MOLECULAR CLONING

OF A

CHINESE HAMSTER MITOCHONDRIAL PROTEIN

MOLECULAR CLONING OF A CHINESE HAMSTER MITOCHONDRIAL PROTEIN RELATED TO THE CHAPERONIN FAMILY OF PROTEINS

By

DAVID JOHN PICKETTS, 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 1989

MASTER OF SCIENCE
(Biochemistry)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Molecular Cloning of a Chinese Hamster Mitochondrial Protein Related to the Chaperonin Family of Proteins.

AUTHOR: David John Picketts, B.Sc. (Queen's)

SUPERVISOR: Dr. R.S. Gupta

NUMBER OF PAGES: xii, 158

ABSTRACT

The complete cDNA sequence of a mitochondrial protein from Chinese hamster ovary cells has been determined. This protein, designated P1, was originally identified in cells resistant to the microtubule inhibitor podophyllotoxin (Gupta, 1981). The mutant cell line contained an alteration of the P1 protein that gave rise to a new, more acidic protein, designated M1 (Gupta et al., 1982). The P1 protein was determined to be microtubule- related based on the cross-resistance pattern of the mutants to other microtubule inhibitors, and corelease with tubulin under conditions which cause microtubule depolymerization (Gupta et al., 1982). Subcellular fractionation studies localized this protein to the matrix of the mitochondria (Gupta and Austin, 1987). Antibodies raised against P1 were used to isolate a cDNA clone from human cells (Jindal et al.,1989).

The human cDNA clone was used as a probe to screen for clones of the P1 protein in bacteriophage λgt10/λgt11 cDNA libraries prepared from CHO cells. The P1 cDNA encodes a protein of 573 amino acids with a relative molecular mass of 60,983 daltons. The first 26 amino acids meet the requirements of a mitochondrial matrix targeting sequence. The mature protein is 547 residues in length with a relative molecular mass of 57,949 daltons.

The deduced amino acid sequence shows 97% identity to the the human P1 protein. More interestingly, the amino

acid sequence shows extensive homology (42 to 55% identical residues and an additional 15 to 20% conservative replacements) to the chaperonin class of molecular chaperones. This class of proteins includes the hsp60 protein of yeast, the groEL protein of Escherichia coli, the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit binding protein of plant chloroplasts, and the 62-65-kDa major antigenic protein of mycobacteria and Coxiella burnetii. The homology of P1 with the above proteins begins after the putative mitochondrial presequence and extends to the C-terminal end. Several regions throughout the protein sequence are highly conserved and are proposed to be functional domains of the protein. Also highly conserved is a Gly-Gly-Met repeating motif at the carboxy-terminus. The function of this sequence is undetermined, as yet. A dendrogram was constructed from the sequence homology data. It suggested that mitochondrial P1 evolved from purple bacteria which is the endosymbiont which gave rise to mitochondria.

The chaperonin class of proteins have been shown to assist in the assembly of oligomeric protein structures. It is suggested that the P1 protein may play a similar role in mammalian cells. The high degree of homology between P1 and the 65-kDa mycobacterial antigen also suggests that P1 may be involved in certain autoimmune diseases.

ACKNOWLEDGEMENTS

I would like to express sincere gratitude to my supervisor, Dr. R.S. Gupta, for allowing me the opportunity to attack this project, and also for keeping me on the right track with guidance, profitable discussions, and valuable criticism. I would also like to thank the other members of my supervisory committee, Dr. R.A. Rachubinski and Dr. K.B. Freeman, who provided constructive and helpful discussions during these studies.

The financial support from the Department of Biochemistry and McMaster University through scholarships and teaching assistantship duties is gratefully acknowledged.

I must also thank the many friends I made here for providing an enjoyable atmosphere for both work and pleasure. The laboratory remained an interesting place thanks to: Mr. B. Eng, for friendship and excellent technical assistance, and Drs. S. Jindal, C.S.K. Mayanil, and S. Ahmad for thoughtful discussions. I would also like to thank Mr. R. Ivimey for help with the figures, Mr. J Mewborn for use of his laser printer, Ms. S. Taylor and Drs. D.P. Lillicrap and P.J. Bridge for reading the manuscript.

Finally, thanks for everything Mom, Dad, Katherine, and Cheryl. Your love and support have helped me immensely.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADP	adenosine diphosphate
AMV	avian myeloblastosis virus
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
CHO	Chinese hamster ovary
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
EMS	ethyl methanesulfonate
g,ug	gram, microgram
GTP	guanosine triphosphate
IPTG	isopropyl-ß-D-thiogalactopyranoside
kb, kbp	kilobase, kilobase pair
KCl	potassium chloride
kDa	kilodalton
LB	Luria-broth
M	molar
Mg^{2+}	magnesium ions
$MgCl_2$	magnesium cloride

min	minutes
ml	milliliter
mM	millimolar
M-MLV	Moloney-murine leukaemia virus
MOPS	(3-[N-morpholino]propane sulfonic acid)
mRNA	messenger RNA
MT	microtubules
NaCl	sodium chloride
NaOH	sodium hydroxide
pfu	plaque forming units
PIPES	[1,4 piperazinebis(ethanesulfonic acid)]
pmol	picomoles
Pod^{Rn}	mutants of CHO cells obtained after n selections in the presence of podophyllotoxin
RNA	ribonucleic acid
rRNA	ribosomal RNA
Rubisco	ribulose-1,5-bisphosphate carboxylase/oxygenase
SDS	sodium dodecyl sulfate
SSC	saline sodium citrate
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
Temed	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
V	Volt
W	Watt
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside

1.0 INTRODUCTION

The cytoplasm is an important and well organized component of the cell. This cellular organization is due, in part, to a complex network of filaments known as the cytoskeleton. As a highly structured system, the cytoskeleton has been implicated in a vast number of cellular functions, including directional cell movement (Anderson et al., 1982), ciliary and flagellar movement (Dustin, 1984), secretion and intracellular transport (Hayden et al., 1983; Posner et al., 1982), movement of intracellular organelles (Schnapp et al., 1986; Vale et al., 1985a,b,c), maintenance of cell shape (Ciesielska-Treska et al., 1982; Ochs 1982), and cytokinesis (for review see Dustin, 1984).

The cytoskeleton consists of three major filament systems: microfilaments, intermediate filaments, and microtubules (MTs). Each system is polar, helical, and comprised of polymers of protein subunits. However, they vary with respect to diameter, length, and protein composition.

Microfilaments are fibers, 7 nm in diameter, that are composed of actin monomers. In most eukaryotic cells, actin is the most abundant protein, typically representing 5-10% of total protein (Schliwa, 1986). Actin is a heterogeneic protein, with at least 6 isoforms found in

vertebrates (Schliwa, 1986).

Intermediate filaments are alpha-helical structures 8-11 nm in diameter. They are composed of different proteins in different cell types and usually provide tensile strength to cells. Some examples include keratin filaments in epithelial cells, neurofilaments in nerve cells, vimentin filaments in fibroblasts, and desmin filaments in muscle (Clarkson et al., 1986).

Microtubules are the largest of the cytoskeletal filaments with an outside diameter of 25 nm, and an inside diameter of 15 nm (see Stryer, 1988). They range in length from less than a micrometer up to several micrometers. However, there have been reports of microtubules several millimeters in length in brain axons (Dustin, 1984).

Although there is a vast amount of literature on all three filament systems, this thesis will only discuss some of the work done on MTs.

1.1 Microtubules: Biochemistry and Structure

Microtubules are hollow cylinders that consist of 13 protofilaments around their circumference (Figure 1a; Lewin, 1980), although there have been reports of MTs composed of 10 to 16 protofilaments (Burton et al., 1975; Pierson et al., 1978; Saito and Hamn, 1982). When MTs are polymerized in vitro, 14 protofilaments are most common (Pierson et al., 1978). Each protofilament is composed of alternating alpha and beta subunits of the protein tubulin.

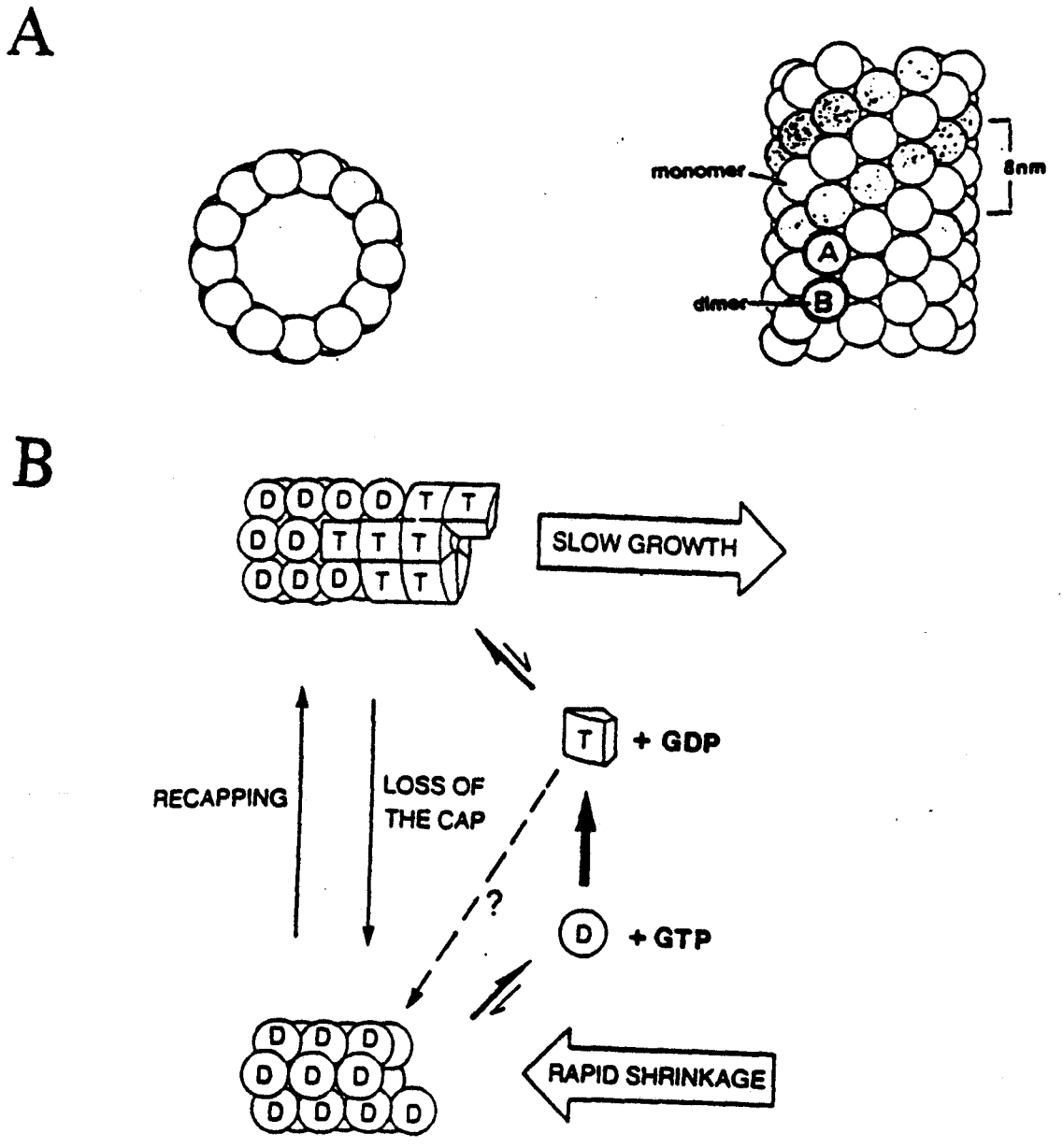


Figure 1 Schematic diagram of the helical pattern and assembly of tubulin subunits into a microtubule. (A) Tubulin dimers are arranged into a helical pattern forming a hollow filament with thirteen protofilaments (taken from Stryer, 1988). (B) GTP cap model for dynamic stability. MTs with a cap of GTP-containing subunits, denoted by T, grow slowly by the addition of T-containing subunits. At some probability, GTP-hydrolysis catches up with assembly, the T cap disappears, and the polymer transits to the rapid shrinkage phase (all D). The GDP-containing subunits released exchange with free GTP to form T subunits. It is not known whether T subunits can recap a depolymerizing end (taken from Kirschner, 1989).

Tubulin was first identified as a colchicine binding protein (Borisy and Taylor, 1967), and later purified by ammonium sulfate precipitation and DEAE cellulose column chromatography (Weisenberg et al., 1968). Four years later, Weisenberg (1972) determined the conditions necessary for in vitro reconstitution of MTs. This provided the basis for the cycle procedure to purify tubulin (Shelanski et al., 1973). This procedure involves warming a tissue extract to 37°C in a glycerol based medium with GTP and magnesium ions. This promotes the assembly of MTs, which are collected by centrifugation at warm temperatures. The MTs are then cooled to 4°C to disassemble the MTs and centrifuged at 4°C to remove undissociated material. This cycle is repeated several times. Tubulin is then separated from other microtubule-associated proteins.

After purification, tubulin appeared as a 110,000 dalton protein on non-denaturing polyacrylamide gels. This was later shown to consist of two similar but non-identical subunits of 55,000 daltons, alpha- and beta-tubulin, with beta-tubulin showing greater mobility (Bryan and Wilson, 1971). The subunits have a one to one stoichiometry (Bibring and Baxandall, 1974), with alpha-beta as the predominant dimer (Timasheff and Grisham, 1980; Wilson and Bryan, 1974).

Two groups independently determined the amino acid

sequences of alpha- and beta-tubulin. Ponstingl and coworkers determined the two sequences via amino acid sequencing of proteolytic fragments of porcine brain tubulin (Krauhs et al., 1981; Ponstingl et al., 1981), whereas Valenzuela and coworkers used cDNA cloning of chick brain tubulin mRNA (Valenzuela et al., 1981). From the nucleotide and deduced amino acid sequences the tubulins were shown to be about 50,000 daltons each, to have acidic pI values, and to have a high proportion of acidic residues at the C-terminal end. Comparison of the sequences of alpha-tubulin from rat and chicken brain showed that only 2 of the 411 amino acids were different (Lemischka et al., 1981). This suggests that tubulin is a highly conserved protein. This has also been shown by several other experiments. Antibodies prepared against bovine brain tubulin cross-react with tubulins from a variety of other species (Dustin, 1984; Lewin, 1980). In addition, tubulin purified from one source will polymerize to MTs from another species (Dustin, 1984).

Although highly conserved, the mammalian alpha- and beta-tubulin gene families each comprise 15 to 20 members (Cleveland et al., 1980; Lee et al., 1983). Many of these genomic copies are pseudogenes (Cowan and Dudley, 1983). In humans only two of the alpha-tubulin genes and three of the beta-tubulin genes examined thus far, have been shown to be functional (Cleveland and Sullivan, 1985). Both the

alpha- and beta-tubulin genes, show substitutions that appear to be restricted to domains at the extreme carboxy-terminal 15 residues and close to the extreme amino-terminus of the protein (Cleveland and Sullivan, 1985). This conservation makes it difficult to distinguish all the different gene copies. However, the 3'-untranslated sequences are quite dissimilar which allows for gene distinction (Cowan et al., 1983). It is well established that beta-tubulin genes contain four exons separated by three introns (Sullivan et al., 1986). Alpha-tubulin also has a unique characteristic, it is the substrate for a posttranslational modification whereby the C-terminal tyrosine residue is removed and then re-added. The precise role of this cyclic process is unknown but it has been suggested that it plays a role in stabilizing a subset of MTs (Pratt et al., 1987).

It has been postulated that the diversity of function of MTs suggests that the tubulin genes are differentially expressed and under tight regulation. For example, in CHO cells it was shown that three alpha-tubulin genes were expressed with frequencies of 71, 24, and 5% by RNA dot blot analysis (Elliott et al., 1985). It has also been proposed that tubulin gene expession is regulated by a feedback control mechanism which is linked to the level of unpolymerized tubulin subunits present in the cytoplasm (Ben Ze'ev et al., 1978). Recently, Cleveland has shown

that actively translated beta-tubulin mRNAs are targeted for degradation by co-translational binding to a tetrapeptide (probably of unpolymerized tubulin subunits), at the amino terminus, as the newly synthesized protein emerges from the ribosome (Cleveland, 1988).

Each tubulin dimer contains two GTP binding sites, one per subunit. The beta-tubulin site has been termed the exchangeable site (E-site) and promotes MT assembly. The alpha-tubulin site was termed the non-exchangeable site (N-site). At the N-site, GTP is irreversibly bound and is thought to play a role in the structural stability of the dimer (Spregelman et al., 1977). This molecule of GTP is not hydrolysed during assembly of the dimer to the MT. However, at the E-site one molecule of GTP is reversibly bound and then hydrolysed with the addition of the dimer to the growing end of the MT (David-Pfeufy et al., 1977, 1978). Hydrolysis is not absolutely required as assembly can occur in the presence of non-hydrolysable GTP analogs (Schliwa, 1986). In fact, it has been shown that hydrolysis actually lags behind elongation (Carlier and Pantaloni, 1981). Thus, hydrolysis is not essential for the polymerization mechanism, but probably occurs within the polymer. This also explains the polymerization with GTP analogs.

