saffer and Singer (1984) have characterized a genomic fragment from monkey DNA which was isolated on the basis of homology to the SV40 control region. Both strands of this "SV40-like" fragment hybridized to transcripts from CV-1 monkey Kidney cells and multiple initiation sites were determined for both orientations of this fragment. All of these bidirectional promoter regions, like the human c-Ha-ras1 promoter, are G + C rich, contain multiple Sp1 binding sites and initiate transcription from multiple initiation sites. Furthermore, an indication that bidirectional promoter activity can also be associated with primitive genes, comes from the observation that both human mitochondrial promoters function bidirectionally (Chang et al., 1986).

Finally, it is worth speculating on putative functions for a divergently transcribed, opposite-strand RNA. The first possibility is that such an RNA may represent a divergently transcribed, opposite strand gene, such as is the case with the SV40 300 bp and the HSV-1 700 bp control regions. This possibilty is less likely in the case of the human c-Ha-ras1 gene, as no opposite-strand transcripts could be detected in poly A^+ RNA isolated from EJ and A431 cells. However, it is possible that such a gene may not be transcribed in these cell lines. Another possibility is that divergent transcripts may help to "protect" the gene

they are associated with, from insertional activation by agents such as viruses and transposons (Wood et al., 1984; Allan and Paul, 1984; Rohrbaugh et al., 1985). In this context, it is interesting that oncogenic activation of the c-Ha-ras1 proto-oncogene by retroviral insertion into adjacent sequences has yet to be reported. Alternatively, divergent transcripts may function as trans regulatory elements either by interfering with gene transcripts coming from further upstream or by facilitating a general domain opening of the region containing the gene.

III. THE EFFECT OF THE SV40 ENHANCER ON TRANSCRIPTION OF THE HUMAN c-Ha-ras1 GENE

When the SV40 enhancer was covalently linked 5' to the human c-Ha-ras1 gene (Figure 3-2) it was found to enhance transcription approximately 10 to 20 fold in transient expression assays performed in Cos7 cells. Furthermore, the enhancer effect was not specific to a subset of transcription initiation sites, rather, transcription was increased proportionately from each initation site (Figure 3-3). It is not known from this study whether a similar effect would be observed if the SV40 enhancer was located 3' to the c-Ha-ras1 gene, or if it was located at a greater distance. This analysis is the first direct demonstration that the SV40 enhancer element can affect transcription of genes with "housekeeping-type" promoters. Furthermore, this result is not surprising as similarities between the promoter regions of the SV40 early transcription unit and the c-Ha-ras1 gene have already been noted (Introduction, section II.a). However, the enhancement effect on each initiation site of the c-Ha-ras1 gene is rather small, compared to the enhancement observed when this enhancer is linked to genes of the "luxury-protein" class (e.g. the β-globin genes). Such CAAT and TATA box containing promoters are typically enhanced at least 50 to 200 fold,

usually at a single transcription initition site (de Villiers et al., 1982; Wasylyk et al., 1983 and Allan et al., 1984). However, as the c-Ha-ras1 gene has multiple transcription initiation sites, the additive effect of a 10 to 20 fold enhancement at each site is potentially very large.

An interesting implication of these results is that mechanisms by which transcription of this gene could be quantitatively increased in vivo, would include not only activation of a putative nearby enhancer (perhaps coupled with inhibition of opposite-strand transcription), but also insertion of an enhancer into the vicinity of the c-Ha-ras1 gene (e.g. by translocation of retroviral insertion). This speculation is important because levels of elevated c-Ha-ras1 transcription,as low as only 10 to 20 fold, have been implicated in carcinogenesis (Introduction, section I.c). Chang et al., (1982) have also shown that elevated expression of the c-Ha-ras1 gene when it was placed under the control of an enhancer from a retroviral long terminal repeat, resulted in tumourigenic transformation of murine NIH 3T3 cells.

IV. THE UPSTREAM INTRON CONTAINS REPETITIVE SEQUENCES

Northern blot analysis of c-Ha-ras1 RNA indicates that sequences within the upstream intron are homologous to abundantly transcribed species of RNA. These sequences map between the Sau3A site and the BstNI site, at positions -1079 and -623 respectively, relative to the ATG codon (the donor splice site of exon -1 is at position -1094). The abundant RNAs observed, produced a strong band slightly larger than the 28 S ribosomal RNA (rRNA) marker and other prominent bands at approximately 26 S, 20 S, 17 S and 14 S, which were superimposed on a relatively uniform smear. These bands do not correspond to the sizes expected for ribosomal RNAs and were observed with human total and poly

A^- RNAs (Figures 4-1 and 4-2), but only to a limited extent with human poly A^+ RNA (Figure 4-2). This suggests that the transcribed sequences homologous to the 5' half of the c-Ha-ras1 upstream intron are mainly present in the poly A^- RNA fraction. This observation may be explained if the human c-Ha-ras1 intronic sequences are fortuitously homologous to ribosomal RNAs. However, the banding pattern was not observed with EJ focus 8.3.5 total RNA, which is essentially mouse NIH 3T3 total RNA containing elevated levels of the human c-Ha-ras1 transcript. Therefore the human c-Ha-ras1 upstream sequences, between positions -1079 to -623, cannot be homologous to mouse ribosomal RNAs. Since ribosomal RNAs are known to be highly homologous between all organisms (Pace et al., 1986), it would be surprising that if the banding pattern is caused by ribosomal RNA species, that a similar pattern is not observed with EJ focus 8.3.5 total RNA. An explanation could be that there are sequence differences between mouse and human ribosomal RNAs and that the homology to the human c-Ha-ras1 intron fortuitously occurs at such a divergent region. Torczynski et al. (1985) have cloned and sequenced a human 18 S ribosomal RNA gene. Comparisons between the human and mouse ribosomal RNAs reveal that the only significant region of divergence between the two occured between the 5' terminal forty nucleotides which corresponds to the external transcribed spacer sequence. A less significant divergent region was also identified between bases 190 to 300, which was due to G or C single base insertions or deletions. Overall, the human and mouse 18 S ribosomal RNAs are 98.8% homologous to each other. Unfortunately no sequence information is available on the human 28 S ribsomal RNA sequence. Therefore, it is possible that if the bands below the major band in Figures 4-1 and 4-2 correspond to degradation products, that differences between mouse and human 28 S ribosomal RNA sequences could explain the differences observed when mouse and human RNAs are probed with the c-Ha-ras1 upstream intron sequences.

However, the bands below the major band (slightly greater than 28 S in Figures 4-1 and 4-2) have been observed with independent RNA preparations, in three separate experiments. This suggests that they indeed correspond to genuine discrete RNA species. Thus not all of these bands could be caused by hybridization to ribosomal RNA species, but rather to other predominantly G + C rich, mainly non-polyadenylated, human specific species of RNA.

