Molecular Cloning of Human and Murine hsp60 Related Sequences

Molecular Cloning of Human and Murine hsp60 Related Sequences

by

Thomas James Venner, B.Sc.

A Thesis Submitted to the
School of Graduate Studies in
Partial Fulfillment of the Requirements
for the Degree

Master of Science
McMaster University

December, 1990

ABSTRACT

Full length P1 cDNA clones have been obtained from human and CHO sources (Jindal *et al.*, 1989; Picketts *et al.*, 1989) which contain sequences that show extensive sequence and structural similarity to the chaperonin family of proteins, including the mitochondrial hsp60 protein. In the studies described here human genomic DNA libraries were screened with human P1 (hsp60) cDNA probes and those clones containing P1 related sequences were isolated. One clone, pGem10, was found to be completely homologous to the human P1 cDNA in both coding and non-coding regions, devoid of intervening sequences, and terminates at a point 24 base pairs upstream of the translation initiation signal (ATG). The other human clones analyzed were all found to be pseudogenes containing numerous additions, deletions and base substitutions, but no introns. A total of six different classes of pseudogenes were identified. Four of these were sequenced completely across the translated region of the functional P1 gene. Sequence homologues of 86.1, 87.4 89.7 and 90.2% were observed.

In addition, rat kidney and mouse 3T3 cell cDNA libraries were screened similarly for P1 sequences. The rat P1 cDNA sequence was obtained by combining the sequence information

from three different clones. The clones obtained lacked the 5'- leader sequence as well as the mitochondrial targeting sequence. However, the entire coding sequence for a mature P1 protein of 547 amino acids could be deduced. The mouse P1 DNA sequence was also obtained from three different clones. These clones contained a portion of the mitochondrial targeting sequence and the entire sequence for the mature P1 protein. The protein sequences of the rat and mouse P1 clones were highly homologous (98-99%) to those obtained from human and CHO sources. The calculated molecular weights of the mature rat and mouse P1 proteins are 57,916 and 57,940 daltons, respectively, which are in close agreement with those predicted for the human (57,939 daltons) and CHO (57,949) proteins (Jindal *et al.*, 1989; Picketts *et al.*, 1989).

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. R.S. Gupta, for the opportunity to pursue this project. You have been both a friend and a mentor, freely offering your opinions, advice and citicisms. I would also like to thank my supervisory committee, Dr. K.B. Freeman, Dr. C.B. Harley and Dr. G. Singh for their input over the duration of these studies.

The financial support I have received from the Department of Biochemistry and McMaster University through both scholarships and teaching assistantship duties is greatly appreciated.

Special thanks to the many friends and colleagues with whom I have had the pleasure of associating with over the years. Special thanks to Mr. B. Rowley for reading the manuscript and offering valuable criticisms. Very special thanks go to Mrs. B. Sweet for typing the text.

I would also like to thank Karen for her patience, understanding and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATP	adenosine triphosphate
APS	ammonium persulphate
BiP	immunoglobulin heavy chain binding protein
bp	base pair
BSA	bovine serum albumin
$CaCl_2$	calcium chloride
cDNA	complementary DNA
CHO	Chinese hamster ovary
cm	centimeter
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DNA	deoxyribonucleic acid
dGTP	deoxyguanosine triphosphate
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
EDTA	ethylenediamine tetraacetic acid
EMS	ethyl methanesulfonate
ER	endoplasmic reticulum
grp	glucose regulated protein
g, mg, μg	gram, milligram, microgram
GTP	guanosine triphosphate
HCl	hydrochloric acid
hr	hours
HSE	heat shock element

hsp	heat shock protein
HSTF	heat shock transcription factor
Ig	immunoglobulin
IPTG	isopropyl-ß-D-thiogalactopyranoside
kb, kbp	kilobase, kilobase pair
kDa	kilodalton
l, ml, µl	liter, milliliter, microliter
LB	Luria-broth
MAP	microtubule associated protein
$MgSO_4$	magnesium sulphate
min	minutes
mM, M	millimolar, molar
Mr	relative molecular mass
mRNA	messenger RNA
MT	microtubule
NaCl	sodium chloride
Na_2HPO_4	disodium hydrogen orthophosphate
NaOH	sodium hydroxide
ng	nanograms
nm	nanometers
pfu	plaque forming units
pmole	picomole
Pod^{Rn}	CHO cell mutants obtained after n selections in the presence of podophyllotoxin
$RbCl_2$	rubidium chloride
RNA	ribonucleic acid

RubBP, RBP	ribulose-1,5'-bisphosphate carboxylase subunit binding protein of plant chloroplasts
SDS	sodium dodecyl sulphate
sec	seconds
SM	suspension medium
SRE	serum responsive element
SSC	saline sodium citrate
TAE	tris-acetate-EDTA
TBE	tris-borate-EDTA
TE	tris-EDTA
Tris	tris(hydroxymethyl)amino methane
V	volt
W	watt
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside

1.0 INTRODUCTION

1.1 Microtubule Structure and Function

Microtubules (MTs) represent one of the three major cytoskeletal protein filament systems of most vertebrate cells. Together, MTs, microfilaments (or actin filaments) and intermediate filaments provide a framework for intracellular organization. This highly structured system plays a role in a wide variety of cellular functions including secretion and intracellular transport (Hayden *et al.*, 1983), ciliary and flagellar movement (Dustin, 1984), directional cell locomotion and chemotaxis (Zigmond, 1989), intracellular organelle transport (Vale *et al.*, 1985a,b,c; Schnapp *et al.*, 1986), maintenance of cell shape (Cieselska-Treska *et al.*, 1982; Ochs, 1982) and cytokinesis (Dustin, 1984).

A large volume of literature has been accumulated on the three major cytoskeletal filament systems, however, the focus of this thesis is on the application of somatic cell genetics to the study of MT structure and function.

1.1.1 Microtubules: General features

Microtubules are long cylindrical filaments measuring approximately 25 nm in diameter with walls about 5 nm thick. Protofilaments are formed from polymers of a globular protein, tubulin, which is present in the cell as a 6S heterodimer of two similar 55,000 dalton subunits, α and ß (Timasheff and Grisham, 1980). Most MTs consist of 13 protofilaments *in vivo* (Lewin, 1980), whereas 14 protofilaments are more commonly observed for MTs polymerized *in vitro* (Pierson, *et al.*, 1978).

The isolation of tubulin (Weisenberg *et al.*, 1968), and subsequent determination of conditions for *in vitro* reconstitution of MTs (Weisenberg, 1972), led to an enormous increase in MT research and literature. The more salient features of these findings will be summarized here.

The cycling procedure adapted by Shelanki *et al.*, (1973) from conditions established by Weisenberg (1972) involves repeated asssembly and disassembly of tubulin polymers, resulting in a MT protein preparation which is 75-90% tubulin. In addition, some 15-30 minor bands have been identified as MT associated proteins or MAPS (see 1.1.2). Assembly of MTs *in vitro* requires incubation of tissue or cell extract at 37°C in a glycerol based (up to 4 M) medium with magnesium ions and GTP. After collecting MTs by centrifugation, they are cooled to 4°C to allow disassembly, and recentrifuged at low temperature to remove undissociated materials. Repeating the cycle number increases the tubulin purity and decreases the relative number of MAPs.

One tubulin heterodimer can bind two guanine nucleotides. While α-tubulin binds GTP tightly and does not exchange the nucleotide with the medium (GTPn), ß-tubulin binds GTP exchangeably (GTPe). This GTP is hydrolyzed to GDP during the assembly process (Mandelkow and Mandelkow, 1989). Higher eukaryotes contain numerous tubulin genes, for both α- and ß-tubulin, which are expressed under complex developmental programs and subject to several types of post-transcriptional modifications (for review see Cleveland and Sullivan, 1985; Cleveland, 1989; MacRae and Langdon, 1989).

Microtubules can be relatively static or highly dynamic, showing alternating phases of growth and shrinkage, referred to as dynamic instability. In the most recent model for MT polymer formation, GTP hydrolysis lags behind the rate of polymerization during growth. This lag results in the formation of a cap of GTP-bound subunits at the terminus of the MT, thus permitting further addition of subunits and slow growth of the MT. When the rate of hydrolysis exceeds that of polymerization, the cap is reduced in size, and the MT begins to depolymerize rapidly (Hill and Chen, 1984).

The relative amounts of each of these populations (growing or shrinking) is a function of tubulin concentration and the number of nucleating sites. Nucleation occurs at preferred sites called microtubule organizing centres (MTOC; Brinkley, 1985). Immunohistochemical techniques have shown the MTs eminate from the perinuclear region, frequently at or

in the vicinity of the centrosome (Weber *et al.*, 1975). The role of centrosomes (now referred to as MTOC) in spindle formation is well documented.

1.1.2 Microtubule associated proteins

A number of associated proteins (MAPs) have been found to co-purify with tubulin in constant stoichiometry during cycles of MT assembly/disassembly. These proteins were initially classified as either high molecular weight MAPs or Tau (τ) proteins. The high molecular weight MAPs have been resolved and identified as the MAP-1 group (330 kDa) and the MAP-2 group (280 kDa) (Murphy and Borisy, 1975; Sloboda *et al.*, 1975). The lower molecular weight τ proteins consist of 4 to 6 closely spaced bands between 55 and 62 kDa (Cleveland *et al.*, 1977). Several other MAPs have recently been identified (for review see Olmstead, 1986) including MAP 3, MAP 4, STOPS and kinesin. Another MAP, ß-internexin has recently been identified as the cognate form of the 70 kDa heat shock protein (Green and Liem, 1989). The HMW and τ proteins have been studied most extensively and therefore will be given a more detailed discussion.

The high molecular weight MAPs are flexible, rodlike structures (approximately 100-200 nm) which project from MT surfaces. The MAP 1 group contains at least three distinct members, MAP 1a, MAP 1b and MAP 1c (Bloom *et al.*, 1984). MAP

1a appears to be the MT binding domain based on immunofluorescence data (Asai *et al.*, 1985). MAP 1c has been shown to be a cytoplasmic dyenin (Vallee *et al.*, 1988) which is an ATPase capable of producing force in the direction corresponding to retrograde organelle transport in the cell (Paschel *et al.*, 1987). The MAP 2 group contains two essentially identical subspecies, MAP 2A and 2B. Limited proteolysis reveals two distinct domains, a MT binding segment and a projection domain (Herman *et al.*, 1985). Extensive phosphorylation of the available sites on MAP 2 inhibits its binding to MTs (Olmstead, 1986).

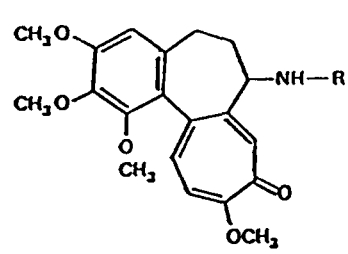
In early studies with the τ proteins, Weingarten *et al.*, (1975) reported that the removal of the associated proteins from tubulin rendered it totally incompetent for self-assembly, and that addition of purified τ protein back to tubulin restored polymerizability. Subsequent purification of τ showed these proteins to be heat resistant (Cleveland *et al.*, 1977). More recently, dephosphorylation of τ was found to promote MT polymerization (Lindwall and Cale, 1984). The application of molecular biology technologies has revealed a common MT binding motif in τ and MAP 2 (Lewis *et al.*, 1988). In addition, both proteins possess a short C-terminal α-helical sequence which is able to crosslink MTs by means of a hydrophobic zipper interaction into dense stable parallel arrays (Lewis *et al.*, 1989).

A number of proteins have been identified which bind to MTs but are localized on different organelles. A 50 kDa lysosome membrane protein appears to mediate an ATP-dependent stable association of lysosomes to MTs (Mithieux and Rousset, 1989). In addition, a 58 kDa Golgi membrane protein has been shown to possess many of the properties of MAPs, including co-sedimentation with tubulin, and stimulation of tubulin polymerization (Bloom and Brashear, 1989).

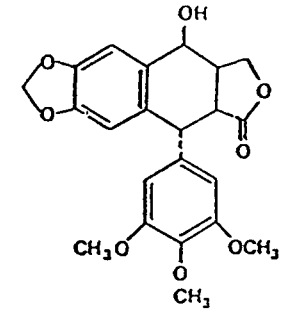
1.2 Drugs Which Affect MT Action

Drugs capable of arresting mammalian cells in mitosis have been known for at least 100 years. The best characterized of these is colchicine, which has been used for centuries to treat gout. Borisy and Taylor (1967) were the first to show colchicine inhibits mitosis by binding to tubulin. Since then, many other compounds have been shown to interact with tubulin in its dimeric or polymeric state. The structures of a number of these drugs are shown in Fig. 1. Many of these drugs have been found useful as chemotherapeutic agents (vincristine, vinblastine, podophyllotoxin), antifungal agents (griseofulvin) and herbicides (isopropyl-N-phenyl carbamate). In addition, these compounds have provided invaluable tools for the study of MT assembly and function (Cabral and Barlow, 1989; Gupta, 1989).

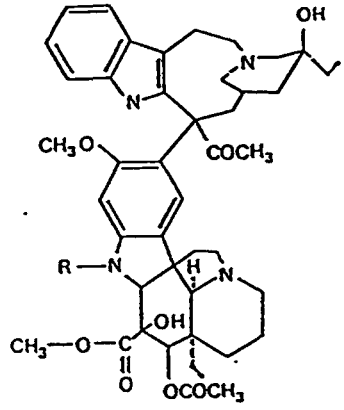
1.2.1. Mechanism of action of MT inhibitors



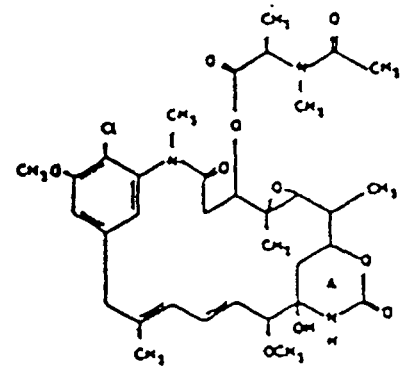
Colchicine: R = $COCH_3$
Colcemid: R = CH_3



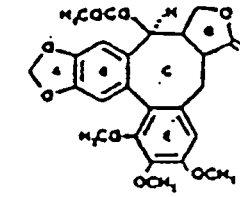
Podophyllotoxin



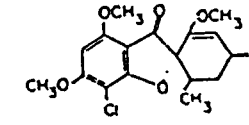
Vinblastine: R = CH_3
Vincristine: R = CHO



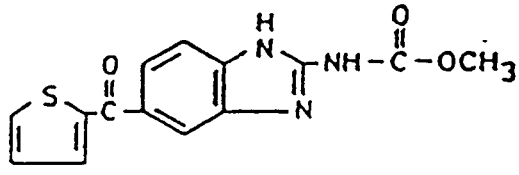
maytansine



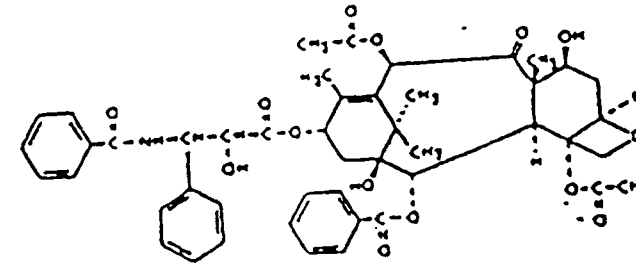
steganacin



griseofulvin



Nocodazole



taxol

Figure 1 **Chemical structures of some common microtubule inhibitors.** Included are the structures for colchicine, podophyllotoxin, vinblastine, maytansine, griseofulvin, nocodazole and taxol. Colcemid and vincristine are derivatives of colchicine and vinblastine, respectively (taken from various sources).

While colchicine binds to tubulin with high affinity in a 1:1 complex, inhibition of MT assembly requires substoichiometric concentrations of drug. This observation led Margolis and Wilson (1977) to postulate that the drug-tubulin complex binds to the growing end of the MT thereby preventing further assembly. While this model has been widely accepted, it cannot explain some of the data obtained from kinetic studies, or the observation of colchicine binding along the length of MTs (Schibler and Cabral, 1985).

Podophyllotoxin and colchicine are structurally related and share at least part of their binding sites on tubulin (Schilstra *et al.*, 1989). At low concentrations, relative to free tubulin concentration, podophyllotoxin, like colchicine, exerts a disproportionally large effect in inhibiting MT growth and suppressing MT dynamics (Schilstra *et al.*, 1989). However, while the colchicine-tubulin complex is a potent inhibitor of MT disassembly (Wilson and Farrel, 1986) elevated podophyllotoxin concentrations result in full disassembly of MTs (Schilstra *et al.*, 1989). Thus, it appears that the precise mechanism of action of podophyllotoxin is likely to be significantly different from that of colchicine.

Steganacin and nocodazole, like podophyllotoxin, are competitive inhibitors of colchicine binding to tubulin (Dustin, 1984). Vinblastine and vincristine are not competitive inhibitors of colchicine-binding. They possess

two binding sites per tubulin dimer, one of high affinity, the other of moderate affinity (Schliwa, 1986). Interaction with these drugs causes tubulin aggregation and paracrystalline formations to appear in cells (Dustin, 1984). Podophyllotoxin, nocodazole, vinblastine and vincristine binding is readily reversible, shows less temperature dependence, and is much faster (approximately 10-fold) than colchicine binding (Mareel and DeMets, 1984). While maytansine is a competitive inhibitor of vinblastine binding, it causes MT depolymerization rather than aggregation of tubulin (Luduena *et al.*, 1986).

Taxol is a novel type of antimitotic drug in that it has the ability to promote MT assembly and prevent depolymerization of MTs to which it binds (Schiff and Horwitz, 1980). By binding to the MT polymer, taxol stabilizes the entire structure rendering the MT resistant to depolymerization induced by exposure to low temperatures or by the addition of calcium (Schiff and Horwitz, 1980). Finally, griseofulvin has a unique mechanism of action, resulting in MT disorganization which occurs as a consequence of a suggested interaction with MAPs rather than tubulin itself (Robol *et al.*, 1977).

1.2.2 Somatic cell genetics

One general approach which has been found useful in the study of various cell structures and their functions, has been to isolate cells which are altered in a particular function, and to subsequently study the phenomenon in both mutant and wild type (WT) cells (Schibler and Cabral, 1985). The availability of a large number of drugs which specifically affect MT function by a diverse variety of mechanisms, has made the use of a combined genetic and biochemical approach ideal for the study of MTs. This approach has been used by many investigators (for review see Dustin, 1984; Schibler and Cabral, 1985; Gupta, 1989).

It is important to distinguish mutants which are specifically affected in MT function, from those whose resistance is related to changes in cell permeability (Cabral and Barlow, 1989; Gupta, 1989). In general, it has been found that permeability mutants show cross resistance to a wide variety of unrelated compounds. Another class of mutants which are specifically affected in MTs are resistant to drugs which depolymerize MTs and show specific cross resistance to other destabilizing drugs,with an increased sensitivity towards taxol (Cabral and Barlow, 1987; Gupta, 1989).

The majority of Chinese hamster ovary (CHO) MT mutants isolated to date have shown an electrophoretic alteration in α or ß tubulin on two dimensional (2-D) gels (Ling *et al.*, 1979; Sawada and Cabral, 1989). While some of these mutants have shown decreased drug binding and revertant phenotypes

have been demonstrated, the precise location of the genetic lesion has thus far remained elusive. This endeavour has been complicated by the observation of numerous tubulin isotypes in CHO cells. Elliot *et al.*, (1986) have published the sequence of 3 distinct α tubulin isoproteins. Recently, the DNA sequence of a fourth α tubulin isoform and three ß tubulin isoforms has been obtained (Ahmad and Gupta, unpublished results).

Microtubule mutants which show changes in proteins other than tubulin include a griseofulvin resistant mutant affected in a 200 kDa protein (Gupta, 1984), a colchicine resistant mutant altered in a 70 kDa protein (Gupta *et al.*, 1985) and podophyllotoxin resistant (Pod^R) mutants in which an electrophoretically altered 60 kDa protein has been detected (Gupta *et al.*, 1982). The observation of this latter class of mutants provided a unique opportunity to study the role of a non-tubulin protein in MT structure and function. These Pod^R mutants are discussed in more detail in Section 1.2.3.

1.2.3 **Studies of podophyllotoxin resistant mutants**

The spontaneous rate of mutation to a podophyllotoxin resistant (Pod^R) phenotype in CHO cells has been estimated to be 1.2×10^{-5} and 4×10^{-7} mutations/cell/generation at 20 ng/ml and 30 ng/ml of the drug (Gupta, 1981). Stable mutants were isolated after a single step selection (Pod^{RI}) in

podophyllotoxin at the concentrations described above. In some cases WT cells were pretreated with ethylmethanesulfonate (EMS) to increase mutation frequency. The Pod^{RI} cells showed 2- to 4-fold increases in resistance relative to WT. Two clones (Pod^{RI} and $Pod^{RI}16$) were subjected to a second round of selection in 50 ng/ml podophyllotoxin. These mutants thus obtained (referred to as Pod^{RII}) show an 8- to 10-fold increase in resistance relative to WT cells (Gupta *et al.*, 1981).

These mutants shows specific cross resistance and collateral sensitivity to MT inhibitors but unchanged resistance to unrelated toxic compounds (e.g. puromycin). Furthermore, these mutants showed increased resistance to podophyllotoxin derivatives which possess MT disrupting abilities (e.g. podophyllic acid, deoxypodophyllotoxin), but not to derivatives which are known inhibitors of DNA synthesis (e.g. VP16-213 and VM-26). Somatic cell hybrids constructed between WT (sensitive) and mutant cells showed intermediate levels of resistance, indicating a codominant drug resistance phenotype (Gupta, 1981). The MT specific nature of the genetic lesion(s) responsible for this drug resistance was strengthened by kinetic studies with ^{3}H-podophyllotoxin. These studies show reduced drug binding by Pod^{RI} and Pod^{RII} cell extracts relative to those obtained from parental (WT) cells (Gupta, 1983).

To identify the cellular component(s) affected in these mutants, whole cell extracts were subjected to 2-D electrophoresis and compared to those obtained for WT cells. No apparent changes were observed in any Pod^{RI} cells, however, 23 out of 26 independent Pod^{RII} mutants contained a novel protein spot (designated as M1; 63 kDa) not previously observed in the parental strains (Gupta *et al.*, 1982). The position of the M1 spot indicated that it may have arisen as a result of a charge alteration in the more basic protein, P1, which was directly adjacent to M1. This was confirmed by comparing P1 and M1 by means of peptide maps generated from partial *Staphylococcus* V8 protease digestion (Gupta *et al.*, 1982).

In Pod^{RII} cells that contained the M1 spot, it was found in a 1:1 ratio with P1. Hybrids formed between these Pod^{RII} and WT cells were found to contain a 3:1 ratio of P1 to M1 consistent with the codominant behaviour seen in resistance studies with these cells. This observation was regarded as an indication that post-translational modification was not involved in the genetic lesion seen in these mutants (Gupta, 1982).

In order to investigate the possible MT-association of the P1 protein, the criteria developed by Duerr *et al.*, (1981) were employed. Briefly, both the P1 and M1 proteins were found to co-release along with tubulin MT containing fractions in response to increased calcium concentration or cold

temperatures (4°C) (Gupta *et al.*, 1982; Gupta and Gupta, 1984). Cells pretreated with colchicine were found to release actin and a few other non-specific proteins (Gupta *et al.*, 1985). Similar results were obtained from cold incubation of crude MT-containing fractions prepared from different cell lines under conditions which preserve MTs (Cabral *et al.*, 1980; Gupta and Gupta, 1984).

Antibodies which specifically react with P1 (and M1) were raised and used to study the cellular location and distribution of the P1 protein (Gupta *et al.*, 1985; Gupta and Venner, 1986). Immunoblot analysis indicated the P1 protein was present in a wide variety of animal species including human, monkey, mouse, rat, dolphin, chicken, bullfrog, snake and mosquito (Gupta *et al.*, 1985; Gupta and Dudani, 1987). Indirect immunofluorescence studies localized P1 structures to the mitochondria, based on their similarity to structures stained by the mitochondrial specific fluorescent dye rhodamine 123, or antibodies to the known mitochondrial protein malate dehydrogenase (Gupta *et al.*, 1985; Gupta and Austin, 1987). Double immunofluorescence studies with both P1 and tubulin antibodies show that mitochondria are codistributed along the length of MTs in interphase cells but not in mitotic cells (Gupta *et al.*, 1985; Gupta and Dudani, 1987).

Subfractionation of mitochondria revealed the association of P1 mainly with the matrix fraction (Gupta and

Austin, 1987). This observation was substantiated by treatment of cells with the K^+ ionophores non actin and valinomycin. These compounds abolish the mitochondrial membrane potential resulting in the loss of uptake of proteins destined for import into mitochondria. As a result, precursor forms of these proteins accumulate in the cytoplasm, including the P1 precursor which is approximately 3 kDa larger than mature P1 (Gupta and Austin, 1987).