The current model for MT polymer formation is the model of dynamic instability (see Kirschner, 1989). In

this model, MTs grow if the hydrolysis of GTP lags behind the rate of polymerization. This lag in hydrolysis causes a cap of GTP-containing subunits to occur at the end of the MT, which permits the addition of more subunits and the slow growth of the microtubule. However, when hydrolysis outruns polymerization the cap is small and the MT starts to rapidly depolymerize (Kirschner, 1989). This theory can be distinguished from the treadmilling theory (Margolis and Wilson, 1978) by examination of single microtubules within a population. At low tubulin concentration a small number of MTs will have GTP-caps, and thus, the fraction of microtubules growing will be small. At high concentrations of tubulin most MTs will contain a GTP cap and thus most microtubules will be growing. At any time in the cell there will be both populations of polymerizing and depolymerizing MTs and only the fraction in the two phases changes (Kirschner and Mitchison, 1986).

Microtubules showing dynamic stability have a critical concentration at which the fraction of MTs polymerizing and the growth rate is equal to the fraction of MTs depolymerizing and the shrinkage rate (Kirschner, 1989). Below the critical concentration the majority of the population is depolymerizing and ultimately the polymer population would become zero. However, the presence of stable nucleating sites allows for polymer assembly. The extent of this assembly depends directly on the number of

nucleating sites (Kirschner, 1989). In the cell nucleation occurs at preferred sites called microtubule organizing centres (MTOC; Brinkley, 1985). Data obtained by immunohistochemical techniques showed that in interphase cells MTs emanate from the perinuclear region, mainly at or near the centrosome (De Brabander et al., 1977; Weber et al., 1975). The centrosomes, now termed the MTOC, are well known for their role in spindle formation during mitosis.

1.2 Microtubule-Associated Proteins

The most common method for isolation of microtubule-associated proteins (MAPs) is by the in vitro assembly/disassembly method for purifying tubulin. Along with tubulin, some 15 to 30 different proteins appear on a polyacrylamide gel, depending on the exact details of the procedure (Stearns and Brown, 1979; Timasheff and Grisham, 1980). These proteins copurify with tubulin in constant stoichiometry and can be divided into two major groups, high molecular weight proteins (250-350 kDa) and proteins ranging from 55 to 68 kDa. The high molecular weight MAPs were shown to form projections on MT surfaces (Kim et al., 1979) and have been resolved into two major components, MAP 1 and MAP 2 (Olmsted, 1986). The lower molecular weight proteins represent a set of 4 to 6 proteins collectively termed tau protein (Cleveland et al., 1977). In the last few years several other MAPs have been reported. These are summarized in Table 1, and include MAP 3, MAP 4, STOPS, and

Protein	Subspecies	Subunit mass (kDa)*	Primary source	Properties
MAP 1	1A, 1B, 1C	350	brain	thermolabile; projection on microtubule
Light chains		28, 30	brain	
MAP 2	2A, 2B	270	brain	thermostable; projection on microtubule; separable into projection (235 kDa) and bindin (35 kDa) domains; phosphorylat binds calmodulin; associated wit MAP 2 projection domain
Type II cAMP-dependent protein kinase		53, 39		
Tau	3–5	55–62	brain	thermostable; number of polypeptides depends on age and species; phosphorylated; binds calmodulin
MAP 3	—	180	brain	
MAP 4; 210-kDa HeLa MAP; 205-kDa *Drosophila* MAP	3–4	200–240 depending on species	cultured mammalian cells; mouse tissues (MAP 4); *Drosophila* (205-kDa)	thermostable
125-kDa MAP	—	125	cultured mammalian cells	
Chartins	—	69, 72, 80	cultured mammalian cells; primary neurons	thermolabile; phosphorylated subspecies
STOPS	—	140, 72, 56	brain	associated with cold-stable microtubules
Sea urchin MAPS	—	37, 75, 80, 150, 200, 235	sea urchin eggs; sea urchin spindles	spindle localization
Kinesin	—	110	squid axoplasm	moves particles on microtubules
		134	sea urchin eggs	

* Denatured mass of major polypeptides in each class.

Table 1 Major classes of microtubule-associated proteins. (taken from Olmsted, 1986)

kinesin, however, only MAP 1, MAP 2, and tau will be discussed in more detail.

MAP 1 is composed of three high molecular weight polypeptides, MAP 1A, 1B, and 1C, and two lower molecular weight proteins (28,000 and 30,000 daltons). These proteins appear as projections on microtubules under thin-section electron microscopy. The gene for MAP 1 has been cloned and Northern blot analysis with this clone has shown that transcripts are prevalent in a wide array of tissues (Lewis et al., 1986a). This is in agreement with immunological data which shows that MAP 1A and 1B is widely distributed but most prevalent in the neuronal cytoskeleton.

MAP 2 has been extensively studied. It is a 280,000 molecular weight doublet, MAP 2A and 2B (Kim et al., 1979). Immunoreactivity and peptide mapping have shown that these two peptides are almost identical (Herrmann et al., 1984, 1985). Gentle proteolysis of MAP 2 cleaves each doublet into two domains, one of which has MT binding capability (Olmsted, 1986). The other domain is the projection domain. Herrmann et al. (1985) also determined that MAP 2 contained 12 phosphorylation sites. The extent of phosphorylation determines its binding to MTs; extensive phosphorylation inhibits MT binding (Olmsted, 1986). Based on homology to spectrin it is thought that MAP 2 may interact with actin (Olmsted, 1986). MAP 2 has been cloned

by Lewis et al. (1986b) and appears to be present in only one copy. Interestingly, transcripts can only be detected in brain tissue which contradicts immunological data (Lewis et al., 1986b).

Tau proteins were initially thought to be breakdown products of a common precursor protein. They have since been shown to be a closely related set of four to six proteins ranging in size from 52,000 to 68,000 daltons (Olmsted, 1986). Tau proteins are heat resistant (Cleveland et al., 1977) and phosphorylated (Lindwall and Cole, 1984). They tend to promote MT polymerization much more rapidly and extensively when dephosphorylated (Lindwall and Cole, 1984). Tau has been cloned and was determined to be a single copy 6 kb gene that is differentially spliced (Drubin et al., 1984). From human and mouse sequences, it appears that the tau protein has a characteristic repeat of 31 or 32 amino acids at the carboxy terminal half (Lee et al., 1988). Each repeat contains a Pro-Gly-Gly-Gly motif and it is thought that the three repeats may represent the tubulin binding region (Lee et al., 1988). Tau has been found in the neurofibrillary tangles seen in the brain tissue of individuals affected with Alzheimer's disease (Kosik et al., 1986).

One group has isolated a MAP, by its association with MT preparations, that was a specific isoform of HSP 70 (Lim et al., 1984; Whatley et al., 1986). The significance of

this finding has yet to be determined.

1.3 Microtubule Inhibitory Drugs and Microtubule Mutants

There is a long list of chemicals that have an affect on MTs. This list has recently been subdivided into over 20 distinct chemical classes (Wilson, 1986). However, only the most common inhibitors will be mentioned in this thesis. These are outlined in Figure 2.

These drugs include the plant alkaloids colchicine, podophyllotoxin, vinblastine, vincristine, maytansine, and taxol; nocodazole, one of many benzimidazole derivatives; the fungal antibiotic griseofulvin; and the close relatives of colchicine, colcemid and steganacin (Dustin, 1984; Mareel and DeMets, 1984; Schliwa, 1986). All these drugs with the exception of taxol and possibly griseofulvin inhibit MT assembly. They prevent microtubule assembly by binding to tubulin. Some of these drugs have proven useful in a clinical setting. For example, griseofulvin has been used as a fungistatic, and colchicine is widely used in the treatment of gout (Dustin, 1984).

Colchicine is by far the best studied of all the MT inhibitors. It has been used for years in the treatment of gout, and it was known since the 1930's to disrupt mitotic spindles (for review see Dustin, 1984). Almost 40 years later colchicine was determined to bind tubulin and then used to purify this protein (Borisy and Taylor, 1967). Colchicine binds tubulin with high affinity in a 1:1

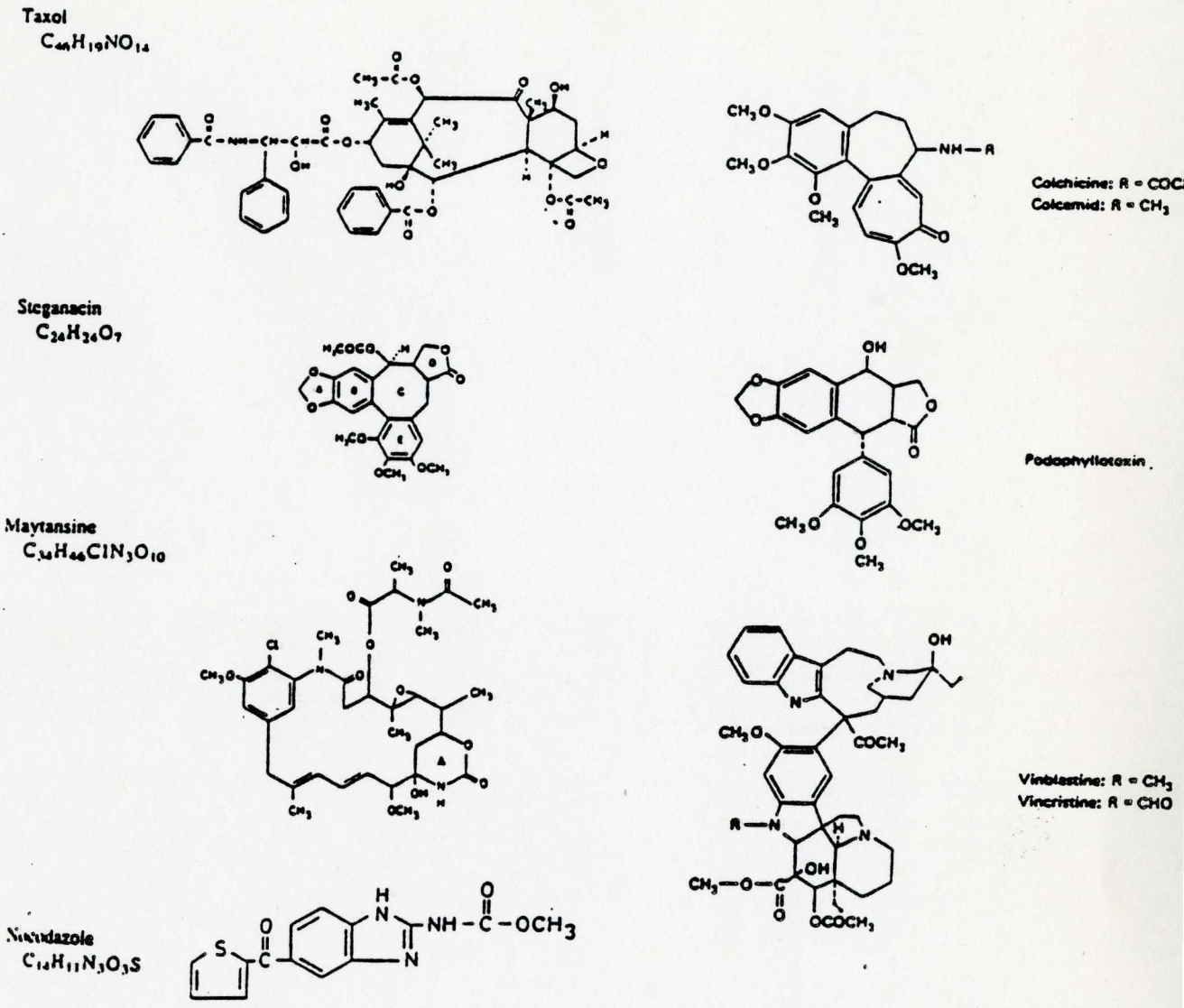


Figure 2 Chemical structures of the major microtubule inhibitors. These microtubule inhibitors include the plant alkaloids colchicine, podophyllotoxin, vinblastine, vincristine, maytansine, and taxol. Nocodazole is a benzimidazole derivative, colcemid is a colchicine derivative and steganacin (taken from Mareel and DeMets, 1984).

complex (Wilson et al., 1974). This causes a conformational change in tubulin, but still allows it to bind to the growing end of the MT (Schliwa, 1986). It is postulated that the binding of the drug-tubulin complex to the polymer prevents further assembly at this end, thereby shifting the equilibrium in favour of disassembly (Margolis and Wilson, 1977). Although this hypothesis has been widely accepted it is probably a more complicated process within the cell.

Nocodazole, steganacin, and podophyllotoxin also bind to tubulin with high affinity. However, binding of these drugs is more rapid, readily reversible, and less temperature dependent (DeBrabander et al., 1976). It was later determined that these drugs are competitive inhibitors of colchicine and presumably bind to the same site (Mareel and DeMets, 1984). It has also been proposed that podophyllotoxin binds at an overlapping, but non-identical site, as colchicine, and that the trimethoxy ring on each molecule is involved in this interaction (Cortese et al., 1977; Mareel and DeMets, 1984).

Vinblastine and vincristine bind to a different site on tubulin. In fact, there are two binding sites per dimer, neither of which is similar to the colchicine binding site. One site has a high affinity, the other only a moderate affinity (Schliwa, 1986). The binding of these drugs is similar to podophyllotoxin in that it is rapid,

reversible and independent of temperature (Mareel and DeMets, 1984). These drugs cause aggregates of tubulin which result in paracrystalline structures to be formed in the cell (Dustin, 1984). Maytansine is a competitive inhibitor of vinblastine binding, but its interaction is not identical to that of vinblastine. This binding is not inhibited by colchicine suggesting another binding site on tubulin (Mareel and DeMets, 1984). Also, it does not initiate tubulin aggregates but causes MT depolymerization.

Taxol is very different from the other inhibitors described in that it promotes MT assembly. It binds to the MT polymer, stabilizing it, and thereby shifting the equilibrium in favour of the polymer, which decreases the critical concentration of tubulin needed for assembly (DeBrabander et al., 1981; Schiff et al., 1979; Schiff and Horwitz, 1980). MTs formed in the presence of taxol do not depolymerize either when the temperature is lowered or by the addition of calcium (Schiff and Horwitz, 1980).

Since these drugs bind to tubulin with high affinity, and are cytotoxic to cells, they have been a principal tool in the study of MTs. Many laboratories, including this one, have utilized these drugs in a genetic and biochemical approach to the study of MTs. There are several ways to obtain mutant cells (for review see Schibler and Cabral, 1985). One approach involves the selection of cells which survive at a concentration of a drug that kills greater

STRAIN	PHENOTYPE	BIOCHEMICAL ALTERATION	REFERENCE
Cmd-4	Colcemid^R	ß-Tubulin	Cabral et al., 1980
Col-2	Colchicine^R	ß-Tubulin	Cabral et al., 1980
Grs-1	Griseofulvin^R	ß-Tubulin	Cabral et al., 1980
Grs-2	Griseofulvin^R	ß-Tubulin	Cabral et al., 1980
Tax-1	Taxol^R	α-Tubulin	Cabral et al., 1981
Tax-2	Taxol^R	ß-Tubulin	see Schibler and Cabral, 1985
Tax-18	Taxol^R	ND	Cabral, 1983
CM^{R}795	Colcemid^R	α-Tubulin	Ling et al., 1979
CM^{R}761	Colcemid^R	ND	Ling et al., 1979
CM^{R}117	Colcemid^R	ND	Ling et al., 1979
Tax_1	Taxol^R	ND	Warr et al., 1982
Ben 1	Benzimidazole carbamate^R	ND	Warr et al., 1982
Pod^{RI}	Podophyllotoxin^R	ND	Gupta, 1981
Pod^{RII}	Podophyllotoxin^R	60,000 protein	Gupta et al., 1982
Col^{R}22a	Colchicine^R	70,000 protein	Gupta et al., 1985
Tax^R-1	Taxol^R	ND	Gupta, 1983
Grs^R-4	Griseofulvin^R	200,000 protein	Gupta, 1984
Noc^R-3	Nocodazole^R	ND	Gupta, 1986
Vin^R-1	Vinblastine^R	ND	Gupta, 1985.

Table 2 Microtubule mutants derived from CHO cells. (taken from Schibler and Cabral, 1985)

than 99% of the initial cell population. Generally, the lowest drug concentration that reduces plating efficiencies to 10^{-5} is used (Schibler and Cabral, 1985). After the drug concentration is determined, cells are mutagenized with EMS or ultraviolet light to 20% survival. The cells are allowed to recover for several days, then they are replated at a concentration of 10^5 to 10^6 cells in the presence of the selecting drug. After approximately one week, colonies of cells can be chosen for further growth and characterization (Schibler and Cabral, 1985). This type of mutant selection is known as a single step selection process. Other methods of selection include temperature-sensitivity selections and a multiple step selection where the drug concentration is increased slightly after each round of selection.

Single step selection procedures usually generate mutant cells that are only 2- to 3-fold resistant to the selecting drug, as compared to wild type cells. These methods have provided experimenters with several MT mutants, many of which have been derived from CHO cells (see Table 2). Most of the cell lines established show an alteration in beta-tubulin, or less often, alpha-tubulin. These cell lines are further characterized by examining their cross-resistance patterns to other microtubule inhibitors. This eliminates mutant cell lines that are of the mdr-phenotype (Ling and Thompson, 1974). Other

experimenters characterize their mutants by examining 2-dimensional gel patterns for an altered tubulin protein (Cabral et al., 1980). However, this can exclude mutations in the MAPs, or mutations which do not show electrophoretic alterations in the tubulins. This thesis will deal with a class of mutants obtained by selection with podophyllotoxin in this laboratory (Gupta, 1981; Gupta et al., 1982; Gupta, 1983).