The Southern blot analysis presented in Figure 4-3 suggested that sequences within the human c-Ha-ras1 upstream intron are also moderately repetitive. The repetitiveness of this region could be caused by fortuitous homologies to other G + C rich regions throughout the human genome. Furthermore, this Southern blot analysis also established that the sequences within the upstream intron of the human c-Ha-ras1 gene are homologous to sequences within the mouse genome, although the previous northern blot analyses established that homologous transcripts are not produced. These data suggest an alternative explanation for the origin of the homology to both genomic DNA and human RNAs. The homology could be due to a family of moderately repetitive G + C rich sequences in the human genome, which are transcribed. This homology could be fortuitous, or the sequences in the 5' portion of the c-Ha-ras1 upstream intron could represent a member of this putative family. However, mouse genomic DNA hybridized nearly as strongly to these sequences, as did human genomic DNA and computer searches of gene banks and the 22 Kb of 5' sequence flanking the human epsilon globin gene did not reveal any strong homologies. Finally, the recent cloning and sequencing of the mouse c-Ha-ras1 gene (Dr. Alan Balmain, personal communication), revealed that while approximately the 3' half of the mouse and human upstream introns are highly homologous, the 5' half of these introns diverge considerably. The lack of sequence homology in the 5' portions of these introns suggests that these sequences may not be functionally significant. However, these

sequence comparisons contrast somewhat with data reported, but not shown, by Reddy (1983). Reddy describes a 305 bp deletion between two XmaIII sites (mapping between positions +19 to +324 relative to the donor splice site of exon -1, whereas the repetitive region has been mapped here between positions +15 to +470 which is believed to drastically reduce the transforming activity of the c-Ha-ras1 oncogene isolated from the human bladder cell line ,T24. It is interesting that the repetitive 456 bp region identified in this thesis, closely corresponds to the 305 bp deletion decribed by Reddy (1983). Additionally, comparisons between the CAT activities produced by constructs described in this thesis and the CAT construct described by Ishii et al. (1986), which assayed a c-Ha-ras1 promoter fragment containing sequences between positions -149 to +90, suggested that sequences between positions +15 to +90 may contain a positive regulatory element(s).

CONCLUSIONS

1) A non-coding exon (termed exon -1) has been identified for the human c-Ha-*ras*1 gene. Exon -1 is located 1049 bp upstream of the ATG codon and consists of an approximately 200 bp region containing multiple transcription initiation sites. The existence of exon -1 and the locations of the transcription initiation sites have been confirmed by a variety of RNA mapping techniques (S1 nuclease mapping, primer extension and northern blot analysis) and functionally by CAT analysis. The implied intron between exon -1 and the ATG codon, is bounded by consensus donor and acceptor splice sites, at positions -1094 and -54 respectively, relative to the ATG codon.

2) Mutational analysis of exon -1 and the region of approximately 400 bp upstream of the most 5' transcription initiation site, using the CAT system, has identified a regulatory element mapping between positions -243 to -196, relative to the donor splice site. This element up-regulates transcription by approximately 4 fold and contains CAAT and GGGCGG sequence motifs, both of which are known to be bound by their respective DNA binding proteins (Ishii et al., 1986; Jones et al., 1987). A second positive regulatory element may also exist between positions -276 and -243, which appears to up-regulate transcription at least 2 fold. However, the evidence for this second regulatory element is more tentative. A putative negative regulatory element has also been identified. This element is located between positions -103 and -34, relative to the donor splice site and appears to down regulate promoter activity by up to 4 fold.

3) A further promoter region has been located by CAT analysis which maps between positions -392 to -196, relative to the donor splice site. This promoter has been shown to direct transcription in the opposite-orientation,

relative to the c-Ha-ras1 gene. The arrangement of the bidirectional promoter, situated on either side of a positive regulatory region (Figure 1-5), is strikingly similar to the 300 bp SV40 control region.

4) Insertion of the SV40 enhancer 5' to the c-Ha-ras1 gene was shown to increase transcription of each c-Ha-ras1 transcript by approximately 10 to 20 fold. This is the first direct demonstration that the SV40 enhancer can influence a "housekeeping-type" gene. Although each initiation site is increased in activity by only 10 to 20 fold, the total activity of the gene in terms of the protein produced is potentially very large. Insertion, or activation of a nearby enhancer could therefore be an important event in the oncogenic activation of this gene.

5) A region corresponding to the 5' half of the intron between exon -1 and exon 1 (between positions +15 to +471, relative to the donor splice site) was found to be moderately repetitive in the human and mouse genomes and homologous to abundant human, non-polyadenylated RNAs. These RNAs are sized between 29 S to 10 S. It is not known if this region of the intron between exon -1 and exon 1 is itself transcribed, or what its function, if any might be. Sequence comparisons with the "equivalent" region of the mouse c-Ha-ras1 gene, did not reveal any homology.

6) In summary, the work reported in this thesis has identified the major region of transcription initiation, for this important oncogene, and has suggested several mechanisms by which quantative regulation of this gene can be achieved. These may have important implications for the regulation of this gene and its oncogenic activation *in vivo*.

BIBLIOGRAPHY

Allan, M. and Paul, J. (1984). Transcription in vivo of an Alu family upstream from the human e-globin gene. Nucl. Acids Res. 12, 1193-1200.

Allan, M., Zhu, J., Montague, P. & Paul J. (1984). Differential response of multiple e-globin cap sites to cis and trans acting controls. Cell 38, 399-407.

Allan, M., Lanyon, G.W. & Paul, J. (1983). Multiple origins of transcription in the 4.5 Kb upstream of the e-globin gene. Cell 35, 187-197.

Atchison, M.L. & Perry, R.P. (1986). Tandem Kappa immunoglobin promoters are equally active in the presence of the Kappa enhancer: Implications for models of enhancer function. Cell 46, 253-262.

Baker, S.M. & Platt, T. (1986). PolI transcription: Which comes first, the end or the beginning. Cell 47, 839-840.

Balmain, A., Ramsden, M., Bowden, G.T. & Smith, J. (1984). Activation of the mouse cellular Harvey-ras gene in chemically induced benign skin papillomas. Nature 307, 658-660.

Balmain, A. & Pragnell, I.B. (1983). Mouse skin carcinomas induced in vivo by chemical carcinogens have a transforming Harvey-ras oncogene. Nature 303, 72-74.

Banerji, J., Rusconi, S. & Schaffner, W. (1981). Expression of a globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Banerji, J., Olson, L. & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobin heavy chain genes. Cell 33, 729-740.

Bar-Sagi, D. & Feramisco, J.R. (1985). Microinjection of the ras oncogene protein into PC12 cells induces morphological differentiation. Cell 42, 841-848.

Beckner, S.K., Hattori, S. & Shih, T.Y. (1985). The ras oncogene product p21 is not a regulatory component of adenylate cyclase. Nature 317, 71-72.

Benoist, C. & Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature 290, 304-310.

Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). The ovalbumin gene-sequence of putative control regions. Nucl. Acids Res. 8, 127-142.

Berk, A.J. & Sharp, P.A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12, 721-732.

Berk, A.J. & Sharp, P.A. (1978). Structure of the adenovirus 2 early mRNAs. Cell 14, 695.

Berridge, M.J. (1984) Inositol triphosphate and diacylglycerol as second messengers. Biochem. J. 220, 345-360.

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

Bienz, M. (1984). Developmental control of the heat-shock response in Xenopus. Proc. Natl. Acad. Sci. USA 81, 3138-3142.

Bienz, M. (1986). A CCAAT box confers cell-type-specific regulation on the Xenopus hsp70 gene in oocytes. Cell 46, 1037-1042.

Bienz, M. & Pelham, H.R.B. (1986). Heat shock regulatory elements function as an inducible enhancer in the Xenopus hsp70 gene and when linked to a heterlogens promoter. Cell 45, 753-760.