In order to gain further insight into the cellular role of the P1 protein, the cloning and sequencing of P1 cDNA was initiated. First, the P1 antibody described earlier was used to screen a human promyelocytic leukemia (HL-60) cDNA expression library. A number of P1 reactive clones were isolated and the largest, containing an insert of 1.4 kb (designated λ22a), was used for further analysis (Jindal *et al.*, 1989). The P1-specific nature of the λ22a cDNA was confirmed immunologically using a ß-galactosidase-P1 fusion protein. The 150 kDa fusion protein (approximately 114 kDa of which is ß-galactosidase) was used to affinity purify antibodies from the original polyclonal serum. These antibodies showed specific binding to mitochondria in immunofluorescence experiments, and to a 63 kDa protein in CHO cell extracts (Jindal *et al.*, 1989).

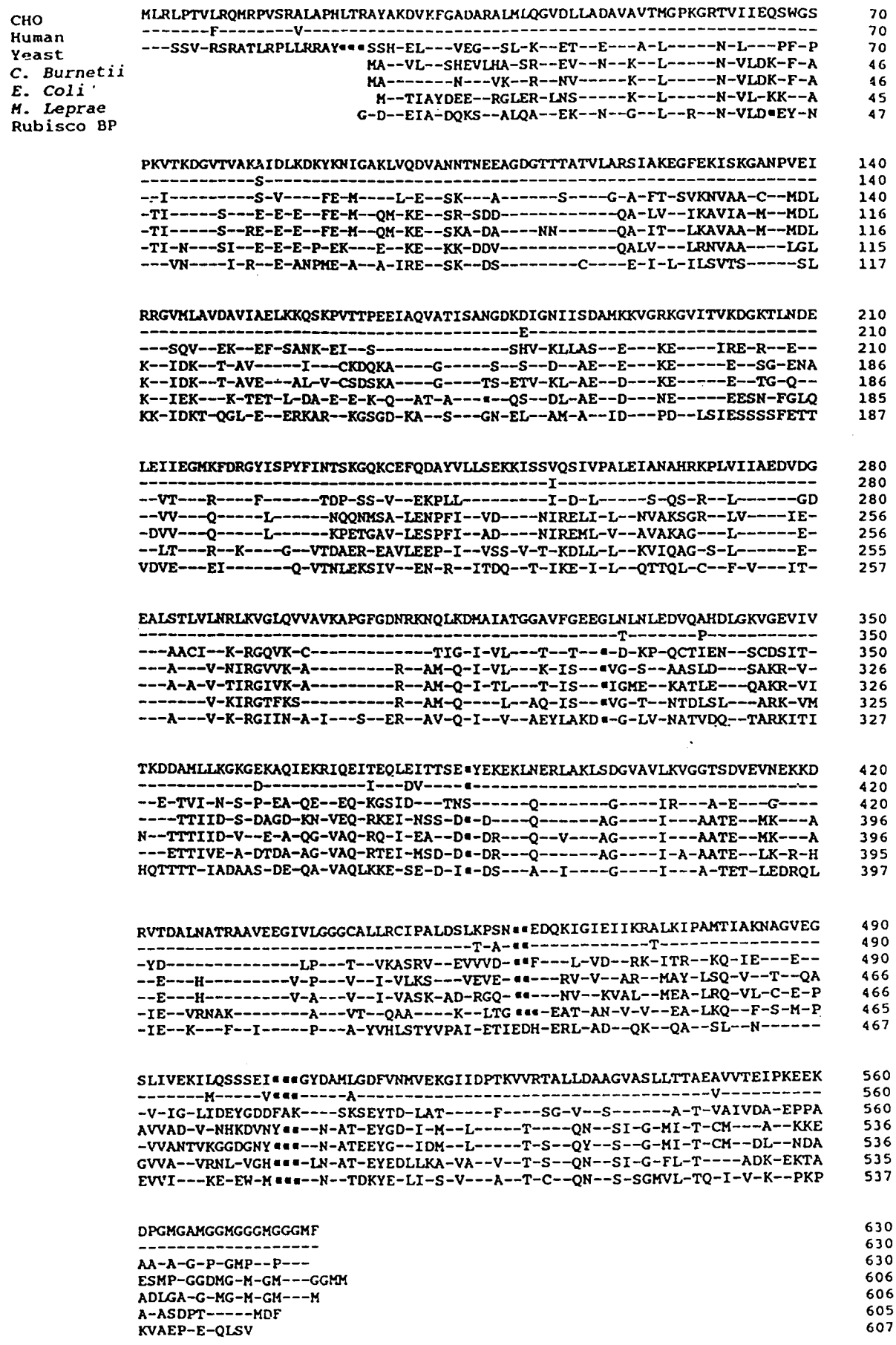
The 1.4 kb probe was used to isolate a full length cDNA clone from another HL-60 library. One of these clones, λC5, contained two EcoRI fragments of approximately 0.8 and 1.4 kb.

Previous sequencing of the 1.4 kb probe revealed a 3'- poly(A) tail. This probe was found to hybridize to the 1.4 kb fragment from λC5. Complete sequencing of the human P1 cDNA showed that the P1 gene encodes a protein of 573 amino acids including a 26 amino acid mitochondrial targeting sequence (Jindal *et al.*, 1989)..

In order to identify the precise location of the genetic lesion in the Pod^{RII} mutants described earlier, it was first necessary to deduce the sequence of P1 from WT CHO cells. The 0.8 kb and 1.4 kb EcoRI fragments obtained from the human cDNA library were used as probes to screen a λgt10 cDNA library, prepared from poly $(A)^+$ RNA from CHO cells. The complete sequence of P1 from CHO cDNA was obtained and the deduced amino acid sequence shows 97% identity to the human P1 protein (Picketts *et al.*, 1989).

The amino acid sequences of these mammalian P1 proteins show extensive sequence homology to the "chaperonin" family of bacterial (GroEL), yeast (hsp60) and plant (ribulose-1, 5'-bisphosphate carboxylase subunit binding protein of plant chloroplasts; RubBP) and to the 60-65 kDa major antigenic protein of mycobacteria and *Coxiella brunetii* (see 1.4) (Jindal *et al.*, 1989; Picketts *et al.*, 1989). Amino acid comparison analysis (summarized in Fig. 2) reveals 42-60% identical residues and an additional 15-25% conservative replacements. Gel filtration elution profiles and sucrose density gradients indicated native P1 exists as a homooligomeric complex of 7

Figure 2 **Comparison of the sequence of mammalian P1 protein with related proteins from other species.** Line 1, CHO P1 sequence; line 2, human P1; line 3, yeast hsp60; line 4, 62 kDa antigen from *C. burnetii*; line 5, GroEL protein of *E. coli*; line 6, 65 kDa antigen from *M. leprae*; line 7, Rubisco large subunit binding protein from wheat chloroplasts (taken from Picketts *et al.*, 1990).


CHO
Human
Yeast
C. Burnetii
E. Coli
M. Leprae
Rubisco BP

subunits (approximately 440-450 kDa) and was found to possess weak ATPase activity (Picketts *et al.*, 1989). These data are consistent with those observed in studies of the chaperonin proteins cited previously (see 1.3.2).

1.3 Heat Shock Proteins

Heat shock proteins (hsp) are induced in most prokaryotic and eukaryotic cells upon transfer to elevated growth temperatures or treatment with a variety of chemical agents (e.g. transition metals, amino acid analogues, teratogens, mutagens, etc.), many of which interfere with mitochondrial function (Lindquist and Craig, 1988; Tanguay, 1988). The induction of hsp, in response to these stresses is rapid, transient and is regulated primarily at the transcriptional level, and to a lesser extent by post-transcriptional and post-translational events (Lindquist, 1986). The hsp's have been grouped into three general classes (80-110, 66-73 and 15-30 kDa). The precise function of the various hsp's remains unclear, however, numerous observations suggest an involvement in cellular protection processes such as thermotolerance, and in the efficiency of cellular recovery following exposure to chemical or environmental stress (Lindquist and Craig, 1988; Tanguay, 1988).

1.3.1 Transcription regulation of hsp's

The highly conserved sequence of some hsp genes and the early observation that the *Drosophila* hsp70 gene retained stress inducibility after transfection into monkey COS cells, indicated the presence of conserved sequence elements. It was felt these sequences were involved in both the recognition of stress signals and the transcriptional activation of this group of genes (Pelham, 1982, 1989). Pelham (1982) performed deletion analysis of the hsp70 promoter, and by comparing it to other heat shock promoter sequences, he was able to confirm the existence of a specific cis-element (heat shock control element; HSE). A heat shock consensus sequence has been defined (Pelham, 1982) and the same regulatory sequence has been observed in the heat shock proteins of a variety of species (Pelham, 1989). The HSE binding protein or heat shock transcription factor (HSTF) has been identified for both *Drosophila* and yeast (Sorger and Pelham, 1987) and the yeast HSTF gene has been cloned (Sorger and Pelham, 1988). The HSTF found in non-shocked HeLa cells is not as effective a transcriptional activator as the equivalent factor isolated from heat shocked cells (Kingston *et al.*, 1987). Interestingly, the phosphorylation level of HSTF appears to have a modulating effect on the level of transcriptional stimulation (Sorger and Pelham, 1988).

Multiple HSE's are not uncommon, and other types of novel upstream and/or downstream regulatory elements have been

identified. These include the GC-rich Spl binding motif (Morgan, 1989) and SRE's or serum responsive elements (for review see Tanguay, 1988). Recently the human genomic nucleotide sequence has been obtained for Hsp89α (Hickey *et al.*, 1989) and Hsp90ß (Rebbe *et al.*, 1989). These heat shock promoters contain multiple HSE's (both upstream and within the first intron), Spl factor binding sites and were induced by serum. Recently, reports have demonstrated the stimulation of hsp70 transcription by both the nuclear oncogene E1A 13S gene product (Williams *et al.*, 1989) and polyoma virus large T antigen (Taylor *et al.*, 1989).

The most definitive information regarding the role of heat shock proteins in both stressed and normal cells has come through the study of polypeptide chain binding (PCB) proteins or molecular chaperones (Ellis *et al.*, 1989; Rothman, 1989).

1.3.2 **Polypeptide chain binding proteins: molecular chaperones**

It has long been assumed that protein folding and assembly is a spontaneous process dictated by amino acid sequence only, thus precluding the role of catalytic agents such as enzymes (Anfinsen, 1973). Recently it has been shown that proper protein folding and/or assembly of oligomeric protein complexes involves a class of proteins termed "molecular chaperones" (Ellis, 1987). This term was first

used by Laskey *et al.*, (1978) to describe nucleoplasmin, a nuclear protein required for the assembly of nucleosomes from DNA and histones in *Xenopus* egg extracts. The term 'chaperone' is appropriate, as nucleoplasmin binds exclusively to histones, promoting histone-histone interaction and minimizing non-specific aggregates, without itself becoming part of the nucleosome (Laskey, 1980).

Much of the evidence that has led to the proposal that various aspects of protein folding are catalyzed by molecular chaperones came from the study of the two major families (see Table 1), hsp70 and GroEL (Rothman, 1989).

1.3.2a **The hsp70 family**

All members of the hsp70 family (which range from approx. 70 to 78 kDa) show ATP binding ability and are closely related in sequence. They have been found in bacteria as well as the cytoplasm, mitochondria, and the lumen of the endoplasmic reticulum (ER) of eukaryotic cells. Most hsp70 members are induced by conditions of stress which promote protein denaturation and misfolding, presumably to aid in protein repair (Rothman, 1989). Other constitutively expressed versions (hsc70) are presumably sufficient under normal cell growth conditions (Pelham, 1986)

The dna k gene was so named because *E. coli* DNA synthesis shuts off when mutant cells are shifted to high

Protein	Location	Family	Subunit Size
GroEL	bacterial cytoplasm	GroEL	58 kd
Subunit binding protein	chloroplast stroma	GroEL	61 kd α chain 60 kd β chain
hsp60	mitochondrial matrix	GroEL	58–64 kd
GroES	bacterial cytoplasm	—	10 kd
hsc70	cytoplasm of eukaryotes	hsp70	69–72 kd
DnaK	bacterial cytoplasm	hsp70	69 kd
BiP	lumen of ER	hsp70	78 kd
SSC1	mitochondria (yeast)	hsp70	~70 kd
Protein disulfide isomerase	lumen of ER	—	57 kd

Table 1 **Polypeptide chain binding proteins implicated in protein folding and/or assembly reactions in cells** (adapted from Rothman, 1989).

temperatures (Lindquist and Craig, 1988). Originally identified as the host gene necessary for lambda DNA replication, Dna K was found to interact with the P protein of lambda. A six protein complex (2 lambda, 4 host) is involved in unwinding the DNA duplex at the origin of replication (Lebowitz *et al.*, 1986). It has been proposed that Dna K and Dna J interrupt the association between Dna B and P, thus allowing Dna B (a helicase) to function (Yamamoto *et al.*, 1987), and thereby permitting Dna G (primer) binding and DNA synthesis.

Mutants of Dna K show inhibition of both RNA and DNA synthesis and a block in cell division at non-permissive temperatures (Bukau and Walker, 1989). Dna K is capable of self phosphorylation on a threonine residue. Temperature-sensitive Dna k mutants which normally phosphorylate 5 protein substrates (including glutamine and threonyl tRNA synthetase) are unable to do so at non-permissive temperatures (Wada *et al.*, 1986).

The hsp70 family in yeast contains at least nine members (Lindquist and Craig, 1988). They have been grouped on the basis of structural and functional similarities as follows: SSA1-4, SSB1 and 2, SSC1, SSD1 and KAR2 genes (Lindquist and Craig, 1988). Both the sequence relationships, and expression patterns in response to temperature exhibited by these family members are complex (for review see Lindquist and Craig, 1988). Recent studies focussing on the SSA1-4

group have suggested that the SSA gene products perform interchangeable functions *in vivo* (Deshais *et al.*, 1988a). SSA1 and SSA2 are expressed constitutively, although SSA1 levels increase dramatically (3 to 10 fold) upon heat shock. SSA3 and SSA4 are only expressed at elevated temperatures. Phenotypic variations depend on the number of SSA genes affected. Mutants with a single SSA gene disruption are indistinguishable from WT, ssa1:ssa2 double mutant strains are not viable at 37°C but will grow slowly at 34°C and ssa1:ssa2:ssa4 triple mutants are completely inviable (Deshais *et al.*, 1988a). The latter mutant can be rescued by a single copy of ssa1 present on a plasmid and fused to a Gal 1 promoter (Deshais *et al.*, 1988a).

Repression of the Gal 1 promoter by addition of glucose allowed examination of the effect of hsp70 depletion (specifically Ssa1 protein) on the biogenesis of secretory (α-factor) and mitochondrial (F_1-ATPase) proteins. Once SSA1 protein dropped below wt levels, accumulation of the precursor proteins, prepro-α-factor (ppαf) and pre-F_1ß could be seen (Deshais *et al.*, 1988b). These results strongly indicated a role for hsp70 in the protein importing mechanism of mitochondria and ER.

Chiraco *et al.*, (1988) used an *in vitro* system (wheat germ extract) to demonstrate a stimulation of ppαf uptake into microsomal vesicles by the ssa 1 and ssa 2 gene products. Both Chiraco *et al.*, (1988) and Deshais *et al.*, (1988b) agree

that the role of hsp70 in translocation likely involves an unfoldase function, and possibly disruption of aggregate formation.

Munro and Pelham (1986), in a search for hsp70 homologues in yeast, identified a novel hsp70 which contained a signal peptide. This ER protein had previously been identified as a glucose regulated protein (grp 78; see Lee, 1987 for review) and an immunoglobulin heaving chain binding protein or BiP (Haas and Wabl, 1983). BiP was found to bind incompletely folded or assembled protein and could be released from its substrates *in vitro* by the addition of ATP (Munro and Pelham, 1986). It has been shown that BiP is essential for normal secretion and hence cell viability (Normington *et al.*, 1989; Rose *et al.*, 1989). According to Pelham (1989), recent evidence indicates normal membrane and secretory proteins, when emerging into the ER, transiently associate with BiP prior to achieving their ultimate conformational state. nI mammalian species, at least three main members of the hsp70 family have been identified. These include the highly inducible hsp70, the 70-86 kDa glucose regulated protein (located in ER; BiP related) and the constitutively expressed hsc70 (Green and Liem, 1989). Recently, a glucose regulated mitochondrial protein, grp 75, which is a member of the hsp70 family, has been identified in rat and HeLa cells (Mizzen *et al.*, 1989). Upon heat shock, both hsp70 and hsc70 have been found to migrate to the nucleus and are associated with

nucleoli (Welch and Suhan, 1985). This binding is readily reversible upon addition of ATP (Lewis and Pelham, 1985). An ATP dependent uncoating of clathrin coated vesicles by hsc70 has also been demonstrated (Chappell *et al.*, 1986).

Numerous reports suggest that protein(s) related or identical to hsc70 may be associated with MTs and intermediate filaments (see Green and Liem, 1989). Napolitano *et al.*, (1985) described a ubiquitous cytoskeletal associated protein, termed ß-internexin, which appeared to be related to hsc70. Recently, this identity was confirmed through immunological evidence and partial amino acid sequence comparisons (Green and Liem, 1989). A colchicine resistant mutant (Col^R22) isolated in this lab shows an electrophoretic alteration in a major 70 kDa protein (P2) which was judged to be MT-related, based on both genetic and biochemical data (Gupta and Gupta, 1984). Recently, imunological comparisons of P2 and hsc70 anti-sera and translation of a hsc70 cDNA clone has provided direct evidence of their identity (Ahmad *et al.*, 1990).

1.3.2b **The GroEL family (chaperonins)**

Sequence determinations of cDNAs for the chloroplast protein involved in the assembly of ribulose bisphosphate carboxylase (rubisco) led to the identification of a subset of molecular chaperones termed 'chaperonins'. These proteins are constitutively expressed, but increase in abundance after

heat shock (Ellis, 1990). Chaperonins are related by sequence and are found in all bacteria, mitochondria and plastids (Hemmingsen *et al.*, 1988).

The best studied chaperonin is the product of the groEL gene in *E. coli*. In conjunction with the smaller chaperonin GroES encoded in the same operon, they were first identified as the host products required for correct assembly of the head of phage lambda T4 and tail of T5 (reviewed by Friedman *et al.*, 1984). Both GroE proteins possess a seven-fold axis of symmetry, GroES being composed of seven identical subunits of 97 amino acids each, and GroEL of 14 identical subunits of 548 amino acids each (Georgopoulos and Ang, 1990). Genetic studies indicate both GroE products are essential for bacterial growth at all temperatures (Fayet *et al.*, 1989). Evidence of a functional interaction between GroEL and GroES includes (a) the ATP dependent co-sedimentation of GroEL and GroES, (b) the specific retention of GroES protein on a GroEL-affinity column, and (c) the inhibition of the weak ATPase activity of GroEL by GroES (Chandrasekhar *et al.*, 1986).

The isolation of temperature-sensitive GroE mutants has been instrumental in identifying cellular functions involving GroEL and GroES. It has been found that GroE$^-$ mutations result in (a) reduction in DNA and RNA synthesis rates at the non-permissive temperature (Wada and Itikawa, 1984), (b) a block in cell division at non-permissive temperatures (Georgeopoulos and Eisen, 1974), and (c) an overall reduction

in protease activity in GroEL$^-$ but not GroES$^-$ mutants (Straus *et al.*, 1988). Van Dyk *et al.*, (1989) have recently demonstrated the ability of overproduced GroE chaperonins to suppress the temperature-sensitive phenotypes of mutations in a variety of genes. One of these genes, dna A, is involved in the initiation of DNA replication. The authors concluded that GroE overproduction allows correct folding or oligomerization of the mutant polypeptide at non-permissive temperatures (Van Dyk *et al.*, 1989).

In an elegant set of experiments, Bochkareva *et al.*, (1988) employed a photoreactive crosslinker on pre-ß-lactamase (pre-ß-lac) or chloramphenicol acetyltransferase (CAT) to study the interaction of GroEL with newly synthesized proteins. Plasmids containing pre-ß-lac or CAT were translated *in vitro* and the irradiated mixture sedimented and analyzed under various conditions. The authors demonstrated that (a) GroEL forms a transient complex with the newly synthesized polypeptide chains, (b) this complex was not observed with mature, folded proteins, (c) the interaction of GroEL with pre-ß-lac resulted in more efficient membrane translocation of pre-ß-lac, and (d) addition of ATP results in dissociation of the protein-GroEL complex. Similar studies were performed using purified chaperones; GroEL, Sec B and trigger factor, and comparisons were made with respect to their abilities to bind the unfolded pro-ompA polypeptide (Lecker *et al.*, 1989). One interesting observation made was

that the relative binding affinities of these chaperones to pro-ompA (Sec B > GroEL > trigger) was exactly opposite to the relative abundance of these proteins.

Recently, Phillips and Silhavy (1990) have shown that overexpression of the GroE operon allows quantitative export of a lam-ß-lac Z mutant polypeptide (which is unable to be properly exported). Similar results were obtained with dna K, a member of the hsp70 family (Phillips and Silhavy, 1990).

The plastid chaperonin was first described by Barraclough and Ellis (1980) during studies on the synthesis and assembly of the chloroplast enzyme ribulose bisphosphate carboxylase-oxygenase (rubisco). In higher plants, rubisco is a soluble chloroplast protein composed of eight large (55 kDa) and eight small (14 kDa) polypeptide subunits (L_8S_8; Barraclough and Ellis, 1980). Time course experiments indicated a long period existed between the onset of synthesis of large subunits and the onset of assembly. When run on non-denaturing polyacrylamide gels, newly synthesized large subunits were found to be associated with a protein of subunit molecular weight 60 kDa. The rubisco subunit binding protein (RBP) was transiently associated with the large subunit and was not itself a component of the assembled holoenzyme (Barraclough and Ellis, 1980).

The purification and characterization of the RBP from pea leaves revealed the following: (a) the binding protein consists of two types of subunits of apparent M_r 61 kDa (α) and M_r 60 kDa (ß), (b) the subunits are encoded by nuclear

genes and are imported into chloroplasts after synthesis as a precursor form by cytosolic ribosomes, (c) the purified oligomeric binding protein has an estimated M_r of approximately 700 kDa based on molecular sieving, and (d) binding protein oligomer dissociated reversibly upon addition of Mg-ATP, demonstrating a weak ATPase activity (Hemmingsen *et al.*, 1988).

The complete coding sequence was obtained from a wheat RBP protein cDNA clone (Hemmingsen *et al.*, 1988). The calculated M_r of the mature α subunit was found to be 57,393 based on the predicted amino acid sequence. Furthermore, a high degree of sequence homology (46% identical with many conservative substitutions) was found between the α subunit binding protein and the GroEL protein (Hemmingsen *et al.*, 1988). An earlier report showed that total leaf extracts of *P. sativum* contained a protein similar to the GroEL protein in that it consists of 14 subunits, arranged in two layers of 7 monomers, and showed weak ATPase activity (Pushkin, 1982). This observation led Hemmingsen *et al.*, (1988) to suggest that RBP was identical to GroEL and that the binding protein consists of seven α and seven ß subunits.

The similarity in function of the GroEL and RBP was demonstrated by Goloubinoff *et al.*, (1989a) who showed that assembly of foreign prokaryotic dimeric (L_2), octomeric (L_8) and hexadecameric (L_8S_8) rubiscos in *E. coli* required both GroEL (cpn60) and GroES (cpn10). The *in vivo* studies involved

transformation of *E. coli* with separate plasmids containing either GroE genes or rubisco genes. Dramatic increases in the amount of assembled L_8 cores were seen when the GroE proteins were overproduced, however, when a mutant ($GroE^-$) plasmid was transformed into *E. coli* with rubisco genes, neither rubisco activity, nor assembled holoenzyme, could be detected (Goloubinoff *et al.*, 1989a).

In a similar study, purified cpn60 and cpn10 were used to demonstrate that the *in vitro* reconstitution of the catalytically functional dimeric form of rubisco from an unfolded biologically inert polypeptide was dependent on both chaperonin proteins and Mg-ATP (Goloubinoff *et al.*, 1989b). These studies involved following the renaturation of guanidine-HCl or urea treated rubisco proteins, utilizing the inherent differences in circular dichroism between active and denatured protein. The absence of the GroE gene products (cpn60-cpn10) resulted in the formation of biologically unproductive aggregates upon dilution of the denatured rubisco, whereas dilution into chaperonin containing solutions under optimum conditions resulted in approximately 90% recovery of active rubisco (Goloubinoff *et al.*, 1989b). The heterologous chaperonin combinations cpn60 (yeast)-cpn10 (*E. coli*) and cpn60 (chloroplast)-cpn10 (*E. coli*) were respectively 10% and 25% as effective as the homologous cpn60-cpn10 (*E. coli*) combination in facilitating recombination of the urea denatured prokaryotic rubisco (Goloubinoff *et al.*, 1989b).