1.4 Characterization of Podophyllotoxin Mutants

Podophyllotoxin mutants were isolated, in a single step, by selecting resistant colonies from 5 x 10^5 cells plated at a drug concentration of 20 ng/ml from non-mutagenized cells or at a concentration of 30 ng/ml from EMS-treated cells (Gupta, 1981). Colonies from both mutagenized and non-mutagenized cells were shown to be two- to four-fold more resistant than wild type cells. These drug resistant phenotypes were determined to be stable since they maintained their resistance after growth in non-selective medium for more than six months (Gupta, 1981). They were designated Pod^{RI} mutants. Some of these Pod^{RI} mutants ($Pod^{RI}2$, and $Pod^{RI}16$) were subjected to a second selection step. These cells were plated in culture medium containing 50 ng/ml podophyllotoxin and mutants were obtained with a frequency of 1 x 10^{-6}. This value was enhanced when the Pod^{RI} cells were EMS-treated prior to selection. These mutants (eg. $Pod^{RII}6$) showed an increased

resistance to podophyllotoxin up to 8- to 10-fold greater than wild type cells (Gupta, 1981).

These mutants were subjected to cross-resistance studies to other MT inhibitors and to unrelated drugs (Gupta,1981; Gupta et al., 1982; Gupta, 1983). The Pod^{RI} mutants showed increased resistance to the MT inhibitors colchicine, colcemid, steganacin, and slight cross-resistance to nocodazole (Gupta et al., 1982; Gupta, 1983). These mutants showed increased sensitivity to taxol and maytansine and were unchanged in their resistance to vinblastine, griseofulvin, and the unrelated drug puromycin. The second step mutants showed similar cross-resistance patterns except that they showed much increased resistance to nocodazole and enhanced sensitivity to colchicine and its derivatives. These results suggested that colchicine and podophyllotoxin do not bind to identical spots on the tubulin dimer (Gupta et al., 1982). Also, both Pod^{RI} and Pod^{RII} mutants showed increased cross-resistance to various podophyllotoxin derivatives that show MT inhibitory activity (eg. deoxypodophyllotoxin), but not to some of the other derivatives which inhibit DNA synthesis (eg. VM-26, and VP-16-213). Finally, somatic cell hybrids were constructed between mutants and sensitive cells. These cells showed a level of resistance that was in between that of the sensitive and the mutant cells (Gupta, 1981). This indicated that the drug-resistant

phenotype was codominant.

Binding studies were performed with [^{3}H]podophyllotoxin on cell extracts from both mutant cell types and the WT cell line. The amount of [^{3}H]podophyllotoxin bound was normalized with respect to a constant amount of [^{3}H]colchicine bound. Four experiments each showed a slightly reduced binding of podophyllotoxin in the mutant lines. The Pod^{RI} mutants averaged a 12% reduction whereas the Pod^{RII} mutants were reduced by almost 50% in their binding (Gupta, 1983). The results of this experiment suggested that the genetic lesion somehow affects the binding of podophyllotoxin to its cellular target site.

The mutants were next examined by two-dimensional gel electrophoresis to see if there were alterations of tubulin or other MT related proteins. The Pod^{RI} mutants proved to have an identical 2-D gel pattern as the WT cells (Gupta et al., 1982). This suggests that the mutation is not one that causes a change in charge or size of the protein. The Pod^{RII} mutants were much more interesting. Although they did not show alterations of the tubulin proteins there was an alteration of a 63 kDa protein, designated P1, which gave rise to a new, more acidic protein spot (Gupta et al., 1982). This spot was designated M1 and was present in a large number (23/26) of independently selected Pod^{RII} mutants (Gupta et al., 1982). To confirm that M1 had

originated from P1 the spots were excised from gels. The proteins were partially digested with Staphylococcus V8 protease and separated by one-dimensional gel electrophoresis. The only difference found was a high molecular weight band in M1 that appeared as two lower bands in P1 (Gupta et al., 1982). This suggests that the mutation in M1 alters a protease sensitive site.

The Pod^{RII} cells contained both P1 and M1 protein which suggests that only one copy of the gene for P1 is mutated and also a codominant nature of expression. Hybrids formed between Pod^{RII} and WT cells also suggested codominance from the intermediate resistance value recorded. To examine this, 2-D gels were analyzed on the cell hybrids and found to contain a 3:1 ratio of P1 to M1 as expected (Gupta et al., 1982).

The alteration of a protein other than tubulin was unexpected and two types of biochemical experiments were performed to determine if it was MT-related. In the first type of experiment, a crude MT-containing fraction was prepared from the various cell lines under conditions which stabilized MTs (Gupta et al., 1982). These fractions were incubated in the cold, which depolymerized MTs and released the MT proteins. The MT proteins were then separated by centrifugation in the cold and analyzed by 2-D gel electrophoresis. Results of this experiment showed that the tubulins, actin, P1, and a number of other proteins

were released. In addition to the above proteins, the Pod^{RII} cell line showed the release of M1 (Gupta et al., 1982).

The second type of experiment involved the preparation of cytoskeletons (Gupta et al., 1982). Cytoskeletons were prepared from cells by treatment with NP-40 under conditions which preserve MTs. The cytoskeletons were then extracted with buffers containing $5x10^{-3}$ M Ca^{2+} to depolymerize MTs and release the MT proteins. Results of this experiment were essentially the same as the first experiment. The mutant form of the P1 protein was released from the Pod^{RII} cell line (Gupta et al., 1982).

Attempts were also made to obtain revertants from Pod^{RII} cells. Since Pod^{RII} cells are more sensitive to colchicine the rationale was that these cells could be reverted to show podophyllotoxin sensitivity by selection with colchicine. Several colchicine mutants were obtained in this manner (Gupta and Gupta, 1984). However, these mutants were not only resistant to colchicine but also more resistant to podophyllotoxin. The cell extracts of these cells contained not only P1 and M1 proteins, but an additional protein, M2, which has been shown to be related to a neighbouring protein, P2 (Gupta and Gupta, 1984). This protein has also been determined to be MT-related (Gupta and Gupta, 1984; Gupta et al., 1985).

To elucidate the cellular role and distribution of P1, antibodies were raised to the eluted protein in rabbits (Gupta et al., 1985; Gupta and Venner, 1986). The antibodies were shown to be specific for P1 and M1 by 2-D immunoblots of WT and Pod^{RII} cells (Gupta et al., 1985; Gupta and Venner, 1986). The antibody preparation was also used to show the species distribution of this protein. The antibody showed cross-reactivity with a protein of similar size in cells of many vertebrate and invertebrate species including human, monkey, mouse, rat, dolphin, chicken, bullfrog, snake, and mosquito (Gupta and Dudani, 1987). However, no cross-reacting activity was observed in plants (corn and mung), yeast, fungi, amoebae, and bacteria (Gupta and Dudani, 1987). The antibodies also proved useful in determining the specific localization and distribution of the P1 protein. By the use of indirect immunofluorescence it was observed that anti-P1 antibodies bound to structures that was later determined to be mitochondria by the use of the dye Rhodamine 123 (Gupta et al., 1985; Gupta and Venner, 1986; Gupta and Dudani, 1987). Double immunofluorescence staining with P1 and tubulin antibodies showed that MTs and mitochondria had distributed in a similar type of pattern in interphase cells (Gupta et al., 1985; Gupta and Venner, 1986; Gupta and Dudani, 1987).

To investigate the location of the P1 protein within mitochondria, subfractionation studies of rat liver

mitochondria were performed (Gupta and Austin, 1987). These studies indicated that P1 protein was mainly associated with the matrix fraction. Further proof of this was obtained when the cells were treated with the K^+ ionophores nonactin and valinomycin. These compounds abolish the membrane potential and thereby prevent the import of proteins into the mitochondria. Proteins destined for import remain in the cytoplasm as larger precursor proteins. The P1 protein appeared as a larger precursor protein, approximately 3000 daltons larger than the mature form, after this treatment (Gupta and Austin, 1987). To further understand the cellular role of the P1 protein it was necessary to isolate the gene which coded for this protein.

Initially, a human promyelocytic leukemia (HL-60) cDNA expression library was screened using the antibody previously prepared. Prior to screening, the antibody was preadsorbed to E. coli proteins to remove any cross-reacting antibodies that may be present. Eight clones were identified and purified, the largest of which was 1.4 kbp in length. This clone was designated P1-22a (Jindal et al., 1989). Lysogenization of this clone gave rise to a beta-galactosidase fusion protein of approximately 150,000 daltons, with a P1 specific portion of 35,000 daltons. The clone was confirmed to be P1 specific by using the fusion protein to affinity purify the P1 antibody from the

original polyclonal serum. These affinity purified antibodies cross-reacted to a protein in CHO cells of approximately 63,000 daltons, and bound to mitochondria in immunofluorescence studies (Jindal et al., 1989). This gene hybridized to a transcript of appoximately 2.4 kbp from both human and CHO cells and was determined to be at the 3' end of the gene by dideoxy sequencing (Jindal et al., 1989).

This clone was then used concomitantly in the screening of another HL-60 library and for the work included in this thesis. While this work was being completed the sequence of the human P1 gene was determined. The human gene for P1 encodes a protein of 573 amino acids with a mitochondrial targetting sequence of 26 amino acids. Comparisons of the P1 sequence with known sequences revealed a high degree of similarity to a class of proteins known as molecular chaperones (Jindal et al., 1989).

1.5 Molecular Chaperones

It has been well established that the primary sequence of a protein is sufficient for its self-assembly into a functional unit. This folding is a result of many non-covalent forces that cause the protein to adopt a structure which is thermodynamically favoured (see Stryer, 1988). The first demonstration of this principle was the denaturation/renaturation experiments with ribonuclease A (Anfinsen, 1973).

In the case of some complex oligomers the structural information of the primary sequence is not sufficient for self- assembly. These protein complexes require additional proteins to assist in the correct post-translational folding event. This group of proteins has been named molecular chaperones.

A functional definition of a molecular chaperone is a protein which prevents and disrupts improper interactions between potentially complementary surfaces (Ellis et al., 1989). This is most notable in the disruption and prevention of protein aggregates and their transient role in the assembly of protein oligomers.

Since correct folding of proteins is fundamental to every cell these molecular chaperones are thought to be ubiquitous (Hemmingsen et al., 1988). With many proteins now being described as chaperones, they have recently been subdivided into three classes (Ellis et al., 1989). Members from each class are related by amino acid homologies and are distinct from the other two classes. The three classes are the nucleoplasmins, the BiP group, and the chaperonins. The P1 protein is a member of the chaperonin class which will be discussed in greater detail than the other two classes.

1.5.1 Nucleoplasmins

Nucleoplasmin and the concept of molecular chaperones were first described in 1978 by Laskey et al. while

studying nucleosome assembly (Laskey et al., 1978). Laskey and coworkers found that adding a small amount of Xenopus laevis egg homogenate to purified Xenopus DNA and histones increased the efficiency of nucleosome assembly. This was most pronounced at physiological ionic strength where normally DNA and histones precipitate after rapid mixing. However, when the Xenopus cell extract was added, nucleosomes formed. The protein was purified from the homogenate using a sucrose gradient and assayed for the ability to insert superhelical turns into relaxed SV40 circular DNA (Laskey et al., 1978).

The purified protein fraction retaining nucleosome assembly activity consisted of one major protein, with a molecular weight of 29,000 daltons. The protein is acidic with a pI of approximately 5.0. The sedimentation and gel filtration properties suggested that this protein consisted of at least four subunits in the native state (Laskey et al., 1978). This was later determined to be a pentamer in solution (Earnshaw et al., 1980).

Further studies showed that nucleoplasmin binds neither DNA nor histones and does not form part of the nucleosome. However, it will assemble nucleosome cores in vitro and it interacts with histones such that they no longer adhere to negatively charged surfaces (Earnshaw et al., 1980; Laskey et al., 1978). They were also unable to find evidence for sterically specific interactions with a

particular histone. In addition, the protein was found to comprise 7.5 to 10% of the total protein in the nucleoplasm and an antigenically related protein has been found in the nucleoplasm of a wide range of vertebrate cell types (Earnshaw et al., 1980).

Laskey suggested from this data that this protein acts as a molecular chaperone (Earnshaw et al., 1980). This seems an appropriate suggestion since it appears that nucleoplasmin promotes histone-histone interactions by minimizing the interactions of the histones to DNA. This allows for the correct formation of nucleosome cores which then bind to DNA while nucleoplasmin does not become a part of the final structure. Also it appears to be abundant, ubiquitous and a homo-oligomeric protein of five subunits.

A protein immunologically related to nucleoplasmin, nucleoplasmin S, has now been found in the nuclei of somatic cells from animals. This protein has a similar role to nucleoplasmin (Ellis et al., 1989).

1.5.2 The BiP Group

The BiP group contains the major heat shock protein, hsp 70, the 78 kDa glucose regulated protein (GRP-78), and the immunoglobulin heavy chain binding protein known as BiP. Cloning of these three proteins has shown that GRP-78 and BiP are identical and are related to the evolutionary conserved HSP-70 protein family (Munro and Pelham, 1986).

Analysis of pre-B cells and hybridomas derived from

these cells led to the identification of a 78 kDa protein that immunoprecipitated with Ig heavy chains (Bole et al., 1986; Haas and Wabl, 1983). This immunoglobulin binding protein (BiP) bound non-covalently to the heavy chains in a 1:1 ratio and was localized to the microsomal fraction after subcellular fractionation (Haas and Wabl, 1983). This group also found that BiP bound noncovalently to the heavy chains in a 1:1 ratio, but not at all when kappa-light chains were present. Bole et al. (1986) extended this work to show that BiP binds incompletely assembled Ig molecules in secreting cell lines. They also found that when they treated cells with tunicamycin, to inhibit glycosylation, there was an increase in the BiP-heavy chain complex. Similarly, Shiu and coworkers induced a 78 kDa protein, as well as a 95 kDa protein, when they inhibited N-linked glycosylation by glucose starvation (Shiu et al., 1977). These proteins were localized to the endoplasmic reticulum and called glucose regulated proteins (GRPs). The sequencing of both BiP and GRP78 has determined that they are identical and closely related to hsp70 (Munro and Pelham, 1986). The authors suggest that this protein may prevent aggregation of newly synthesized heavy chains in the endoplasmic reticulum. It would then be able to release itself from the heavy chain by hydrolyzing ATP. Since this protein is induced by inhibitors of glycosylation they suggest that it prevents all abnormal

proteins from exiting the endoplasmic reticulum (Munro and Pelham, 1986). This is very similar to the function proposed for hsp70 in heat-shocked cells. Hsp70 travels to the nucleus and nucleoli where it is thought to solubilize aggregated, heat-damaged proteins in an ATP-dpendent manner (Lewis and Pelham, 1985).

It has also been speculated that hsc70 (heat shock cognate) may play a role in protein folding and assembly by disrupting inappropriate protein-protein interactions, and preventing aggregation problems while oligomeric complexes assemble (Pelham, 1986).

Currently four members of the human hsp70 protein family have been identified (Lindquist and Craig, 1988). These proteins are listed in Table 3, which also shows other nomenclature that has been used in the literature. Aside from their roles in heat shock and the endoplasmic reticulum, other functions include disrupting clathrin-clathrin interactions, thereby dismantling coated vesicles in the cell (Pelham, 1986); and a role in DNA replication through studies on the bacterial hsp70 homolog, dnaK (Bardwell and Craig, 1984; Lindquist and Craig, 1988).

The dnaK protein is essential for lambda DNA replication. Biochemical experiments have elucidated the role of this protein in the phage DNA replication process. At the origin of replication six proteins are needed for unwinding the DNA duplex, prior to the binding of DnaG

Protein[1] name	Other name(s)	pI	Regulation
hsp70	72K[2]; hsx70[3], SP71[4]; hsp68[5]	5.8-6.3	Major heat-inducible 70K; basal expression; serum stimulated; cell cycle regulated; E1a inducible
hsp72	hsp70	5.6-5.8	No basal expression; heat inducible
p72[6]	73K[2]; hsc70[5] hsc73[3]	5.5-5.6	High basal expression; slightly heat inducible
grp78	BIP; hsp80	5.2-5.3	High basal expression (especially in secretory cells; expression enhanced by glucose deprivation, calcium ionophores, glycosolation inhibitors, etc.

1. Using the nomenclature of Watowich & Morimoto (234) 2. Welch & Feramisco (236); 3. Pelham (165); 4. White & Currie (241); 5. Lowe & Moran (130); 6. a major heat-inducible protein in primates, but not found in rodents.

Table 3 Proteins of the hsp70 family.
(taken from Lindquist and Craig, 1988)

primase. Two phage proteins and one host protein (a helicase) form at the origin of replication. After binding of the other three host proteins to the complex it is thought that DnaK loosens the association between one phage protein and the helicase to allow unwinding to begin (Lindquist and Craig, 1988).

DnaK has been found to be essential for growth of normal bacterial cells at high (42°C) and low (16°C) temperatures, but not at 30°C (Bukau and Walker, 1989). At 37°C the protein is still highly expressed although the function is not well understood. At high temperatures it is induced as part of the heat shock response, has an ATPase activity, is capable of autophosphorylation at one or more threonine residues, and is involved in the phosphorylation of aminoacyl-tRNA synthetases (Bukau and Walker, 1989). Mutants of dnaK show an inhibiton in DNA and RNA synthesis, and a block in cell division when grown at the non-permissive temperature. From these observations they concluded that dnaK has multiple functions that change with growth temperature (Bukau and Walker, 1989). An additional study with dnaK mutants and other E. coli heat shock genes indicate that one general function of these proteins is to recognize abnormal polypeptides so they can be degraded (Straus et al., 1988).