Birnboim, H.C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7, 1513.

Birchmeier, C., Broek, D. & Wigler, M. (1985). Ras proteins can induce meiosis in Xenopus oocytes. Cell 43, 615-621.

Bird, A.P. (1985). DNA methylation and the frequency of CpG in animal DNA. Nucl. Acids Res. 12, 4385-4396.

Bird, A. Taggart, M., Frommer, M., Miller, D.J. & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40, 91-99.

Bishop, J.M. (1985). Viral oncogenes. Cell 42, 23-38.

Bohmann, D., Keller, W., Dale, T., Scholer, H.R., Tebb, G. & Mattaj, I.W. (1987). A transcription factor which binds to the enhancers of SV40 , immunoglobin heavy chain and U2 snRNA genes. Nature 325, 268-272.

Bos, J.L., Verlaan-deVries, M., Marshall, C.J., Veeneman, G.H., vanBoom, J.H. & van der Eb, A.J. (1986). A human gastric carcinoma contains a single mutated and an amplified normal allele of the Ki-ras oncogene. Nucl. Acids Res. 14, 1209-1217.

Boyer, H.N. & Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. 41, 459-472.

Brady, J., Bolen, J.B., Radonovich, M., Salzman, N. & Khoury, G. (1984). Stimulation of Simian virus 40 late gene expression by Simian virus 40 tumor antigen. Proc. Natl. Acad. Sci. USA 81, 2840-2044.

Breathnach, R. & Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349-383.

Broek, D., Samily, N., Fasano, O., Fujiyama, A., Tamanoi, F., Northup, J. & Wigler, M. (1985). Differential activation of yeast adenylate cyclase by wild-type and mutant RAS proteins. Cell 41, 763-769.

Brown, K., Quintanilla, M., Ramsden, M., Kerr, I.B., Young, S. & Balmain, A. (1986). V-ras Genes from Harvey and BALB murine sarcoma viruses can act as initiators of two-stage mouse skin carcinogenesis. Cell 46, 447-456.

Burns, F.J., Vanderlaan, M., Snyder, E. & Albert, R.E. (1978). Induction of progression kinetics of mouse skin papillomas. In. Carcinogenesis, ed. Slaga, T.S., Sivak, A. & Butwell, R.K. Vol. 65, pp 91. Raven Press, New York.

Cairns, J. (1975) Mutation selection and the natural history of cancer. Nature 255, 197-200.

Campisi, J., Gray, H.E., Pardee, A.B., Dean, M. & Sonenshein, G. (1984). Cell-cycle control of c-myc but not c-ras expression is lost following chemical transformation. Cell 36, 241-247.

Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H. & Goeddel, D.V. (1983a). Complete nucleotide sequence of the T24 human bladder carcinoma oncogene and its normal homologue. Nature 302, 33-37.

Capon, D.J., Seeburg, P.H., McGrath, J.P., Hayflick, J.S., Edman, L.L., Levinson, A.D. & Goeddel, D.V. (1983b). Activation of Ki-ras2 gene in human colon and lung carcinomas by two different point mutations. Nature 304, 507-513.

Carter, A.D., Felber,B.K., Jubier,M.F., Schmidt,C.J. & Hamer, D.H. (1984). Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc. Natl. Acad. Sci. USA 81, 7392-7396.

Casey, J. & Davidson, N. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high

concentrations of formamide. Nucl. Acids Res. 4, 1539-1552.

Casey, J. & Davidson, N. (1978). RNA:DNA hybrids are more stable than DNA:DNA duplexes in concentrated perchlorate and trichlorate solutions. Nucl. Acids Res. 5, 1627-1637.

Chang, E.H., Furth, M.E., Scolnick, E.M. & Lowy, D.R. (1982). Tumorigenic transformation of mammalian cells induced by a normal human gene homologous to the oncogene of Harvey murine sarcoma virus. Nature 297, 479-483.

Chang, D.D., Hixson, J.E. & Clayton D.A. (1986). Minor transcription initiation events indicate that both human mitochondrial promoters function bidirectionally. Mol. Cell. Biol. 6, 294-301.

Charnay, P., Mellon, P. & Maniatis, T. (1985). Linker-scanning mutagenesis of the 5'-flanking region of the mouse b-major globin genes: Sequence requirements for transcription in erythroid and non-erythroid cells. Mol. Cell. Biol. 5, 1498-1511.

Chase, J.N. & Richardson, C.C. (1964). Exonuclease VII of Escherichia coli. J. Biol. Chem. 249, 4545.

Chichutek, K. & Duesberg, P.H. (1986). Harvey ras genes transform without mutant codons apparently activated by truncation of a 5' exon (exon -1). Proc. Natl. Acad. Sci. USA 83, 2340-2344.

Chirgwin, J.M., Przybyla, A.E, MacDonald, R.J. & Rutter, W. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299.

Choi, O.R. & Engel, J.D. (1986). A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult b-globin gene. Nature 323, 731-734.

Cockerill, P.N. & Garrard, W.T. (1986). Chromosomal loop anchorage of the Kappa immunoglobin gene occurs next to the enhancer in a region continuing topoisomerase II sites. Cell 44, 273-282.

Cohen, R.B., Sheffrey, M. & Kim, G.C. (1986). Partial purification of a nuclear protein that binds to the CCAAT box of the mouse a-1 globin gene. Mol. Cell. Biol. 6, 821-832.

Cooper, D.N., Taggart, M.H. & Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. Nucl. Acids Res. 11, 647-658.

Crouse, G.F., Leys, E.J., McEwan, R.N., Frayne, E.G. & Kellems, R.E. (1985). Analysis of the mouse dhfr promoter region: Existence of a divergently transcribed gene. Mol.

Cell Biol. 5, 1847-1838.

Davidson, B.L., Elgy, J.M., Mulvihill, E.R. & Chambon, P. (1983). Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature 301, 680-686.

DeFeo, D., Gonda, M.A., Young, H.A., Chang, E.H., Lowy, D.R., Scolnick, E.M. & Ellis, R.W. (1981). Analysis of two divergent rat genomic clones homologous to the transforming gene of Harvey murine sarcoma virus. Proc. Natl. Acad. Sci. USA 78, 3328-3332.

DeFeo-Jones, D., Scolnick, E.M., Koller, R. & Dhar, R. (1983). Ras-related gene sequences identified and isolated from Saccharomyces cerevisiae. Nature 306, 707-709.

DeFoe-Jones, D., Tatchell, K., Robinson, L.C., Sigal, I.S., Vass, W.C., Lowy, D.R. & Scolnick, E.M. (1985). Mammalian and yeast ras gene products: biological function in their heterologous systems. Science 228, 179-184.

Der, C.J., Finkel, T. & Cooper, G.M. (1986). Biological and biochemical properties of human ras genes mutated at codon 61. Cell 44, 167-176.

DeVilliers, J., Olson, L., Banerji, J. & Schaffner, W. (1982). Analysis of the transcriptional enhancer effect. Cold Spring Harbour Symp. Quant. Biol. 47, 911-921.

Dhar, R., Ellis, R.W., Shih, T.Y., Droszlan, S., Shapiro, B., Maizel, et. al. (1982). Nucleotide sequence of the p21 transforming protein of harvey murine Sarcoma virus. Science 217, 934-937.