Several other proteins have been found to form stable complexes with the RBP upon import into chloroplasts, including the ß subunits of ATP synthase, glutamine synthase, the light-harvesting chlorophyll a/b binding protein, chloramphenicol acetyltransferase, and pre-ß-lactamase (Lubben *et al.*, 1989). The observed interaction of several disparate imported proteins with cpn60 indicates a general role for this oligomer at some point in the folding or assembly pathway of imported chloroplast proteins (Lubben *et al.*, 1989).

An abundant heat shock protein of 58 kd that is structurally related to GroEL was purified from Tetrahymena mitochondria, and was also shown to consist of a 14 subunit double ring with weak ATPase activity (McMullin and Hallberg, 1987, 1988). Antibodies to this protein displayed cross reactivity to related proteins of 58-64 kd in yeast, plant, amphibian and mammalian mitochondria (McMullin and Hallberg, 1988). The sequence of this protein (hsp60) was obtained from yeast and found to possess 54% and 43% amino acid identity with GroEL and RBP, respectively (Reading *et al.*, 1989).

Cheng *et al.*, (1989) have described the isolation of a temperature-sensitive lethal nuclear mutation in yeast (mif 4). At the restricted temperature, subunits of the mitochondrial enzymes ornithine transcarbamylase and ß subunits of F_1-ATPase are processed proteolytically but are not assembled into functional oligomeric enzymes. In addition, the import pathway of cytochrome b2, which passes

through the matrix compartment *en route* to the intermembrane space, was affected. This resulted in a build-up of cytochrome b2 in the matrix at non-permissive temperatures. Wild-type hsp60 was shown to fully rescue the defects of mif 4 when introduced by homologous recombination, indicating mif 4 and hsp60 are one and the same gene (Cheng *et al.*, 1989).

Ostermann *et al.*, (1989) devised an *in vitro* system for studying the import and folding of proteins targeted to the mitochondria. Dihydrofolate reductase (DHFR) linked at its N-terminus to a matrix-targeting signal peptide was incubated with mitochondria, imported into the matrix where it was cleaved to yield a mature DHFR product (< 1 min at 25°C). The protein was then folded into a compact, protease-resistant monomeric conformation with a half-time of approximately 5 min. The use of ATP depleted mitochondria, or non-hydrolyzable ATP analogs, resulted in DHFR import and cleavage, but the folded form was no longer observed. The protease-sensitive unfolded form was found to (a) co-elute with hsp60 protein upon gel filtration and (b) co-precipitate with antibodies directed against hsp60. Mutant mif 4 mitochondria failed to fold DHFR into compact monomers. Authentic mitochondrial proteins, such as the Rieske iron sulphur protein of complex III and the ß-subunit of F_1-ATPase, were also found to accumulate in the matrix in high molecular weight complexes with hsp60 when imported under conditions of ATP depletion (Ostermann *et al.*, 1989).

These recent observations have enhanced the understanding of the processes involved in the import, sorting, folding and assembly of mitochondrial proteins. Current speculations include the concept of a direct role for hsp60 in the translocation of precursor proteins (Hartl and Neupert, 1990) although no direct evidence for this hypothesis exists.

1.4 The 65 kDa Mycobacterial Antigen, Chaperonins, and Autoimmunity

The characterization of rabbit antisera raised against *Pseudomonas aeruginosa* led to the observation of a cross reactive protein antigen which was widely distributed amongst bacteria, termed the 'common antigen' (Hoiby, 1975). Biochemical studies of the *Pseudomonas* protein revealed features which are now known to be characteristic of the chaperonin family (Young, 1990a). Studies on the murine antibody response following immunization with extracts of pathogenic mycobacteria revealed a high proportion of antibodies recognizing a 65 kDa protein which was cross reactive with proteins of similar size present in a wide range of bacteria (Engers *et al.*, 1985, 1986; Young *et al.*, 1987). Cloning and sequencing of the gene for the 65 kDa protein from different strains of mycobacteria (*M. leprae*, *M tuberculosis* and *M. bovis*) show that its deduced amino acid sequence is extensively

homologous to the *E. coli* GroEL gene product (Young *et al.*, 1985; Shinnick, 1987; Thole *et al.*, 1987).

Anti-mycobacterial monoclonal antibodies have been used in conjunction with recombinant DNA and synthetic peptide techniques to map the specific antigenic determinants involved in antibody recognition (Young, 1990b). These antibodies recognize multiple short, linear, peptide epitopes on the 65 kDa protein (Young, 1990b). Antibody and T-cell immune responses to chaperonins are also induced during infection by a wide variety of bacteria including *Legionella pneumophila*, *Mycobacterium tuberculosis*, *Bordetella pertusis*, *Treponema pallidum*, *Coxiella burnetii* and *Borrelia burgdorferi* (Young, 1990b).

The structural similarity and highly conserved nature of the chaperonin protein in both nonviral pathogens and host cells increases the risk for an autoimmune reaction (Winfield, 1989). Recently, considerable evidence has accumulated which suggests that the immune response to the 65 kDa 'common antigen' may play an important role in the development of the chronic autoimmune condition of adjuvant arthritus in rats, and possibly the human autoimmune disease rheumatoid arthritis (RA) (for review see Van Eden *et al.*, 1989; Winfield, 1989; Young, 1990a; Young, 1990b).

Pearson (1956) first demonstrated the induction of adjuvant arthritis in rats upon intracutaneous injection of heat-killed mycobacteria suspended in mineral oil. The autoimmune nature of the disease was suggested by the

observation that the disease could be transferred by T lymphocytes (Pearson and Wood, 1964). Holoshitz *et al.*, (1984) selected T-cell lines from Lewis rats immunized with crude mycobacteria. One of the lines (A2) was found to induce arthritis when innoculated in irradiated syngeneic rats but not in non-irradiated recipients. Cloning of A2 revealed an arthritogenic subclone, A2b, and an exclusively protective subclone, A2c, which acts as a suppressor-inducer T lymphocyte (Cohen, 1985). In addition to mycobacterial antigens, clone A2b was found to respond to cartilage extracts enriched for proteoglycans (Van Eden *et al.*, 1985). Both clones A2b and A2c were found to recognize the same mycobacterial hsp60 epitope (180-188). A large number of epitopes recognized by CD4 helper T cells have been mapped using recombinant DNA and synthetic peptide techniques (Lamb *et al.*, 1987; Thole *et al.*, 1988). However, the exact nature and location of the rat autoantigen epitopes has yet to be defined (Young, 1990a).

Evidence for the hypothesis that molecular mimicry causes autoimmune diseases in man has been increasing in recent years (Oldstone, 1987). Elevated levels of IgG and IgA anti-65K have been reported in rheumatoid arthritis patients (Tsoulfa *et al.*, 1989). Recently, Karlsson-Parva *et al.*, (1990) showed a monoclonal antibody with cross reactivity between hsp60 and 65 kDa antigen possessed increased reactivity to material from rheumatoid joints and rheumatic nodules, as compared to normal joints and other types of chronic

inflammations. Current research includes efforts to incorporate data regarding protective T-cell subclones from the rat model into a strategy for vaccination against T-cells that recognize 65 kDa related epitopes (Van Eden *et al.*, 1989).

1.5 Objectives of the Current Study

Studies in this laboratory with CHO cells selected for resistance to antimitotic drugs have identified a protein, designated P1 (Mr 60 kDa), which showed a specific electrophoretic alteration in a large number of resistant mutants (Gupta *et al.*, 1982; Gupta and Gupta, 1984). The MT-associated nature of P1 was initially established by its co-release with tubulin from crude MT fractions under destabilizing conditions (Gupta *et al.*, 1982, 1985). However, subsequent immunofluorescent and subcellular fractionation studies localized P1 to the mitochondrial matrix (Gupta *et al.*, 1985; Gupta and Austin, 1988). The complete DNA sequence of P1 from both human (Jindal *et al.*, 1990) and CHO (Picketts *et al.*, 1990) sources has been obtained. The P1 cDNA shows extensive sequence and structural similarities to the bacterial (GroEL), yeast (hsp60) and plant (RBP) chaperonin family of proteins, as well as to the 65 kDa 'common' bacterial antigen (Jindal *et al.*, 1990).

Genomic clones have been isolated for a number of hsp's including hsp70 and 90. Analysis and comparisons of various

heat shock proteins have provided valuable insight into the factors involved in transcriptional activation of these proteins (Tanguay, 1988; Lindquist and Craig, 1988). Information regarding the copy number, structure and regulation of the P1 (hsp60) gene has not been reported for higher eukaryotes. Therefore, one of the objectives of this project was to screen a human genomic library for clones containing the hsp60 gene using probes obtained from the cloning of the human P1 cDNA.

A number of studies have linked the onset of AA in rats to the immune response to the 65 kDa 'common' antigen of bacteria (Winfield, 1989; Young, 1990a; Young, 1990b). The autoimmune nature of AA was indicated when T-cell clones (directed against 65 kDa specific epitopes) were transferred to non-infected, irradiated rats resulting in the onset of AA (Holoshitz *et al.*, 1983, 1984). As a result, hsp60 has been considered as a likely target for autoimmune response given the highly conserved nature of this family of proteins (Young, D.B., 1990). The isolation and sequencing of cDNA clones encoding the rat hsp60 gene would be useful in identifying epitopes on hsp60 which are stimulatory for arthritogenic T-cell clones in proliferation assays. This could be accomplished by synthesizing peptides which represent small portions of rat hsp60, based on sequence similarities to epitopes which have previously been mapped to the 65 kDa 'common' antigen. Therefore, the second part of this project

involves the characterization of rat (and mouse) hsp60 cDNA clones.

2.0 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Source of chemicals and reagents

Yeast extract, bacto-agar and bacto-tryptone were purchased from Difco Laboratories, Detroit, MI. Tris-HCl, tris-base, ampicillin, salmon sperm DNA, SDS, ethidium bromide, dithiothreitol, beta-mercaptoethanol, maltose and bromphenol blue were purchased from Sigma Chemical Co., St. Louis, MO. Nitrocellulose disks and rolls, acrylamide, bis-acrylamide, ammonium persulfate and TEMED were purchased from Bio-Rad Laboratories, Mississauga, Ont. DNA molecular weight markers (1 kb ladder), IPTG, phenol and X-gal were purchased from Bethesda Research Laboratories (BRL), Burlington, Ont. High molecular weight nucleic acid standards (Lambda DNA Hind III fragments) and agarose NA were purchased from Pharmacia P-L Biochemicals, Dorval, Que. Kodak XAR-5 and XK-1 negative film was purchased from Picker International, Ottawa, Ont.

2.1.2 Enzymes

Large fragments of *E. coli* DNA polymerase I (Klenow) and T_4 DNA ligase were purchased from BRL. Restriction endonucleases Eco RI, Bam HI, Sst I, Taq 1, Alu I, Sma I, Ava II, Xba I and Kpn I were purchased from BRL. Buffers were provided for each enzyme by the manufacturer. Deoxyribonuclease I and lysozyme were purchased from Sigma Chemical Co. Calf intestinal phosphatase and ribonuclease A were obtained from Promega Biotech.

2.1.3 Genomic and cDNA bacteriophage libraries and plasmid vectors

Information regarding the nature and source of the phage libraries used in this study is summarized in Table 2. The general structure of plasmid pGem-7zf(+) is shown in Fig. 3. Lysogen λ P1 22a, containing a 1.4 kb insert specific for the 3'-end of the human P1 cDNA, was constructed in this laboratory (Jindal *et al.*, 1989).

Plasmid P1-C5 contains a 4.4 kbp Kpn-I/Sst-I insert (encompassing the complete human P1 cDNA and some additional λgt11 flanking sequence) in pTZ18U (Jindal *et al.*, 1989). Digestion of P1-C5 with Eco RI yields P1-specific fragments of 0.8 (5'-end) at 1.4 (3'-end). Plasmid CHP1A-1.1 contains a 1.1 kbp Eco RI fragment (representing the 3'-end of the CHO P1-cDNA) (Picketts *et al.*, 1989). Plasmid pGem-7zf(+) was purchased from Promega Biotech .

Library	Vector	Source of insert	Insert size (Avg.)	Catalogue number	Supplier
Human Genomic	Charon 4A	Partial Hae III, Alu I, digestion of fetal liver DNA	15-20 kb	ATCC 37333	American Type Culture Collection
Human Genomic	Charon 4A	Partial EcoRI digestion	16-22 kb	ATCC 37385	"
Human Genomic	EMBL-3	Partial Sau 3A digestion of human leukocyte DNA	16 kb	HJ 1006d	Clonetech Inc.
Rat Kidney cDNA	λgt11	Entire rat kidney from Sprague-Dawley	0.6-3.8 kb (1.1 kb)	RL 1007b	"
Mouse 3T3 cell cDNA	λgt11	Embryonic balb/3T3 cells	0.5-2.5 kb (0.94 kb)	ML 1004b	"

Table 2 **Sources and description of bacteriophage libraries used in screening procedures.**

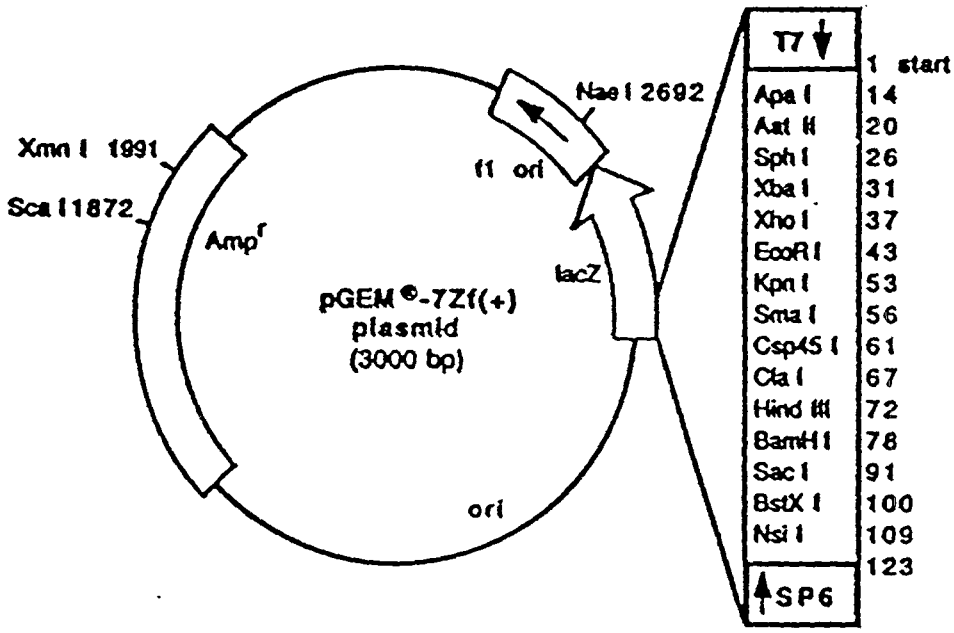


Figure 3 Plasmid vector pGem-7zf (Promega Inc.).

2.1.4 Bacterial strains

A number of *E. coli* K-12 strain derivatives were used including:

JM107, genotype endA1, gyrA96, thi, hsdR17, supE44, nelA1, Δ(lac-proAB); F' [traD36, proA+, proB+, $lacI^q$, lacZΔM15], was provided by Dr. K.B. Freeman, McMaster University.

JM109, genotype recA1 and A, sbcB15, hsdR4, (lac-proAB), [F',traD36, proAB, $lacI^qZM15$], was purchased from Pharmacia LKB Technology.

LE392, genotype F^-, hsdR515(r_k^-,m_k^-, supE44, supF58, lacY1 or Δ(lac12Y)6, galK2, galT22, metB1, trp55, λ^-, was purchased from Clonetech, Inc.

Y1090, genotype Δlac^U169, proAT, Δlon, araD139, strA, supF, trpC22:Tn10; pMC9 was purchased from Progema Biotech.

2.1.5 Molecular biology kits

DNA probes were labelled with the Random Primer Labelling System purchased from BRL. DNA dideoxy sequencing was performed using the Sequenase kit obtained from United States Biochemicals. DNA fragments were recovered from agarose gels using a Gene Clean kit obtained from Bio-Can, Mississauga, Ont.

2.1.6 Radiochemicals

[Alpha-^{32}P]dATP (3000 Ci/mmol, 10 μCiμl) was purchased from Amersham Radiochemicals Co., Oakville, Ont.

2.1.7 Synthetic oligonucleotide primers

A number of oligonucleotide primers specific for the human and CHO P1 gene were custom synthesized by the Institute for Molecular Biology and Biotechnology, McMaster University. These include forward primers; AB729(5'-GAGCCTTAATGCTTCAAG-3'), AB730 (5'-GGCACTACCACTGCTACT-3'), AB731 (5'-ATCATCTCTGATGCAATGA-3'), AB728 (5'-ACCTGCTCTTGAAATTGC-3'), AB725 (5'-AAGTTGGAGAGGTCATTGT-3'), AB726 (5'-AGACAGAGTTACAGATGC-3'), AB727 (5'-GGATCTTTGATAGTTGAG-3') and reverse primers: AB634 (5'-GTTCTTCCCTTTGGCCCC-3'), AB674 (5'-AATTACAGCATCAACAGC-3'), AB308 (5'-CTCAACAGAACATAGGCAT-3'), AB277 (5'-AACTGCTCAATGATTTCTTG-3'), AB675 (5'-CAACACCTGCATTCTTAGC-3'),AB537 (5'-CCCATTCCAGGGTCCTTC-3'), AB357 (5'-TGTCAACTGAAACCAGTAAC-3'), AB538 (5'-CCAGTATCAGGAATGTAC-3') and reverse universal primer AB172 (5'-AACAGCTAGTACCATG-3'). Forward universal primer (5'-GTAAACGACGGCCAGT-3') was provided with the Sequenase kit.

2.2 METHODS

2.2.1 Random primer labelling

DNA probes were labelled using a random primer labelling kit (BRL) using a modified methodology based on the protocol provided with the kit. Probe DNA (250-500 ng) was mixed with water to a final volume of 20 µl and denatured in a boiling water bath for 5 min. The sample was quick cooled on ice and 2 µl each of dCTP, dGTP and dTTP were added, followed by 15 µl of random primer buffer mix, 5 µl of water, 2 µl [α-^{32}P]dATP], (10 µCi/µl, 3000 Ci/mmol) and 1 µl of Klenow fragment. Following a 2-3 hr room temperature incubation, 5 µl of stop buffer, 25 µl 7.5 M ammonium acetate and 160 µl of absolute ethanol was added. After mixing, the sample was chilled at -70°C for 15 min and then centrifuged for 15 min in a microcentrifuge to pellet the precipitated DNA. The pellet was washed once with 70% ethanol, once with absolute ethanol, air dried and resuspended in 100 µl TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). To estimate the number of counts incorporated into the probe, a 1 µl aliquot was spotted on a DE81 filter paper disc. The filter was dried, placed into a scintillation vial with 3 µl of toluene containing 4 g omnifluor/l and counted in a scintillation counter.

2.2.2 Titering of phage libraries

Phage library stocks were diluted in SM medium (0.1 M

NaCl, 10 mM $MgSO_4$, 50 mM Tris-HCl (pH 7.5), and 0.1% gelatin). Two or three small aliquots (between 5 and 50 μl) were added to 200 μl of an overnight culture of either LE392 (for Charon 4A or EMBL-3) or Y1090 (λgt11) cells, grown in LB broth (10 g NaCl, 10 g bacto-tryptone, 5 g yeast extract per liter) supplemented with 0.2% maltose. Incubation at 37°C for 20 min allowed phage absorption to bacteria which were then added to 3 ml of top agarose (0.7 g/l agarose in LB broth) which had been maintained at 45°C in dry bath incubator. After mixing the top agarose was plated onto agar plates (10 ml of LB broth containing 15 g/l bacto-agar). The plates were incubated at 37°C (Charon 4A, EMBL-3) or 42°C (λgt11) for 5-7 hr. The plaques were counted on each plate and an average pfu value calculated.

2.2.3 **Screening of human genomic and cDNA libraries**

Phage were plated on LB agar at concentrations varying from 5 x 10^3 to 1 x 10^4 and incubated at the permissive temperatures described above. When plaques were clearly evident the plates were cooled at 4°C for 1 hr. Nitrocellulose disks were placed on the plates for 4-5 min during which time the orientation of the filter on the plate was recorded by stabbing a 20-G needle (which had previously been dipped in Indian ink) through the filter at several asymmetric points around the edge of the plate. Duplicate filters were placed on plates for 10-12 min and marked at the

same position as the original. Upon removal the filters were air dried on the benchtop for 20 min and the plates stored at 4°C.

Denaturation, neutralization and hybridization were carried out essentially as described by Quartermous (1987) and Strauss (1987). Briefly, the filters were placed on 3 MM paper saturated with 0.2 M NaOH, 1.5 M NaCl, followed by neutralization on 0.4 M Tris-HCl (pH 7.6), 2 x SSC, and finally on 2 x SSC, each for a duration of 2 min. The filters were then drained on a dry sheet of 3 MM paper and baked at 80°C for 2 hr in a vacuum oven.

After carefully soaking the filters in 65°C hybridization solution (1% BSA, 1 mM EDTA, 0.5 M $NaHPO_4$, pH 7.2, 7% SDS), 10-20 filters were sealed in plastic bags and prehybridized at 65°C in 20 ml of hybridization solution. Radiolabelled probes (1-2 x 10^6 cpm) were boiled in 1 ml of salmon sperm DNA (2 mg/ml) for 10 min, cooled on ice and added directly to the bag.

Following overnight incubation the filters were transferred from the bags into 250 ml of low stringency wash buffer (0.5% BSA, 40 mM $NaHPO_4$ (pH 7.2), 1 mM EDTA, 5% SDS). After 15 min the above wash was repeated, the filters were quickly washed several times (5-8) with high stringency wash buffer (40 mM $NaHPO_4$ (pH 7.2), 1% SDS, 1 mM EDTA, 65°C), followed by a final 15 min wash at 65°C. After drying, the filters were exposed on fast film (XAR-5) at -70°C for 2 days.

Phage clones which hybridized to probe on duplicate

filters were adjudged to be positives. These clones were excised from the agar as plugs using a pipette tip which had a portion of the tip cut off. The plugs were placed in a tube containing 1 ml of SM medium and a small drop of chloroform. The tube was rotated at 4°C for 2 hr to overnight, allowing diffusion of the phage into the SM. The resulting mixed population was then plated out at lower density in a second round of plaque purification. This procedure was repeated until a single plaque could be isolated which upon replating demonstrated hybridization to every plaque on the dish.

2.2.4 **Phage DNA isolation**

Two different plate lysate procedures were used to isolate phage DNA. Initially, the method of Ziai *et al.*, (1988) was employed for phage clones containing human genomic DNA. Briefly, plate lysates were prepared by infecting 200 μl of an overnight culture of LE392 with approximately 1 x 10^5 pfu of purified phage and plating it on an LB agarose plate (100 mm diameter). The plates were allowed to proceed to confluent lysis. The phage was then eluted by adding 5 ml of 100 mM Tris-HCl, 10 mM $MgSO_4$ (pH 7.4) and shaking for 2 hr . The lysate was cleared of bacterial debris by centrifugation (SS 34 rotor; 20,000 x g; 5 min at 4°C). The cleared lysate was mixed with an equal volume of saturated ammonium sulfate solution, incubated on ice for 30 to 60 min, and precipitates

collected by centrifugation (20,000 x g; 20 min at 4°C). The pellet was resuspended in 0.5 ml of TE 0.1% SDS buffer and treated sequentially with RNase A (20 μg/ml, 20 min at 60°C) and proteinase K (200 μg/ml). Phage particles were disrupted by addition of NaOH (one-tenth volume, 10 min) followed by neutralization with 10 M ammonium acetate, pH 6.0. After a brief centrifugation to remove debris, the supernatant was extracted twice with phenol/chloroform. The phage DNA was then precipitated with an equal volume of isopropanol and pelleted in a microcentrifuge. The pellet was washed in 70% ethanol, vacuum dried and resuspended in 50 μl of TE.

Phage DNA samples from λgt11 clones were prepared essentially as described by Sambrook *et al.*, (1989). Briefly, plates which had reached confluent lysis were eluted with 5 ml SM. After centrifugation and RNAse A/DNAse-I-treatment, the intact phage particles were precipitated with 20% w/v polyethylene glycol and 2 M NaCl in λ diluent (10 mM Tris-HCl, 10 mM $MgSO_4$, pH 7.5). Phage particles were disrupted by incubation at 68°C in the presence of SDS. Phage DNA was precipitated and pelleted as previously described.

2.2.5 **Subcloning DNA fragments into pGem-7zf(+)**

(a) Restriction Enzyme Digests

Phage or plasmid DNA digestions were performed in a final volume of 10-50 μl including 5-20 units of the

appropriate enzyme (depending on DNA concentration) and one-tenth the volume of the 10 X buffer recommended by the manufacturer. After incubation at 37°C for 1-3 hr the enzyme was either heat denatured at 70°C and removed by phenol/-chloroform extraction or the sample was directly analyzed on a 0.8 to 1% agarose gel (see 2.2.7).