In yeast there are at least nine genes related to hsp70 of higher eucaryotes (Lindquist and Craig, 1988).

Four of these genes SSA1, SSA2, SSA3, and SSA4 are related on the basis of structure and function. All four are expressed in the unstressed cells, yet each reacts differently to heat shock. SSA3 and SSA4 are expressed at very low levels in the unstressed cells but increase dramatically at higher temperatures. SSA2 expression is quite steady, showing only a small increase at higher temperatures. SSA1 is abundant during normal growth and increases 3-10 times on an upshift in temperature (Lindquist and Craig, 1988). Yeast strains with mutations in a single gene are indistinguishable from the wild type strain; an SSA1, SSA2 double mutant grows slowly at 24°C and not at all at 37°C; and an SSA1, SSA2, SSA4 triple mutant is inviable (Deshaies et al., 1988a). This triple mutant can be rescued by inserting SSA1 on a single copy plasmid under control of the GAL1 promoter (Deshaies et al., 1988a).

Recently two independent groups utilizing yeast as a model system have shown that hsp70 is involved in translocating proteins across membranes (Chirico et al., 1988; Deshaies et al., 1988b). Chirico et al. (1988) used a cell free translocation system to examine the uptake of yeast prepro-alpha-factor which they synthesized in a wheat-germ extract. They purified a factor that stimulated the uptake of prepro-alpha-factor into microsomal vesicles. Further characterization showed that the purified factor

was composed of two closely related proteins of 70,000 daltons. Both proteins cross-reacted with a monoclonal antibody against Drosophila hsp70. The proteins were shown to be the SSA1 and SSA2 gene products of yeast. They confirmed this by illustrating that these two proteins increased translocation efficiency (Chirico et al., 1988). From their data the authors suggest that hsp70 is responsible for unfolding prepro-alpha-factor for translocation or to break up aggregates of the protein to facilitate translocation (Chirico et al., 1988).

Deshaies and coworkers had a similar objective but they utilized an in vivo yeast system (Deshaies et al., 1988b). They used a yeast strain with a triple hsp70 mutation (SSA1, SSA2, SSA4) harboring a plasmid containing SSA1 under control of the GAL1 promoter. The strain was grown in the presence of galactose, after which, glucose was added to repress the GAL1 promoter in order to examine the effect of hsp70 depletion on proteins targeted to different regions of the cell. They found an accumulation of prepro-alpha-factor, and the precursor of the beta-subunit of the mitochondrial F_1 ATPase (Deshaies et al., 1988b). They also observed a slight accumulation of the vacuolar enzyme carboxypeptidase Y and invertase, a secreted enzyme. From their results they concluded that hsp70 is involved in importing proteins into the endoplasmic reticulum and the mitochondria by acting as an

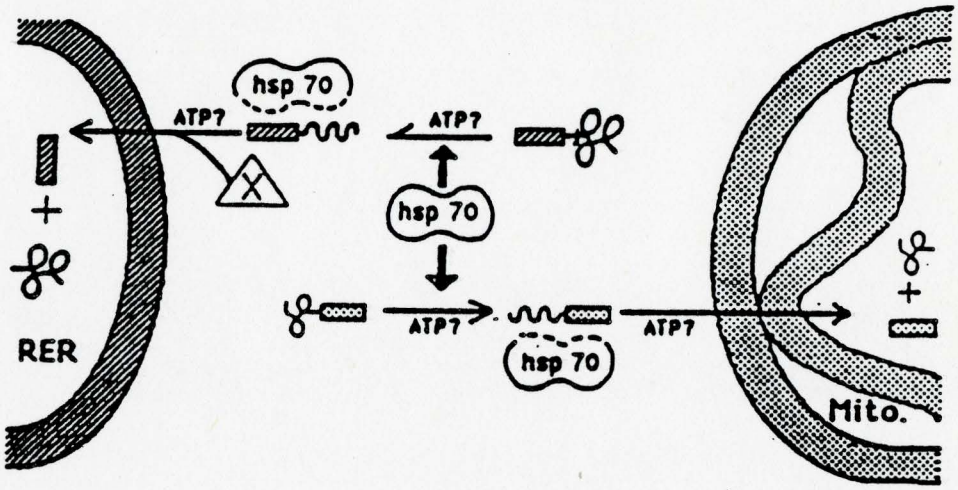


Figure 3 **A model of hsp70 function in protein translocation.** One or more hsp70 molecules may bind to proteins destined for the ER and mitochondria. Binding may occur before or after completion of translation. Hsp70 prepares the protein for import which may involve the hydrolysis of ATP. However, ATP hydrolysis may facilitate dissociation of hsp70 from the protein (taken from Deshaies et al., 1988).

unfoldase. A schematic diagram of this model is presented in Figure 3. They also suggested that carboxypeptidase Y and invertase did not accumulate as greatly because they are secreted into the ER co-translationally (Deshaies et al., 1988b).

Other proteins that can be grouped into the BiP class of chaperones are the hsp90 proteins. They are thought to play similar functions in the cytoplasm in their association with steroid-hormone receptors and tyrosine kinases (Lindquist and Craig, 1988). The BiP class of proteins can be summarized as stress proteins that are induced during periods of stress to prevent the aggregation of proteins and to help restore functioning of proteins. They also appear to target proteins to destinations within the cell by unfolding them for translocation. Finally they seem to play a role in the recognition of abnormal proteins and the regulation of others both in the endoplasmic reticulum and the cytoplasm. An interesting finding in this laboratory was that the P2 protein (and its mutant form M2), identified as a MT-related protein, is an hsp70 homolog.

1.5.3 Chaperonins

The third class of molecular chaperones are a group of proteins found in mitochondria, chloroplasts, and prokaryotes which appear to be involved in the assembly of oligomeric protein complexes. This group includes the

GroEL protein of Escherichia coli, and the ribulose bisphosphate carboxylase-oxygenase binding protein. The 65 kDa major antigenic protein of mycobacteria, the bacterial common antigen, and the P1 protein, isolated in our laboratory, have been included in this class of proteins based on sequence analysis.

The best studied protein of this class is the GroEL protein of E. coli. The GroEL gene is one of two genes in the GroE operon. The operon was first identified when mutants were obtained that were unable to assemble the heads of bacteriophages lambda and T4 (Georgopolous et al., 1973). It was also found to be necessary for the assembly of bacteriophage T5 tails (Ellis et al., 1989). The GroE operon consists of two genes, GroEL and GroES, which encode proteins of 65,000 and 15,000 daltons, respectively. The operon is under heat shock control and the proteins are among the most abundant in the cell, as well as being essential for cell viability (Hemmingsen et al., 1988).

Both the GroEL and GroES proteins are present in the cell as oligomeric structures. GroEL is present as a 14 subunit homo-oligomer in a double ring, each ring containing seven subunits (Ellis et al., 1989; Hendrix, 1979). The GroES protein resides in the cell as a single ring comprising 6-8 subunits (Ellis et al., 1989; Hemmingsen et al., 1988). The GroEL protein is an ATPase and forms a complex with GroES in the presence of Mg^{2+} and

ATP (Chandrasekhar et al., 1986). Further evidence was obtained when GroES was shown to bind to a GroEL-affinity column (Chandrasekhar et al., 1986). Sequencing of the GroE operon showed two open reading frames of approximately 10,000 bp and 58,000 bp, separated by 42 bp (Hemmingsen et al., 1988). Northern analysis showed that both genes hybridized to a single transcript of approximately 2.1 kb and are thus co-transcribed (Hemmingsen et al., 1988). The transcript has been shown to increase up to five-fold upon heat shock treatment, while protein levels increase from about 1% of total protein synthesis to 10% of total synthesis (Ellis et al., 1989; Hemmingsen et al., 1988).

In the formation of bacteriophage heads GroEL has been implicated at the earliest stage. GroEL complex helps assemble glycoprotein B into a 12 subunit oligomer. This oligomer forms the link between the head and tail portions of the bacteriophage and is the site for head-shell assembly (Hemmingsen et al., 1988). Cells with mutations in the GroE operon contain amorphous lumps of the major head protein on the cell membrane (Hemmingsen et al., 1988). The GroE complex does not form part of the assembled bacteriophage.

The role of GroEL in the normal cell is unclear at present. However, evidence is accumulating that suggests a role for GroEL in DNA replication. Temperature-sensitive GroEL and GroES mutants show an inhibition of both cellular

DNA and RNA synthesis at the nonpermissive temperature (Wada and Itikawa, 1984). Also, when GroEL and GroES are inserted into a plasmid and overproduced, they suppress mutations in the dnaA gene (Ellis et al., 1989; Ruben et al., 1988). DnaA is a gene essential for the initiation of DNA synthesis in E. coli (Lindquist and Craig, 1988). Finally, when spontaneous revertants of the single stranded binding protein (ssb) were examined, two types were found. One type showed an overexpression of the ssb gene, the other an overexpression of the GroEL gene (Ruben et al., 1988). This data suggests that GroEL interacts with the ssb gene which encodes another protein involved in DNA replication.

Recently, there has been a report that identified GroEL as the major cytosolic component that binds unfolded proteins and assists in their correct assembly (Bochkareva et al., 1988). This group used a photo-crosslinking approach in an in vitro translation system. They photolabelled the tRNA synthetase carrying the initiator methionine so that proteins would carry the photolabel at the N-terminal formyl methionine. They used an E. coli cell extract and a transcript from a plasmid containing either the pre-beta-lactamase gene or the chloramphenicol acetyltransferase gene of bacteria. After translation the mixture was irradiated to cross-link the newly synthesized protein to one of the extract proteins. They then

sedimented this complex and determined that GroEL was the main acceptor of the newly synthesized protein (Bochkareva et al., 1988). This paper also reported that interactions with GroEL could be abolished by ATP hydrolysis but not with ATP analogs. This data suggests that GroEL is not only involved in DNA replication but that it also plays a role in the folding and assembly, or in the secretion pathways of bacterial cells.

Another well studied protein in this group is the binding protein of the large subunit of the enzyme ribulose bisphosphate carboxylase-oxygenase (Rubisco). Rubisco, found in the stroma of chloroplasts, catalyzes the addition of CO_2 to ribulose 1,5-bisphosphate (Darnell et al., 1986). This is the first step of the Calvin cycle in higher plants. The addition of CO_2 to ribulose 1,5-bisphosphate generates two molecules of 3-phosphoglycerate. One-sixth of all 3-phosphoglycerate molecules are transported to the cytosol where they are converted to glucose. The remaining molecules are recycled into ribulose 1,5-bisphosphate. The enzyme also catalyzes a reaction with O_2 and ribulose 1,5-bisphosphate. This gives rise to one molecule of 3-phosphoglycerate and one molecule of phosphoglycolate. The molecule of phosphoglycolate is hydrolysed to glycolate, which is transported to peroxisomes and oxidized to CO_2 (Darnell et al., 1986). This latter reaction is termed photorespiration and is wasteful for plants because plant

productivity is limited by CO_2 fixation. Thus, it was of interest to study Rubisco to increase suppression of the oxygenase reaction, thereby enhancing photosynthetic rates in important crop plants (Roy and Cannon, 1988).

In higher plants, Rubisco is an oligomer consisting of eight large and eight small subunits. In photosynthetic prokaryotes, a similar complex is observed as well as a dimer of two large subunits in some species (Goloubinoff et al., 1989). The large subunits are encoded in the chloroplast DNA while the small subunits are nuclear encoded. The small subunits contain a signal sequence for the chloroplast which is cleaved upon entry into the organelle (Ellis et al., 1989; Hemmingsen et al., 1988). Both subunits associate transiently with the Rubisco binding protein (Ellis and Van Der Vies, I988; Ellis et al., 1989). This is depicted in Figure 4.

The binding protein was first discovered when isolated intact chloroplasts were incubated with K^+ ions, ^{35}S-methionine and light as an energy source. The chloroplasts synthesized a large number of polypeptides, the most abundant being the Rubisco large subunit (Ellis and Van Der Vies, 1988). However, time course studies showed that synthesis was much faster than assembly into the holoenzyme. They also observed that the large subunits migrated on polyacrylamide gels with a protein with a molecular weight of about 720,000 daltons (Ellis and Van

Der Vies, 1988; Roy and Cannon, 1989). Initially it was thought that this high molecular weight form was a result of aggregation of large subunits. However, when this protein band was excised and electrophoresed under denaturing conditions it migrated as a 60 kDa protein, somewhat larger than the large subunit (52 kDa). Partial digestion of this protein by proteases showed no relation to the large subunit of Rubisco (Ellis and Van Der Vies, 1988). It is now known that the binding protein consists

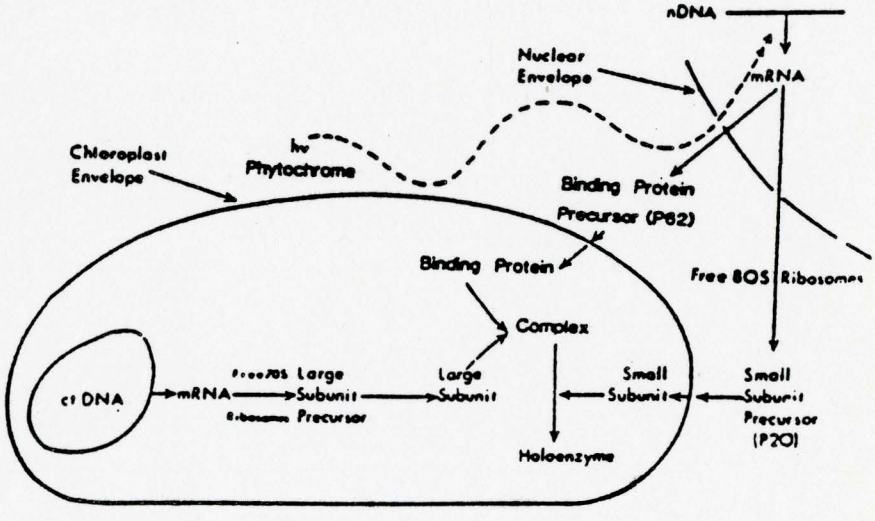


Figure 4 **Assembly of Rubisco in higher plants.** Rubisco large subunits are synthesized from chloroplast genes and assembled with small subunits by the Rubisco binding protein. Both the binding protein and the small subunits are synthesized from nuclear genes (taken from Ellis, 1987).

of two types of subunit of apparent molecular mass 61,000 daltons (alpha) and 60,000 daltons (beta). The two subunits are present in equal amounts and the binding protein likely has the composition $alpha_6beta_6$ (Ellis and Van Der Vies, 1988).

The subunits are encoded by nuclear genes and are imported into the chloroplast after synthesis on cytosolic ribosomes. The oligomer dissociates reversibly in the presence of Mg-ATP, and has been shown to have some ATPase activity (Hemmingsen et al., 1988). Clones encoding castor-bean (Ricinus communis) and wheat (Triticum aestivum) binding proteins were isolated from cDNA libraries (Hemmingsen et al., 1988). These clones, encoding the alpha subunit, contain a single open reading frame encoding a polypeptide of approximately 58,000 daltons. The protein sequence showed similarity to the GroEL and the 65-kDa mycobacterium protein (Hemmingsen et al., 1988).

Evidence that Rubisco binding protein is needed for correct assembly of Rubisco in higher plants comes from immunological studies. Experiments with stromal extracts showed that there was an inhibition in the formation of large subunits into the holoenzyme when antiserum to the binding protein was included in the mixture (Ellis and Van Der Vies, 1988; Roy and Cannon, 1989). It has also been shown that purified cyanobacterial Rubisco can reassemble

in vitro into a functional enzyme following dissociation into subunits, whereas Rubisco purified from higher plants does not retain this property (Ellis *et al.*, 1989). Similar observations have been made *in vivo*. The cyanobacterial Rubisco genes can be expressed and the proteins assembled into a functional enzyme in *E. coli*, but not the Rubisco gene products from higher plants. Recently, Goloubinoff and coworkers have shown that prokaryotic Rubisco assembly in *E. coli* utilizes the GroE gene products (Goloubinoff *et al.*, 1989). The GroE operon was placed in a multicopy plasmid which increased the level of GroEL and GroES proteins from 1-2% to 10%. This enhanced level of the GroE proteins increased the activity of Rubisco. This activity was abolished, as was the detection of holoenzyme, when mutations were introduced into the GroE genes (Goloubinoff *et al.*, 1989). They showed that the GroE genes are essential for large subunit dimer formation, and could not be excluded in the formation of tetramers and octamers. They also found that the GroE genes were not necessary for the addition of small subunits to the large subunit octamer. They could not exclude the possibility that the GroE gene products are necessary for small subunit folding prior to oligomerization (Goloubinoff *et al.*, 1989). Since the GroEL protein and the Rubisco binding protein are related, it seems plausible that Rubisco from higher plants should be assembled when the

GroE genes are overexpressed. It also seems that the Rubisco binding protein is essential for the correct assembly of this enzyme complex. It may also play a role in assisting other protein complexes within the chloroplast to assemble.