Dierks, P., van Oogen, A., Cochran, M.D., Dobkin, C., Reiser, J. & Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit b-globin gene in mouse 3T6 cells. Cell 32, 695-706.

Dush, M.K., Sikela, J.M., Khan, S.A., Tischfield, J.A. & Stambrook, P.S. (1985). Nucleotide sequence and organization of the mouse adenine phosphoribosyltransferase gene: Presence of a coding region common to animal and bacterial phosphoribosyltransferase that has a variable intron/exon arrangement. Proc. Natl. Acad. Sci. USA 82, 2731-2735.

Dynan, W.S., & Tjian, R. (1983a). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell 32, 669-680.

Dynan, W.S. & Tjian, R. (1983b). The promoter-specific transcription factor Sp1 binds to upstream sequences in the

SV40 early promoter. Cell 35, 79-87.

Dynan, W.S. & Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. Nature 316, 774-778.

Ellis, R.W., DeFeo, D., Maryak, J.M., Young, H.A., Shih, T.Y., Chang, E.H., Lowy, D.R. & Scolnick, E.M. (1980). Dual evolutionary origin for the rat genetic sequences of Harvey murine sarcoma virus. J. Virol. 36, 408-420.

Ellis, R.W., Defeo, D., Shih, T.Y., Grouda, M.A., Young, H.A., Tsuchida, N., Lowy, D.R. & Scolnick, E.M. (1981). The p21 src genes of Harvey and Kirsten sarcoma viruses originate from divergent members of family of normal vertebrate genes. Nature 292, 506-511.

Enver, T. (1985). A pulling out of fingers. Nature 317, 385-386.

Farnham, P.J., Abrams, J.M. & Schimke, R.T. (1985). Opposite-strand RNAs from the 5' flanking region of the mouse dihydrofolate reductase gene. Proc. Natl. Acad. Sci. USA 82, 3978-3982.

Favaloro, J., Treisman, R.H. & Kamen, R.J. (1980). Transcription maps of poloma virus-specific RNA:analysis by two-dimensional nuclease S1 gel mapping. Meth. Enzymol. 65, 718-749.

Feinberg, A.P. & Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Analytical Biochemistry 132, 6-13.

Feinberg, A.P. & Vogelstein, B. (1983a). Analytical Biochemistry 137, 266-267.

Ferramisco, J.R., Gross, M., Kamata, T., Rosenberg, M. & Sweet, R.W. (1984). Microinjection of the oncogene form of the human H-ras (T24) protein results in rapid proliferation of quiescent cells. Cell 38, 109-117.

Fleischman, L.F., Chahwala, S.B. & Cantley, L. (1986). Ras-transformed cells: altered levels of phosphatidylinositol 4-,5-biphosphate and catabolities. Science 231, 407-410.

Foulds, L. (1969). Neoplastic Development Vol. 1, Academic Press, London.

Fromm, M. & Berg, P. (1982). Deletion mapping of DNA regions required for SV40 early region promoter function in vivo. J. Mol. Appl. Genet. 1, 457-481.

Fromm, M. & Berg, P. (1983). Simian virus 40 early- and late- region promoter functions are enhanced by the 72 base

pair repeat inserted at distant locations and inverted orientations. Mol. Cell. Biol. 3, 991-999.

Fukui, Y. & Kaziro, Y. (1985). Molecular cloning and sequence analysis of a ras gene from Schizosaccharomyces pombe. EMBO J. 4, 687-691.

Fukui, Y., Kozasa, T., Takeda, T. & Yamamoto, M. (1986). Role of a ras homolog in the life cycle of Schizosaccharomyces pombe. Cell 44, 329-336.

Gaffney, D.F., McLauchlan, J., Whitton, J.L. & Clements, J.B. (1985). A modular system for the assay of transcription regulatory signals: the sequence TAATGARAT is required for herpes simplex virus immediate early gene activation. Nucl. Acids Res. 13, 7847-7863.

Gallick, G.E., Kurzrock, R., Kloetzer, W.S., Arlinghaus, R.B. & Gutterman, J.U. (1985). Expression of $p21^{ras}$ in fresh primary and metastatic human colorectal tumors. Proc. Natl. Acad. Sci. USA 82, 1795-1799.

Gallwitz, D., Donath, C. & Sander, C. (1983). A yeast gene encoding a protein homologous to the human c-has/bas proto-oncogene product. Nature 306, 704-707.

George, D.L., Scott, A.F., Trusko, S., Gilick, B., Ford, E. & Dorney, D.J. (1985). Structure and expression of amplified c-Ki-ras gene sequences in Y1 mouse adrenal tumour cells. EMBO J. 4, 1199-1203.

Gerhard, D.S., Dracopoli, N.C., Bale, S.J., Houghton, A.N., Watkins, P., Payne, C.E. et al. (1987). Evidence against Ha-ras1 involvement in sporadic and familial melanoma. Nature 325, 73-75.

Ghosh, P.K., Lebowitz, P., Frisque, R.J. & Gluzman, Y. (1981). Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. USA 78, 100-104.

Gibbs, J.B., Sigal, I.S. & Scolnick, E.M. (1985). Biochemical properties of normal and oncogenic ras p21. Trends Biochem. Sci. 10, 350-353.

Gidoni, D., Dynan, W.S. & Tjian, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature 312, 409-413.

Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P. & Tjian, R. (1985). Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science 230, 511-517.

Gillies, S.D., Morrison, S.L., Oi, V.T. & Tonegawa, S. (1983). A tissue-specific transcription enhancer element is

located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33, 717-728.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23, 175-182.

Goldfarb, M., Shimizu, K., Perucho, M. & Wigler, M. (1982). Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. Nature 296, 404-409.

Goyette, M., Petropoulos, C.J., Shank, P.R. & Fausto, N. (1983). Expression of a cellular oncogene during liver regeneration. Science 219, 510-512.

Goyette, M., Petropoulos, C.J., Shank, P.R. & Fausto, N. (1984). Regulated transcription of c-Ki-ras and c-myc during compensatory growth of rat liver. Mol. Cell. Biol. 4, 1493-1498.

Graves, B.J., Johnson, P.F. & McKnight, S.L. (1986). Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell 44, 565-576.

Grosschedl, R. & Birnsteil, M.L. (1981). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. USA 77, 1432-1436.

Grosveld, G.C., deBaer, E., Shewmaker, C.K. & Flavell, R. (1982). DNA sequences necessary for transcription of the rabbit b-globin gene in vivo. Nature 295, 120-126.

Hall, A. & Brown, R. (1985). Human N-ras: cDNA cloning and gene structure. Nucl. Acids Res. 13, 5255-5268.

Hall, A., Marshall, C., Spurr, N. & Weiss, R. (1983). Identification of transforming genes in two human cell lines as a new member of the ras gene family located on chromosome 1. Nature 303, 396-400.

Haslinger, A. & Karin, M. (1985). Upstream promoter element of the human metallothionein-II_A gene can act like an enhancer element. Proc. Natl. Acad. Sci. USA 82, 8572-8576.