(b) Dephosphorylation of Vector DNA

The foreign DNA fragments in the phage vectors had been inserted at a unique site (Eco RI or Bam HI). In order to subclone the entire insert in one fragment the same unique cloning site was utilized in the plasmid vector. Plasmid self-ligation was minimized by pre-treatment with calf intestinal phosphatase (CIP), thereby removing terminal 5' phosphates as described by Maniatis *et al.*, (1982).

(c) Ligation

The ratio of insert to vector used was 5:1. Phage DNA (250-500 ng) was digested with one of the following enzymes or combinations; Eco RI, Bam HI, Bam HI/Eco RI or Kpn-I/Sst-I. Plasmid pGem-7zf(+) DNA (50-100 ng) was digested with the appropriate enzyme(s) (dephosphorylated where necessary) mixed with insert DNA, 5 X ligase buffer, 1.0 unit of T_4 DNA ligase and incubated overnight at 14°C to allow ligation to occur.

(d) Preparation of Competent Cells

Competent cells were prepared by a modification of the method described by Kushner (1978). A single colony of JM109 or JM107 cells was inoculated into 3 ml of LB medium and grown

overnight with vigorous shaking at 37°C. An aliquot (0.5 ml) of this culture was added to 100 ml of LB broth in a 500 ml flask. After 2.5 hr of growth (approximate OD 600 = 0.6) the culture was transferred to 40 ml plastic centrifuge tubes and the cells harvested at 5500 rpm (SS34) for 10 min (4°C). The cell pellets were washed with ice cold 10 mM PIPES buffer (pH 6.8) containing 10 mM $RbCl_2$. The cells were pelleted and resuspended in 10 mM PIPES buffer (pH 6.8), 10 mM $RbCl_2$, 75 mM $CaCl_2$, and incubated on ice for 30 min. After repelleting, the cells were gently resuspended in the latter buffer containing 15% glycerol, aliquoted into microcentrifuge tubes (700 µl/tube), and frozen by quickly placing in liquid nitrogen and stored at -70°C.

(e) Bacterial Cell Transformation

The protocol used for cell transformation was modified from that described by Mandel and Higa (1970). The frozen competent cells were thawed on ice and 150 µl added to the ligation mix. The cells were shocked by incubation at 37°C for two min and placed on ice for 30 min. After addition of LB broth (to a total volume of 1 ml) the cells were incubated for 1 hr at 37°C with constant agitation. The cells were pelleted (30 sec in a microfuge) and 750 µl of supernatant was removed. The suspended mixture was spread onto LB plates containing 100 µg/ml ampicillin, 50 µl of 2% X-Gal, and 10 µl of 100 mM IPTG.

2.2.6 **Isolation of plasmid DNA**

Initially plasmid DNA was prepared by the method of Birnboim and Daly (1979). A single bacterial colony was inoculated into 3 ml of LB broth supplemented with 100 μg/ml ampicillin. The culture was grown overnight at 37°C with vigorous shaking. A 1.5 ml aliquot was centrifuged for 1 min in a microcentrifuge. The supernatant was decanted and the pellet resuspended in 100 μl of lysis buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris HCl (pH 8.0). After 5 min at room temperature, 200 μl of 0.2 N NaOH/1% SDS was added and the tube incubated on ice for 5 min. Finally, the contents were incubated on ice for 5 min after the addition of 150 μl of ice cold potassium acetate solution (pH 5.2). After centrifugation in a microcentrifuge to remove debris, the supernatant was extracted once with phenol/chloroform then ethanol precipitated. The DNA was pelleted by centrifugation and the pellet washed with 70% ethanol. After air-drying, the pellet was resuspended in TE buffer containing RNase A (20 μg/ml).

More recently, plasmid DNA has been isolated by the method of Morelle (1989). This procedure is basically the same as above, however, substitution of 7.5 M ammonium acetate for potassium acetate eliminates the need for a phenol/-chloroform extraction step.

2.2.7 **Agarose gel electrophoresis**

Restriction digests of phage and plasmid DNA

preparations were analyzed on 0.8 to 1% agarose gels. The gels were prepared by dissolving agarose in 1 X TAE buffer (0.04 M Tris-acetate(pH 8.3), 0.001 M EDTA) in a microwave oven for 3-4 min. After cooling to 50°C, solutions were poured into gel forms and allowed to set. The gels were placed in 1 X TAE buffer containing 0.5 μg/ml ethidium broimide. Samples were mixed with one-fifth a volume of loading dye (0.25% bromphenol blue, 0.25% Xylene cyanol, and 30% glycerol) and applied to the gel. Electrophoresis was performed for 1-3 hr at 100 V or overnight at 30 V. The gel was then visualized on an ultraviolet light box and photographed.

2.2.8 Purification of DNA from agarose gels

Agarose gels prepared as described in 2.2.6 were visualized on an ultraviolet light box (preparative setting) and the band of interest excised with a razor blade. DNA was recovered with a Gene Clean Kit following the procedure supplied by the manufacturer (Bio-Can). Recoveries were generally in the 50 to 80% range depending on the size of the DNA fragmment.

2.2.9 Southern analysis

Southern (1975) analysis of DNA fragments were performed as described by Maniatis *et al.*, (1982). DNA gels were electrophoresed and photographed as described (2.2.6). The gels were denatured for 1 hr in 1.5 M NaCl/0.5 M NaOH, neutralized in 1.5 M NaCl, 1 M Tris-HCl (pH 8.0) for 1 hr and briefly soaked in 2 X SSC prior to transfer to nitrocellulose.

The method of transfer employed results in two identical copies of the gel which are reverse images. This allows the sample to be analyzed with more than one probe. Briefly, 3 MM paper, nitrocellulose sheets and paper towels were cut to the same size as the gel. The nitrocellulose was soaked in 2 X SSC until saturated and placed on either side of the gel. Two pieces of 2 X SSC saturated 3 MM paper were placed on either side of the nitrocellulose and 4 to 5 inches of the pre-cut paper towels were stacked on either side of the 3 MM paper. Finally, a glass plate and small weight (approximately 500 grams) was placed on top of the sandwich to enhance transfer efficiency. The transfer was completed within 3 hr.

Upon completion of transfer, the nitrocellulose filters were retrieved, soaked 5 min in 6 X SSC, and the well positions marked with a pencil. The filters were then baked at 80°C for 2 hr in a vacuum oven. The baked filters were again soaked in 6 X SSC for 5 min before prehybridization at 68°C in a seal-a-meal bag with 0.2 ml of hybridization solution (6 X SSC, 0.5% SDS, 5 X Denhardt's solution (0.1%

Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin) and 100 μg/ml of denatured salmon sperm DNA) for each square cm of filter. After 1 hr, the probe (1-2 x 10^6 cpm) which had been previously denatured (10 min in a boiling water bath, followed by cooling on ice) was added to the bag. Hybridization proceeded overnight at 68°C in a shaking water bath.

The filters were removed from the bag and placed in wash solution 1 (2 X SSC and 0.5% SDS) for 5 min at room temperature. This was followed by 15 min at room temperature in wash solution 2 (2 X SSC and 0.1% SDS) and two washes of 30 min each at 68°C in wash solution 3 (0.1% SSC and 0.5% SDS). After air-drying the filter was exposed to film at -70°C for 8 to 16 hr.

2.2.10 **DNA sequencing**

(a) Template Preparation and Sequencing Reactions

Double stranded DNA templates were used for sequencing and were prepared by the method described by Zhang *et al.*, (1988). Plasmid DNA (1-2 μg), isolated as described in 2.2.6, was adjusted to 20 μl with sterile distilled water then denatured for 5 min by the addition of 2 μl of 2 M NaOH, 2 mM EDTA. The reaction was neutralized by adding 3 μl of 3 M sodium acetate (pH 5.2) and 7 μl of sterile distilled water. After mixing, the DNA was precipitated by the addition of 75

μl of absolute ethanol, stored for 10 min at -70°C, and recovered by centrifugation for 10 min. The pellet was washed with ice-cold 70% ethanol then absolute ethanol, air-dried and dissolved in 7 μl of sterile distilled water.

Sequencing buffer (2 μl) and the appropriate primer (1 μl, approximately 0.5 pmol) were mixed with template DNA and annealed for 2 min at 65°C and slow cooled to below 37°C. Oligonucleotide primers used were previously described in 2.1.7. Sequencing reactions were performed with the annealed template and primer according to the protocol provided with the Sequenase kit. This method was modified from the chain termination protocol developed by Sanger *et al.*, (1977) and employs a modified DNA polymerase (Tabor and Richardson, 1987).

Direct sequencing of phage clones containing human genomic DNA inserts was performed using a protocol developed in this laboratory. Phage DNA (purified as described in 2.2.4) was prepared for sequencing by pre-digesting with a frequent cutter such as Alu I or Taq I, and then boiling 10 μg of DNA and 1.0 pmol of primer for 5 min. The primer/template mixture was immediately placed into liquid nitrogen. The labeling reaction mixture (containing labeling mix, ^{32}P-dATP, DTT and Sequenase enzyme) was added directly to the frozen primer/template mixture and spun briefly in a microcentrifuge, allowing the solution to melt and mix. Termination reactions were carried out immediately as

described in the protocol provided with the Sequenase kit.

(b) Preparation and Electrophoresis of Sequencing Gels

Sequencing gels were prepared on a Bio-Rad integral plate/chamber sequencing apparatus (38 cm x 50 cm). Prior to setting up the apparatus, the glass plates were washed with soap and water, rinsed with deionized water and washed again with absolute ethanol. The plates were then assembled according to the instructions provided by the manufacturer.

The acrylamide gel solution was prepared by dissolving 84 g of urea into 33.3 ml of a 30% acrylamide stock solution (29:1 acrylamide:bis-acrylamide) and 20 ml of 10 X TBE buffer. The volume was adjusted to 200 ml with deionized water and deaerated for 10-15 min.

Eighty milliliters was removed to a flask, 300 μl of a fresh 25% APS solution and 80 μl of TEMED were added, the solution was mixed and poured into the gel casting tray to seal the bottom of the gel apparatus. After polymerization (5 to 10 min), the apparatus was placed on an angle and the remaining gel solution poured following the addition of 105 μl of 25% APS and 90 μl TEMED. The gel was fitted with a well forming comb and left to polymerize overnight.

The gel apparatus was inserted into the tank buffer after removal of the casting tray and pre-electrophoresed at 72 W for 1 hr in 1 X TBE buffer. The samples were heated at 75°C for 2 min, quickly cooled on ice, and loaded. Electrophoresis was at 72 W until the first dye ran off the gel. Two or 3 loadings were applied for each sequencing

reaction. Upon completion, the gel apparatus was disassembled and the gel transferred onto a pre-cut sheet of 3 MM paper. The gel was dried on a Bio-Rad gel dryer at 80°C for 30 min and exposed on Kodak XK-1 film overnight at -70°C.

2.2.11 Computer analysis

DNA and protein sequences were compared with the IBI DNA/Protein Sequence Analysis System (IBM XT version) obtained from International Biotechnologies, Inc., New Haven, CT.

3.0 RESULTS

3.1 Analysis of P1 Specific Human Genomic DNA Clones Isolated from λ Charon 4A and EMBL-3 Phage Libaries

3.1.1 Screening of phage libraries

Initially, the λ Charon 4A genomic libraries ATCC 37333 and ATCC 37385 were plated at a concentration of 0.5 to 1 x 10^4 pfu/plate on 20 dishes. Duplicate plaque lifts were performed prior to hybridization. In addition, a filter was lifted containing plaques from λP1-22a for use as a positive control. One filter from each set was probed with either the 5'- end (0.8 kb) or 3'- end (1.4 kb) human P1 specific fragments. Plaques which produced a strong signal, in identical positions, with each probe were adjudged to be positives. Three clones were found to fulfill these criteria. Figure 4 shows clone D3 as an example of one of the positive clones, which was obtained in duplicate (Figure 4a,b), in comparison to the same clone after further purification (Figure 4c), and the positive control (Figure 4d). The positive plaques (D3, C2, C3) were plaque purified as described in Materials and Methods.

After plaque purification of the positive clones, small

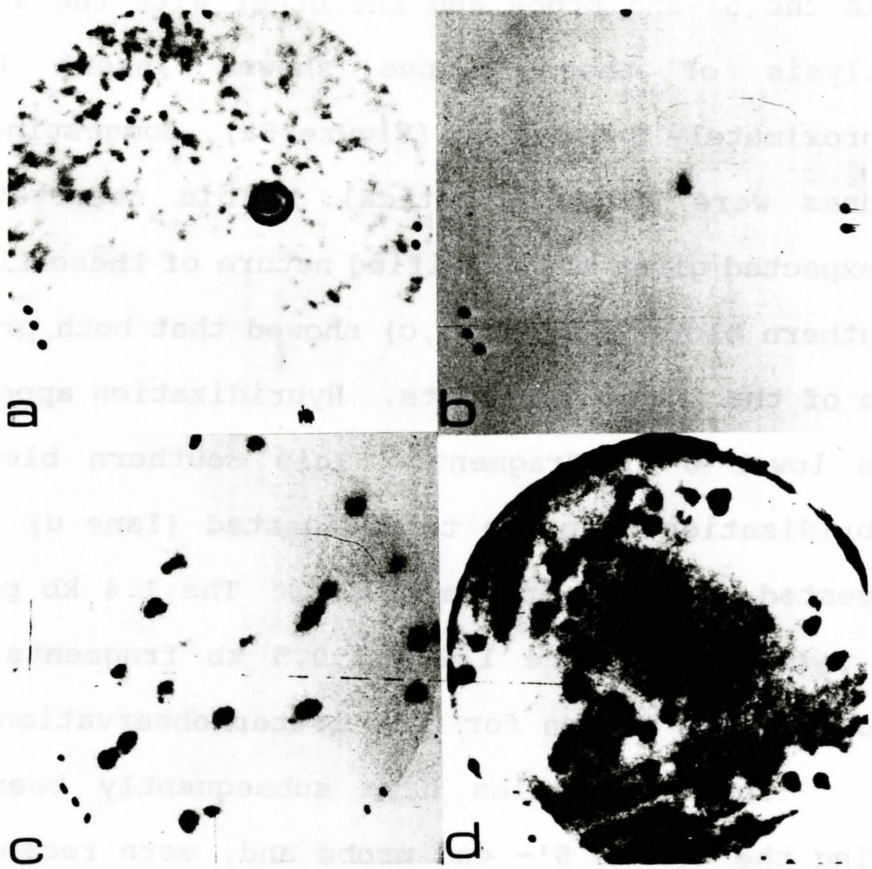


Figure 4 **Detection of a positive clone.** Clone D3 was identified as a hybridizing signal found on duplicate filters (a,b). Secondary screening of clone D3 (c). Positive control filter (d).

scale phage DNA preparations were performed to examine the inserts. The DNA was digested with Eco RI and electrophoresed on a 0.8% agarose gel. The gels were transferred to nitrocellulose in duplicate such that one blot could be probed with the 5'-end probe and the other with the 3'- end probe. Analysis of these clones showed insert fragments of approximately 8 and 9 kb (Figure 5a), suggesting these three clones were likely identical. This observation was not unexpected given the amplified nature of these libraries. The Southern blots (Figure 5b,c) showed that both probes bound to one of the insert fragments. Hybridization appeared to be to the lower 8 kb fragment. This Southern blot also shows hybridization of probe to undigested (lane u) and partially digested (lane 1) λphage C2 DNA. The 1.4 kb probe appeared to hybridize to the 1.6 and 0.5 kb fragments of the 1 kb ladder. The reason for this latter observation is unknown.

These libraries have subsequently been re-screened using the 0.8 kb 5'- end probe and, more recently, a 0.5 kb 3'- end probe which consists largely of untranslated human P1 cDNA sequence. Screening with the former probe resulted in the isolation of two new clones, C-II and D-II. Employing the latter probe led to the isolation of 7 additional clones. These clones were analyzed by direct sequencing of phage DNA with internal primers (see 3.1d).

The EMBL-3 phage library was screened as above, however, some clones which hybridized to the 0.8 kb 5'- end probe only were also kept for further analysis. A total of

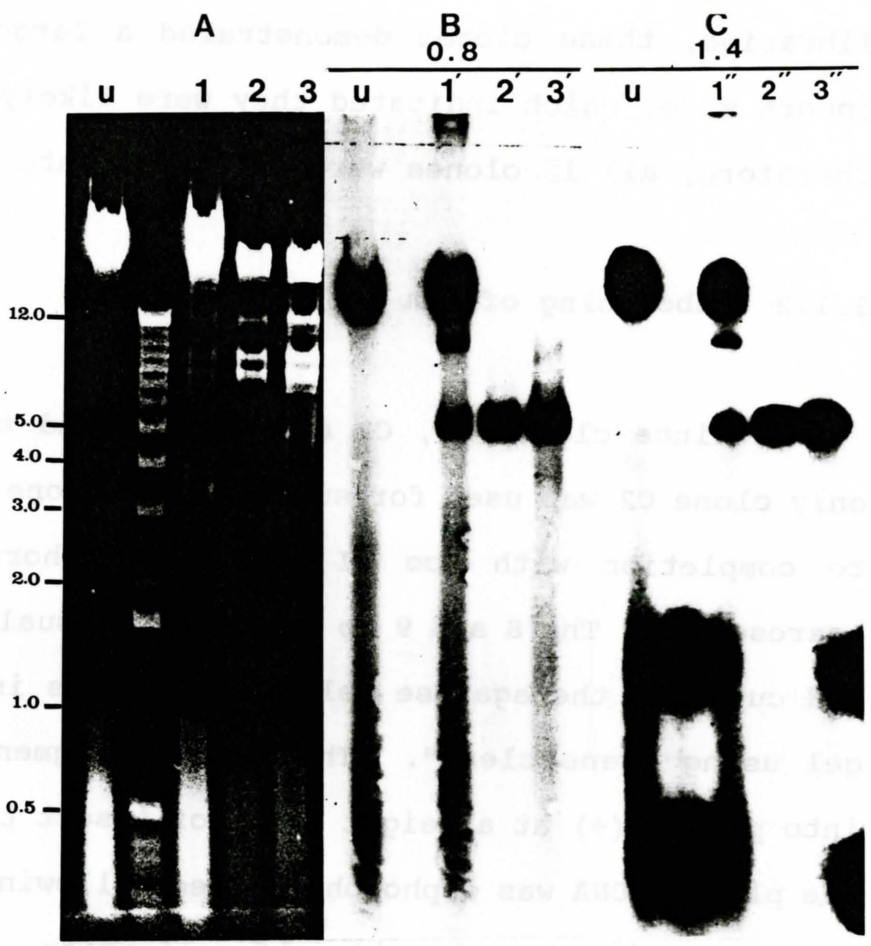


Figure 5 **Southern blot analysis of inserts from Charon 4A positive clones.** DNA from clones C2 (lanes 1, 1', 1"), C3 (lanes 2, 2', 2", D3 (lanes 3, 3', 3") was digested with Eco RI and electrophoresed in 0.8% agarose gels. Lane u contains undigested phage C2 DNA. Ethidium bromide stained gel is shown in (A). Gel was blotted in duplicate and probed with either the ^{32}P-labelled 0.8 kbp (B) or 1.4 kbp (C) human P1 fragments. Molecular weight size markers are given in kbp.

12 clones were isolated and plaque purified. The phage DNA was isolated and the inserts were analyzed by digestion with Bam HI followed by agarose gel electrophoresis (data not shown). Unlike the phage clones isolated from the Charon 4A libraries, these clones demonstrated a larger variation in insert size, which indicated they were likely not identical. Therefore, all 12 clones were subcloned into plasmid vector.

3.1.2 Subcloning of DNA into pGem7zf(+)

Since clones C2, C3 and D3 appeared to be identical, only clone C2 was used for subcloning. Clone C2 was digested to completion with Eco RI and electrophoresed on a 0.8% agarose gel. The 8 and 9 kb bands were visualized on a UV box and cut from the agarose gel. The DNA was isolated from the gel using "gene clean". The Eco RI fragments were ligated into pGem7zf(+) at a weight ratio of insert to vector of 5:1. The plasmid DNA was dephosphorylated following digestion with Eco RI to minimize plasmid self-religation.

Following ligation and transformation into JM107, the mixture was plated onto agar plates containing ampicillin, X-gal and IPTG (blue/white screening). Bacteria containing recombinant plasmids appeared as white colonies and represented approximately 70-80% of all colonies. Sixteen white colonies were initially picked for small scale plasmid preparations. The plasmids were then digested with Eco RI and run on 1.0% agarose gels. None of the plasmids appeared to

be digested by this enzyme (results not shown). Digestions were repeated with Xba I and Sst I. Sites for these enzymes are found on either side of the Eco RI site in the multiple cloning region of pGem7zf(+) (see Figure 3d). Some of the recombinant clones were found to release inserts with this combination of enzymes (Figure 6a). These subclones were analyzed by Southern analysis with the 5'- end probe (Figure 6b). A number of these clones, including pGem10, were kept for further analysis.

The digestion of pGem10 shown in Figure 6 is incomplete. Upon complete digestion with Xba I/Sst I these clones revealed an insert of approximately 2.2 kb corresponding to the lower band in the partial digestion pattern (results not shown). An aditional 32 white colonies were picked and their plasmid DNA was digested as described previously. Those plasmids which hybridized to the 0.8 kb probe were kept for upstream sequencing (data not shown).

Subcloning of the EMBL-3 clones was done by two different methods. Initially, clones 4, 7 and 10 were digested with Bam HI and Eco RI, and 'shotgun' ligated into pGem7zf(+). This method was thought to offer two advantages: (1) forced cloning into the Bam HI/Eco RI cut vector would reduce plasmid self-religation, thereby eliminating the necessity of the dephosphorylation step. (2) The functional P1 gene should contain an internal Eco RI site, therefore, Bam HI/Eco RI digestion should reduce the insert size, thereby decreasing plasmid instability and structural rearrangement.

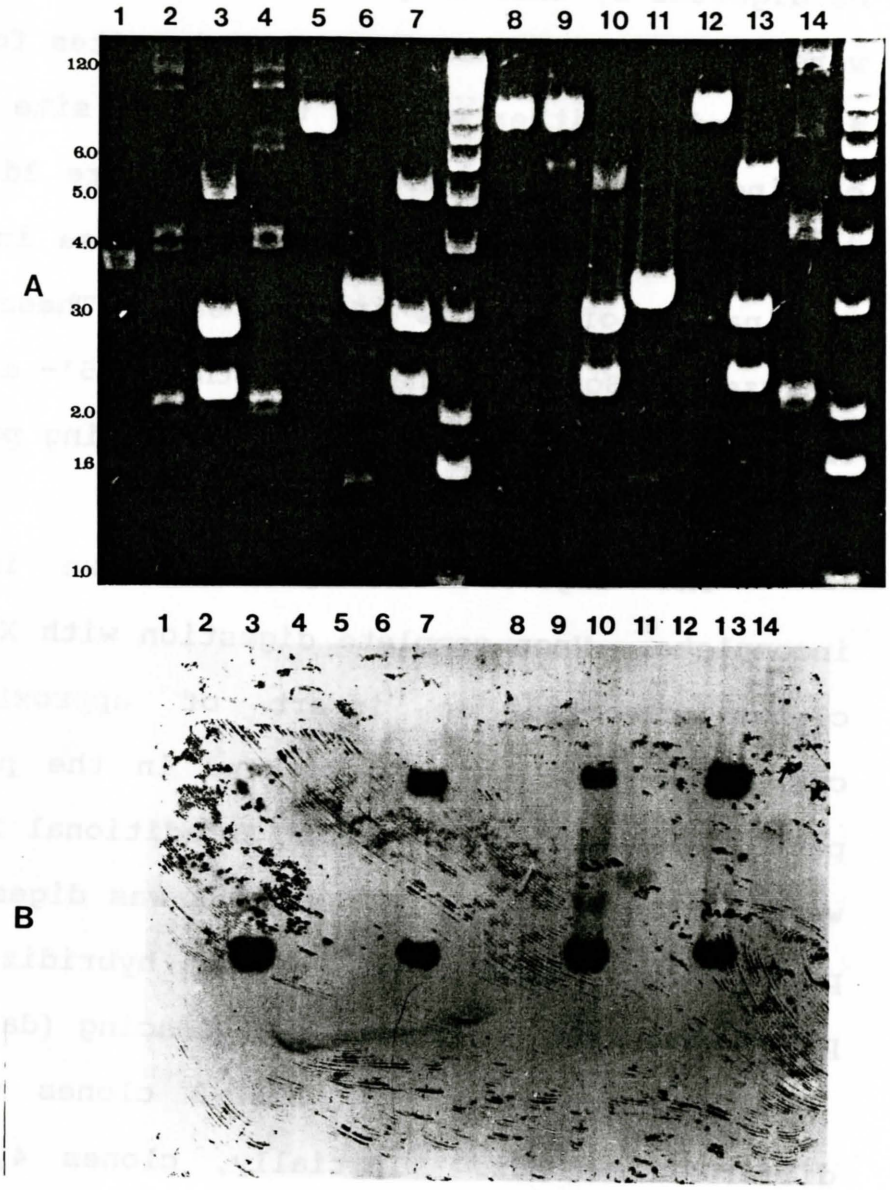


Figure 6 **Digestion and Southern analysis of some recombinant clones obtained by 'shotgun' ligation of Eco RI fragments from human genomic P1 clone λ C2.** (A) Agarose gel electrophoresis of 14 recombinant clones digested with Xba I/Sst I, stained with ethidium bromide. (B) Southern analysis of the same gel, after transfer to nitrocellulose, with ^{32}P-labelled 0.8 kbp human P1 fragment. Molecular weight size markers are given in kbp.