The third member, based solely on sequence similarity, is the 65 kDa antigen of mycobacteria. The mycobacteria are a collection of gram-positive bacteria, some of which cause lethal human diseases (Shinnick, 1987). The two most common diseases caused by mycobacterium species in humans are tuberculosis and leprosy which result from infection with M. tuberculosis and M. leprae, respectively. These diseases are still prevalent throughout the world and thus have elicited a large amount of research into the understanding of the immune response to these infectious bacteria so a better vaccine can be developed.

Currently, M. bovis BCG is used as a vaccine against tuberculosis, and it also provides immunological protection against leprosy (Shinnick et al., 1987). The strain of M. bovis used for vaccination is avirulent to humans and is closely related to M. tuberculosis and M. leprae. The vaccine has a wide range of efficiency, as results show that it provides from 0 to 80% protection from tuberculosis, and 20 to 80% protection against leprosy (Shinnick et al., 1987). It is believed that immune

responses to antigens shared by the three strains causes the resistance to the diseases.

Animal studies have indicated that resistance to mycobacteria is a cell-mediated process and that antibodies alone contribute little to resistance (Boom et al., 1987). Macrophages elicit the immune response by providing surface antigens for recognition by T-cells. Macrophages, however, are also the major target of the pathogenic mycobacterium species for growth and proliferation (Shinnick, 1987). In order to develop better vaccines and skin tests the major antigens of mycobacterium species have been examined.

More than fifty monoclonal antibodies have been raised to mycobacterial antigens (Britton et al., 1985; Buchanan et al., 1987; Shinnick et al., 1987; Young et al., 1985a). These antibodies recognized protein and carbohydrate antigens shared by the three species (mentioned above), as well as antigens found only in M. leprae or M. tuberculosis (Shinnick et al., 1987). Fourteen of the antibodies recognized proteins in M. leprae with sizes of 12, 18, 28, 36, 65, and 70 kilodaltons, and 19 recognized proteins in M. tuberculosis with sizes of 10, 14, 19, 23, 38, 65, and 71 kilodaltons. One monoclonal antibody reacted with both the M. leprae 28 kDa protein and the M. tuberculosis 23 kDa protein. Similarly, one monoclonal antibody reacted to the two 70 kDa species.

Interestingly, of all 33 antibodies 10 reacted with the 65 kDa antigen and seven of those recognized the 65 kDa antigen from all three species. These three proteins are thought to be the major antigenic proteins in mycobacteria with the 65 kDa protein being the most antigenic. In fact, in patients affected with either tuberculosis or leprosy, researchers have been able to isolate antibodies and T-cell clones towards the 65 kDa antigen (Buchanan et al., 1987).

These monoclonal antibodies have also been used to clone the genes for the 65 kDa protein from different strains of mycobacteria. Genomic DNA from M. tuberculosis was fragmented and cloned into λgt11, an expression vector (Shinnick, 1987). The library was screened using a mixture of three monoclonal antibodies to the 65 kDa antigen. A number of overlapping clones were used to deduce the nucleotide sequence of 4,380 base pairs. The clones contained two long open reading frames, one of 540 amino acids preceding one of 517 amino acids. By fusing the 540 amino acid open reading frame to the lacZ gene it was determined that this fragment coded for the protein (Shinnick, 1987). A similar library was produced from M. leprae and screened with a mixture of seven monoclonal antibodies to the 65 kDa antigen to clone this protein from M. leprae (Young et al., 1985b). This group used the same library to find clones to the 36, 28, 18 and 12 kDa proteins as well (Young et al., 1985b). Similarly, this

approach was also used to clone this gene from M. bovis BCG (Thole et al., 1987). Comparison of the amino acid sequences for these genes show that the M. tuberculosis and the M. bovis BCG proteins are identical and they show greater than 95% homology to the M. leprae protein (Lu et al., 1987; Shinnick et al., 1987).

Further studies with polyclonal and monoclonal antibodies to the 65-kilodalton antigen showed that they cross-reacted to a variety of bacteria (Thole et al., 1988). The protein recognized by the 65 kDa antisera ranged in size from 59 to 65 kDa and is referred to as the common antigen due its presence in a large variety of bacteria (Thole et al., 1988). This bacterial common antigen is found in Gram-positive and Gram-negative eubacteria, archaebacteria, and is the GroEL gene product in E. coli (Shinnick et al., 1988; Thole et al., 1988a; Young et al., 1988a). It was also determined that the 71-kilodalton mycobacterial antigen was related to the E. coli dnaK gene product (Young et al., 1988a). After the sequence was determined for the 65-kilodalton antigen it was used to map epitopes of the various antibodies and T-cell clones (Boom et al., 1987; Mehra et al., 1986; Thole et al., 1988). The approach used was to make a sublibrary containing fragments of the 65-kilodalton gene and then screening this with the monoclonal antibodies (Mehra et al., 1986). These defined regions were then used by other

groups to define T- and B-cell epitopes (Thole et al., 1988). In a different approach Boom et al. (1987) stimulated lymph nodes, from sensitive mice, with both recombinant and purified mycobacterial antigens to obtain T-cell clones, which they then characterized (Boom et al., 1987).

The 65-kDa mycobacterial antigen has been used to develop an adjuvant arthritis model in the rat (van Eden et al., 1988). This disease is inducible in susceptible rats by immunization with heat-killed M. tuberculosis in oil, and is related to rheumatoid arthritis in humans. From such rats a T-cell clone was isolated, that induced arthritis in irradiated recipient rats by passive transfer (Holoshitz et al., 1984). The T-cell clone was found to react with the 65-kDa mycobacterial antigen and cartilage proteoglycans (van Eden et al, 1985). The rat epitope was characterized and shown to react with T-cells from patients with rheumatoid arthritis (van Eden et al., 1988). It has been suggested that if patients are immunized with this epitope it may induce therapeutic suppression of the disease process, similar to the rat model (van Eden et al., 1988).

Recently, other proteins from the chaperonin family have been identified. These include the 58-kilodalton protein from Tetrahymena thermophila (McMullin and Hallberg, 1987), the 60-kilodalton mitochondrial heat shock

protein from yeast (Reading et al., 1989), and the 62-kilodalton antigen of Coxiella burnetii (Vodkin and Williams, 1988).

1.6 Objectives of the Current Study

A large number of mutants of Chinese hamster ovary cells have been selected in this laboratory for resistance to the MT inhibitor podophyllotoxin (Gupta et al., 1982). The second step mutants, Pod^{RII}, contained a specific electrophoretic alteration in a 63 kDa protein, designated P1. This genetic lesion was determined to be related to the cellular action of the drug by several criteria. First, the mutants exhibited highly specific cross-resistance and collateral sensitivity to other MT inhibitors, such as colchicine, nocodazole, and taxol (Gupta et al., 1982). Second, the mutants showed reduced binding of the drug in cell extracts (Gupta et al., 1985). In addition, P1 appears to be MT-related since it is coreleased with tubulin from subcellular fractions of crude MTs or detergent extracted cytoskeletons under conditions which depolymerize MTs (Gupta et al., 1985; Gupta et al., 1982). However, immunofluorescence studies with an antibody raised against P1 and subcellular fractionation of rat liver mitochondria have localized the P1 protein to the matrix of the mitochondria (Gupta and Dudani, 1987; Gupta and Austin, 1987).

The aim of this project was to construct a cDNA

library from CHO cells in order to clone and sequence the P1 protein. Obtaining the sequence from the wild type cells provides a starting point for determining where the mutation lies in P1 that results in the change to M1, in Pod^{RII} cells. The cDNA can be used in an RNA:DNA cleavage assay using RNA from Pod^{RII} cells. This should reveal the approximate location of the mutation. The mutation can then be pinpointed by sequencing the relevant area of the gene. The sequence provides a template for the construction of oligonucleotide primers which are important for sequencing.

The sequence can also be used to determine if additional expression of P1 will enhance resistance to wild type or Pod^{RI} cells. This can be determined by inserting the cDNA for P1 into a mammalian expression vector and transfecting wild type or Pod^{RI} cells. Once the system is stable, resistance studies can be performed to determine if there is a change in resistance.

Therefore, the construction of a cDNA library and sequencing of the P1 gene will provide a valuable tool for future research in this laboratory. It is also an important first step in understanding the cellular role of the P1 protein.

2.0 MATERIALS AND METHODS

2.1 Materials

2.1.1 Source of Chemicals and Reagents

Ampicillin, bromophenol blue, DEPC, dithiothreitol, ethidium bromide, SDS, beta-mercaptoethanol, salmon sperm DNA, Tris-HCl, Tris-base, and xylene cyanol were purchased from Sigma Chemical Co., St.Louis, Moussouri. Bacto-agar, bacto-tryptone, and yeast extract were purchased from Difco Laboratories, Detroit, Michigan. Acrylamide, bis-acrylamide, ammonium persulfate, TEMED, nitrocellulose disks (82.5 mm) and nitrocellulose rolls were purchased from Bio-Rad Laboratories, Mississauga, Ontario. Cesium chloride, guanidium thiocyanate, IPTG, DNA molecular weight markers (1 kb ladder), oligo d(T) cellulose, phenol, and X-Gal were purchased from Bethesda Research Laboratories (BRL), Burlington, Ontario. Alpha-minimal essential medium, fetal calf serum, and tryptone phosphate broth were purchased from Gibco Canada Inc., Burlington, Ontario. Agarose NA, and EcoRI linkers were purchased from Pharmacia P-L Biochemicals, Dorval, Quebec. Kodak XAR-5 and Kodak XK-1 negative film was purchased from Picker International, Ottawa, Ontario. S-Adenosyl methionine, and DNA packaging extracts were purchased from Promega Biotech, Mississauga, Ontario.

2.1.2 Enzymes

Restriction endonucleases Bam HI, Dde I, Eco RI, Kpn I, Pst I, Rsa I, Sma I, Sph I, Sst I, and Xba I were purchased from BRL and used under the conditions provided by the supplier. E. coli exonuclease III, E. coli and T4 DNA polymerase I, Large fragment of E. coli DNA polymerase I (Klenow), T4 polynucleotide kinase, S1 nuclease, and T4 DNA ligase were also purchased from BRL. AMV reverse transcriptase, and ribonuclease A were obtained from Pharmacia P-L Biochemicals. Dam methylase (EcoRI) was purchased from Promega Biotech. Deoxyribonuclease I and lysozyme were purchased from Sigma Chemical Co..

2.1.3 Plasmids and Bacteriophage Vectors

Bacteriophage λgt11 DNA was purchased from Promega Biotech. Plasmids pTZ18U, and pTZ19U were obtained from United Staes Biochemicals (USB). Plasmid P1-22a, is a derivative of pTZ18U containing a 1.4 kb HL-60 cDNA clone of P1 inserted at the EcoRI site. It was provided by Dr. Anil Dudani, of this laboratory. Plasmid P1-C5 contains a 4.4 kbp Kpn I - Sst I insert (containing the complete human P1 cDNA plus some flanking sequences of λgt11) in pTZ18U. Upon Eco RI digestion, it yields two P1-specific fragments of which the larger fragment, 1.4 kbp, is identical to the P1-22a insert. The other fragment is 0.8 kbp and corresponds to the 5'-end of the human P1 gene. This plasmid was

obtained from Dr. Satish Jindal of this laboratory. The λgt10 cDNA library from CHO cells was kindly provided by Dr. Wayne Flintoff from the Department of Microbiology, University of Western Ontario, London, Ontario.

2.1.4 Molecular Biology Kits

DNA probes were initially labelled using a Nick Translation kit, and later with a Random Primer Labelling System, both purchased from BRL. DNA dideoxy sequencing was performed using the Sequenase kit purchased from United States Biochemicals. The cDNA was synthesized using a cDNA synthesis kit purchased from BRL.

2.1.5 Radiochemicals

[Alpha-^{32}P]dATP and [γ-^{32}P]dATP (3000 Ci/mmol, 10 uCi/ul) were purchased from Amersham Radiochemicals Co., Oakville, Ontario.

2.1.6 Bacterial Strains

The E. coli K-12 strain derivatives used were as follows:

C600hfl, genotype C600hflA150 chr::Tn10, was provided by Dr. C.B. Harley, McMaster University;

JM107, genotype endA1, gyrA96, thi, hsdR17, supE44, relA1, Δ(lac-proAB); F′ [traD36, proA+, proB+, lacI^q, lacZ ΔM15], was provided by Dr. K.B. Freeman, McMaster

University;

Y1090, genotype Δlac^{U}_{169}, proA+, Δlon, araD139, strA, supF, trpC22::Tn10; pMC9, was purchased from Promega Biotech.

2.1.7 Cell Culture and Cell Lines

The parental Chinese hamster ovary (CHO) cell line is referred to as wild type (WT) (Gupta et al., 1982). $Col^{R}22a$ is a cell line derived from $Pod^{R}II6$ cells and has been described earlier (Gupta and Gupta, 1984; Gupta et al., 1985). These cell lines were routinely grown as monolayers at 37^{o}C in alpha-minimal essential medium supplemented with 3% fetal calf serum by procedures described earlier (Gupta et al., 1982).

2.2 METHODS

2.2.1 Isolation of RNA from Mammalian Cells

(a) Preparation of Solutions

Distilled water was treated for ribonucleases by adding DEPC to a final concentration of 0.01%. The DEPC-water was then incubated at 37^{o}C for 30 minutes before autoclaving. A 4 M guanidium isothiocyanate solution (50 ml) was prepared by dissolving 11.82 g of guanidium

thiocyanate in 30 ml of DEPC-water. After adjusting the pH to 7.0, 0.625 ml of 1 M sodium citrate (pH 7.0) and 0.35 ml of 2-mercaptoethanol was added. The solution was filter sterilized and then stored at 4°C, for up to 2 weeks. A 5.7 M cesium chloride solution in 0.1 M EDTA (pH 7.5) was prepared fresh and stored on ice until use.

(b) Total RNA Extraction

Total RNA was extracted, from 20-25 nearly confluent 100 mm petri dishes of CHO cell monolayers (approximately 4-5 x 10^7 cells) by a method modified from Chirgwin et al (1979). The cells were washed three times with PBS (0.8% sodium chloride, 0.115% disodium hydrogen orthophosphate, 0.02% potassium chloride, and 0.008% potassium dihydrogen phosphate) then lysed by adding 10 ml of 4 M guanidium isothiocyanate solution for every 10 dishes. The dish was gently swirled until all cells were lysed and the homogenate was transferred to the next dish. After all dishes were lysed, the homogenates were pooled and layered onto a 3 ml cushion of 5.7 M cesium chloride (Beckman SW41 polypropylene tubes, soaked overnight in 1% SDS, 1 mM EDTA). Centrifugation (SW41 rotor, 30,000 rpm) was at room temperature for 16 hours.

Following centrifugation, the supernatant was discarded and the sides of the tube wiped dry. The RNA pellet was redissolved in 0.5% SDS, heated for 5 minutes at 65°C, then

cooled quickly on ice. The supernatant was transferred to a fresh tube and precipitated with ethanol. The pellet was collected by centrifugation in a microcentrifuge at 18,000 rpm. It was then washed once with 70% ethanol, once with absolute ethanol, and air-dried. The sample was redissolved in 250 ul of DEPC-treated water. Before use the RNA concentration of the sample was determined spectrophotometrically by measuring absorbance at 260 and 280 nm.

(c) Isolation of poly $(A)^+$ RNA

A 1.5 ml oligo (dT) cellulose column was prepared in a Pasteur pipet for the separation of polyadenylated RNA from total RNA according to the procedure developed by Aviv and Leder (1972). The only modification to the method was the reapplication of the sample flow through to the column 4 more times to maximize poly $(A)^+$ RNA binding. The eluted RNA was stored at $-70^{o}C$ as an ethanol precipitate, and was redissolved in DEPC-treated water before use.

2.2.2 Construction of the CHO cell cDNA library

(a) First and Second Strand cDNA Synthesis

Double stranded cDNA was constructed using the procedure outlined in a cDNA synthesis kit (Bethesda

Research Laboratories). This method was a modification of the procedure that was first described by Gubler and Hoffman (1983). The synthesis is performed in a single tube and it utilizes the properties of M-MLV reverse transcriptase. These include a short half-life, inefficiency in hydrolyzing mRNA for hairpin loop formation, and endogenous RNase H activity.

Briefly, 5 or 10 ug of poly $(A)^+$ RNA was heated at 65°C for 10 minutes then cooled quickly on ice. This was mixed with, first strand buffer (final concentration: 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, and 10 mM dithiothreitol), 500 uM each of dATP, dCTP, dGTP, and dTTP, and 50 ug/ml oligo (dT) in a final volume of 50 ul. The reaction was incubated at 37°C for one hour after the addition of 2.5 ul of M-MLV reverse transcriptase (10 000 U/ml).

In order to calculate the yield and efficiency of the reaction, a tracer reaction was set up, immediately after adding reverse transcriptase, by removing one-fifth of the first strand reaction to a tube containing 5 uCi [alpha-^{32}P] dATP. The tracer reaction was terminated with the addition of 1 ul of 0.25 M EDTA (pH 8.0), and then the volume was increased to 100 ul with water. Duplicate 5 ul aliquots were spotted on glass fiber filters. One filter was dried under a heat lamp (filter 1). The other filter was washed sequentially with 50 ml of 10% trichloroacetic acid (3

times), water, and absolute ethanol, and then dried (filter 2). Both filters were counted in standard scintillant. The remaining tracer sample was precipitated with ethanol for alkaline agarose electrophoresis (see 2.2.7b).