Hess, J., Perez-Stable, C., Deisseroth, A. & Shen, C.K. (1985). Characterization of an unique RNA initiated immediately upstream from the human alpha-1 globin gene in vivo and in vitro: polymerase II-dependence, tissue specificity and subcellular location. Nucl. Acids Res. 13, 6059-

Honkawa, H., Masahashi, S., Hashimoto, S. & Hashimoto-Gotoh, T. (1987). Identification of the principal promoter sequence of the c-H-ras transforming oncogene: Deletion analysis of the 5'-flanking region by focus

forming assay. Molec. Cell. Biol. 7, 2933-2940.

Horan-Hand, P.H., Thor, A., Wunderlich, D., Muraro, R., Caruso, A. & Sehlom, J. (1984). Monoclonal antibodies of predefined specificity detect activated ras gene expression in human mammary and colon carcinomas. Proc. Natl. Acad. Sci. USA 81, 5227-5231.

Hurley, J.B., Simon, M.I., Teplow, D.B., Robishaw, J.D. & Gilman, A.G. (1984) Homologies between signal transducing G-proteins and ras gene products. Science 226, 860-862.

Ishii, S., Kadonaga, J.T., Tjian, R., Brady, J.N., Merlino, G.T. & Pastan, I. (1986) Binding of the Sp1 transcription factor by the human Harvey ras1 proto-oncogene promoter. Science 232, 1410-1413.

Ishii, S., Merlino, G.T. and Pastan, I. (1985) Promoter region of the human Harvey ras proto-oncogene: Similarity to the EGF receptor proto-oncogene promoter. Science 230, 1378-1381.

Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J. & Tjian, R. (1987) A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48, 79-89.

Jones, K.A. & Tjian, R. (1985) Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate-early' gene transcription in vitro. Nature 317, 179-182.

Jones, K.A., Yamamoto, K.R. & Tjian, R. (1985) Two distinct transcription factors bind to the HSV thymidine Kinase promoter in vitro. Cell 42, 559-572.

Kadonaga, J.T., Jones, K.A. & Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. 11, 20-23.

Kadonaga, J.T. & Tjian, R. (1986) Affinity purification of sequence-specific DNA binding proteins. Proc. Natl. Acad. Sci. USA 83, 5889-5893.

Kagimoto, M.K., Miyoshi, J., Tashiro, K., Naito, Y., Sakaki., Y., Sueishi, K., et al., (1985). isolation and characterization of an activated c-H-ras1 gene from a sequamous cell lung carcinoma line. Int. J. Cancer 35, 809-812.

Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P. Westphal, H.M. & Beato, M. (1984) Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-II_A gene. Nature 308, 513-519.

Karlsson, S. & Neinhuis, A.W. (1985) Developmental regulation of human globin genes. Ann. Rev. Biochem. 54, 1071-1108.

Kataoka, T., Powers, S., Cameron, S., Fasano, O., Goldfarb, M., Broach, J. & Wigler, M. (1985) Functional homology of mammalian and yeast RAS genes. Cell 40, 19-26.

Kataoka, T., Powers, S., McGill, C., Fascino, O., Strathern, J., Broach, J. & Wigler, M. (1984) Genetic analysis

Keshet, I., Yhisraeli, J. & Cedar, H. (1985) Effect of regional DNA methylation on gene expression. Proc. Natl. Acad. Sci. USA 82, 2560-2564.

Klein, S., Sablitzky, F. & Radbruch, A. (1984) Deletion of the I_gH enhancer does not reduce immunoglobulin heavy chain production of a hybridoma IgD class switch variant. EMBO J. 3, 2473-2476.

Krontiris, T.G., DiMartino, N.A., Colb, M. & Parkinson, D.R. (1985) Unique allelic restriction fragments of the human Ha-ras locus in leukocyte and tumour DNAs of cancer patients. Nature 313, 369-374.

Land, H., Parada, L.F. & Weinberg, R.A. (1983) Tumorigenic conversion of primary embryo fibroblast requires at least two cooperating oncogenes. Nature 304, 596-602.

Land, H., Parada, L.F. & Weinberg, R.A. (1983) Cellular oncogenes and multistep carcinogenisis. Science 222, 771-778.

Lee, W., Haslinger, A., Karin, M. & Tjian, R. (1987) Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature 325, 368-327.

Lefkowitz, R.J.M Stadd, J.M. & Caron, M.G. (1983) Adenylate cyclase-coupled beta-adrenergic receptors: structure and mechanisms of activation and desensitization. Ann. Rev. Biochem. 52, 159-186.

Levinson, A.D. (1986) Normal and activated ras oncogenes and their encoded products. Trends in Genetics 2, 81-85.

Levanon, D., Lieman-Harwitz, J., Dafni, N., Wigderson, M., Sherman, L., Bernstein, Y., Laver-Rudich, Z., et al. (1985) Architecture and anatomy of the chromosomal locus in human chromosome 21 encoding the Cu/Zn superoxide dismutase. EMBO J. 4, 77-84.

Lochrie, M.A., Hurley, J.B. & Simon, M.I. (1985) sequence of the alpha subunit of photoreceptor G protein: Homologies between transducin, ras and elongation factors. Science 228, 96-99.

Loomis, W.F. (1975) The development of Dictyostelium discoideum (New York: Academic Press).

Lov, Z., Kinchie, Z., Hessel, R. & Seger, O. (1985) Expression of ras cellular oncogenes during development of Drosophila melanogaster. Mol. Cell. Biol. 5, 1540-1542.

Lowe, D.G., Capon, D.J., Delwart, E., Sakaguchi, A.Y., Naylor, S.L. & Goeddel, D.V. (1987) Structure of the human and murine R-ras genes, novel genes closely related to ras proto-oncogenes. Cell 48, 137-146.

Luchnik, A.N. (1985) Long-distance signal transfer in transcriptionally active chromatin. How does it occur? BioEssays 3, 249-252.

Luskey, M. & Botchan, M. (1981) Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature 293, 79-81.

Madaule, P. & Axel, R. (1985) A novel ras-related gene family. Cell 41, 31-40.

Magag, N., Halegona, S. & Viola, M. (1986). Inhibition of growth factor-induced differentiation of PC12 cells by microinjection of antibody to ras p21. Nature 319, 680-682.

Maniatis, T., Fritch, E.F. & Sambrook, J. (1982) In Molecular Cloning, A Laboratory Manual, Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Manley, J.L., Fire, A. Cano, A., Sharp, P. & Gefter, M.L. (1980) DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA 77, 3855-3859.

Marshall, C.J., Franks, L.M. & Carbonell, A.N. (1977) Markers of neoplastic transformation in epithelial cell lines derived from human carcinomas. J. Nat. Cancer Inst. 58, 1743-1747.

Matsumoto, K., Uno, I., Toh-e, A., Ishikawa, T. & Oshima, Y. (1982) Cyclic AMP may not be involved in catabolic repression in Saccharomyces cervisiae: evidence from mutants capable of utilizing it as an adenine source. J. Bacteriol. 150, 277-285.

Mason, J.O., Williams, G.T. & Neuberger, M.S. (1985) Transcription cell type specificity is conferred by an immunoglobulin V_H promoter that includes a functional concensus sequence. Cell 41, 479-487.

Maxam, A.M. & Gilbert, W. (1980) Sequencing end-labelled DNA with base-specific chemical cleavages. In Methods in Enzymology, ed. Grossman, L. & Moldave, K. vol. 65,

pp521-560. Academic Press, New York.