The ligation mixtures were used to transform JM109 cells. Transformants were grown on ampicillin containing plates for blue/white screening as before and colony hybridization was performed to identify the bacterial clones bearing either 0.8 kb or 1.4 kb-like sequences. Colony hybridization was performed in the same manner as phage hybridization. Positive colonies were grown in ampicillin medium and the plasmid DNA was isolated. The DNA was then digested with Bam HI and Eco RI, run on agarose gels, blotted onto nitrocellulose and subjected to Southern analysis prior to sequencing (data not shown).

The other positive EMBL-3 clones were simply digested with Bam HI and 'shotgun' ligated into Bam HI digested, dephosphorylated pGem7zf(+) vector. Transformants were identified as before and plasmid DNA was isolated from a number of white colonies. Those recombinant plasmids containing P1 related sequences were identified by Southern analysis (for example, see Figure 7). Restriction digests and Southern analysis of P1 related sequences subcloned from EMBL-3 phage DNA are summarized in Figure 8.

3.1.3 **Sequencing of pGem7zf(+) subclones**

The subclone pGem10 was sequenced in its entirety using forward and reverse universal primers and the internal oligonucleotide primers designed for sequencing the human P1 cDNA (see Figure 9). This subclone was found to contain the

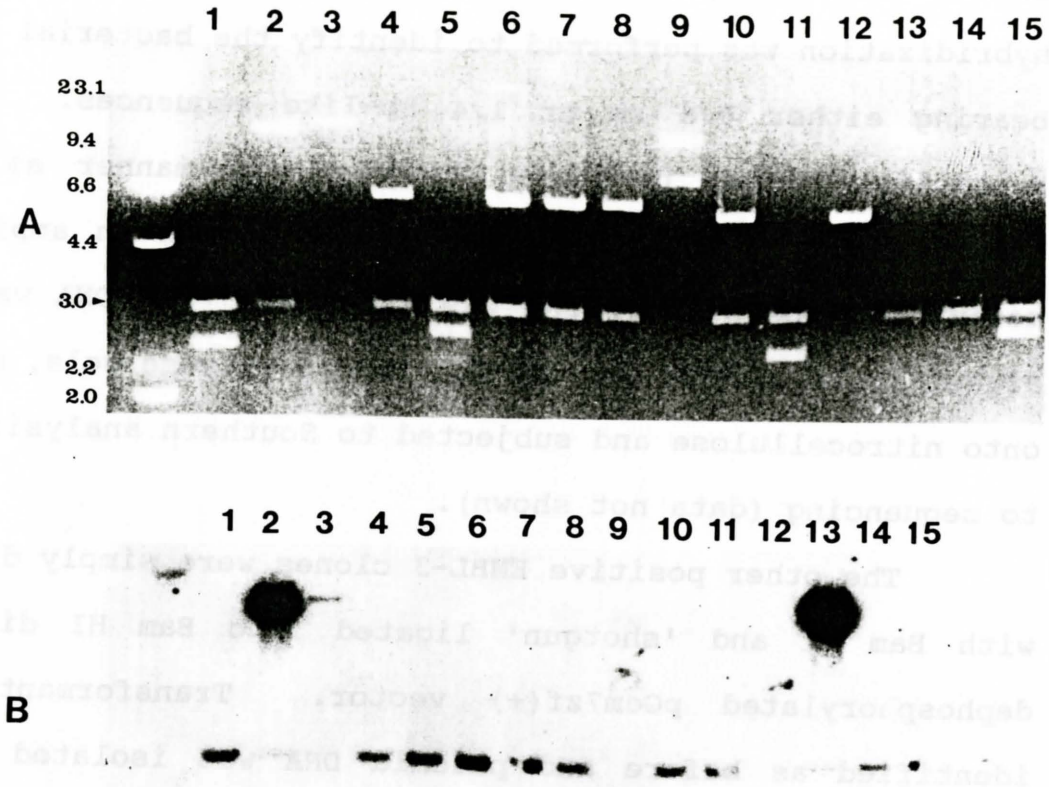


Figure 7 **Identification of recombinant plasmids containing P1 related sequences.** (A) Agarose gel of Bam HI digested plasmid DNA from colonies picked after 'shotgun' cloning of λ PS1-2 DNA, stained with ethidium bromide. (B) Southern analysis of same gel after transfer onto nitrocellulose with ^{32}P-labelled 0.8 kbp human P1 probe. Molecular weight size markers are given in kbp. Plasmid DNA is 3.0 kbp, marked with arrowhead.

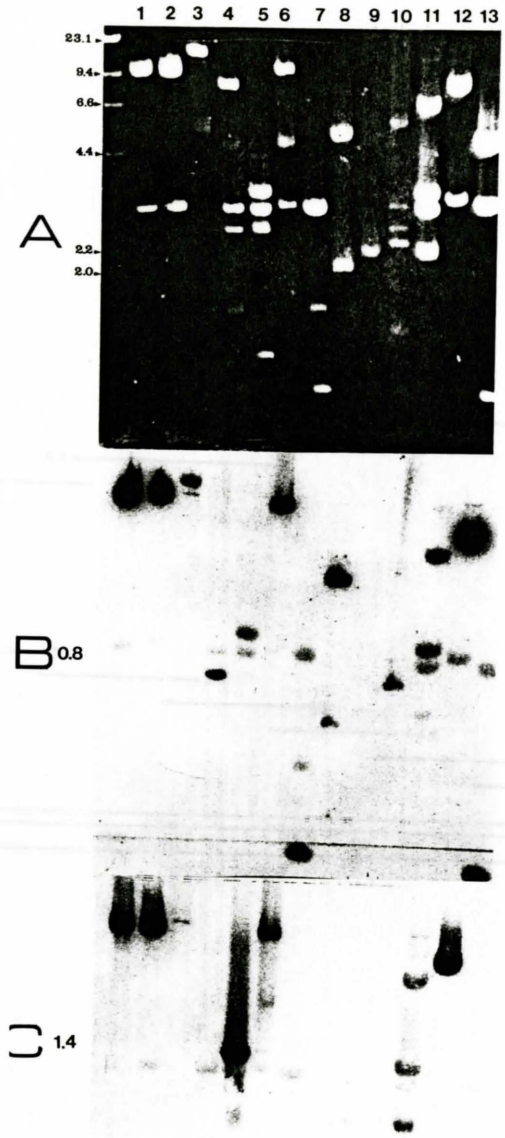


Figure 8 **Southern analysis of EMBL-3 subclones.** Plasmids were digested with Bam HI (lanes 1, 2, 3, 6, 7, 12, 13) or Bam HI/Eco RI (lanes 4, 5, 8-11) and run on an agarose gel. (A) The gel was transferred to nitrocellulose in duplicate and probed with either ^{32}P-labelled 0.8 kbp, (B) 1.4 kbp or (C) human P1 fragments. Lanes 1 to 13; subclones containing ψ's 1-2, 2-1, 3-2, 4-1a, 4-2a, 5-4, 6-1, 7-1a, 7-2b, 10-1a, 11-3b, 11-1a and 12-1, respectively.

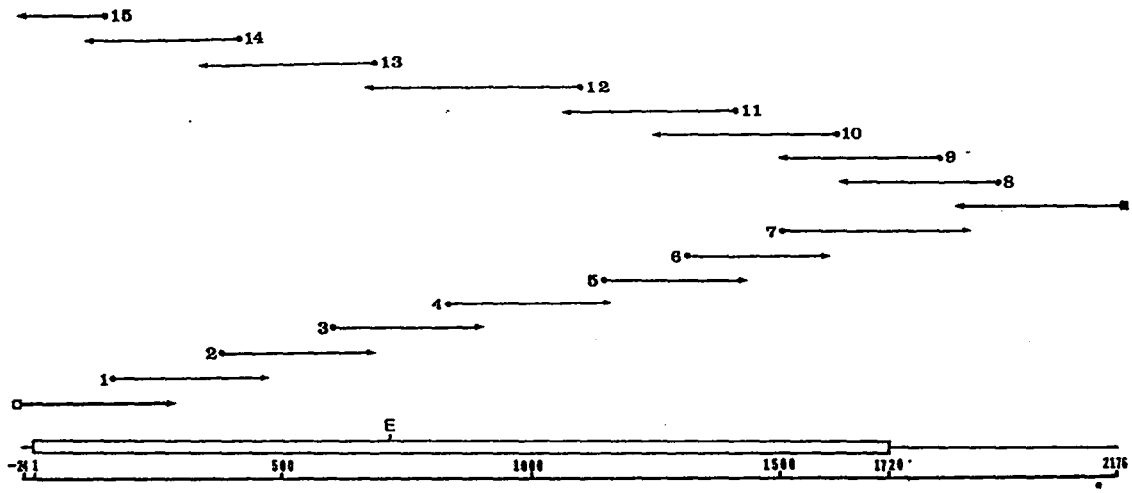


Figure 9 **Strategy employed for sequencing genomic P1 DNA pGem10.** The start points are indicated by circles, for internal primers, and squares for forward universal (open) and reverse universal (closed) primers. Primers 1; AB729; 2, AB730; 3, AB731; 4, AB728; 5, AB725; 6, AB726; 7; AB727; 8, AB538; 9, AB357; 10, AB537; 11, AB675; 12, AB277; 13, AB308; 14, AB674; 15, AB634. Arrows indicate direction of sequencing. E; internal Eco RI restriction site. Protein coding region is represented by large open rectangle. Map contains numbering for reference and is drawn to scale.

sequence corresponding to the human P1 cDNA from a position 24 bp upstream of the translation start site to a position 2176 bp downstream of this site. Since one of the goals of this project was to obtain sequence information regarding upstream regions of the P1 gene, 12 other subclones which were positive for 0.8 kb probe binding were sequenced with the oligonucleotide AB634. This primer binds to a position between bases 165 and 182 of the human P1 cDNA translated region, and allows the sequence to be read in the reverse (upstream) direction. All of the subclones analyzed were found to terminate at the -24 bp location (data not shown).

The subclones obtained from the EMBL-3 library were sequenced using the internal primers and the forward universal primer (in the case of Bam HI/Eco RI subclones). It was observed that many of the internal primers were unable to bind to these clones. This was explained when partial sequencing revealed numerous differences (insertions, deletions, replacements) between the human P1 cDNA sequence and these sequences. These observations also indicated that these P1 related sequences were probably non-functional homologues of P1 or pseudogenes. The partial sequences indicated that six distinct classes of pseudogenes (ps) had been isolated. Four of the pseudogenes were sequenced to completion. The sequencing strategy used for PS4-1 is shown in Figure 10.

3.1.4 Analysis and comparison of human P1 related genomic DNA sequences

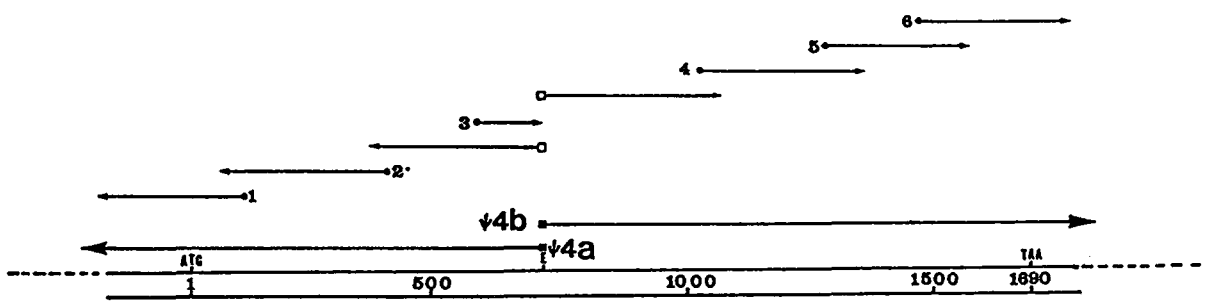


Figure 10 **Strategy employed for sequencing the P1 related genomic sequence PS4-1.** This sequence was contained in two Eco RI (E) subclones 4a and b. Start points are indicated by closed circles; internal primers, and open squares; forward universal primer. Primers 1, AB634; 2, AB674; 3, AB731; 4, AB725; 5, AB726; 6, AB727. Translation start (ATG) and stop (TAA) site relative to functional P1 sequence are indicated. Map contains numbering for reference and is drawn to scale.

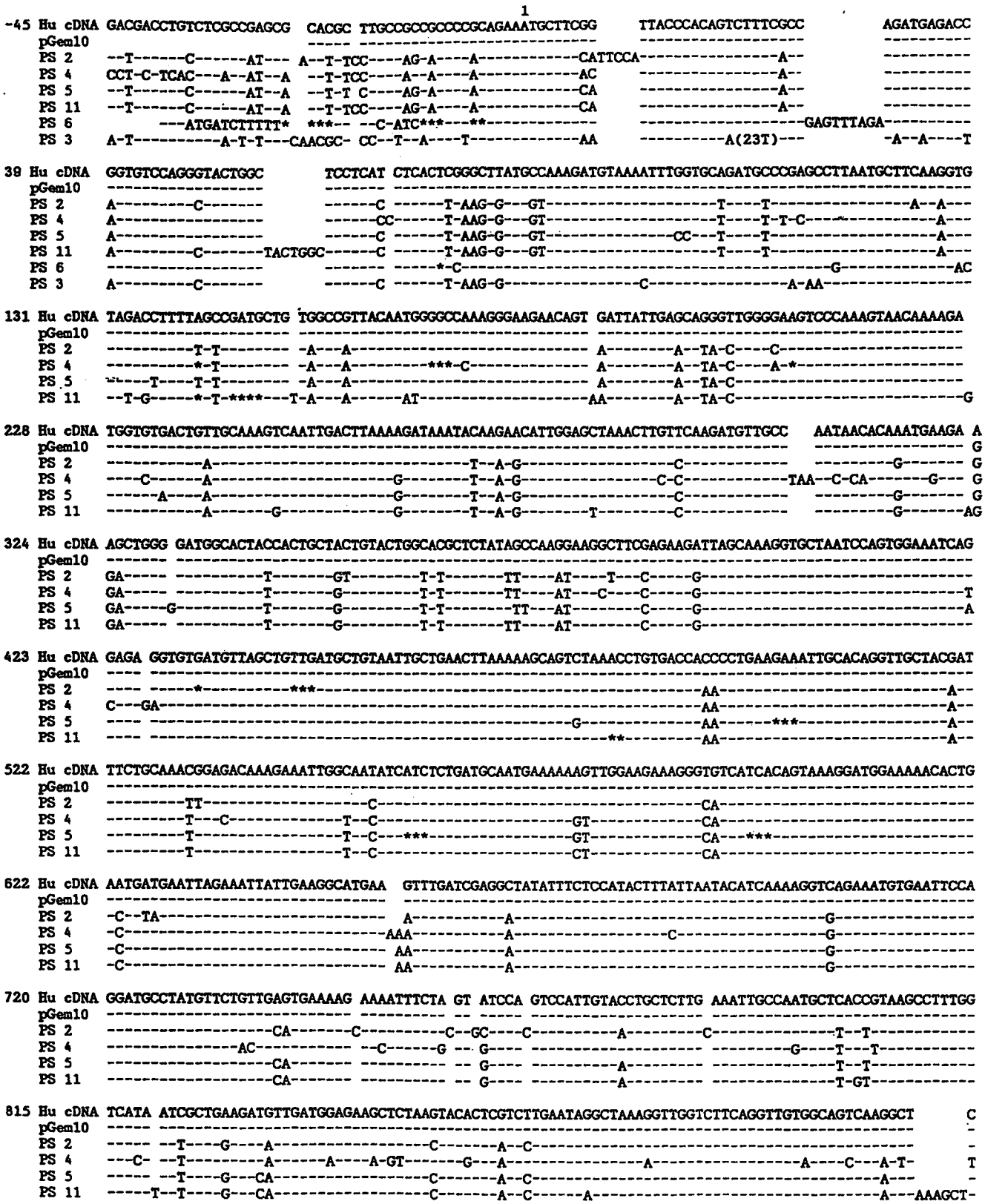
The sequence of the P1 subclone pGem10 matches exactly with that of the human P1 cDNA (see Figure 11) in both the coding and non-coding region, indicating that it likely constitutes the functional gene. The putative human P1 gene lacks intervening sequences based on the sequence information obtained from pGem10.

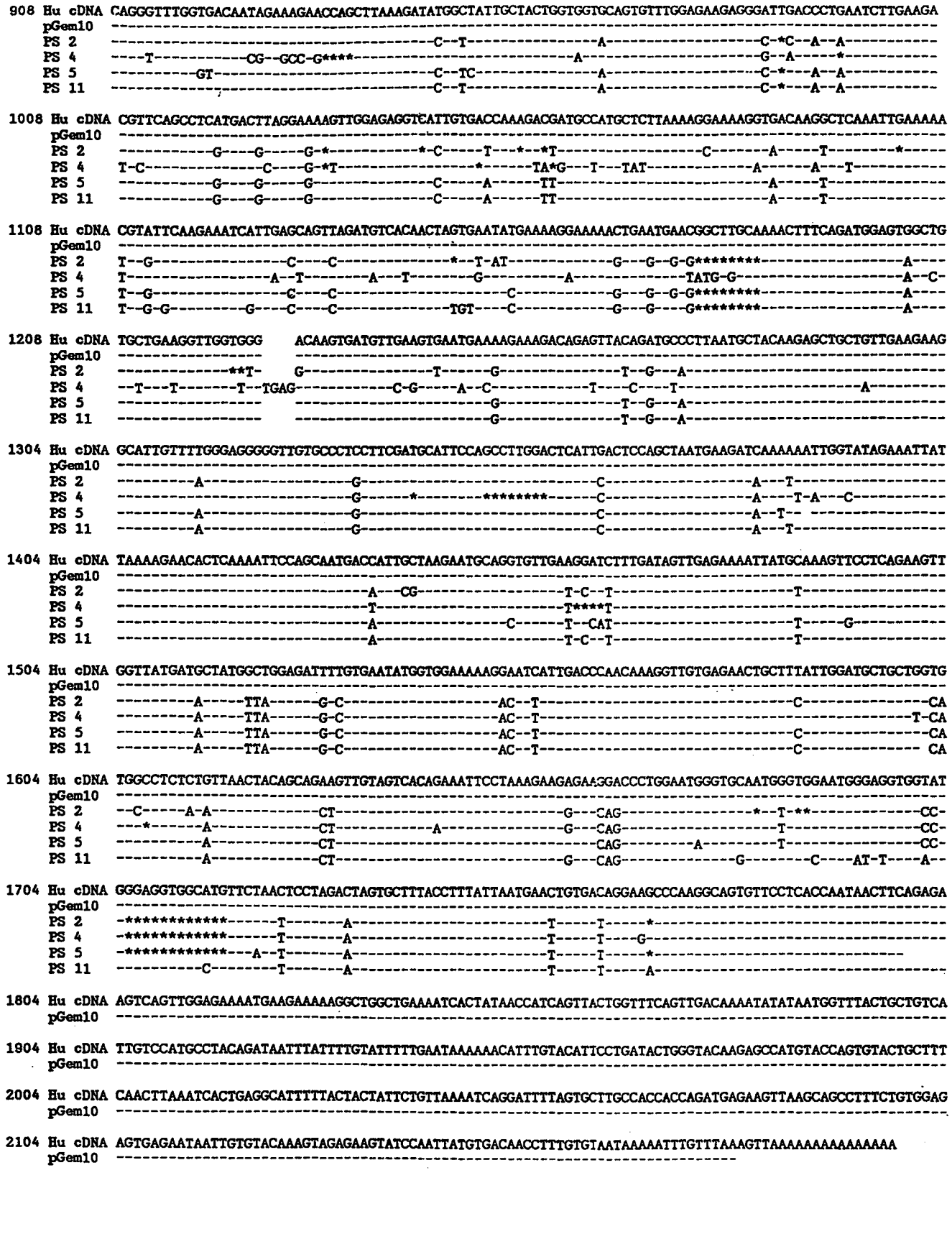
The complete sequence of four of the pseudogenes (PS2, 4, 5 and 11) and partial sequence of two other clones (PS3 and 6) are compared to the human P1 sequences pGem10 and Hu cDNA (Figure 11). The different pseudogenes contained numerous additions, deletions and base substitutions, some of which were unique for a given pseudogene, whereas others were shared among some of the other members. These changes were particularly prevalent across the 5'- untranslated region. The DNA sequences of the four pseudogenes which have been completely sequenced were highly homologous to the cDNA sequence with sequence homologies in the translated region of 86.1, 87.4, 89.7 and 90.2%, respectively. The pseudogenes, like the functional gene, were found to be devoid of introns.

3.1.5 Direct sequencing of P1 positive phage DNA

Re-screening of the λ Charon 4A library ATCC 37333 with the 0.8 kb 5'- end probe resulted in the isolation of two new clones, C-II and D-II. Direct sequencing of these clones with primer AB634 indicated that they contained P1 pseudogene

Figure 11 **Comparison of the nucleotide sequence of human hsp60 cDNA with that of various human genomic clones.** The nucleotides have been numbered assuming the first nucleotide of the initiation codon to be 1. Identical residues are denoted by a dash (-). The gaps in the cDNA sequence and other sequences correspond to additions in one or more of the clones. The deletions in clones are indicated by asterisks (*). The sequence of clone pGem-10 matches exactly with the cDNA sequence. For clones PS3 amd PS6, only partial sequence information is available.