The second strand was synthesized by E. coli DNA polymerase I (250 U/ml) at 16°C for 2 hours in a reaction mixture with the final concentration of 25 mM Tris-HCl, 100 mM KCl, 5 mM $MgCl_2$, 250 uM of dATP, dCTP, dGTP, and dTTP, RNase H (8.5 U/ml) and 5mM dithiothreitol (final volume of 320 ul). This reaction was stopped by adding 25 ul of 0.25 M EDTA (pH 8.0). The sample was extracted with an equal volume of phenol and then precipitated with ethanol after first removing two 10 ul aliquots for the determination of yield.

The first 10 ul aliquot was diluted to 100 ul with water then 5 ul was spotted on a glass fiber filter and dried (filter 3). The other aliquot was directly spotted onto a filter and washed and dried similar to filter 2 (filter 4). These two filters were also counted and the four values were used to determine the yield and efficiency of the synthesis (see Appendix).

(b) Preparation of Blunt Ends

The cDNA pellet was redissolved in 15 ul of TE buffer in preparation for making the ends blunt. This is a safety measure since the cDNA synthesis should have made the

cDNA blunt-ended already. The method was slightly modified from one described by Maniatis et al (1982). The following were added, in order, to the pellet: 4.5 ul of water, 2 ul of dNTP mix (2 mM each of dATP, dCTP, dGTP, and dTTP), 2.5 ul of 10 x repair buffer (0.5 M Tris-HCl, 0.1 M $MgCl_2$, 50 mM dithiothreitol, and 50% glycerol), and 1 ul of Klenow fragment of DNA polymerase I. The mixture was incubated at 37°C for 30 minutes. An additional 1 ul of dNTP mix was added and the incubation continued for another 30 minutes. After increasing the volume to 0.5 ml, the sample was extracted once with an equal mixture of phenol and chloroform and precipitated with ethanol.

(c) Eco RI Methylation Reaction

The pellet from the blunt-end reaction was redissolved in 18 ul of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and added to 4 ul of 10x Eco RI methylase buffer (supplied by Promega Biotech with Eco RI methylase). S-Adenosyl-L-methionine and BSA were added to a final concentration of 0.5 mM and 0.1 mg/ml, respectively. Eco RI methylase was added (twenty units per microgram of cDNA) and the volume adjusted to 40 ul with water. The reaction mix was incubated for 1.5 hours at 37°C, then at 65°C for 10 minutes. Following a phenol/chloroform extraction, the methylated cDNA was ethanol precipitated. This reaction was modified from the one described by Huynh and coworkers

(Huynh et al., 1985).

(d) Preparation of Linkers

Eco RI decamers were dissolved in TE buffer (pH 7.5) to a final concentration of 2 ug/ul. The linkers were then labelled with T4 polynucleotide kinase as described by Maniatis et al (1982). In a reaction volume of 30 ul there was 3 ug of linkers, 100 uCi [γ-^{32}P] dATP, 10 mM ADP, 3 ul of 10 x kinase buffer (0.5 M Tris-HCl (pH 7.6), 0.1 M $MgCl_2$, 0.1 M DTT, and 10 mM ATP) and 10 units of T4 polynucleotide kinase. The reaction was incubated at 37^{o}C for 1 hour. The phosphorylated linkers were then precipitated with ethanol and redissolved in 6 ul of TE buffer. A two microliter aliquot was removed for analysis.

The 2 ul aliquot was analyzed by polyacrylamide gel electrophoresis. The sample was first self-ligated by incubating the tube at 15^{o}C overnight, following the addition of 5x ligase buffer (BRL) and 1.0 ul T4 DNA ligase (0.5 units). One half of the sample was digested with Eco RI and the other half was saved. Bromophenol blue loading dye was added to both samples before loading on the 10% gel. Electrophoresis was at 100 V until the dye had migrated half way. The gel was then dried and exposed to film overnight.

(e) Addition of Linkers

Phosphorylated linkers and methylated cDNA were

mixed at a 1:1 weight ratio with 1.0 unit of T4 DNA ligase and 5x ligase buffer in a final volume of 35 ul. The reaction mixture was incubated overnight at room temperature, then stopped with 0.5 M EDTA (pH 8.0). Following a phenol/chloroform extraction the sample was ethanol precipitated. The sample was redissolved in 88 ul of water for digestion (see below).

(f) Digestion and Separation of Excess Linkers

A method was devised from those described by Maniatis et al (1982) and Huynh et al (1985). One hundred and twenty units of Eco RI and 15 ul of the manufacturer's recommended 10x buffer (BRL) were added to the resuspended cDNA. The linkers were cleaved by incubating the mixture at 37^{o}C. After 4 hours an additional 30 units of Eco RI was added and the incubation continued for 1 hour. An equal volume of 2x column buffer and one-tenth volume of loading buffer (0.25% bromophenol blue, 50% glycerol) was added before loading the sample on the column.

Enough Sepharose CL-4B was pre-swollen in column buffer to pour a 5 ml bed volume in a disposable column with an inside diameter of 0.4 mm. The column buffer was 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 0.3 M NaCl. The column was washed with 50 ml of column buffer before applying the sample. After loading the sample 5 drop fractions were collected in microcentrifuge tubes until the loading dye

exited the column. These tubes were counted in a scintillation counter to determine the radioactive peaks. Tubes from the first peak were pooled and precipitated with ethanol for ligation.

One microgram of dephosphorylated Eco RI cut λgt11 arms (Promega Biotech) were mixed with the remaining purified linkered cDNA and ligated overnight at 15°C as described in (d).

(g) Phage Packaging of Ligated cDNA and Vector

A packaging extract containing phage heads and tails was thawed on ice. The packaging protocol supplied by the manufacturer was followed. One-half of the ligated cDNA was added and the tube gently mixed by tapping the bottom of the tube. The tube was incubated at room temperature for 2 hours. The phage were diluted by the addition of 0.5 ml of SM medium (0.1 M NaCl, 10mM $MgSO_4$, 50 mM Tris-HCl (pH 7.5), and 0.01% gelatin). After gentle mixing, 25 ul of chloroform was added and the tube vortexed gently. The packaged phage were titred to determine yield and the percent recombinants.

(i) Titering and Amplification of Library

Five, ten or fifteen microliters of phage were added to 200 ul of an overnight culture of Y1090 (or C600hfl) cells, grown in LB broth (10 g NaCl, 10 g Bacto-

Tryptone, 5 g yeast extract per liter) supplemented with 0.2% maltose. The phage were adsorbed to the bacteria by incubation at 37°C for 20 minutes, then added to 3 ml of top agarose (LB broth containing 7.0 g/l of agarose) at 48°C, containing 50 ul of 2% X-Gal and 10 ul 100 mM IPTG, mixed and plated onto agar plates (25 ml of LB broth containing 15 g/l Bacto-agar). The plates were incubated at 42°C for 4-6 hours. The plaques were counted on each dish and an average pfu value, the percentage of recombinants, and the library size was calculated.

Once the phage titre was known the library was amplified. The phage were plated at a concentration of $1x10^4$ pfu per plate. After the plaques appeared, 5 ml of SM medium was added and the plates rotated on an orbital shaker at 4°C for 2 hours as described (Maniatis et al., 1982). The bacteriophage suspension was collected and the plates rinsed with an additional 2 ml of SM medium. Chloroform was added to 5% and incubated at room temperature for 15 minutes before centrifuging (4000g, SS34 rotor, 4°C) for 5 minutes. The supernatant was recovered and chloroform added to 0.3%. This was stored at 4°C indefinitely.

2.2.3 Random Primer Labelling

DNA probes were labelled using a random primer labelling kit (BRL) and a method modified from the protocol

provided with the kit. Approximately 25 ng of probe DNA was mixed with water to a final volume of 20 ul and heat denatured for 5 minutes in a boiling water bath. After quick-cooling on ice 2 ul each of dCTP, dGTP, and dTTP were added, followed by 15 ul of random primer buffer mixture, 3 ul [alpha-^{32}P]dATP (10 uCi/ul; 3000 Ci/mmol), 5 ul of water, and 1 ul Klenow fragment. The reaction mix was incubated at room temperature for 2 hours. Following incubation, 5 ul of stop buffer, 1 ul tRNA (10 mg/ml), 50 ul TE buffer, 10 ul 3M sodium acetate, and 300 ul of absolute ethanol was added. The sample was mixed, chilled at -70^{o}C for 20 minutes and then centrifuged for 15 minutes in a microcentrifuge to pellet the precipitated DNA. After washing once with 70% ethanol, and once with absolute ethanol the pellet was dried and redissolved in 100 ul TE buffer. Duplicate 1 ul aliquots were spotted on DE81 filter paper. One filter was dried, the other was first washed 6x in 0.5 M Na_2HPO_4 for 5 minutes, 2x in water, 2x in absolute ethanol each for 1 minute, and then dried. The filters were counted in a scintillation counter after the addition of 3 ml of toluene containing 4g omnifluor/l to the vials.

2.2.4 Screening of cDNA Libraries

The amplified cDNA library was plated at a concentration of approximately 1 x 10^4 pfu/plate onto 10 LB agar plates and incubated at 42^{o}C. After the appearance of

plaques, the plates were cooled at 4°C for 1-2 hours or sometimes overnight. Nitrocellulose disks were labelled in duplicate (or triplicate) with a ballpoint pen and placed on the plates for 2-3 minutes (first filter) or 7-10 minutes (additional filters). Denaturation, neutralization, and hybridization was as described by Quertermous (1987) and Strauss (1987).

The filters were placed on 3MM paper saturated with 0.2 M NaOH, 1.5 M NaCl, followed by, neutralization on 0.4 M Tris-HCl, pH 7.6, 2x SSC, and finally placement on 2x SSC, each for 2 minutes. Follwing the 2x SSC treatment the filters were air dried for 1 hour, then baked at 80°C for 1.5-2 hours in a vacuum oven.

Pre-hybridization was started after wetting the filters in warmed (65°C) hybridization solution (1% BSA, 1 mM EDTA, 0.5 M $NaHPO_4$, pH 7.2, 7% SDS) and placing in hybridization bags (20 filters/bag, 20 ml hybridization solution). After 1 hour at 65°C the radiolabelled probes and salmon sperm DNA (2 mg) were denatured and added directly to the bag. The hybridization was continued overnight.

Following hybridization, the filters were removed from the bags into 200 ml of wash buffer 1 (0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 5% SDS, room temperature). After a second quick wash in buffer 1 the filters were transferred to wash buffer 2 (1 mM EDTA, 40 mM $NaHPO_4$ (pH 7.2), 1% SDS, 65°C) for several (5-8) quick washes and a final 20 minute

wash at 65°C. The filters were then air dried and exposed to film at -70°C for 1 to 5 days.

Positive clones were identified as plaques which hybridised to duplicate filters. These were then excised from the agar as plugs with a pipette tip that had been trimmed back a little. The plaques were removed into a tube containing 1 ml of SM medium and a drop of chloroform. The tube was rotated at 4°C overnight to diffuse the phage out of the agar plug. The partially purified phage were then plated out again, at a lower density, for a further round of plaque purification. This was repeated until a single plaque could be chosen which when replated showed hybridisation to every plaque on the dish.

2.2.5 Isolation of DNA

(a) Isolation of Phage DNA

Initially, phage DNA was isolated by the plate lysate procedure of Maniatis *et al* (1982). Later preparations were obtained by a similar method described by Ziai *et al* (1988). Phage lysates were prepared by infecting an overnight culture of the appropriate bacteria with 1 x 10^5 plaque-forming units of purified phage and plating it on an LB plate (100 mm diameter). When the bacterial lawn was near confluency, 5 ml of SM medium was poured on the plate and the phage were eluted with gentle shaking at room

temperature for 1 hour. The lysate was collected and bacterial debris was removed by centrifugation (SS34 rotor; 20,000g; 5 minutes at 4°C).

The supernatant was mixed with an equal volume of saturated ammonium sulfate solution and precipitated on ice for 30 minutes prior to centrifugation (20,000g; 20 minutes). The pellet was resuspended in 0.5 ml of TE/0.1% SDS buffer and treated with RNase A and proteinase K, as described. The phage particles were broken open by the addition of NaOH followed 10 minutes later by neutralization with 10 M ammonium acetate (pH 6). The debris was pelleted and the supernatant was precipitated with isopropanol following several phenol/chloroform extractions. The DNA pellet was redissolved in 50 ul of TE buffer.

(b) Small Scale Plasmid Preparation

Plasmid DNA was prepared according to the method of Birnboim and Doly (1979). A single bacterial colony was inoculated into 3 ml of LB broth supplemented with 50 ug/ml ampicillin. The culture was grown overnight at 37°C with vigorous shaking. One-half of the culture was centriuged for one minute in a microcentrifuge. The supernatant was decanted and the pellet was resuspended in 100 ul of lysis buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl (pH 8.0), and 4 mg/ml lysozyme). After 5 minutes at room temperature, 200 ul of 0.2 N NaOH/1% SDS was added and the tube incubated

on ice for 5 minutes. Finally, the contents were neutralized on ice for 5 minutes after the addition of 150 ul of an ice-cold solution of potassium acetate (pH 4.8). The debris was removed by centrifugation in a microcentrifuge for 5 minutes. The supernatant was extracted once with phenol/chloroform then ethanol precipitated. The DNA pellet was recovered by centrifugation for 5 minutes, washed once, each with 70% ethanol and absolute ethanol, and air died. The dried pellet was redissolved in 50 ul of TE buffer containing RNase A (20 ug/ml).

(c) Large Scale Plasmid Preparation

Larger and purer preparations of plasmid DNA were isolated by the method of Birnboim and Doly (1979) with slight modifications. An overnight culture (3 ml) containing ampicillin (50 ug/ml) was inoculated into 500 ml of LB broth with the same antibiotic. When the A_{600} reached 0.4 the cells were harvested by centrifugation (4000g for 10 minutes) and washed once with ice-cold STE buffer (0.1 M NaCl, 10 mM Tris-Cl (pH 7.8), 1 mM EDTA). The plasmid DNA was then isolated as in (b) except that 10 ml of lysis buffer, 20 ml of 0.2 N NaOH/1% SDS solution, and 15 ml of 5 M potassium acetate was used. The final DNA pellet was redissolved in 10 ml of TE buffer. Ten grams of cesium chloride were dissolved in the tube and 0.8 ml of ethidium

bromide (10 mg/ml) was added. The solution was mixed well and loaded into Beckman quick seal centrifuge tubes. Centrifugation was for 36 hours at 40,000 rpm in a Ti80 rotor in a Beckman ultracentrifuge at room temperature.

The lower plasmid band was removed from each tube with a 5 ml syringe equipped with an 18 gauge needle. The sample was then extracted with an equal volume of water-saturated butanol until all traces of ethidium bromide were removed. The sample was then dialysed overnight against several changes of TE buffer (pH 8.0).

2.2.6 Subcloning DNA Fragments into pTZ18U

(a) Restriction Enzyme Digests

Restriction enzyme digests were usually performed in a final volume of 20 ul on approximately 1-2 ug of phage or plasmid DNA. This included 10-20 units of the appropriate enzyme and one-tenth the volume of the 10x buffer reccomended by the manufacturer. The mixture was incubated at the temperature specified (usually 37^{o}C) for 1-3 hours. Finally, the enzyme was either heat denatured at 70^{o}C or inhibited by the addition of 1 ul of 0.5 M EDTA. Usually half of a digest was examined on a 0.8% agarose gel (see 2.2.7.a), and the other half used for ligation into the vector.

(b) Ligation

Weight ratios of 10:1 (insert:vector) were used in a "shotgun" ligation approach. Phage DNA (500 ng), digested with Eco RI, was mixed with Eco RI digested pTZ18U (50 ng) and ligated overnight at 14°C, following the addition of 5x ligase buffer and 1.0 unit of T4 DNA ligase. Some phage DNA was also directionally ligated into pTZ18U after digesting both the insert and the vector with Kpn I and Sst I.

(c) Competent Cell Preparation

Competent cells were prepared by a modification of the method detailed by Kushner (1978) and also described by Maniatis et al (1982). A single colony of JM107 cells was inoculated into 5 ml of LB medium and grown overnight at 37°C with vigorous shaking. The following morning the 5 ml culture was added to 500 ml of LB broth in a 2 liter flask. This culture was grown for 2.5 hours or until the A_{600} was 0.6. The culture was transferred to 250 ml centrifuge bottles and the cells harvested at 6000 rpm (GSA rotor) for 10 minutes (4°C). The cell pellets were suspended in a total of 250 ml of chilled 10 mM Pipes buffer (pH 6.8) containing 10 mM $RbCl_2$. The cells were again pelleted and then resuspended in 10 mM Pipes buffer (pH 6.8) containing 10 mM $RbCl_2$ and 75 mM $CaCl_2$. The suspension was left on ice for 30 minutes and then pelleted a final time. The pellet

was gently resuspended in a solution containing 10 mM Pipes buffer (pH 6.8), 10 mM $RbCl_2$, 75 mM $CaCl_2$, and 15% glycerol. The suspended cells were aliquoted into microcentrifuge tubes (500 ul/tube), quickly frozen in liquid nitrogen, and stored at -70°C until use.

(d) Bacterial Cell Transformation

The procedure used for cell transformation was simplified from that described by Mandel and Higa (1970). Frozen competent cells were thawed for 1 hour on ice. The cells (200 ul) were then added to the ligation mix and left on ice or 15 minutes. Next, the cells were incubated at room temperature for 10 minutes. Following this, 500 ul of LB broth was added and the cells were incubated for 30 minutes at 37°C with constant shaking. The cells were pelleted (1 minute in a microfuge) and 500 ul of supernatant was removed. Finally, the mixture was spread onto LB plates containing 50 ug/ml ampicillin, 50 ul of 2% X-Gal, and 10 ul of 100 mM IPTG.