McGrath, J.P., Capon, D.J., Smith, D.H., Chen, E.Y., Seeburg, P.H., Goeddel, D.V. & Levinson, A.D. (1983) Structure and organization of the human Ki-ras proto-oncogene and a related processed pseudogene. Nature 304, 501-506.

McGrogan, M., Simonsen, C.C., Smouse, D.T., Farnham, P.J. & Schimke, R.T. (1985) Heterogeneity at the 5' termini of mouse dihydrofolate reductase mRNAs. J. Biol. Chem. 260, 2307-2314.

McKean, C., OhKubo, H., Pastan, I. & de Crombruggie, B. (1982). Unusual methylation pattern of the alpha 2(I) collagen gene. Cell 29, 203-210.

McKnight, S.L. (1982) Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell 31, 355-365.

McKnight, S.L., Gravis, E.R., Kingsbury, R. & Axel, R. (1981) Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell 25, 385-398.

McKnight, S.L. & Kingsbury, R.C. (1982) Transcriptional control signals of a eukaryotic protein-coding gene. Science 217, 316-324.

McKnight, S.L., Kingsbury, R.C., Spence, A. & Smith, M. (1984) The distal transcription signals of the herpes virus tk gene share a common hexanucleotide control sequence. Cell 37, 253-262.

McKnight, S. & Tjian, R. (1986) Transcriptional selectivity of viral genes in mammalian cells. Cell 46, 795-805.

Medynski, D.C., Sullivan, K., Smith, D., VanDop, C., Chang, F., Fung, B.K.K., Seeburg, P.H. & Bourne, H.R. (1985) Amino acid sequence of the alpha subunit of transducin deduced from the cDNA sequence. Proc. Natl. Acad. Sci. USA 82, 4311-4315.

Mellon, P., Parker, V., Gluzman, Y. & Maniatis, T. (1981) Identification of DNA sequences required for transcription of the human a1-globin gene in a new SV40 host-vector system. Cell 27, 279-288.

Melton, D.N., Kaneeki, D.S., Brennard, J. & Laskey, T.C. (1984) Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA 81, 2147-2151.

Miesfeld, R., Rusconi, S., Godowski, P.J., Maler, B.A., Okret, S., Wikstrom, A.C., Gustafsson, J.A. & Yamamoto,

K.R. (1986) Genetic complementaion of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. Cell 46, 389-399.

Mizusawa, S., Nishimura, S. & Seela, F. (1986) Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucl. Acids Res. 14, 1319-1324.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. & Chambon, P. (1981) The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucl. Acids Res. 9, 6047-6067.

Moser, B., Marlor, R., Parkhurst, S. & Corces, V. (1985) Characterization and developmental expression of a Drosophila ras oncogene. Mol. Cell. Biol. 5, 885-889.

Mulcahy, L.S., Smith, M.R. & Stacey, D.W. (1985) Requirement for ras proto-oncogene function during serum stimulated growth of N1H3T3 cells. Nature 313, 241-243.

Muller, R., Slamon, D.J., Adamson, E.D., Tremblay, J.M., Miller, D., Cline, M.J. & Verma, C.M. (1983) Transcription of c-onc genes c-ras^{Ki} and c-fms during mouse development. Mol. Cell. Biol. 3, 1062-1069.

Muller, R., Slamon, D.J., Tremblay, J.M., Cline, M.J. & Verma, C.M. (1982) Differential expression of cellular oncogenes during pre- and postnatal development of the mouse. Nature 299, 640-644.

Myers, R.M., Tilly, K. & Maniatis, T. (1986) Fine structure genetic analysis of a b-globin promoter. Science 232, 613-618.

Nadin-Davis, S.A., Yang, R.C.A., Narang, S.A. & Nasim, A. (1986) The cloning and characterization of a ras gene from Schizosaccharomyces pombe. J. Mol. Evol. 23, 41-51.

Neuman-Silberberg, F.S., Schejter, E., Hoffman, F.M. & Shilo, B.Z. (1984) The Drosophila ras oncogenes: structure and nucleotide sequence. Cell 37, 1027-1033.

Nevins, J.R. (1983) The pathway of eukaryotic mRNA formation. Ann. Rev. Biochem. 52, 441-466.

Newark, P. (1985) Events at the surface of the cell. Nature 317, 380.

Newbold, R.F. & Overall, R.W. (1983) Fibroblast immortality is a prerequisite for transformation by EJ c-Ha-ras oncogene. Nature 304, 648-651.

O'Brien, S.J., Nash, N.G., Goodwin, J.L., Lowy, D.R. & Chang, E.H. (1983) Dispersion of the ras family of

transforming genes to four different chromosomes in man. Nature 302, 837-841.

O'Toole, C.M., Povey, S., Hepburn, P. & Franks, L.M. (1983) Identity of some human baldder cancer cell line. Nature 301, 429-430.

Pace, N.R., Olsen, G.J. & Woese, C.R. (1986) Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell 45, 325-326.

Padua, R., Barrass, N. & Currie, G. (1984) A novel transforming gene in a human malignant melanoma cell line. Nature 311, 671-673.

Parada, L.F., Tabin, C.J., Shih, C. & Weinberg, R.A. (1982) Human EJ bladder carcinoma oncogene is a homologue of Harvey sarcoma virus ras gene. Nature 297, 474-479.

Pardee, A.B., Dubrow, R., Hamlin, J.L. & Kletzien, R.F. (1978) Animal Cell Cycle. Ann. Rev. Biochem. 47, 715-750.

Parker, C.S. & Topol, J. (1984) A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp70 gene. Cell 37, 273-283.

Pawson, T., Amiel, T., Hinge, E., Auersperg, N., Neave, N., Sobolewski, A. & Weeks, G. (1985) Regulation of a ras-related protein during development of Dictyostelium discoideum. Mol. Cell. Biol. 5, 33-39.

Peto, R. (1977) Epidemiology, multistage models and short term mutagenicity tests. In Origins of Human Cancer pp 1403-1428, Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Plon, S.E. & Wang, J.C. (1986) Transcription of the human b-globin gene is stimulted by an SV40 enhancer to which it is physically linked but topologically uncoupled. Cell 45, 575-580.

Powers, S., Kataoka, T., Fasano, O., Goldfarb, M. Strathern, J. Broach, J. & Wigler, M. (1984) Genes in S. cerevisiae encoding proteins with domains homologous to the mammalian ras proteins. Cell 36, 607-612.

Preston, C.M., Cordingley, M.G. & Stow, N.D. (1984) Analysis of DNA sequences which regulate the transcription of a Herpes Simplex virus immediate early gene. J. Virol. 50, 708-716.

Proudfoot, N.J., Rutherford, T.R. and Partington, G.A. (1984). Transcriptional analysis of human zeta globin genes. EMBO J. vol 3, 1533-1540.

Ptashe, M. (1986) Gene regulation by proteins acting nearby

and at a distance. Nature 322, 697-701.

Pulciani S., Santos, E., Lauver, A.V., Long, L.K., Robbins, K.C. & Barbacid, M. (1982) Oncogenes in human tumour cell lines: Molecular cloning of a transforming gene from human bladder carcinoma cells. Proc. Nat. Acad. Sci. USA 79, 2845-2849.