1
-45 Hu cDNA GACGACCTGTCTCGCCGAGCG CACGC TTGCCGCCGCCCCGCAGAAATGCTTCGG TTACCCACAGTCTTTCGCC AGATGAGACC
pGem10
PS 2
PS 4
PS 5
PS 11
PS 6
PS 3
39 Hu cDNA GGTGTCCAGGGTACTGGC TCCTCAT CTCACTCGGGCTTATGCCAAAGATGTAAAATTTGGTGCAGATGCCCGAGCCTTAATGCTTCAAGGTG
pGem10
PS 2
PS 4
PS 5
PS 11
PS 6
PS 3
131 Hu cDNA TAGACCTTTTAGCCGATGCTG TGGCCGTTACAATGGGGCCAAAGGGAAGAACAGT GATTATTGAGCAGGGTTGGGGAAGTCCCAAAGTAACAAAAGA
pGem10
PS 2
PS 4
PS 5
PS 11
228 Hu cDNA TGGTGTGACTGTTGCAAAGTCAATTGACTTAAAAGATAAATACAAGAACATTGGAGCTAAACTTGTTCAAGATGTTGCC AATAACACAAATGAAGA A
pGem10
PS 2
PS 4
PS 5
PS 11
324 Hu cDNA AGCTGGG GATGGCACTACCACTGCTACTGTACTGGCACGCTCTATAGCCAAGGAAGGCTTCGAGAAGATTAGCAAAGGTGCTAATCCAGTGGAAATCAG
pGem10
PS 2
PS 4
PS 5
PS 11
423 Hu cDNA GAGA GGTGTGATGTTAGCTGTTGATGCTGTAATTGCTGAACTTAAAAAGCAGTCTAAACCTGTGACCACCCCTGAAGAAATTGCACAGGTTGCTACGAT
pGem10
PS 2
PS 4
PS 5
PS 11
522 Hu cDNA TTCTGCAAACGGAGACAAAGAAATTGGCAATATCATCTCTGATGCAATGAAAAAAGTTGGAAGAAAGGGTGTCATCACAGTAAAGGATGGAAAAACACTG
pGem10
PS 2
PS 4
PS 5
PS 11
622 Hu cDNA AATGATGAATTAGAAATTATTGAAGGCATGAA GTTTGATCGAGGCTATATTTCTCCATACTTTATTAATACATCAAAAGGTCAGAAATGTGAATTCCA
pGem10
PS 2
PS 4
PS 5
PS 11
720 Hu cDNA GGATGCCTATGTTCTGTTGAGTGAAAAG AAAATTTCTA GT ATCCA GTCCATTGTACCTGCTCTTG AAATTGCCAATGCTCACCGTAAGCCTTTGG
pGem10
PS 2
PS 4
PS 5
PS 11
815 Hu cDNA TCATA ATCGCTGAAGATGTTGATGGAGAAGCTCTAAGTACACTCGTCTTGAATAGGCTAAAGGTTGGTCTTCAGGTTGTGGCAGTCAAGGCT C
pGem10
PS 2
PS 4
PS 5
PS 11


908 Hu cDNA CAGGGTTTGGTGACAATAGAAAGAACCAGCTTAAAGATATGGCTATTGCTACTGGTGGTGCAGTGTTTGGAGAAGAGGGATTGACCCTGAATCTTGAAGA
pGem10
PS 2
PS 4
PS 5
PS 11
1008 Hu cDNA CGTTCAGCCTCATGACTTAGGAAAAGTTGGAGAGGTCATTGTGACCAAAGACGATGCCATGCTCTTAAAAGGAAAAGGTGACAAGGCTCAAATTGAAAAA
pGem10
PS 2
PS 4
PS 5
PS 11
1108 Hu cDNA CGTATTCAAGAAATCATTGAGCAGTTAGATGTCACAACTAGTGAATATGAAAAGGAAAAACTGAATGAACGGCTTGCAAAACTTTCAGATGGAGTGGCTG
pGem10
PS 2
PS 4
PS 5
PS 11
1208 Hu cDNA TGCTGAAGGTTGGTGGG ACAAGTGATGTTGAAGTGAATGAAAAGAAAGACAGAGTTACAGATGCCCTTAATGCTACAAGAGCTGCTGTTGAAGAAG
pGem10
PS 2
PS 4
PS 5
PS 11
1304 Hu cDNA GCATTGTTTTGGGAGGGGGTTGTGCCCTCCTTCGATGCATTCCAGCCTTGGACTCATTGACTCCAGCTAATGAAGATCAAAAAATTGGTATAGAAATTAT
pGem10
PS 2
PS 4
PS 5
PS 11
1404 Hu cDNA TAAAAGAACACTCAAAATTCCAGCAATGACCATTGCTAAGAATGCAGGTGTTGAAGGATCTTTGATAGTTGAGAAAATTATGCAAAGTTCCTCAGAAGTT
pGem10
PS 2
PS 4
PS 5
PS 11
1504 Hu cDNA GGTTATGATGCTATGGCTGGAGATTTTGTGAATATGGTGGAAAAAGGAATCATTGACCCAACAAAGGTTGTGAGAACTGCTTTATTGGATGCTGCTGGTG
pGem10
PS 2
PS 4
PS 5
PS 11
1604 Hu cDNA TGGCCTCTCTGTTAACTACAGCAGAAGTTGTAGTCACAGAAATTCCTAAAGAAGAGAAGGACCCTGGAATGGGTGCAATGGGTGGAATGGGAGGTGGTAT
pGem10
PS 2
PS 4
PS 5
PS 11
1704 Hu cDNA GGGAGGTGGCATGTTCTAACTCCTAGACTAGTGCTTTACCTTTATTAATGAACTGTGACAGGAAGCCCAAGGCAGTGTTCCTCACCAATAACTTCAGAGA
pGem10
PS 2
PS 4
PS 5
PS 11
1804 Hu cDNA AGTCAGTTGGAGAAAATGAAGAAAAAGGCTGGCTGAAAATCACTATAACCATCAGTTACTGGTTTCAGTTGACAAAATATATAATGGTTTACTGCTGTCA
pGem10
1904 Hu cDNA TTGTCCATGCCTACAGATAATTTATTTTGTATTTTTGAATAAAAAACATTTGTACATTCCTGATACTGGGTACAAGAGCCATGTACCAGTGTACTGCTTT
pGem10
2004 Hu cDNA CAACTTAAATCACTGAGGCATTTTTACTACTATTCTGTTAAAATCAGGATTTTAGTGCTTGCCACCACCAGATGAGAAGTTAAGCAGCCTTTCTGTGGAG
pGem10
2104 Hu cDNA AGTGAGAATAATTGTGTACAAAGTAGAGAAGTATCCAATTATGTGACAACCTTTGTGTAATAAAAATTTGTTTAAAGTTAAAAAAAAAAAAAAA
pGem10

sequences (see Figure 12). This library was also re-screened with a 0.5 kb 3'- end probe, composed largely of untranslated human P1 sequence, and 7 additional clones were isolated. However, no direct sequence information was obtained for any of these clones using the 5'- end internal primer AB634 or the 3'- end primers AB725, 533 or 538.

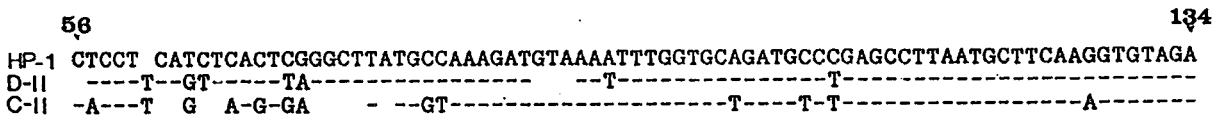


Figure 12 **Comparison of partial nucleotide sequence of human P1 cDNA with the P1 related genomic clones λ DII and λ CII.** A portion of the sequence read from clones λ CII and DII with the internal primer AB634 is compared to the human P1 cDNA sequence (HP-1) across the region 56 to 134 bases downstream of the translation initiation site. Dashes (-) indicate identical residues.

3.2 Analysis of P1 Clones from Rat and Mouse cDNA Libraries

3.2.1 Screening of rat and mouse λgt11 cDNA libraries

Initially, the rat and mouse libraries were plated (approximately 5 x 10^3 pfu/plate; 10 plates each) duplicate filters lifted and hybridized to 0.8 and 1.4 kb probes as described earlier. Two positive rat clones (λRK -6 and 7) and two positive mouse clones (λ3T3 -7 and 9) were kept for further analysis. Subsequently, these libraries were re-screened using a 1.1 kb probe which was obtained by Eco RI digestion of a plasmid vector containing the 3'- terminal portion of the CHO P1 cDNA. One positive rat clone, λRK1, and one positive mouse clone, λ3T3-M1, were kept for further analysis.

The positive phage clones were plaque purified and the phage DNA was isolated as described in Materials and Methods. The DNA was digested with Eco RI and analyzed by agarose gel electrophoresis (data not shown). Insert fragments generated by Eco RI digestion were of the following approximate sizes; λRK6, 0.6 and 0.2 kb; λRK7, 0.5 and 0.3 kb; λRK-1, 1.2 kb; λ3T3-7, 0.7 and 0.6 kb; λ3T3-9, 0.5 and 0.3 kb; and λ3T3-M1, 1.4 kb.

3.2.2 Subcloning of P1 cDNA inserts into pGem7zf(+)

The λgt11 vector contains two sites, Kpn I and Sst I,

that can be used for cloning the entire cDNA insert into a plasmid vector when Eco RI sites are found within the insert. Clones λ3T3-7 and λRK-6 were digested with Kpn I and Sst I and a small aliquot of this DNA was analyzed on an agarose gel to ensure that there were no internal sites for these enzymes. These sites were lacking in both clones and the remaining digested DNA was 'shotgun' ligated into plasmid pGem7zf(+) as described earlier. This method results in the insertion of approximately 1.0 kb of phage DNA on either side of the insert. The remaining phage clones were each digested with Eco RI and 'shotgun' ligated into the same vector. Transformants were identified and treated as previously described and subclones containing the appropriate insert were identified on agarose gels (summarized in Figure 13a). Southern analysis was used to identify the relative location (i.e. 3' or 5') of the subcloned DNA (see Figure 13b and c).

3.2.3 **Sequencing of pGem7zf(+) subclones**

The subclones described here were sequenced using a combination of internal oligonucleotide and universal primers (both forward and reverse). The sequencing strategies used are shown in Figure 14 (mouse) and Figure 15 (rat). In order to increase the length of sequence read in the forward direction for both rat and mouse clones, the 3'- end fragment was digested with both Ava II and Eco RI. The resulting 0.5 kb terminal fragment was subcloned into pGem7zf(+) and read

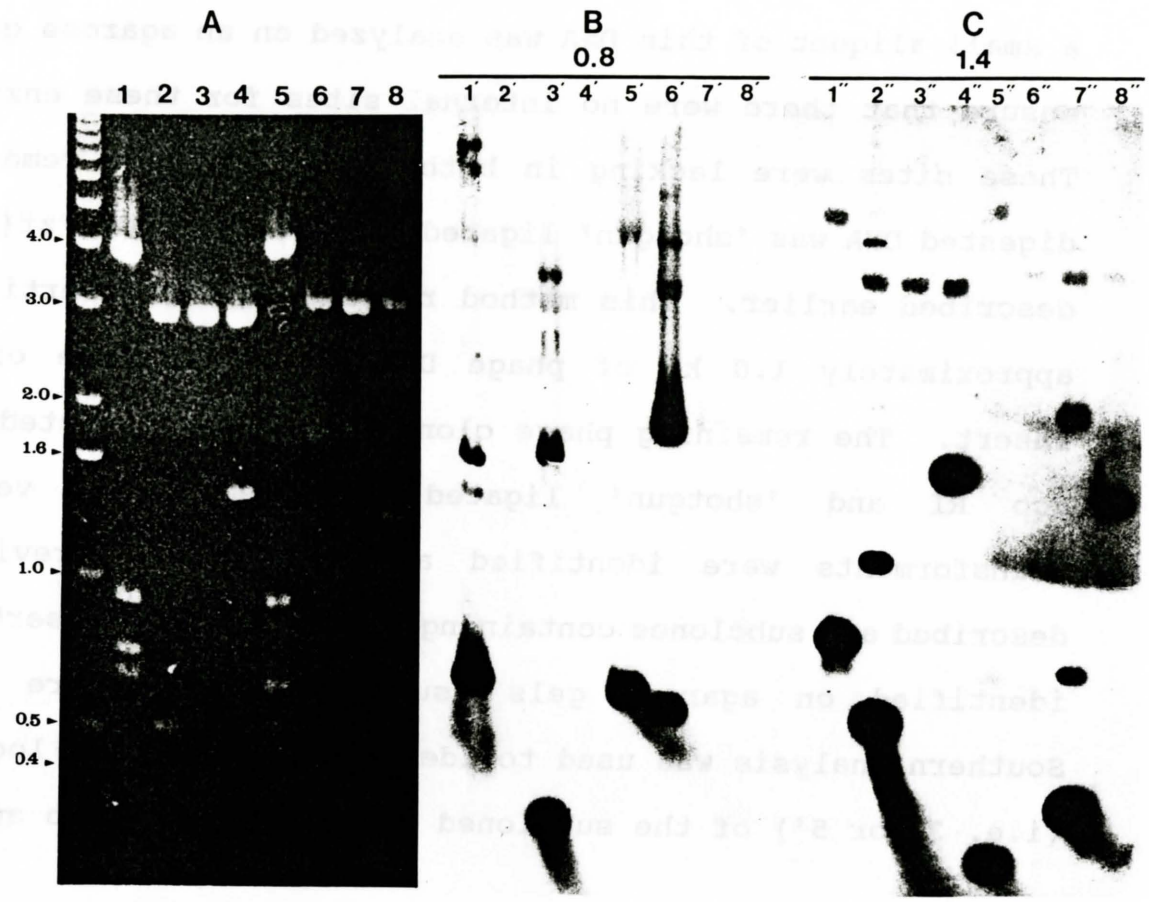


Figure 13 **Digestion patterns and Southern analysis of rat and mouse P1 cDNA subclones.** (A) Agarose gel of Eco RI digested subclones containing mouse (lanes 1-4) and rat (lanes 5-8) P1 cDNA sequences. The same gel was transferred in duplicate to nitrocellulose and probed with either the ^{32}P-labelled 0.8 kbp (B) or 1.4 kbp (C) human P1 fragments. Lane 1, (1' and 1"); λ 3T3-7; lane 2, 3T3-0 (0.5 kbp); lane 3, 3T3-9 (0.3 kbp); lane 4, λ 3T3-M1; lane 5, λ RKL-6; lane 6, λ RK-7 (0.5 kbp); lane 7, λ RK-7 (0.3 kbp); lane 8, λ RK-1. Molecular weight size markers are given in kbp.

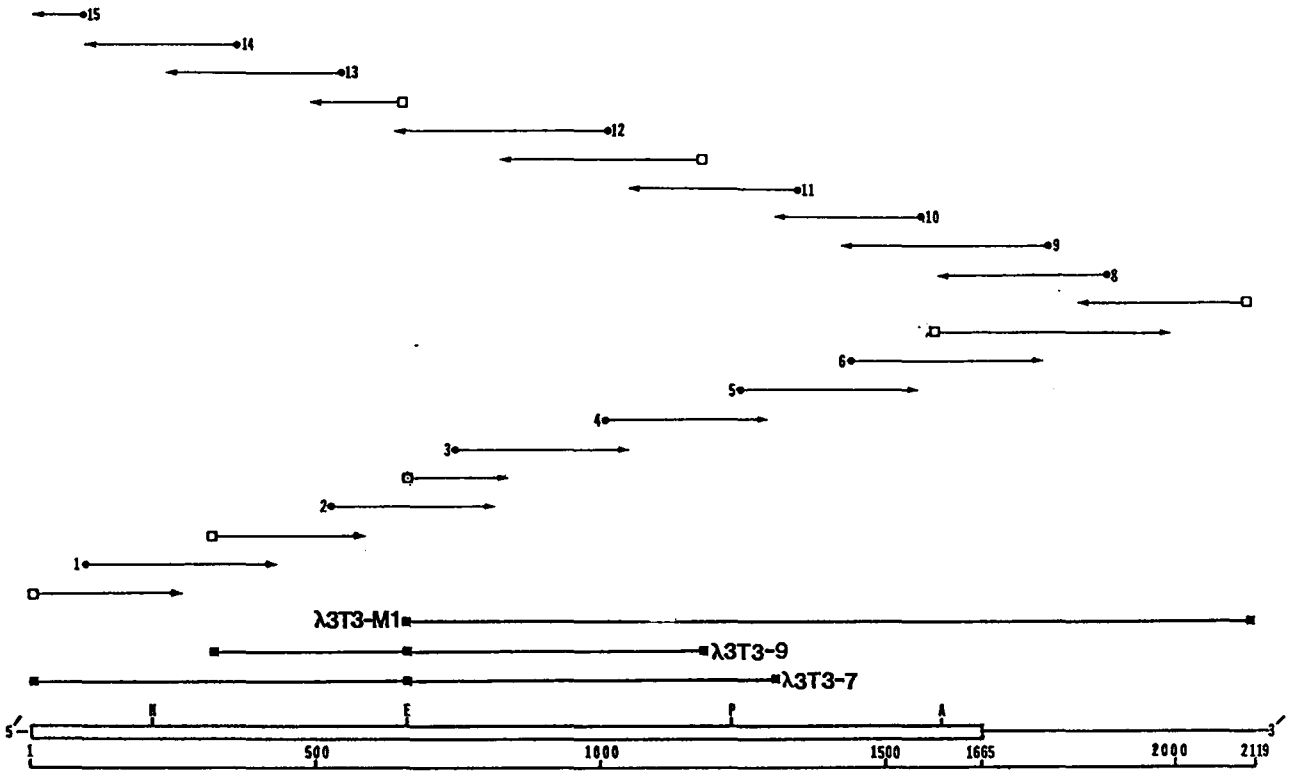


Figure 14 **Partial restriction map and strategy used for sequencing the P1 cDNA from mouse cells.** Clones λ 3T3-7, λ 3T3-9 and λ 3T3-M1 are from λ gt11 library prepared from mouse 3T3 cells. Restriction sites indicated are: H, Hind III; E, Eco RI; P, Pvu II; and A, Ava II. Closed squares represent location of Eco RI sites in subclones. Sequencing start points are indicated by closed circles (internal primers) or open squares (forward and reverse universal primers). Arrows indicate the direction of sequencing. Primers 1, AB729; 2, AB731; 3, AB728; 4, AB725; 5, AB726; 6, AB727; 8, AB538; 9, AB357; 10, AB537; 11, AB675; 12, AB277; 13, AB308; 14, AB674; 15, AB634; 16. Protein coding region is indicated by a large open rectangle. The map is drawn to scale and contains numbering for reference.

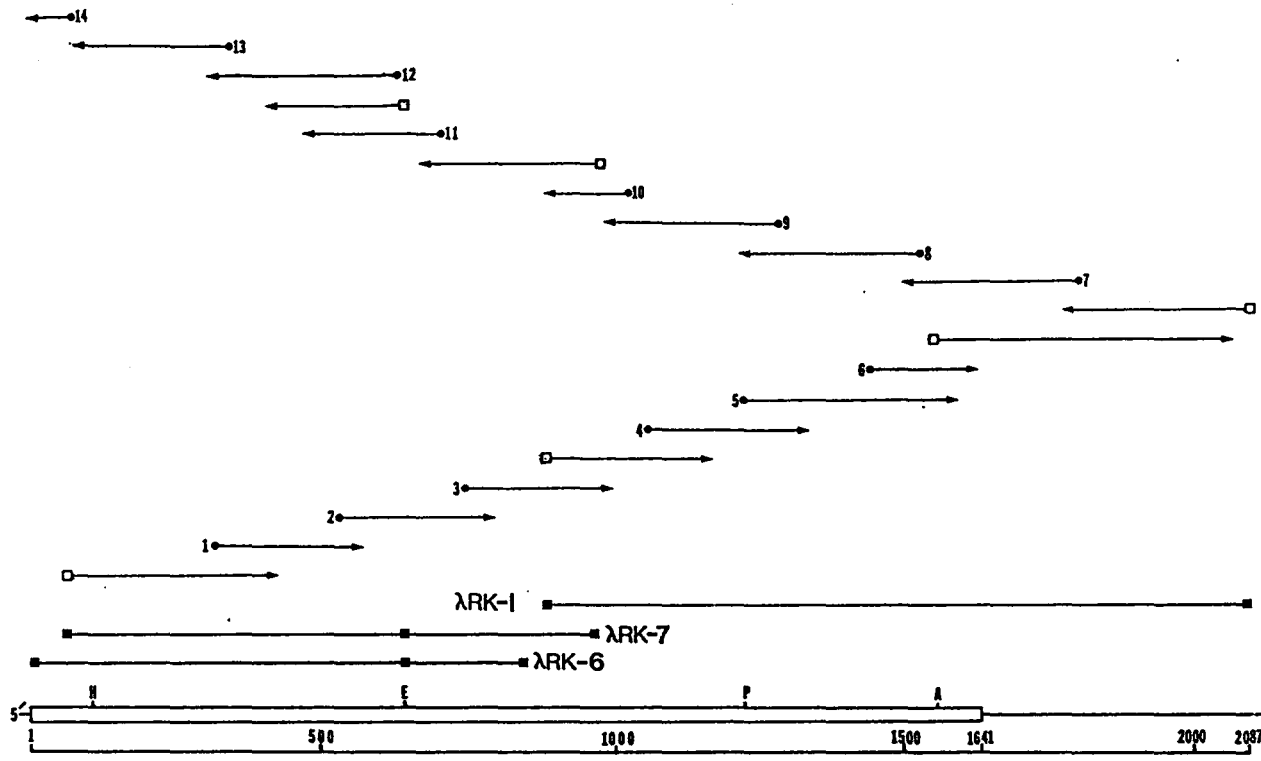


Figure 15 **Partial restriction map and strategy used for sequencing the P1 cDNA from rat cells.** Clones λ RK-6, λ RK-7 and λ RK-1 are from λ gt11 library prepared from rat kidney cells. Restriction sites indicated are: H, Hind III; E, Eco RI; P, Pvu II; and A, Ava II. Closed squares represent location of Eco RI sites in subclones. Sequencing start points are indicated by closed circles (internal primers) or open squares (forward and reverse universal primers). Arrows indicate the direction of sequencing. Primers 1, AB730; 2, AB731; 3, AB728; 4, AB725; 5, AB726; 6, AB727; 7, AB538; 8, AB537' 9, AB675; 10, AB277; 11, AB513; 12, AB308; 13, AB674; 14, AB634. Protein coding region is indicated by a large open rectangle. The map is drawn to scale and contains numbering for reference.

with the forward universal primer from site A (Figures 14 and 15) towards the Eco RI site at the 3'- terminus.

3.2.4 Nucleotide sequence of mouse and rat P1 cDNA

The complete DNA and amino acid sequences for mouse and rat P1 cDNA are shown in Figures 18 and 19. A rat P1 specific sequence of 2087 nucleotides was obtained by combining the sequences of clones λRK-1, -6 and -7. Similarly, combining the sequence information from clones λ3T3-M1, -7 and -9 revealed a mouse P1 specific sequence of 2119 nucleotides. The open reading frames of these sequences were identified by alignment with the P1 cDNA sequences obtained from CHO and human sources. The rat and mouse sequences consist of single open reading frames of 1641 and 1665 nucleotides, respectively. Both sequences have been truncated at the 5'-end, and as a result they lack the 5'-leader sequence and a small portion of the coding region (see Section 3.2e). Neither the rat, nor mouse, sequence possess poly(A) tails, however, the consensus polyadenylation signal AATAAA is found at nucleotides 2009-2014 (mouse) and 2069-2074 (rat) (Proudfoot and Brownlee, 1976).

The relative homology between rat, mouse, CHO and human cDNA sequences across the region coding for mature P1 (nts 1-1641 for rat) was calculated. The rat sequence was found to show 98% identity to both CHO and mouse sequences, and 95% identity with the human sequences. The mouse sequence shows

97% and 94% identity with CHO and human sequences, respectively. The changes in these sequences are mainly restricted to the third position of a codon, resulting in relatively few amino acid changes (see 3.2.5). Differences in nucleotide sequence in the 3'- untranslated regions were more frequent than in the coding region and included small insertions and deletions.