2.2.7 Agarose Gel Electrophoresis

(a) Standard Agarose Gel Electrophoresis

DNA from plasmid preparations, restriction digests, and recombinant phage DNA was analyzed on 0.8% agarose gels. The gel was prepared by dissolving 1.2 g of

agarose in 150 ml of 1x TAE buffer (0.04 M Tris-acetate (pH 8.3), 0.001 M EDTA) in a microwave oven for 4-6 minutes. After cooling to 50°C the solution was poured into a gel form containing either a 15 or 20 well comb. The gel was equilibrated for 5 minutes in 1 liter of 1x TAE buffer containing 0.5 ug/ml ethidium bromide. Samples were applied to the gel following the addition of one-fifth a volume of loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, and 30% glycerol). Electrophoresis was at 30 V overnight or 100 V for 2-3 hours. The gel was then visualized on an ultraviolet light box and photographed. Smaller gels (50 ml) were used for similar types of analysis with fewer samples (8 maximum).

(b) Alkaline Agarose Gel Electrophoresis

The procedure used for alkaline agarose gels was based on the method of McDonnell et al (1977). This type of gel was used to analyze first and second strand cDNA synthesis. A 0.8% agarose gel was prepared in 50 mM NaCl, 1 mM EDTA similar to the standard gel (above). The gel was then equilibrated in 30 mM NaOH, 1 mM EDTA for at least 30 minutes. Before loading the samples, the buffer was removed until it was only 1 mm above the gel. Samples were ethanol precipitated and redissolved in 10-20 ul of alkaline loading buffer (50 mM NaOH, 1 mM EDTA, 2.5% Ficoll (type 400), and 0.025% bromocresol green) before applying to the gel.

Electrophoresis was at 100 V until the dye had migrated 75% down the gel. It should be noted that a glass plate was placed on the gel, during electrophoresis, to help prevent diffusion of the dye. These gels were then dried and exposed to x-ray film.

(c) Formaldehyde-Agarose Gel Electrophoresis

Formaldehyde-agarose gels for RNA separation were prepared by the method of Fourney et al (1988). Agarose (1.8 g), 15 ml of 10x MOPS/EDTA buffer (0.2 M MOPS, 50 mM sodium acetate, 10 mM EDTA) and 130 ml of DEPC-treated water were mixed and heated in a microwave oven until the agarose dissolved. When the solution had cooled to 50°C, 7.7 ml of 37% formaldehyde was added and the gel poured (in a fume hood). The gel was allowed to solidify for 1 hour, then equilibrated in 1x MOPS/EDTA buffer for an additional hour. Samples were prepared by dissolving 10-30 ug total RNA in 10 ul of DEPC-treated water. The samples were heated at 65°C for 15 minutes after the addition of 25 ul electrophoresis sample buffer (0.75 ml deionized formamide, 0.15 ml 10x MOPS/EDTA buffer, 0.24 ml formaldehyde, 0.1 ml DEPC-treated water, 0.1 ml glycerol, and 0.08 ml 10% (w/v) bromophenol blue). Finally, the sample wells were flushed with buffer and then 1 ul of 1.0 mg/ml ethidium bromide was added to the samples and they

were loaded. Electrophoresis was overnight at 30 V or for 3-4 hours at 75-85 volts. The gel was visualized on an ultraviolet light box and photographed. These gels were then used for Northern analysis (2.2.8).

(d) Purification of DNA from Agarose Gels

Agarose gels (0.8%) were prepared in 1x TAE buffer and DNA was separated as described in 2.2.7(a). After electrophoresis, gels were visualized on an ultraviolet light box and the band of interest was excised with a razor blade. The DNA was eluted from the agarose band as described by Maniatis et al (1982). The gel piece was placed into dialysis tubing, presoaked in 0.5x TAE buffer. A minimal amount of buffer was added to cover the gel slice. The bag was then secured in an electrophoresis tank filled with 0.5x TAE buffer. Electrophoresis was for 3-4 hours at 100 volts followed by electrophoresis in the reverse direction for 2-3 minutes, also at 100 volts. The eluted DNA was recovered by ethanol precipitation and examined on a standad gel. Later recoveries of DNA from agarose gels was with a Gene Clean kit, utilizing the procedure supplied by the manufacturer (Bio-Can).

2.2.8 Northern and Southern Analysis

Northern and Southern analysis of RNA and DNA,

respectively, were performed as described by Maniatis et al. (1982) after transferring the nucleic acid to nitrocellulose using the technique described by Southern (1975).

RNA gels were electrophoresed as described (2.2.7.c) and then soaked twice in 10x SSC for 20 minutes (Fourney et al., 1988) before transfer to nitrocellulose. DNA gels were denatured for 1 hour in denaturing solution (1.5 M NaCl, 0.5 M NaOH), then neutralized in 1.5 M NaCl, 1 M Tris-HCl (pH 8.0) for one hour before transfer to nitrocellulose.

Apart from preparing the gels for transfer, all other steps are identical for both Northern and Southern analysis. Two pieces of 3MM paper soaked in 10x SSC were wrapped around a glass plate and placed over a tray of 10x SSC such that the ends of the 3MM paper were submerged in the tray. The gel was placed upside down on the 3MM paper and Saran wrap was used as a border to prevent contact of other layers with the wick. A nitrocellulose sheet, of the same size, was soaked in 2x SSC for 5 minutes then placed on the gel, followed by 2 layers of 3MM paper soaked in 2x SSC, 5 layers of dry 3MM paper, and a stack (10-15 cm) of paper towels. It was important to ensure that all air bubbles were removed between layers. Finally another glass plate and a weight (usually two thick books) were added to enhance transfer efficiency. The transfer was allowed to proceed overnight.

After transfer, the paper towels and 3MM paper were removed and the gel and nitrocellulose filter removed

together. The positions of the wells were marked with a pen and the gel was then discarded. The nitrocellulose filter was allowed to air-dry and then it was baked at 80°C for 1.5-2 hours in a vacuum oven.

The baked filter was soaked in 6x SSC for 5 minutes before prehybridization at 68°C (60°C for Northern) in a seal-a-meal bag with 0.2 ml of hybridization solution (6x SSC, 0.5% SDS, 5x Denhardt's solution, and 100 ug/ml denatured salmon sperm DNA), for each square centimeter of filter. After at least one hour, the probe (approximately 2 x 10^6 cpm, usually 20-30 ul of a labeling reaction) was denatured for 7 minutes in a boiling water bath, quickly cooled on ice and added to the bag. Hybridization was overnight at 68°C (60°C for Northern) in a shaking water bath.

The filter was removed from the hybridization bag into wash solution 1 (2x SSC and 0.5% SDS) for 5 minutes at room temperature. For Southern analysis this was followed by 15 minutes at room temperature in solution 2 (2x SSC and 0.1% SDS) and three washes of 45 minutes each at 68°C in solution 3 (0.1x SSC and 0.5% SDS). For Northern analysis it was followed by two 45 minute washes at 50°C under mild stringency (2x SSC and 0.5% SDS).

The filter was then air-dried and exposed to film at-70°C for the appropriate length of time (usually overnight).

2.2.9 DNA Sequencing

(a) Template Preparation and Sequencing Reactions

Sequencing was performed using double stranded DNA templates prepared by the method described by Zhang et al (1988). Cesium chloride purified plasmid DNA or phage DNA (1-2 ug) was adjusted to 20 ul with sterile distilled water then denatured for 5 minutes by the addition of 2 ul of 2 M NaOH, 2 mM EDTA. The reaction was then neutralized by the addition of 3 ul 3M sodium acetate (pH 5.2) and 7 ul sterile distilled water. After vortexing, the DNA was precipitated by the addition of 75 ul of chilled absolute ethanol (-20°C). After 5 minutes at -70°C, the DNA was recovered by centrifugation for 5 minutes. The pellet was washed with 70% then absolute ethanol, and air-dried before dissolving in 7 ul of sterile distilled water.

Sequencing buffer (2 ul) and the appropriate primer (0.5 pmol of universal or oligonucleotide primer) were added to the template and annealed for two minutes at 65°C with slow cooling to below 35°C. Oligonucleotide primers were purchased from the Central Facility of the Institute for Molecular Biology and Biotechnology, McMaster University. These included AB308 (5'-CTCACAGAAATAGGCAT-3'), AB513 (5'-GTGATTTCTGCTGCTCAG-3'), AB537 (5'- CCCATTCCAGGGTCCTTC-3'), and AB538 (5'- CCAGTATCAGGAATGTAC-3'). Lambda gt11 forward (5'-GACTCCTGGGCCCG-3') and reverse (5'-GGTAGCGACCGGCGC-3')

primers used with Kpn-Sst inserts were purchased from Sin-Can Inc. (Calgary, Alberta).

The sequencing reactions were then set up with the annealed primer and template according to the Sequenase protocol provided with the kit. This method is a modification of the chain termination protocol developed by Sanger et al., (1977) by its use of a modified DNA polymerase (Tabor and Richardson, 1987).

Phage DNA, purified as described in 2.2.5a, was prepared for sequencing by boiling 1 ug of DNA and 1.0 pmol of primer for 4 minutes then cooling slowly to less than 35^{o}C (similar to Liu et al., 1985). After template and primer were annealed the Sequenase protocol was followed for the sequencing reactions.

(b) Preparation and Electrophoresis of Sequencing Gel

A Bio-Rad integral plate/chamber sequencing apparatus (38 cm x 50 cm) was used for all sequencing gels. Before setting up the apparatus the glass plates were washed with soap and water, rinsed with distilled water and allowed to air dry. Then the plates were wiped 3 times with absolute ethanol and assembled according to the manufacturer's guide.

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 10x TBE

buffer. The volume was adjusted to 200 ml with deionized water and deaerated for 15 minutes.

Twenty-five millilters was removed to a flask, 160 ul of a fresh 10% APS solution and 140 ul of TEMED were added and the solution mixed and poured into the gel casting tray to seal the bottom of the gel apparatus. After 5 minutes, the gel was poured on a 45 degree angle after mixing 195 ul of 10% APS and 160 ul of TEMED into the remaining gel solution. The gel was left to polymerize overnight.

The casting tray was removed and the gel apparatus was inserted into the buffer tank. The gel was pre-electrophoresed at 55 W for 30-60 minutes in 1x TBE buffer before loading samples. The samples were heated at 75^{o}C for two minutes, quickly cooled on ice, then loaded. Electrophoreses was at 70 W until the first dye ran off the gel. For most sequencing gels three loadings were used. After all loadings were completed the gel apparatus was disassembled and the gel was removed. The gel was dried on a gel dryer at 68^{o}C for 30-45 minutes prior to exposure to Kodak XK-1 film.

(c) Preparation of Deletions for Sequencing

Deletions were prepared for subclones CHP1A-1.1a, CHP1A-1.1b, CHP1A-0.75a, and CHP1A-0.75b according to the procedure of Henikoff (1984). Approximately 10 ug of plasmid DNA was digested with Sph I, to leave a 3' overhang,

and Sma I, from which exonuclease III digestion was initiated. Exonuclease III digestion and S1 nuclease cleavage was as described by Henikoff (1984). An aliquot was removed at this point for gel analysis. The DNA fom each time point was then blunt-ended and ligated overnight at 14°C. The samples were used to transform competent JM107 cells and plated onto LB ampicillin plates supplemented with X-gal and IPTG. Two colonies from each time point were picked for analysis. Those clones exhibiting an appropriately sized deletion were sequenced.

2.2.10 Computer Analysis

The DNA Sequence Analysis Package, version 2.4, distributed by the Molecular Biology Computer Research Resource (Harvard School of Public Health, Dana-Farber Cancer Institute - D1154, 44 Binney St., Boston, Mass., 02115) was used to compare protein sequences. This package was also used for hydropathy plots which were calculated according to the method of Kyte and Doolittle (1982). A window size of 7 amino acids was used for the analysis.

Helical wheel plots were calculated and drawn using the PC gene DNA/protein sequence analysis program.

3.0 RESULTS

3.1 Construction of a cDNA Library

The cDNA library was constructed from RNA isolated from CHO cells. The yield of total RNA varied with the preparation, but was usually 0.7-1.0 mg when starting with 4 x 10^7 cells. Usually, 500 ug was loaded onto an oligo (dT) cellulose column, and the remainder saved for Northern analyses. The poly $(A)^+$ RNA was eluted from the column with yields of 1.5 to 3.5% of the total RNA loaded.

Using a method similar to that of Gubler and Hoffman (1983), and a BRL cDNA synthesis kit, 10 ug of poly $(A)^+$ was used to construct the first strand cDNA. The first strand reaction mixture was prepared in a final volume of 50 ul. After addition of the reverse transcriptase a tracer reaction was set up by removing one-fifth the volume and adding it to 5 uCi of [alpha-^{32}P]dCTP. This tracer was then used to calculate the yield of first strand product using the equations the Appendix. The first strand yield was 22% which is slightly lower than the 30-40% yields expected for the enzyme (BRL cDNA manual). The first strand synthesis was also performed with 5 ug of 2.3 kb control RNA to ensure that the enzyme was active. The yield for the control RNA was 21% suggesting that the quality of the RNA isolated from the CHO cells was at least as good as the control RNA.

An aliquot of [alpha-^{32}P]dCTP was added to the second strand buffer along with the first strand mixture. After

four hours the reaction was terminated and an aliquot was removed to calculate the yield of cDNA. Of the 2.2 ug of single stranded DNA synthesized in the first strand reaction, 94.5% was converted to double stranded DNA. The control experiment was only slightly more efficient.

The remaining amount of the tracer for both the first and second strands was precipitated and analyzed on an alkaline agarose gel. The samples were redissolved in loading buffer and a small aliquot was counted in a scintillation counter. The gel was loaded with 20,000 cpm in each lane, and radiolabelled markers were run alongside to determine the size of the cDNA.

Figure 5 is an autoradiograph of the first and second strand cDNA synthesis for both the 2.3 kb control RNA and the sample RNA. The cDNA sample prepared from RNA from CHO cells showed that the size of the cDNA ranged from less than 200 bp to greater than 3 kbp with the majority approximately 1 kbp in length (Figure 5, lane 1,2). Some of the difference in intensity between the first and second strands can be attributed to a gel loading error. Lanes 3 and 4 show that the control RNA was synthesized efficiently into a 2.3 kbp cDNA. However, there are a lot of smaller cDNA products resulting in a smear below the major band at 2.3 kbp. It appears that the majority of first strand product converted to double stranded cDNA was full length in size, although there was also a gel loading error for this control

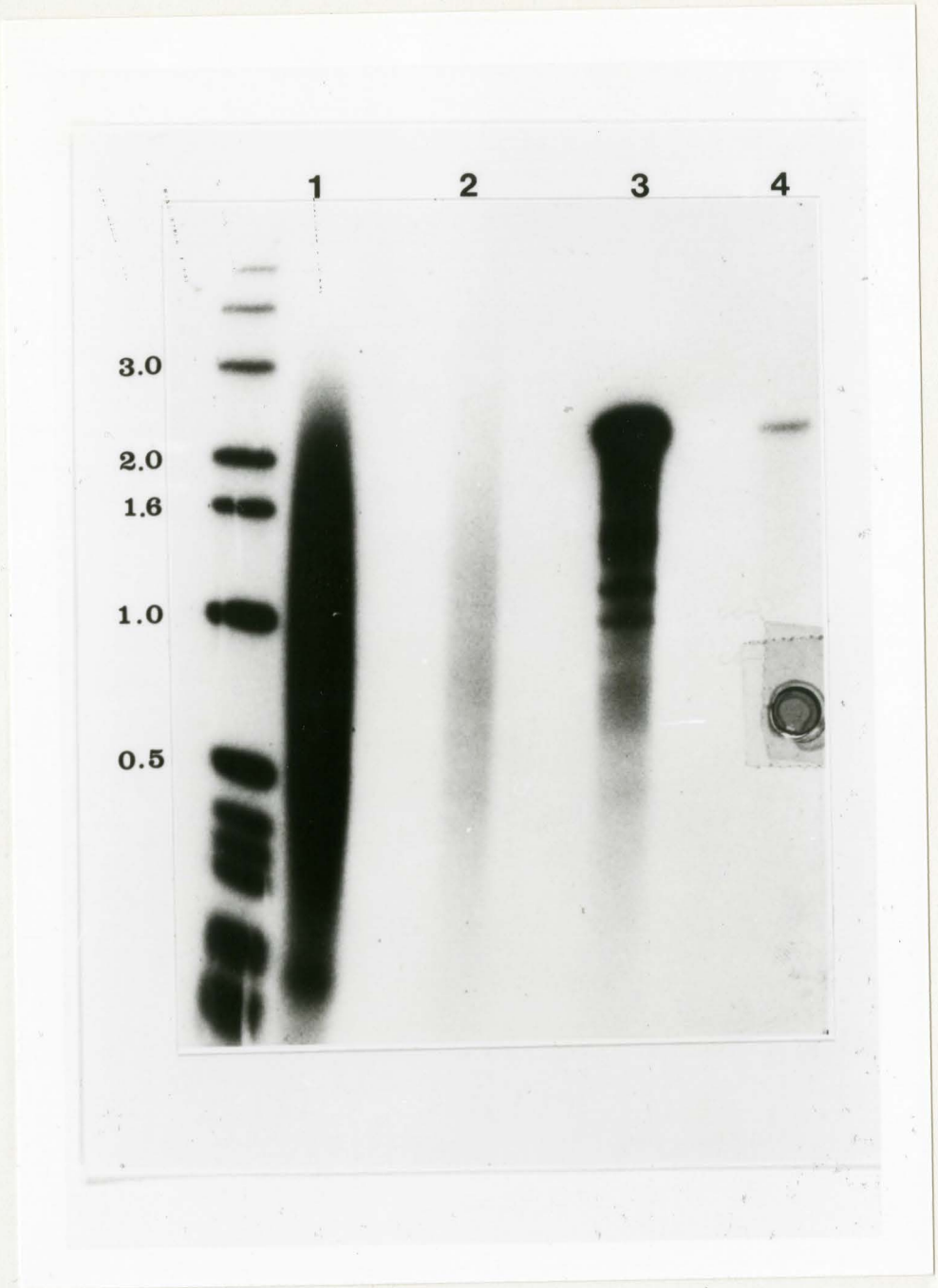


Figure 5 **Analysis of first and second strand cDNA synthesis.** Approximately 20,000 cpm from each cDNA synthesis tracer reaction was loaded onto an alkaline agarose gel. Lane 1, aliquot of first strand cDNA synthesis from CHO RNA. Lane 2, aliquot of second strand synthesis of the same preparation. Lanes 3 and 4 are aliquots from the first and second strand synthesis of the control RNA (2.3 kb), respectively. Molecular weight size markers are labelled in kbp.

experiment.