Pulciani, S., Santos, E., Long, L.K., Sorrentino, V. & Barbacid, M. (1985) ras gene amplification and malignant transformation. Mol. Cell. Biol. 5, 2836-2841.

Rabitts, P.H., Watson, J.V., Lamond, A., Forster, A., Stinton, M.A., Evan, G., Fisher, W., Atherton, E., Sheppard, R. & Rabbitts, J.H. (1985) Metabolism of c-myc gene products: c-myc RNA and protein expression in the cell cycle. EMBO J. 4, 2009-2015.

Reddy, E.P. (1983) Nucleotide sequence analysis of the T24 human bladder carcinoma oncogene. Science 220, 1061-1063.

Reymond, C.D., Gomer, R.H., Mehdy, M.C. & Firtel, R.A. (1984) Developmental regulation of a Dictyostelium gene encoding a protein homologous to mammalian ras protein. Cell 39, 141-148.

Reynolds, G.A., Basu, S.K., Osborne, T.S., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. & Luskey, K.L. (1984) HMG CoA reductase: A negatively regulated gene with unusual promoter and 5' untranslated regions. Cell 38, 275-285.

Richards, R.I., Hegny, A. & Karin, M. (1984) Structural and functional analysis of the human metallothionein-I_A gene: Differential induction by metal ions and glucocorticoids. Cell 37, 263-272.

Rigby, P.W.J., Dieckmann, M., Rhoquades, C. & Berg, P. (1977) Labelling deoxyribonucleic acid to high specific activity in vitro with DNA polymerase I. J. Mol. Biol. 113, 521-560.

Rohrbaugh, M.L., Johnson, J.E., James, M.D. and Hardison, R.C. (1985) Molec. Cell. Biol. 5, 147-160.

Ruley, E. (1983) Adenovirus early region IA enables viral and cellular transforming genes to transform primary cells in culture. Nature 104, 602-606.

Ryan, J., Barker, P.E., Shimizu, K., Wigler, M. & Ruddle, F.H. (1983) Chromosomal assignment of a family of human oncogenes. Proc. Natl. Acad. Sci. USA 80, 4460-4463.

Saffer, J. & Singer, M. (1984) Transcription from SV40-like monkey DNA sequences. Nucl. Acids Res. 12, 4769-4788.

Sakaguchi, A.Y., Naylor, S.L., Shows, T.B., Toole, J.J.,

McCoy, M. & Weinberg, R.A. (1983) Human c-Ki-ras2 proto-oncogene on chromosome 12. Science 219, 1081-1083.

Salser, W. (1977) Globin mRNA sequences: Analysis of base pairing and evolutionary implications. Cold Spring Harbour Symp. Quantit. Biol. 42, 985-1002.

Sassone-Corsi, P., Wildeman, A. & Chambon, P. (1985) A trans-acting factor is responsible for the simian virus 40 enhancer activity in vitro. Nature 313, 458-463.

Schejter, E.D. & Shilo, B.Z. (1985) Characterization of functional domains of p21ras by use of chimeric genes. EMBO J. 4, 407-412.

Schmitt, H.D., Wagner, P. Pfaff, E. & Gallwitz, D. (1986) The ras- related YPT1 gene product in yeast: A GTP-binding protein that might be involved in microtubule organization. Cell 47, 401-412.

Schwab, M., Alitalo, K., Varmus, H.E., Bishop, M. & George, D. (1983) A cellular oncogene (c-Ki-ras) is amplified, overexpressed and located within Karyotypic abnormalities in mouse adrenocortical tumour cells. Nature 303, 497-501.

Scolnick, E.M., Papageorge, A.G. & Shih, T.Y. (1979) Guanine nucleotide-binding activity as an assay for src protein of rat-derived murine sarcoma viruses. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Seeburg, P.H., Colby, W.W., Capon, D.J. Goeddel, D.V. & Levinson, A.D. (1984) Biological properties of human c-Ha-ras1 genes mutated at codon 12. Nature 312, 71-75.

Selleck, S.B. & Majors, J. (1987) Photofootprinting in vivo detects transcription-dependent changes in yeast TATA boxes. Nature 325, 173-177.

Serfling, E., Jasin, M. & Schnaffer, W. (1985) Enhancers and eukaryotic gene transcription. Trends Genet. 1, 224-230.

Shenk, T. (1981) Transcriptional control regions: Nucleotide sequence requirements for initiation by RNA polymerase II and III. Curr. Top. Microbiol. Immunol. 93, 25-46.

Shih, C. & Weinburg, R.A. (1982) Isolation of a transforming sequence from a human bladder carcinoma cell line. Cell 29, 161-169.

Shilo, B-Z. & Weinberg, R.A. (1981) DNA sequences homologous to vertebrate oncogenes are conserved in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 78, 6789-6792.

Shimizu, K., Birnbaum, D., Ruley, M., Fasano, O., Suard, Y., Edlund, L., Taparowsky, E., Goldfarb, M. & Wigler, M. (1983a) The structure of the K-ras gene of the human luung cell carcinoma cell line Calu-1. Nature 304, 497-500.

Shimizu, K., Goldfarb, M., Suard, Y., Perucho, M., Li., Y., Kamata, T., et al. (1983b) Three human transforming genes are related to the viral ras oncogenes. Proc. Natl. Acad. Sci. USA 80, 2112-2116.

Singh, H., Sen, R., Baltimore, D. & Sharp, P.A. (1986) A nuclear factor that binds to a conserved sequence motif in a transcriptional control element of immunoglobulin genes. Nature 319, 154-158.

Sive, H.L. & Roeder, R.G. (1986) Interaction of a common transcription factor with a conserved sequence in histone H2B, immunoglobulin and U2 small nuclear (snRNA) genes. Proc. Natl. Acad. Sci. USA 83, 6382-6386.

Slamon, D.J., deKernion, J.B., Verma, I.M. & Cline, M.J. (1984) Expression of cellular oncogenes in human malignancies. Science 224, 256-262.

Smith, H.D. (1980) Recovery of DNA from gels. In Methods in Enzymology, ed. Grossman, L. & Moldave, K. Vol. 65, pp 371-380. Academic Press, New York.

Southern, E.M. (1977) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517.

Spandidos, D.A. & Kerr, I.B. (1984) Elevated expression of the human ras oncogene family in premalignant and malignant tumours of the colorectum. Br. J. Cancer 49, 681-688.

Spandidos, D.A. & Wilkie, N.M. (1984) In vitro malignant transformation of early passage rodent cells by a single mutated human oncogene. Nature 310, 469-475.

Stein, R., Sciaky-Gallili, N. Razin, A. & Cedar, H. (1983) Pattern of methylation of two genes coding for housekeeping functions. Proc. Natl. Acad. Sci. USA 80, 2422-2426.

Stryer, L. (1983) Transducin and the cyclic GMP phosphodiesterase: Amplifier proteins in vision. Cold Spring Harbour Symp. Quant. Biol. 48, 841-852.

Stuart, G.W., Searle, P.F., Chen, H.Y., Brinster, R.L. & Palmiter, R.D. (1984) A 12-base pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Natl. Acad. Sci. USA 81, 7318-7322.

Sweet, R.W., Yokoyama, S., Kamata, T., Feramisco, J.R., Rosenburg, M. & Gross, M. (1984) The product of ras is a

GTPase and the T24 oncogenic mutant is deficient in this activity. Nature 311, 273-275.