3.2.5 **Deduced amino acid sequence of rat and mouse P1**

The single open reading frames assigned to the rat and mouse sequences (based on alignment with the corresponding CHO and human cDNA sequences) encode proteins of 547 and 555 amino acids, respectively (see Figures 16 and 17). The full-length human and CHO P1 cDNA sequences both encoded proteins of 573 amino acids, the first 26 of which were determined to correspond to mitochondrial presequence (Jindal *et al.*, 1989; Picketts *et al.*, 1989). The truncated rat P1 protein sequence described here lacks the mitochondrial presequence, starting at the site corresponding to codon 27. Thus the 547 amino acid protein corresponds to the mature form of the rat P1 protein, with a calculated molecular weight of 57,916 daltons. The mouse sequence encodes a protein of 555 amino acids with a calculated molecular weight of 58,994 daltons. This sequence was truncated such that only the last eight amino acids of the mitochondrial presequence remain. Cleavage after codon 8 (Tyr) would result in a mature P1 protein of 547 amino

```
1    GCT CCT CAT CTC ACT CGG GCC TAT GCC AAA GAT GTA AAA TTT GGT GCG GAT GCT CGA GCC TTA ATG CTT CAA GGT GTA GAC CTT TTA GCC
      A  P  H  L  T  R  A  Y  A  K  D  V  K  F  G  A  D  A  R  A  L  M  L  Q  G  V  D  L  L  A                          30
91   GAT GCT GTA GCT GTT ACA ATG GGG CCA AAG GGA AGA ACA GTG ATT ATT GAA CAG AGT TGG GGA AGT CCC AAA GTA ACA AAA GAT GGG GTC
      D  A  V  A  V  T  M  G  P  K  G  R  T  V  I  I  E  Q  S  W  G  S  P  K  V  T  K  D  G  V                          60
181  ACT GTT GCA AAG TCA ATT GAT TTA AAG GAT AAA TAC AAA AAT ATC GGA GCT AAG CTT GTT CAG GAT GTT GCC AAT AAC ACA AAT GAA GAG
      T  V  A  K  S  I  D  L  K  D  K  Y  K  N  I  G  A  K  L  V  Q  D  V  A  N  N  T  N  E  E                          90
271  GCT GGG GAT GGC ACC ACC ACT GCC ACT GTT CTG GCA CGG TCT ATT GCC AAG GAG GGC TTT GAG AAG ATC AGC AAA GGG GCT AAT CCA GTG
      A  G  D  G  T  T  T  A  T  V  L  A  R  S  I  A  K  E  G  F  E  K  I  S  K  G  A  N  P  V                          120
361  GAA ATC CGG AGA GGT GTG ATG TTG GCT GTG GAT GCT GTA ATT GCT GAA CTT AAG AAA CAG TCT AAA CCT GTG ACA ACC CCT GAA GAA ATT
      E  I  R  R  G  V  M  L  A  V  D  A  V  I  A  E  L  K  K  Q  S  K  P  V  T  T  P  E  E  I                          150
451  GCT CAG GTT GCT ACA ATT TCT GCA AAC GGA GAC AAA GAC ATT GGG AAC ATC ATT TCT GAT GCA ATG AAG AAG GTT GGA AGA AAG GGT GTC
      A  Q  V  A  T  I  S  A  N  G  D  K  D  I  G  N  I  I  S  D  A  M  K  K  V  G  R  K  G  V                          180
541  ATC ACA GTG AAG GAT GGA AAA ACC CTG AAT GAT GAG CTA GAA ATT ATT GAA GGC ATG AAG TTT GAT AGA GGA TAT ATT TCC CCA TAT TTT
      I  T  V  K  D  G  K  T  L  N  D  E  L  E  I  I  E  G  M  K  F  D  R  G  Y  I  S  P  Y  F                          210
631  ATT AAC ACA TCA AAA GGT CAA AAA TGT GAA TTC CAA GAT GCC TAT GTT TTG TTG AGT GAA AAG AAA TTT TCC AGT GTT CAG TCC ATT GTC
      I  N  T  S  K  G  Q  K  C  E  F  Q  D  A  Y  V  L  L  S  E  K  K  F  S  S  V  Q  S  I  V                          240
721  CCT GCT CTT GAA ATT GCT AAT GCT CAT CGG AAG CCA TTG GTC ATA ATC GCC GAA GAT GTT GAC GGA GAA GCT CTA AGC ACG CTG GTT TTG
      P  A  L  E  I  A  N  A  H  R  K  P  L  V  I  I  A  E  D  V  D  G  E  A  L  S  T  L  V  L                          270
811  AAC AGG CTA AAA GTT GGT CTT CAG GTT GTA GCA GTC AAA GCT CCA GGG TTT GGG GAC AAC AGG AAG AAC CAG CTT AAA GAT ATG GCT ATC
      N  R  L  K  V  G  L  Q  V  V  A  V  K  A  P  G  F  G  D  N  R  K  N  Q  L  K  D  M  A  I                          300
901  GCT ACT GGT GGT GCG GTG TTT GGA GAA GAG GGT TTG AAT CTA AAT CTT GAA GAT GTT CAA GCT CAT GAT TTA GGG AAA GTT GGA GAG GTC
      A  T  G  G  A  V  F  G  E  E  G  L  N  L  N  L  E  D  V  Q  A  H  D  L  G  K  V  G  E  V                          330
991  ATC GTC ACC AAA GAT GAT GCC ATG CTT TTG AAA GGA AAA GGT GAC AAA GCT CAC ATT GAA AAA CGT ATT CAA GAA ATC ACT GAG CAG CTA
      I  V  T  K  D  D  A  M  L  L  K  G  K  G  D  K  A  H  I  E  K  R  I  Q  E  I  T  E  Q  L                          360
1081 GAC ATC ACA ACT AGT GAA TAT GAA AAA GAA AAG CTG AAC GAG CGA CTT GCT AAA CTT TCA GAT GGA GTA GCT GTG TTG AAG GTT GGA GGA
      D  I  T  T  S  E  Y  E  K  E  K  L  N  E  R  L  A  K  L  S  D  G  V  A  V  L  K  V  G  G                          390
1171 ACA AGT GAT GTT GAA GTG AAT GAG AAG AAA GAC AGA GTT ACT GAT GCT CTC AAT GCT ACA AGA GCA GCT GTT GAA GAA GGC ATT GTT CTA
      T  S  D  V  E  V  N  E  K  K  D  R  V  T  D  A  L  N  A  T  R  A  A  V  E  E  G  I  V  L                          420
1261 GGA GGG GGC TGC GCT CTG CTT CGG TGC ATC CCA GCC TTG GAT TCA TTA AAG CCT GCT AAT GAA GAC CAG AAA ATA GGT ATA GAA ATT ATT
      G  G  G  C  A  L  L  R  C  I  P  A  L  D  S  L  K  P  A  N  E  D  Q  K  I  G  I  E  I  I                          450
1351 AAA AGA GCA CTT AAA ATT CCT GCA ATG ACG ATT GCT AAG AAT GCA GGT GTT GAA GGA TCT TTG ATA GTT GAG AAA ATT CTG CAG AGT TCC
      K  R  A  L  K  I  P  A  M  T  I  A  K  N  A  G  V  E  G  S  L  I  V  E  K  I  L  Q  S  S                          480
1441 TCA GAA GTT GGT TAT GAC GCC ATG CTT GGA GAT TTT GTG AAC ATG GTG GAA AAA GGG ATC ATT GAT CCA ACA AAG GTT GTG AGA ACT GCC
      S  E  V  G  Y  D  A  M  L  G  D  F  V  N  M  V  E  K  G  I  I  D  P  T  K  V  V  R  T  A                          510
1531 TTA CTG GAT GCT GCT GGG GTG GCC TCC TTG CTA ACT ACA GCC GAA GCT GTA GTG ACA GAA ATT CCT AAA GAA GAG AAG GAC CCT GGA ATG
      L  L  D  A  A  G  V  A  S  L  L  T  T  A  E  A  V  V  T  E  I  P  K  E  E  K  D  P  G  M                          540
1621 GGT GCA ATG GGT GGC ATG GGA GGG GGT ATG GGA GGC GGC ATG TTC TAA CTCCTAGAGTAGTGCTTTGCCCTTATCAATGAACTGTGACAGGAAGCTCAAGGCA
      G  A  M  G  G  M  G  G  G  M  G  G  G  M  F  *                                                                    555
1725 GGTTCCTCACCAATAACTTCAGAGAAGTCACCTGGAGAAAATGACTGAAGAGAAGGCTGGCTGACCACTGTAATCATCAGTTACTGGTTTCCTTTGACGATATATAATGGTTTACTGCT
1844 GTCATTGTCCATGCCTACAGATAATTTATTTTGTATTTTTGAATAAAGAACATTTGTACATTCCTGATGCTGGTTGCAAGAGCCATATACCAGTGTCCTGCTTTCAACTTAAATCACTG
1963 AGGCATCTCTACTCTTCTGTGAGTCATCAGGACTGTAGCGCTGTGTCAACAAACATAGAGAGTTCAGAAGACAGCCTTTCTGTGGAAGGGTGGGAATGATTGTGTACAAAGTAGAGAAG
2082 TATCCAATTATGTGACAACCTTTGTGTAATAAAATTTT 2119
```

Figure 16 **Nucleotide and deduced amino acid sequences of mouse P1 cDNA.** The numbers on the left correspond to the nucletide sequence. The numbers on the right correspond to the deduced amino acid sequence. The remaining eight amino acid residues of the putative mitochondrial targeting presequence at the N-terminal and the polyadenylation signal at the 3'-end are underlined.

```
1     GCC AAA GAT GTA AAA TTT GGT GCG GAT GCT CGA GCC TTA ATG CTT CAA GGT GTA GAC CTT TTA GCC GAT GCT GTA GCT GTT ACA ATG GGG
        A   K   D   V   K   F   G   A   D   A   R   A   L   M   L   Q   G   V   D   L   L   A   D   A   V   A   V   T   M   G  30
91    CCA AAG GGA AGA ACA GTG ATT ATT GAA CAG AGT TGG GGA AGT CCC AAA GTA ACA AAA GAT GGG GTC ACT GTT GCA AAG TCA ATT GAT TTA
        P   K   G   R   T   V   I   I   E   Q   S   W   G   S   P   K   V   T   K   D   G   V   T   V   A   K   S   I   D   L  60
181   AAG GAT AAA TAC AAA AAT ATC GGA GCT AAG CTT GTT CAG GAT GTT GCC AAT AAC ACA AAT GAA GAG GCT GGG GAT GGC ACC ACC ACT GCC
        K   D   K   Y   K   N   I   G   A   K   L   V   Q   D   V   A   N   N   T   N   E   E   A   G   D   G   T   T   T   A  90
271   ACT GTT CTG GCA CGG TCT ATT GCC AAG GAG GGC TTT GAG AAG ATC AGC AAA GGG GCT AAT CCA GTG GAA ATC CGG AGA GGT GTG ATG TTG
        T   V   L   A   R   S   I   A   K   E   G   F   E   K   I   S   K   G   A   N   P   V   E   I   R   R   G   V   M   L 120
361   GCT GTT GAT GCT GTA ATT GCT GAA CTT AAG AAA CAA TCT AAA CCT GTG ACA ACC CCT GAA GAA ATT GCT CAG GTT GCT ACA ATT TCT GCA
        A   V   D   A   V   I   A   E   L   K   K   Q   S   K   P   V   T   T   P   E   E   I   A   Q   V   A   T   I   S   A 150
451   AAC GGA GAC AAA GAC ATT GGG AAC ATC ATT TCT GAT GCA ATG AAG AAG GTT GGA AGA AAG GGT GTC ATC ACA GTG AAG GAT GGA AAA ACC
        N   G   D   K   D   I   G   N   I   I   S   D   A   M   K   K   V   G   R   K   G   V   I   T   V   K   D   G   K   T 180
541   CTG AAT GAT GAG CTA GAA ATT ATT GAA GGC ATG AAG TTT GAT AGA GGA TAT ATT TCC CCA TAT TTT ATT AAC ACA TCA AAA GGT CAA AAA
        L   N   D   E   L   E   I   I   E   G   M   K   F   D   R   G   Y   I   S   P   Y   F   I   N   T   S   K   G   Q   K 210
631   TGT GAA TTC CAA GAT GCC TAT GTT TTG TTG AGT GAA AAG AAA ATT TCT AGT GTT CAG TCC ATT GTA CCT GCT CTT GAA ATT GCC AAT GCT
        C   E   F   Q   D   A   Y   V   L   L   S   E   K   K   I   S   S   V   Q   S   I   V   P   A   L   E   I   A   N   A 240
721   CAC CGG AAG CCC TTG GTC ATA ATT GCT GAA GAT GTT GAT GGA GAA GCT CTT AGC ACA CTG GTT TTG AAC AGG CTA AAA GTT GGT CTT CAG
        H   R   K   P   L   V   I   I   A   E   D   V   D   G   E   A   L   S   T   L   V   L   N   R   L   K   V   G   L   Q 270
811   GTT GTA GCA GTC AAA GCT CCA GGG TTT GGG GAC AAC AGG AAG AAC CAG CTT AAA GAT ATG GCT ATC GCT ACT GGT GGT GCG GTG TTT GGA
        V   V   A   V   K   A   P   G   F   G   D   N   R   K   N   Q   L   K   D   M   A   I   A   T   G   G   A   V   F   G 300
901   GAA GAG GGT TTG AAT CTA AAT CTT GAA GAT GTT CAA GCT CAT GAT TTA GGG AAA GTT GGA GAG GTC ATC GTC ACC AAA GAT GAT GCC ATG
        E   E   G   L   N   L   N   L   E   D   V   Q   A   H   D   L   G   K   V   G   E   V   I   V   T   K   D   D   A   M 330
991   CTT TTG AAA GGA AAA GGT GAC AAA GCT CAC ATT GAA AAA CGT ATT CAA GAA ATC ACT GAG CAG CTA GAC ATC ACA ACT AGT GAA TAT GAG
        L   L   K   G   K   G   D   K   A   H   I   E   K   R   I   Q   E   I   T   E   Q   L   D   I   T   T   S   E   Y   E 360
1081  AAG GAA AAG CTG AAC GAG CGA CTT GCT AAA CTC TCA GAT GGA GTA GCT GTG TTG AAG GTT GGA GGG ACA AGT GAT GTT GAA GTG AAT GAG
        K   E   K   L   N   E   R   L   A   K   L   S   D   G   V   A   V   L   K   V   G   G   T   S   D   V   E   V   N   E 390
1171  AAG AAA GAC AGA GTT ACA GAT GCT CTC AAT GCT ACA AGA GCA GCT GTT GAA GAA GGC ATT GTT CTA GGA GGG GGC TGT GCT CTA CTT CGG
        K   K   D   R   V   T   D   A   L   N   A   T   R   A   A   V   E   E   G   I   V   L   G   G   G   C   A   L   L   R 420
1261  TGC ATC CCA GCC TTG GAT TCA TTA AAG CCT GCT AAT GAA GAT CAG AAG ATA GGT ATA GAA ATT ATT AAA AGA GCA CTT AAA ATT CCT GCA
        C   I   P   A   L   D   S   L   K   P   A   N   E   D   Q   K   I   G   I   E   I   I   K   R   A   L   K   I   P   A 450
1351  ATG ACA ATT GCT AAG AAT GCA GGT GTT GAA GGA TCT TTG ATA GTT GAA AAA ATT CTG CAG AGT TCC TCA GAG GTT GGC TAT GAT GCC ATG
        M   T   I   A   K   N   A   G   V   E   G   S   L   I   V   E   K   I   L   Q   S   S   S   E   V   G   Y   D   A   M 480
1441  CTT GGA GAT TTT GTG AAC ATG GTG GAA AAG GGA ATC ATT GAT CCA ACA AAG GTT GTA AGA ACT GCT TTA CTG GAT GCT GCT GGG GTG GCC
        L   G   D   F   V   N   M   V   E   K   G   I   I   D   P   T   K   V   V   R   T   A   L   L   D   A   A   G   V   A 510
1531  TCC CTG CTA ACT ACA GCC GAA GCT GTA GTG ACA GAA ATT CCT AAA GAA GAG AAG GAC CCT GGA ATG GGT GCA ATG GGT GGA ATG GGA GGG
        S   L   L   T   T   A   E   A   V   V   T   E   I   P   K   E   E   K   D   P   G   M   G   A   M   G   G   M   G   G 540
1621  GGT ATG GGA GGT GGC ATG TTC TAA CTCCTAGAATAGTGCTTTGCCCTTATCAATGAACTGTGGCAGGAAGCTCAAGGCAGGTTCCTCACCAATAACTTCAGAGAAGTCACC
        G   M   G   G   G   M   F   *  547
1732  TGAAGAAAATGACTGAAGAGAAGGCTGGCTGATCACTGTAACCATCAGTTACTGGTTTCCTTTGACAATACATAATGGTTTACTGCTGTCATTGTCCATGCCTACAGATAATTTATTTT
1850  GTATTTTTGAATAAAGACATTTGTACATTCCTGATGCTGGGTGCAAGAGCCAGTGTCCTGCTTTCAACTTAAATCACTGAGGCATCTCTACTGTTCTGTTAGCATCAGGACTGTAGCGC
1969  TGTGTCACCACATGAGAAGTTCAGAAGCAGCCTTTCTGTGGAGGGTGAGAATGATTGTGTACAGAGTAGAGAAGTATCCAATTATGTGACAACCTTTGTGTAATAAAATTTTGTTTAAA
```

Figure 17 **Nucleotide and deduced amino acid sequences of rat P1 cDNA.** The numbers on the left correspond to the nucleotide sequence. The numbers on the right correspond to the amino acid sequence. The 3'- end polyadenylation signal is underlined.

acids with an estimated molecular weight of 57,940 daltons. The molecular weights of the mature rat and mouse P1 proteins are in close agreement with those predicted for the human (57,939 daltons) and CHO (57,949) proteins (Jindal *et al.*, 1989; Picketts *et al.*, 1989).

3.2.6 Amino acid alignment of P1 sequences from rat, mouse, CHO, and human cDNAs

The protein sequences deduced from the rat and mouse P1 cDNAs were compared, at the amino acid level, to those previously obtained for human and CHO P1 proteins. The sequence alignment shown in Figure 18 demonstrates the protein homology between these species.

The rat P1 protein differs from CHO and human P1 by 6 and 13 amino acids, respectively. The mature mouse P1 sequence (amino acids 9-555) differs from the corresponding rat, CHO and human sequences in 1, 7 and 14 residues, respectively.

In a comparison of the human and CHO P1 sequences, Picketts *et al.*, (1989) previously showed that of 18 amino acid differences, 9 were conservative substitutions, and two changes were within the mitochondrial presequence. Relative to the CHO sequence, 6 of 7 mouse and 5 of 6 rat amino acid differences were conservative in nature.

Figure 18 **Comparison of protein sequences obtained from CHO, rat, mouse and human sources.** Line 1, CHO P1; line 2, rat P1; line 3, mouse P1; and line 4, human P1. The amino acid residues that are identical to the CHO P1 sequence are shown by a dash (-). Numbers are for reference to CHO and human sequences. Amino acids 1 to 26 represent the putative mitochondrial presequence of the CHO and human P1 sequences.

```
                1                  26
CHO     MLRLPTVLRQMRPVSRALAPHLTRAYAKDVKFGADARALMLQGVDLLADAVAVTMGPKGRTVIIEQSWGS
rat                                ----------------------------------------
mouse                              ----------------------------------------
human   -------F--------V-----------------------------------------------G---

        PKVTKDGVTVAKAIDLKDKYKNIGAKLVQDVANNTNEEAGDGTTTATVLARSIAKEGFEKISKGANPVEI
        ----------S-----------------------------------------------------------
        ----------S-----------------------------------------------------------
        ----------S-----------------------------------------------------------

        RRGVMLAVDAVIAELKKQSKPVTTPEEIAQVATISANGDKDIGNIISDAMKKVGRKGVITVKDGKTLNDE
        ------------------------------------------------------------------------
        ------------------------------------------------------------------------
        ----------------------------------------E-------------------------------

        LEIIEGMKFDRGYISPYFINTSKGQKCEFQDAYVLLSEKKISSVQSIVPALEIANAHRKPLVIIAEDVDG
        ------------------------------------------------------------------------
        ----------------------------------------F-------------------------------
        ---------------------------------------------I--------------------------

        EALSTLVLNRLKVGLQVVAVKAPGFGDNRKNQLKDMAIATGGAVFGEEGLNLNLEDVQAHDLGKVGEVIV
        ------------------------------------------------------------------------
        ------------------------------------------------------------------------
        ------------------------------------------------------T-------P-----------

        TKDDAMLLKGKGEKAQIEKRIQEITEQLEITTSEYEKEKLNERLAKLSDGVAVLKVGGTSDVEVNEKKDF
        ----------D--H----------D--------------------------------------------
        ----------D--H----------D--------------------------------------------
        ----------D----------I---DV-------------------------------------------

        VTDALNATRAAVEEGIVLGGGCALLRCIPALDSLKPSNEDQKIGIEIIKRALKIPAMTIAKNAGVEGSL
        ------------------------------------A-----------------------------------
        ------------------------------------A-----------------------------------
        ----------------------------------T-A----------T-----------------------

        VEKILQSSSEIGYDAMLGDFVNMVEKGIIDPTKVVRTALLDAAGVASLLTTAEAVVTEIPKEEKDPGMG
        ----------V------------------------------------------------------------
        ----------V------------------------------------------------------------
        ----M-----V-----A-------------------------------------V-----------------

                573
        MGGMGGGMGGGMF
        -------------
        -------------
        -------------
```

4.0 DISCUSSION

4.1 Screening of Human Genomic Phage Libraries

One of the objectives of this project was to identify the regulatory sequences upstream of the human P1 gene. One clone, pGem10, contained an insert of approximately 2.2 kb which hybridized to both the 0.8 and 1.4 kb human P1 probes. This clone was sequenced and found to be identical to the human P1 cDNA sequence, from a position -24 bp upstream of the translation start site to the position 11 bp downstream from the consensus polyadenylation signal. This sequence consists of one open reading frame and lacks intervening sequences. The absence of intervening sequences has been observed for a number of hsp genes, including the genes that encode the heat-inducible forms of hsp70, which appear to lack introns in all organisms including humans (Lindquist, 1986). Evidence exists from experiments with *Drosophila*, which indicate the absence of introns may be crucial for their expression since RNA splicing is inhibited at elevated temperatures (Yost and Lindquist, 1988).

Clones containing inserts which hybridize to the 0.8 kb (5'- end) human P1 cDNA probe were sequenced in the upstream direction, using the internal primer AB634. All of

these clones were found to terminate at the same upstream position (-24 bp) as observed with pGem10. Since this clone appeared to lack upstream sequences, the screening of an EMBL-3 human genomic library was initiated.

4.2 Human P1 Gene Family Contains Numerous Pseudogenes

Twelve P1 positive phage clones were isolated from the EMBL-3 library, and upon subcloning and partial sequencing they were classified into six distinct groups. These P1-related sequences all contained numerous insertions, deletions and base substitutions, resulting in frame shifts and premature stop signals. Therefore, these sequences were adjudged to represent pseudogenes. The identification of a large family of P1 related sequences is consistent with data from Southern blot analysis of genomic DNA digested with unique restriction enzymes, which indicates the presence of 10-12 genes in human and mouse cells, and a slightly lower number (about 6-7) in CHO cells (Venner *et al.*, 1990).

Previous examination of a CHO P1 related genomic clone revealed a pseudogene which, like the human pseudogenes, was highly homologous in the coding region and part of the 3'-non-coding sequence, but differed in the 5'- upstream region as well as in the distal part of the 3'- non-coding sequence (Venner *et al.*, 1990). However, in contrast to the human pseudogenes, the CHO pseudogene was found to contain an 87

nucleotide intron, which is flanked at the 5'- and 3'- end by the consensus splice sequences (Padgett *et al.*, 1986).

Large pseudogene families have previously been reported for a number of different genes including rat tubulin, human actin, mouse and human immunoglobulin variable regions and many others (see Firtel, 1981; Cowan and Dudley, 1983; Weiner *et al.*, 1986; Wilde, 1986). Data presented in this study show a large number of pseudogenes are present for the P1 protein in humans. The DNA sequence data obtained from the four human pseudogenes which were completely sequenced show that they contain many common changes, indicating that they may have evolved from a common ancestor. Pseudogenes most probably result from either the accumulation of deleterious mutations following gene duplication, or genomic insertion of retrotranscribed copies of cytoplasmic mRNA molecules (processed pseudogenes) which subsequently accumulate deleterious genetic changes (Vanin, 1985; Wilde, 1986). The pseudogenes described here are likely of the latter class given their abundance and common features, however, this family could have been derived from a common ancestral processed pseudogene by gene duplication and mutation.

4.3 Analysis of P1-specific Clones from Rat and Mouse cDNA Libraries

Initial screening of rat and mouse cDNA libraries with probes specific for the 3'- and 5'- end of human P1 cDNA resulted in the isolation of partial clones with average lengths of less than 1 kb. Sequencing revealed the absence of sequences corresponding to the 3'- end region of the P1 cDNA. Rescreening of the same libraries with a 3'- end specific CHO P1 cDNA probe led to the isolation of clones containing inserts of larger sizes (>1 kb). These clones overlapped those obtained earlier in their 5'- end, and contained the consensus polyadenylation signal (AATAAA) near the 3'- terminus.

Examination of the sequences obtained for both the rat and mouse P1 cDNA reveals truncations at the 5'- end. The open reading frames were identified through comparison to the cDNA sequence previously obtained from human and CHO sources. These truncations result in the loss of 18 of the 26 amino acids which should comprise the putative mitochondrial pre-sequence of the mouse clone. The rat clone lacks presequence information entirely, encoding the mature P1 sequence only.

Protein sequence comparison analysis between the rat, mouse, CHO, and human P1 proteins reveals a high degree of homology. Only a small number of amino acid differences can be observed, and many of these are conservative in nature. The rat and mouse mature P1 amino acid sequences differ in only one amino acid. Previous comparison of mammalian (human and CHO) P1 sequences to those obtained for the chaperonin proteins from a variety of sources (viz. bacterial GroEL,

yeast mitochondria hsp60, plasmid chloroplast RSBP) showed 40-50% identity with an additional 20-25% similarity based on conservative amino acid substitutions (refer to Fig. 2; Picketts *et al.*, 1989). Similarly, all four of the mammalian P1 proteins show extensive sequence homology to the 65 kDa antigen of *M. leprae*.

4.4 Possible Functions of the P1 Protein

In this laboratory, mutants of CHO cell lines have been selected for increased resistance to the antimitotic drug podophyllotoxin (Gupta *et al.*, 1982; Gupta and Gupta, 1984). Many of these mutants are altered in a protein designated P1 which has subsequently been localized to the mitochondrial matrix (Gupta *et al.*, 1985; Gupta and Austin, 1987). This observation is somewhat surprising given that the known mechanism of podophyllotoxin action requires binding of the drug to free tubulin dimers (Shilstra *et al.*, 1989).

It has been suggested that P1 plays a direct role in the *in vivo* assembly of MTs (Gupta, 1990). This argument is based on the genetic and biochemical studies described earlier, which indicate the MT-related nature of the P1 lesion and the observed sequence homology of P1 to the chaperonin proteins which are known to facilitate the assembly of oligomeric protein complexes. Furthermore, a number of studies describe specific interactions between MTs and

mitochondria, including (1) a close association of mitochondria with MTs and intermediate filaments in interphase cells (Ball and Singer, 1982; Gupta *et al.*, 1985; Schnapp *et al.*, 1985), (2) mitochondrial binding of fluorescently labelled colchicine and podophyllotoxin derivatives in interphase cells (Moll *et al.*, 1982; Gupta and Dudani, 1989), and (3) specific association of free tubulin with mitochondrial membranes (Bernier-Valentin *et al.*, 1983; Hargreaves and Avila, 1985).

Recently, another protein, P2, which is altered in some of the mutants resistant to MT inhibitors (and previously identified as a MT-related protein) has been shown to correspond to the constitutive form of the major heat shock protein hsp70. A main function of hsc70 involves the transport of cytoplasmic protein to organelles such as mitochondria.

Comparisons between the P1 protein and yeast hsp60 have revealed a high degree of sequence homology and similar structural properties (i.e. high molecular weight complex, weak ATPase activity). Studies with yeast mutants defective in hsp60 function have demonstrated a role for hsp60 in the processing of mitochondrial import proteins such as the ß subunit of F_1-ATPase, cytochrome b_2 and the Rieske F_e/S protein of complex III (Cheng et al., 1989). It would seem logical that the P1 protein should also play a role in the assembly of mitochondrial enzyme complexes and the retranslocation of import proteins into the intermembrane space.

4.5 Future Research Prospectives

4.5.1 Identification of Regulatory Sequences Upstream of the P1 gene

Any strategy for obtaining sequence information upstream of the protein coding region of the P1 gene will have to address the inherent problems created by the presence of multiple pseudogene sequences. One approach would be to simply design human P1 probes that correlate to stretches of sequence over which the largest deviations to P1 are seen in the pseudogenes (such as the 3' and 5' untranslated sequences). These probes should preferentially bind to functional P1 genes under stringent washing conditions. Another approach would involve the application of the polymerase chain reaction (PCR). Genomic DNA would be digested with a frequent cutter such as Acc I or Taq I and circularized by ligation under the appropriate conditions (Ochman *et al.*, 1990). The precise location of the Taq I and Acc I sites are known for the human P1 gene and primers can be synthesized which (a) are incapable of binding to the pseudogenes, and (b) read in opposite directions. The upstream sequence of interest could then be amplified using this "inverse PCR" technique.

4.5.2 Identification of Antigenic Determinants Involved in Adjuvant Arthritis in Rats

The rat P1 sequence described here has a number of useful applications with respect to the identification of the rat autoantigen involved in adjuvant arthritis. First, comparison of the rat P1 protein sequence to those used previously for identification of antibody and T cell epitopes on the mycobacterial chaperonin should provide a starting point for the design of synthetic peptides which could then be tested for their antigenicity, or ability to induce adjuvant arthritis, in rats. Secondly, biochemical purification of rat hsp60 could be facilitated by recombining the rat P1 cDNA clones in such a way as to produce a full length mature P1 sequence which could in turn be used for protein synthesis. In such a system the number of contaminants would be kept to a minimum, providing a relatively clean source of rat hsp60 for use in a variety of immunological procedures.

4.5.3 Biological Significance of P1

It has been suggested that P1 plays a direct role in the <u>in vivo</u> assembly of MTs (Gupta, 1990). One tenet of this hypothesis involves the uptake of α and ß tubulin by mitochondria. Recently, a number of tubulin cDNAs have been isolated from CHO cell libraries in this laboratory. These

clones can be transcribed and translated *in vitro* and then used for mitochondrial binding and/or uptake studies employing similar methodology to that described for the mitochondrial uptake of the human P1 protein (Singh *et al.*, 1990).

The availability of full-length P1 cDNA clones which can be modified to produce mature P1 protein could be a useful tool in identifying the conditions necessary for formation of homooligomeric P1 complexes. Such a system would provide an opportunity to perform *in vitro* binding studies under a variety of controlled conditions (i.e. type of substrate, substrate concentration, ionic strength of medium, temperature, etc.). The ability to produce a functional P1 complex *in vitro* would in turn provide an easily manipulated and controlled system in which the P1 complex mediated assembly of α and β tubulin monomers (which can be independently synthesized *in vitro*) into dimers could be examined.

The P1 mutants isolated in this laboratory are the first to be described for hsp60 in mammalian systems. Therefore, they provide a unique opportunity to study the role of this protein in <u>in vivo</u> microtubule function and assembly/-import of mitochondrial proteins.

Other experiments which will help clarify the nature of the mutation in the Pod^{RII} cell lines include (1) placing the WT CHO P1 gene into a mammalian expression vector and transfecting it into Pod^{RII} cells. Overexpression of the WT P1 should result in a decline in resistance to podophyllotoxin in

the transfected cells. (2) Identification of the DNA lesion responsible for the electrophoretic alteration seen in the M1 protein. One simple approach is to use a CHO cDNA library generated from mutant cell mRNA as a template for PCR amplification of the P1 sequence. Genetic studies indicate the mutation is present on only one copy of the gene and, therefore, upon sequencing, two bases should be read on the same level in the gel at the point of mutation. In fact, this project has been initiated in this laboratory employing a CHO cDNA library from the Col^{R22} cell line described previously.

It would also be interesting to assess the effect of various manipulations of the P1 gene on its function both *in vitro* and *in vivo*. Methodologies such as deletion analyses and site directed mutagenesis should help to answer a variety of questions including: What are the structurally significant domains of the P1 protein? What is the nature of the subunit interaction within the P1 oligomer? What is the role of ATP hydrolysis in P1 function? Where are the substrate binding sites and what are their specific sequence requirements?

5.0 REFERENCES

Ahmad, S., Ahuja, R., Venner, T.J. and Gupta, R.S. (1990) Mol. Cell. Biol. 10: in press.

Anfinsen, C.B. (1973) Science 181: 223-230.

Asai, D.J., Thompson, W.C., Wilson, L., Dresden, C.F., Schulman, H. and Purich, D.L. (1985) Proc. Natl. Acad. Sci. USA 82: 1434-1436.

Ball, E.H. and Singer, S.J. (1982) Proc. Natl. Acad. Sci. USA 79: 123-126

Barraclough, R. and Ellis, R.J. (1980) Biochim. Biophys. Acta 608: 19-31.

Bernier-Valentin, F., Aunis, D. and Rousset, B. (1983) J. Cell Biol. 97: 209-216.

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

Bloom, G.S., Schoenfeld, T.A. and Vallee, R.B. (1984) J. Cell. Biol. 98: 320-330.

Bloom. G.S. and Brashear, T.A. (1989) J. Biol. Chem. 264: 16083-16092.

Bochkareva, E.S., Lissin, N.M. and Girshovich, A.S. (1988) Nature 336: 254-257.

Borisy, G.G. and Taylor, E.W. (1967) J. Cell. Biol. 34: 525-534.

Brinkley, B.R. (1985) Ann. Rev. Cell. Biol. 1: 145-172.

Bukau, B. and Walker, G.C. (1989) J. Bacteriol. 171: 2337-2346.

Cabral, F., Sobel, M. and Gottesman, M.M. (1980) Cell 20: 29-36.

Cabral, F. and Barlow, S.B. (1989) FASEB J. 3: 1593-1599.

Chandrasekhar, G.N., Tilly, K., Woolford, C., Hendrix, R.W. and Georgopoulos, C.P. (1986) J. Biol. Chem. 261: 12414-12419.

Chappell, T.G., Welch, W.J., Schlossman, D.M., Palter, K.B., Schlesinger, M.J. and Rothman, J.E. (1986) Cell 45: 3-13.

Cheng, M.Y., Hartl, F.-U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Hallberg, E.M., Hallberg, R.L. and Horwich, A.L. (1989) Nature 337: 620-625.

Chirico, W.J., Waters, M.G. and Blobel, G. (1988) Nature 332: 805-810.

Ciesielski-Treska, J., Bader, M.F. and Aunis, D. (1982) Neurochem. Res. 7: 275-286.

Cleveland, D.W., Ilwo, S.-Y. and Kirschner, M.W. (1977) J. Mol. Biol. 16: 207-225.

Cleveland, D.W. and Sullivan, K.F. (1985) Ann. Rev. Biochem. 54: 331-365.

Cleveland, D.W. (1989) Cur. Op. Cell. Biol. 1: 10-14.

Cohen, I.R., Holoshitz, J., Van Eden, W. and Frenkel, A. (1985). Arthritis Rheum. 28: 841-845.

Cowan, N.J. and Dudley, L. (1983) Int. Rev. Cytology 85: 147-173.

Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A. and Schekman, R. (1988a) Nature 332: 800-805.

Deshaies, R.J., Koch, B.D. and Schekman, R. (1988b) Trends Biochem. Sci. 13: 384-388.

Dustin, P. (1984) *Microtubules*, Second Ed., Springer-Verlag, Berlin.

Elliott, E.M. Henderson, G., Sarangi, F. and Ling, V. (1986) Mol. Cell. Biol. 6: 906-913.

Ellis, R.J. (1987) Nature 728: 378-379.

Ellis, R.J., Van Der Vies, S.M. and Hemmingsen, S.M. (1989) Biochem. Soc. Symp. 55: 145-153.

Ellis, R.J. (1990) Sem. Cell. Biol. 1: 1-9.

Engers, H.D. and Workshop Participants (1985) Infect. Immun. 48: 603-605.

Engers, H.D. and Workshop Participants (1986) Infect. Immun. 51: 718-720.

Fayet, O., Ziegelhoffer, T. and Georgopoulos, C. (1989) J. Bacteriol. 171: 1379-1385.

Firtel, R.A. (1981) Cell 24: 6-7.

Friedman, D.I., Olsen, E.R., Tilly, K., Georgopoulos, C., Herkowitz, I. and Banuett, F. (1984) Microbiol. Rev. 48: 299-325.

Georgopoulos, C.P. and Eisen, H. (1974) J. Supramol. Struct. 2: 349-359.

Georgopoulos, C. and Ang, D. (1990) Sem. Cell. Biol. 1: 19-25.

Goloubinoff, P., Gatenby, A.A. and Lorimer, G.H. (1989a) Nature 337: 44-437.

Goloubinoff, P., Christeller, J.T., Gatenby, A.A. and Lorimer, G.H. (1989b) Nature 342: 884-888.

Green, L.A.D. and Liem, R.K.H. (1989) J. Biol. Chem. 264: 15210-15215.

Gupta, R.S. (1981) Som. Cell. Genet. 7: 59-71.

Gupta, R. S., Ho, T.K.W., Moffat, M.R.K. and Gupta, R. (1982) J. Biol. Chem. 257: 1071-1078.

Gupta, R.S. (1983) Cancer Res. 43: 505-512.

Gupta, R.S. (1984) Mol. Cell. Biol. 4: 1761-1768.

Gupta, R.S., Venner, T.J. and Chopra, A. (1985) Can. J. Biochem. Cell. Biol. 53: 489-502.

Gupta, R.S. (1986) Mol. Pharmacol. 30: 142-148.

Gupta, R.S. and Venner, T.J. (1986) Ann. NY Acad. Sci. 466: 854-857.

Gupta, R.S. and Dudani, A.K. (1987) Eur. J. Cell. Biol. 44: 278-285.

Gupta, R.S. and Austin, R.C. (1987) Eur. J. Cell. Biol. 45: 170-176.

Gupta, R.S. (1989) in *Drug Resistance in Mammalian Cells*, CRC Press, Boca Raton, FL, pp. 109-137.

Gupta, R.S. and Dundani, A.K. (1989) Eur. J. Cell. Biol. 44: 278-285.

Gupta, R.S. (1990) Trends in Biochem. Sci. (in press).

Haas, I.G. and Wabl, M. (1983) Nature 306: 387-389.

Hargreaves, A.J. and Avila, J. (1985) J. Neurochem. 45: 490-496.

Hartl, F.-U. and Neupert, W. (1990) Science 247: 930-938.

Hayden, J.H., Allen, R.D. and Goldman, R.D. (1983) Cell. Motil. 3: 1-19.

Hemmingsen, S.M., Wooldford, C., Van Der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, G.P., Hendrix, R.W. and Ellis, R.J. (1988) Nature 333: 330-334.

Hermann, H., Dalton, J.M. and Wiche, G. (1985) J. Biol. Chem. 260: 5797-5803.

Hickey, E., Brandon, S.E., Smale, G., Lloyd, D. and Weber, L.A. (1989) Mol. Cell. Biol. 9: 2615-2626.

Hill, D. and Chen, C.H. (1984) Proc. Natl. Acad. Sci. USA 81: 5772-5776.

Hoiby, N. (1975) Scand. J. Immunol. 4 (Suppl.): 187-196.

Holoshitz, J., Matitiau, A. and Cohen, I.R. (1984) J. Clin. Invest. 73: 211-215.

Jindal, S., Dudani, A.K., Singh, B., Harley, C.B. and Gupta, R.S. (1989) Mol. Cell. Biol. 9: 2279-2283.

Karlsson-Parva, A., Söderström, K., Ferm, M., Ivanyi, J., Kiessling, R. and Klareskoq, L. (1990) Scand. J. Immunol. 31: 283-288.

Kingston, R.E., Schuetz, T.J. and Larin, Z. (1987) Mol. Cell. Biol. 7: 1530-1534.

Kushner, S.R. (1978) in *Genetic Engineering*. Editors H.B. Boyer and S. Nicosia, Elsevier/North-Holland, Amsterdam, p. 17.

Lamb, J.R., Ivanyi, J., Rees, A.D.M., Rothbard, J.B., Howland, K., Young, R.A. and Young, D.B. (1987) EMBO J. 6: 1245-1249.

Laskey, R.A., Honda, B.M., Mills, A.D. and Finch, J.T. (1978) Nature 275: 416-420.

Laskey, R.A. and Earnshaw, W.C. (1980) Nature 286: 763-767.

Lebowitz, J., Roberts, J D. and McMacken, R. (1986) Proc. Natl. Acad. Sci. USA 83: 7638-7642.

Lecker, S., Lill, R., Ziegelhoffer, T., Georgopoulos, C.,

Bassford, P.J., Kumamoto, C.A. and Wickner, W. (1989) EMBO J. 8: 371-380.

Lee, A.S. (1987) Trends Biochem. Sci. 12: 20-23.

Lewin, B. (1980) Gene Expression 2. John Wiley and Sons, New York.

Lewis, M.J. and Pelham, H.R.B. (1985) EMBO J. 4: 3137-3143.

Lewis, S.A., Ivanov, I.E., Lee, G.-H. and Cowan, N.J. (1989) Nature 342: 498-505.

Lindquist, S. (1986) Ann. Rev. Biochem. 55: 1151-1191.

Lindquist, S. and Craig, E.A. (1988) Ann. Rev. Genet. 22: 631-677.

Ling, V., Aubin, J.E., Chase, A. and Sarangi, F. (1979) Cell 18: 423-430.

Lubben, T.J., Donaldson, G.K., Vitanen, P.V. and Gatenby, A.A. (1989) Plant Cell 1: 1223-1230.

Luduena, R.F., Anderson, W.M., Prasad, V., Jordan, M.A., Ferrigini, K.C., Roach, M.C., Horowitz, P.M., Murphy, D.B. and Fellows, A. (1986) Ann. NY Acad. Sci,. 466: 718-726.

MacRae, T.H. and Langdon, C.M. (1989) Biochem. Cell. Biol. 67: 770-790.

Mandel, M. and Higa, A. (1970) J. Mol. Biol. 53: 154.

Mandelkow, E. and Mandelkow, E.-M. (1989) Cur. Op. Cell. Biol. 1: 5-9.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mareel, M.M. and Demets, M. (1984) Intl. Rev. Cytol. 90: 125-168.

Margolis, R.L. and Wilson, L. (1977) Proc. Natl. Acad. Sci. USA 74: 3466-3470.

McMullin, T.W. and Hallberg, R.L. (1987) Mol. Cell. Biol. 7: 4414-4423.

McMullin, T.W. and Hallberg, R.L. (1988) Mol. Cell. Biol. 8: 371-380.

Mithieux, G. and Rousset, B. (1989) J. Biol. Chem. 264: 4664-4668.

Mizzen, L.A., Chang, C., Garrels, J.I. and Welch, W.J. (1989) J. Biol. Chem., 264: 20664-20675.

Morelle, G. (1989) Focus 11:1; BRL/Life Technologies Inc., pp. 7-8.

Morgan, W.D. (1989) Mol. Cell. Biol. 9: 4099-4104.

Mose-Larsen, P., Bravo, R., Fey, S.J., Small, J.V. and Celis, J.E. (1982) Cell 31: 681-692.

Munro, S. and Pelham, H.R.B. (1986) Cell 46: 291-300.

Murphy, D.B. and Borisy, G.G. (1975) Proc. Natl. Acad. Sci. USA 72: 2696-2700.

Napolitano, E.W., Pachter, J.S., Chin, S.S.M. and Liem, R.K.H. (1985) J. Cell. Biol. 101: 1323-1331.

Normington, K., Kohn, K., Kozutsumi, Y., Gething, M.-J. and Sambrook, J. (1989) Cell 57: 1223-1236.

Ochman, J., Medhosa, M.M., Garza, D. and Hartl, O.L. (1990) in *PCR Protocols: A guide to methods and applications*, Academic Press, San Diego, CA, pp. 219-227.

Ochs, R.L. (1982) Eur. J. Cell. Biol. 2: 161-170.

Oldstone, M.B.A. (1987) Cell 50: 819-820.

Olmsted, J.B. (1986) Ann. Rev. Cell. Biol. 2: 421-457.

Ostermann, J., Horwich, A.L., Neupert, W. and Hartl, F.-U. (1989) Nature 341: 125-130.

Padget, R.A., Grabowski, P.J., Konerska, M.M., Seiler, S. and Sharp, P.A. (1986) Ann. Rev. Biochem. 55: 1119-1150.

Paschal, B.M., Shpetner, H.S. and Vallee, R.B. (1987) J. Cell. Biol. 105: 1273-1282.

Pearson, C.M. (1956) Proc. Soc. Exp. Biol. Med. 91: 95-101.

Pearson, C.M. and Wood, F.D. (1964) J. Exp. Med. 120: 547-560.

Pelham, H.R.B. (1982) Cell 30: 517-528.

Pelham, H.R.B. (1986) Cell 46: 959-961.

Pelham, H.R.B. (1988) EMBO J. 7: 913-918.

Pelham, H.R.B. (1989) EMBO 8: 3171-3176.

Phillips, G.J. and Silhavy, T.J. (1990) Nature 344: 882-884.

Picketts, D.J., Mayanil, C.S.K. and Gupta, R.S. (1989) J. Biol. Chem. 264: 12001-12008.

Pierson, G.B., Burton, P.R. and Himes, R.H. (1978) J. Cell. Biol. 76: 223-228.

Pushkin, A.V. (1982) Biochim. Biophys. Acta 704: 379-384.

Quartermaus, T. (1987) in *Current Protocols in Molecular Biology*. Green Publishing Associates and Wiley-Interscience, pp. 6.1.1-6.1.4.

Reading, D.S., Hallberg, R.L. and Myers, A.M. (1989) Nature 337: 655-659.

Rebbe, N.F., Hickman, W.S., Ley, T.J., Stafford, D.W. and Hickman, S. (1989) J. Biol. Chem. 264: 15006-15011.

Robol, A., Gull, K. and Pogson, C.I. (1977) FEBS Letts. 75: 149-153.

Rose, M.P., Misra, L.M. and Vogel, J.P. (1989) Cell 57: 1211-1221.

Rothman, J.E. (1989) Cell 59: 591-601.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A laboratory manual.* (2nd Ed.), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

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

Sawada, T. and Cabral, F. (1989) J. Biol. Chem. 264: 3013-3020.

Schibler, M.J. and Cabral, F. (1985) *Molecular Cell Genetics*. M.M. Gottesman, editor, John Wiley and Sons.

Schiff, P.B. and Horwitz, S.B. (1980) Proc. Natl. Acad. Sci. USA 77: 1561-1565.

Schiliwa, M. (1986) *The Cytoskeleton.* Springer-Verlag, New York.

Schilstra, M.J., Martin, S.R. and Bayley, P.M. (1989) J. Biol. Chem. 264: 8827-8834.

Schnapp, B.J., Vale, R.D., Sheetz, M.P. and Reese, T.S. (1985) Ann. NY Acad. Sci. 466: 909-918.

Shelanski, M.L., Gaskin, F. and Cantor, C.R. (1973) Proc. Natl. Acad. Sci. USA 70: 765-768.

Shinnick, T.M. (1987) J. Bacteriol. 169: 1080-1088.

Singh, B., Patel, H.V., Ridley, R.G., Freeman, K.B. and Gupta, R.S. (1990) Biochem. Biophys. Res. Commun. 169: 391-396.

Sloboda, R.D., Rudolph, S.A., Rosenbaum, J.L. and Greengard, P. (1975) Proc. Natl. Acad. Sci. USA 72: 177-181.

Sorger, P.K. and Pelham, H.R.B. (1988) Cell 54: 855-864.

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

Strauss, W.M. (1987) in *Current Protocols in Molecular Biology*. Green Publishing Associates and Wiley-Interscience, pp. 6.3.1-6.3.6.

Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 84: 4767-4771.

Tanguay, R.M. (1988) Biochem. Cell. Biol. 66: 584-593.

Taylor, I.C.A., Solomon, W., Weiner, B.M., Paucha, E., Bradley, M. and Kingston, R.E. (1989) J. Biol. Chem. 264: 16160-16164.

Thole, J.E.R., Keulen, W.K., Kolk, A.H.J., Groothuis, D.G., Berwald, L.G., Tiesjema, R.H. and Van Embden, J.D.A. (1987) Infect. Immun. 55: 1466-1475.

Thole, J.E.R., van Schooten, C.A., Keulen, W.J., Hermans, P.W.M., Janson, A.A.M., de Vries, R.R.P., Kolk, A.H.J. and van Embden, J.D.A. (1988) Infec. Immun. 56: 1633-1640.

Timasheff, S.N. and Grisham, L.M. (1980) Ann. Rev. Biochem. 49: 565-591.

Tsoulfa, G., Rook, G.A.W., van Embden, J.D.A., Young, D.B., Metilert, A., Isenberg, D.A., Hay, F.C. and Lydyard, P.M. (1989) Ann. Rheum. Dis. 48: 118-123.

Vale, R.D., Reese, T.S. and Sheetz, M.P. (1985a) Cell 40: 39-50.

Vale, R.D., Schnapp, B.J., Reese, T.S. and Sheetz, M.P. (1985b) Cell 40: 449-454.

Vale, R.D., Schnapp, B.J., Reese, T.S. and Sheetz, M.P.

(1985c) Cell 40: 559-569.

Vallee, R.B., Wall, J.S., Paschal, B.M. and Shpetner, H.S. (1988) Nature 332: 561-563.

Van Dyk, T., Gutenby, A. and LaRossa, R. (1989) Nature 342: 451-453.

Van Eden, W., Holoshitz, J., Nevo, Z., Frenkel, A., Klajman, A. and Cohen, I.R. (1985) Proc. Natl. Acad. Sci. USA 82: 5117-5120.

Van Eden, W., Hogervorst, E.J.M., Van der Zee, R., Van Embden, J.D.A., Hensen, E.J. and Cohen, I.R. (1989) Rheumatol. Int. 9: 187-191.

Vanin,. E.F. (1985) Ann. Rev. Genet. 19: 253-272.

Venner, T.J., Singh, B. and Gupta, R.S. (1990) J. DNA and Cell. Biol. (In press).

Wada, M. and Itikawa, H. (1984) J. Bacteriol. 157: 694-696.

Wada, M., Sekine, K. and Itikawa, H. (1986) J. Bacteriol. 168: 213-220.

Weber, K., Bibring, T.H. and Osborn, M. (1975) Exp. Cell. Res. 95: 111-1120.

Weiner, A.M., Deininger, P.L. and Efstratidis, A. (1986) Ann. Rev. Biochem. 55: 631-666.

Weisenberg, R.C., Borisy, G.G. and Taylor, E.W. (1968) Biochemistry 7: 4466-4479.

Weisenberg, E.C. (1972) Science 177: 1104-1105.

Welch, W.J. and Suhan, J.P. (1985) J. Cell. Biol. 101: 1198-1211.

Wilde, C.D. (1986) CRC Crit. Rev. Biochem. 19: 323-352.

Williams, G.T., McClanahan, T.K. and Morimoto, R.I. (1989) Mol. Cell. Biol. 9: 2574-2587.

Wilson, L. and Farrell, K.W. (1986) Ann. N.Y. Acad. Sci. 466: 690-708.
Winfield, J.B. (1989) Arthrit. Rheumat. 32: 1497-1504.

Yamamoto, T., McIntyre, J., Sell, S.M., Georgopoulos, C., Skowyra, D. and Zylicz, M. (1987) J. Biol. Chem. 262: 7996-7999.

Yost, H.J. and Lindquist, S. (1988) Science 242: 1544-1548.

Young, D.B., Fohn, M.J., Khanolkar, S.R. and Buchanan, T.M. (1985) Clin. Exp. Immunol. 60: 546-552.

Young, D.B., Ivanyi, J., Cox, H.K. and Lamb, J.R. (1987) Immunol. Today 8: 215-219.

Young, D.B. (1990a) Sem. Cell. Biol. 1: 27-35.

Young, R.A. (1990b) Ann. Rev. Immunol. 8: 401-420.

Ziai, M.R., Giordano, A., Armandola, E. and Ferrone, S. (1988) Anal. Biochem. 171: 192-196.

Zigmond, S.H. (1989) Cur. Op. Cell. Biol 1: 80-86.