Tanabe, T., Nukada, T., Nishikawa, Y., Sugimoto, K., Suzuki, H., Takahashi, H., Noda, M., et al. (1985) Primary structure of the a-subunit of transducin and its relationship to ras proteins. Nature 315, 242-245.

Taparowsky, E., Shimizu, K., Goldfarb, M. & Wigler, M. (1983) Structure and activation of the human N-ras gene. Cell 34, 581-586.

Tatchell, K., Chalaff, D.T., DeFeo-Jones, D. & Scolnick, E. (1984) Requirement of either of a pair of ras- related genes of Saccharomyces cerevisiae for spore viability. Nature 309, 523-527.

Thompson, C.B., Challaner, P.B., Neiman, P.E. & Groudine, M. (1985) Levels of c-myc oncogene mRNA are invariant throughout the cell cycle. Nature 314, 363-366.

Tinoco, I., Borer, P.N., Dengler, B., Levine, M.D. Uhlenbeck, O.C., Crothers, D.M. & Gralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. 246, 40-41.

Toda, T., Uno, I., Ishikawa, T., Powers, S., Kataoka, T., Broek, D., Cameron, S., Broach, J., Matsumoto, K. & Wigler, M. (1985) In yeast, RAS proteins are controlling elements of adenylate cyclase. Cell 40, 27-36.

Torczynski, R.M., Fuke, M. & Bollon, A.P. (1985) Cloning and sequencing of a human 18S ribosomal RNA gene. DNA 4, 283-291.

Treisman, R. & Maniatis, T. (1985) Simian virus 40 enhancer increases the number of RNA polymerase II molecules on linked DNA. Nature 315, 72-75.

Triesman, R., Proudfoot, N.J., Schnyder, M. & Maniatis, T. (1982) A single-base change at a splice site in a B^0-thalassemic gene causes abnormal RNA splicing. Cell 29, 903-911.

Tykoinski, M.L. & Max, E.E. (1984) CG dinucleotide clsuters in MHC genes and in 5' demethylated genes. Nucl. Acids Res. 14, 5869-5877.

Valerio, D., Dugvesteyn, M.G.C., Dekker, B.M.M., Weeda, G., Berkrens, T.M., van der Voorn, L., van Ormondt, H. & van der Eb, A.J. (1985) Adenosine deaminase: Characterization and expressin of a gene with a remarkable promoter. EMBO J. 4, 437-443.

Varmus, H.E. (1984) The molecular genetics of cellular oncogenes. Am. Rev. Genet. 18, 553-612.

Voss, S.D., Schlokat, U. & Gruss, P. (1986) The role of enhancers in the regulation of cell-type-specific transcriptional control. Tr. Biochem. Sci. 11, 287-289.

Wabl, M.R. & Burrows, P.D. (1984) Expression of immunoglobulin heavy chain at a high level in the absence of a proposed immunoglobulin enhancer element in cis. Proc. Natl. Acad. Sci. USA 81, 2452-2455.

Wakelam, M.J.O., Davies, S.A., Hanslay, M.D., McKay, I., Marshall, C.J. & Hall, A. (1986) Normal p21 $^{H\text{-}ras}$ couples bombesin and other growth factors receptors to inositol phosphate production. Nature 323, 173-176.

Wang, X.-F., & Calame, K. (1986) SV40 enhancer-binding factors are required at the establishment but not the maintenance step of enhancer-dependent transcriptional activation. Cell 47, 141-247.

Wasylyk, B., Wasylyk. N, Augereau, P. & Chambon, P. (1983) The SV40 72 bp repeat preferentially potentiates transcription starting from proximal natural or substitute promoter elements. Cell 32, 503-514.

Weaver, R.F. & Weissman, C. (1979) Mapping of RNA by a modification of the Berk-Sharp Procedure: The 5'termini of 15 S B-globin mRNA precursor and mature 10S B-globin mRNA have identical map co-ordinates. Nucl. Acids Res. 6, 1175-1193.

Weber, F. & Schaffner, W. (1985) Simian Virus 40 enhancer increases RNA polymerase density within the linked DNA. Nature 315, 75-77.

Weinberg, R.A. (1985) The action of oncogenes in the cytoplasm and nucleus. Science 230, 770-776.

Weintraub, H. (1985) High-resolution mapping of S1 and DNase I-hypersensitive sites in chromatin. Mol. Cell. Biol. 5, 1538-1539.

Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacy, E., Maniastis, T., Silverstein, S. & Axel, R. (1979) Transformation of mammalian cells with genes from prokaryotes and eukaryotes. Cell 16, 777-785.

Willingham, M.C., Pastan, I., Shih, T.Y. & Scolnick, E.M. (1980) Localization of the src gene product of the Harvey strain of HSV to the plasma membrane of transformed cells by electron microscope immunocytochemistry. Cell 19, 1005-1014.

Wolf, S.F., Jolly, D.S., Lunnen, K.D., Fredmann, T. & Migeon, B.R. (1984). Methylation of the hypoxantine phosphoribosyltransferase locus on the human X chromosome:

Implications for X-chromosome inactivation. Proc. Natl. Acad. Sci. USA 81, 2806-2810.

Wolfman, A. & Macara, I.G. (1987) Elevated levels of diacylglycerol and decreased phorbol ester sensitivity in ras-transformed fibroblasts. Nature 3425, 359-381.

Wolf, S.F. & Migeon, B.R. (1985) Clusters of CpG dinucleotides implicated by nuclease hypersensitivity as control elements of housekeeping genes. Nature 314, 467-469.

Wood, T.G., McGeady, M.L., Barandy, B.M., Blair, D.G. and Vande Woude, G.F., (1984) Mouse c-mos oncogene activation is prevented by upstream sequences. Proc. Natl. Acad. Sci. USA 81, 7817-7821.

Wu, C. (1984) Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. Nature 309, 229-234.

Wu, C. (1985) An exonuclease protection assay reveals heat-shock element and TATA box DNA-binding proteins in crude nuclear extracts. Nature 317, 84-87.

Yuasa, Y., Srivastava, S., Dunn, C., Rhim, J., Reddy, E. & Aaronson, S. (1983) Acquisition of transforming properties by alternative point mutations within c-bas/has human proto-oncogene. Nature 303, 775-779.

Yoakum, G.H., Lechree, J.F., Gabrielson, E.W., Korba, B.E., Malan-Shibley, L., Willey, J.C., et al. (1985) Transformation of human bronchial epithelial cells transfected by Harvey ras oncogene. Science 227, 1174-1179.

Yen, P.H., Patel, P., Chinault, A.G., Mohnolas, T. & Shapiro, L.J. (1984) Differential methylation of hypoxanthine phosphoribosyltransferase gene on active and inactive human X-chromosomes. Proc. Natl. Acad. Sci. USA 81, 1759-1763.

Yatsunami, K. & Khorana, H.G. (1985) GTPase of bovine rod outer segments: The amino acid sequence of the x subunit as derived from the cDNA sequence. Proc. Natl. Acad. Sci. USA 82, 4316-4310.

Zalles, D.M. & Eckhardt, L.A. (1985) Deletion of a B-cell-specific enhancer effects transfected, but not endogenous, immunoglobulin heavy-chain gene expression. Proc. Natl. Acad. Sci. USA 82, 5088-5092.