After blunt-ending the product with T4 DNA polymerase the cDNA was methylated with Eco RI methylase in preparation for the addition of end-labelled Eco RI linkers. To determine if the linkers were functional, after end-labelling with [gamma-^{32}P]dATP, an aliquot of labelled linkers were ligated to themselves. The ligated linkers were examined on a 10% polyacrylamide gel alongside ligated linkers that had been digested with Eco RI (Figure 6). The presence of linker multimers (Figure 6, lane A) showed that they were functional. These were then ligated to the cDNA sample.

After ligation of the linkers to the cDNA, the cDNA was digested with Eco RI to create the Eco RI sticky ends. Finally, before ligation into the vector, the excess linkers were removed from the cDNA by passing the mixture over a Sepharose CL-4B column. Five-drop fractions were collected in microcentrifuge tubes and counted in a scintillation counter. Figure 7 is a graph of the counts in each fraction and it shows two large peaks of radioactivity. The first peak is representative of the cDNA attached to the linkers and the second peak contains excess linkers. The graph demonstrates that there was good separation of the excess linkers from those linkers attached to the cDNA.
Fractions 4 to 8 were pooled and precipitated, prior to ligation into λgt11.

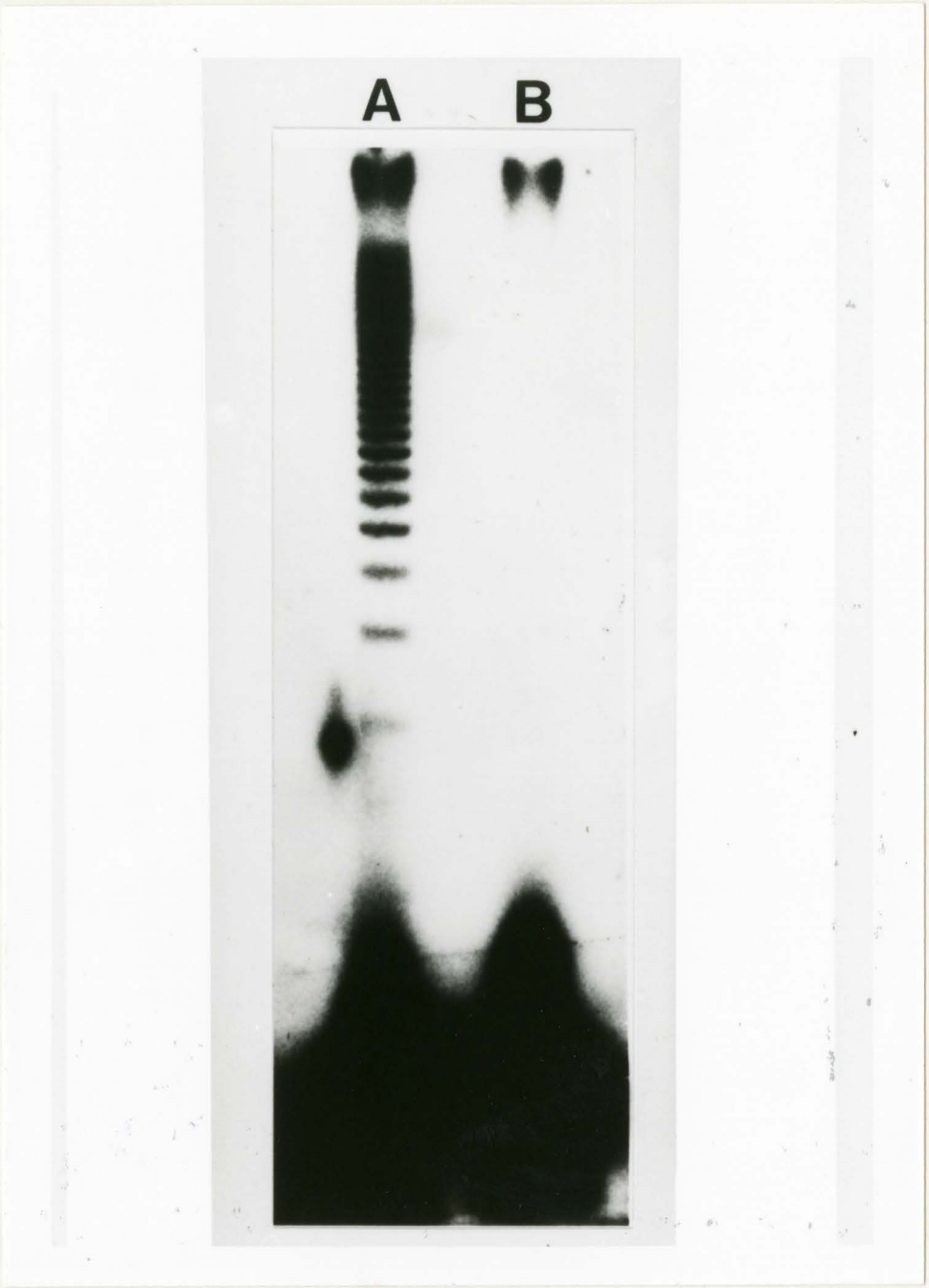


Figure 6 **Linker ligation experiment.** Eco RI linkers were end-labelled with [γ-^{32}P]dATP then ligated. The ligated linkers (A) were electrophoresed on a polyacrylamide gel beside an aliquot of Eco RI digested linkers (B).

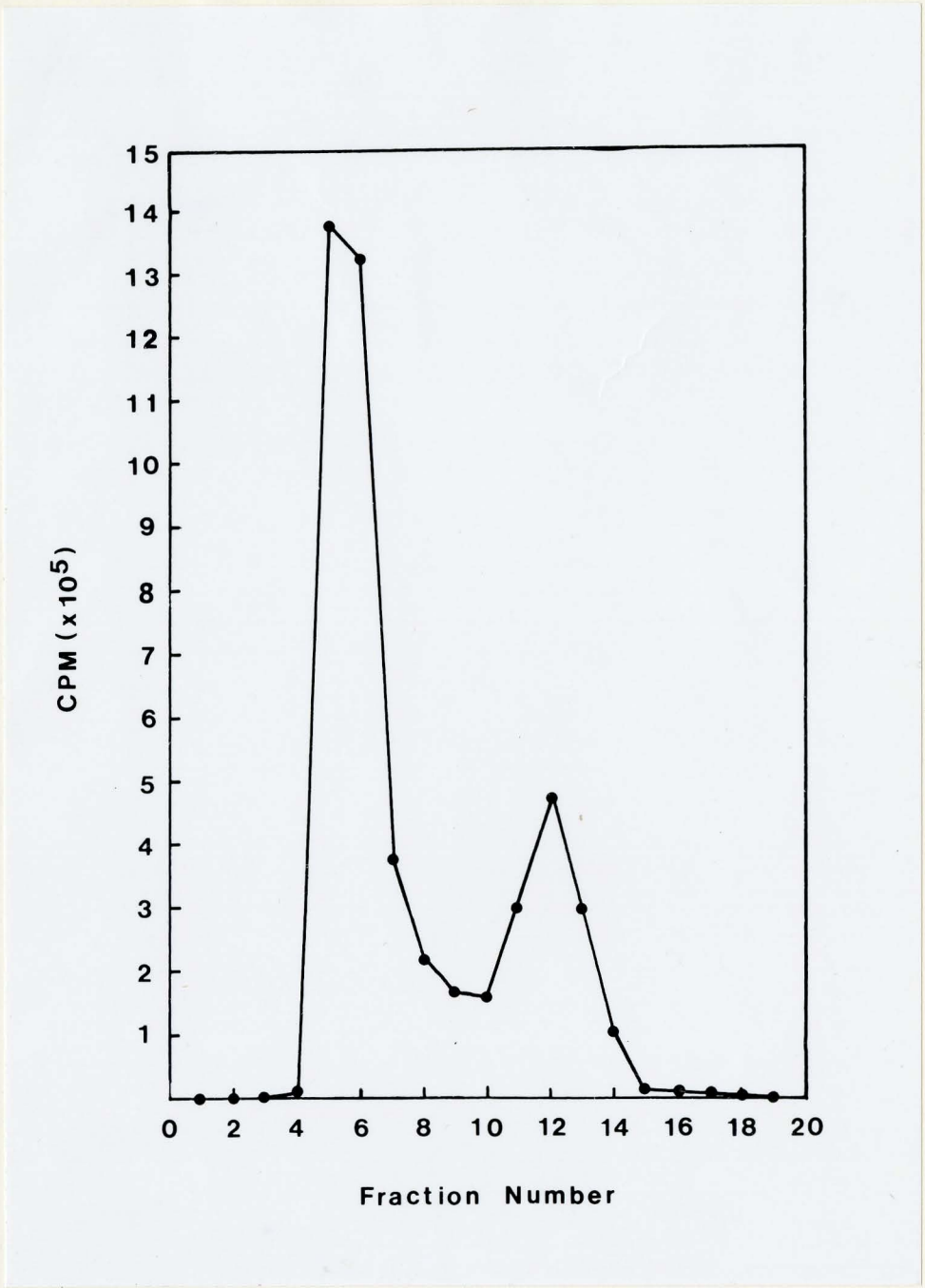


Figure 7 **Separation of excess linkers from cDNA.** Five-drop aliquots were collected from a CL-4B column and counted in a scintillation counter. The counts per minute of each fraction was plotted against the fraction number. Fractions 4 to 8 were pooled and ethanol precipitated.

Lambda gt11 is a vector containing an Eco RI site for cloning and was the vector used for the formation of the cDNA library (Figure 8a). The Eco RI site lies within the beta-galactosidase gene, such that cDNAs ligated into this site can be expressed as fusion proteins if they are in the same reading frame. This vector can accomodate inserts of up to 7 kbp in size for proper packaging into functional bacteriophage. Also, cloning into the beta-galactosidase gene allows for easy recognition of recombinant phage. When phage are plated onto LB plates containing X-gal and IPTG a non-recombinant will have a functional beta-galactosidase gene (via complementation with the host) which will break down the substrate and grow as a blue plaque. The recombinant phage will not be able to break down the substrate and will grow as a white plaque.

The cDNA was ligated into λgt11, packaged using a commercial packaging extract (Promega Biotech), and then plated to determine the size of the library as well as the number of recombinants. The library contained 1.5×10^5 pfu/ug of cDNA with approximately 85% of the clones as recombinants. The library was then amplified to a titer of 1.2×10^9 pfu/ml and screened for positive clones.

The other library used in this work was a cDNA library constructed from CHO cells in λgt10. Figure 8b depicts the features of this vector, which are very similar to λgt11. The cDNA was cloned into the Eco RI site and an aliquot from

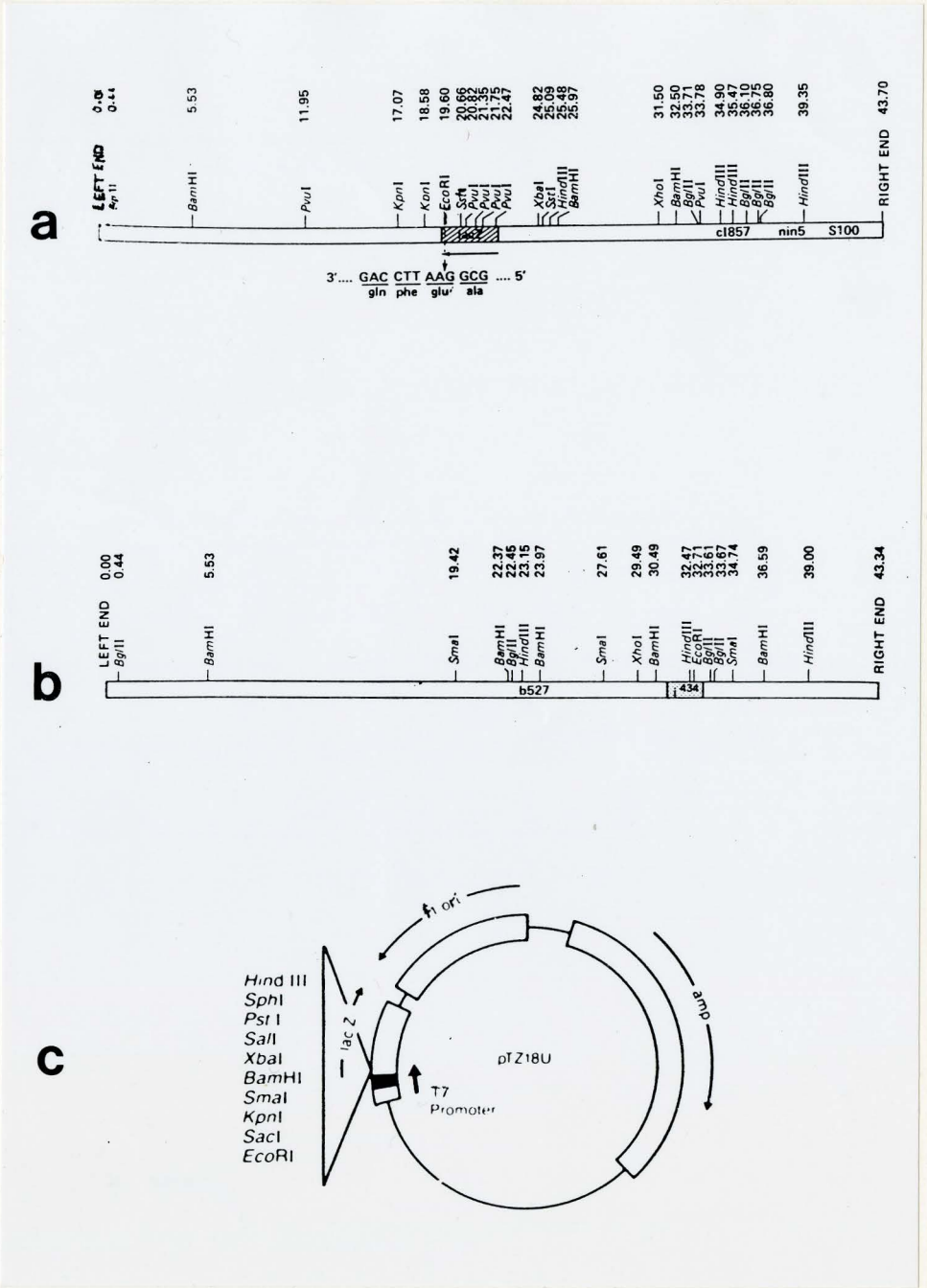


Figure 8 **Bacteriophage and plasmid cloning vectors.** Lambda gt11 (a) and lambda gt10 (b) were used for construction of the cDNA libraries. Plasmid pTZ18U (c) was used for subcloning the bacteriophage inserts.

the amplified library was a gift from Dr. W. Flintoff.

3.2 Isolation of a P1 clone from the cDNA Libraries

Approximately 1 x 10^4 plaque forming units from the amplified cDNA library were plated on 15 dishes for screening with the human 1.5 kbp and 0.8 kbp P1 probes. Triplicate plaque lifts were performed for hybridization. Since, a cDNA library has a selective advantage for 3'-end clones most positive signals should hybridize to the 3'-probe. Therefore two filters were probed with the 1.5 kbp fragment and the other filter was probed with the 5'-end 0.8 kbp fragment. In addition to these filters, a filter containing plaques from P1-22a was probed as a positive control. It was postulated that by this method a positive plaque should have signals in an identical position on two filters if it was a clone of 1.5 kbp or less and on all three filters if it was larger than 1.5 kbp. Screening of the λgt11 library yielded eight clones, six of which hybridized to only the 1.5 kbp fragment. The other two clones did not hybridize to the 1.5 kbp clone, but did show a strong signal when hybridized to the 0.8 kbp fragment. Figure 9 shows clone WT-1, as an example of one of the positive clones, which was obtained in duplicate (Figure 9 a,b), in comparison to the same clone partially purified (Figure 9c), and the positive control (Figure 9d). The other clones that hybridized to the 1.5 kbp fragment were

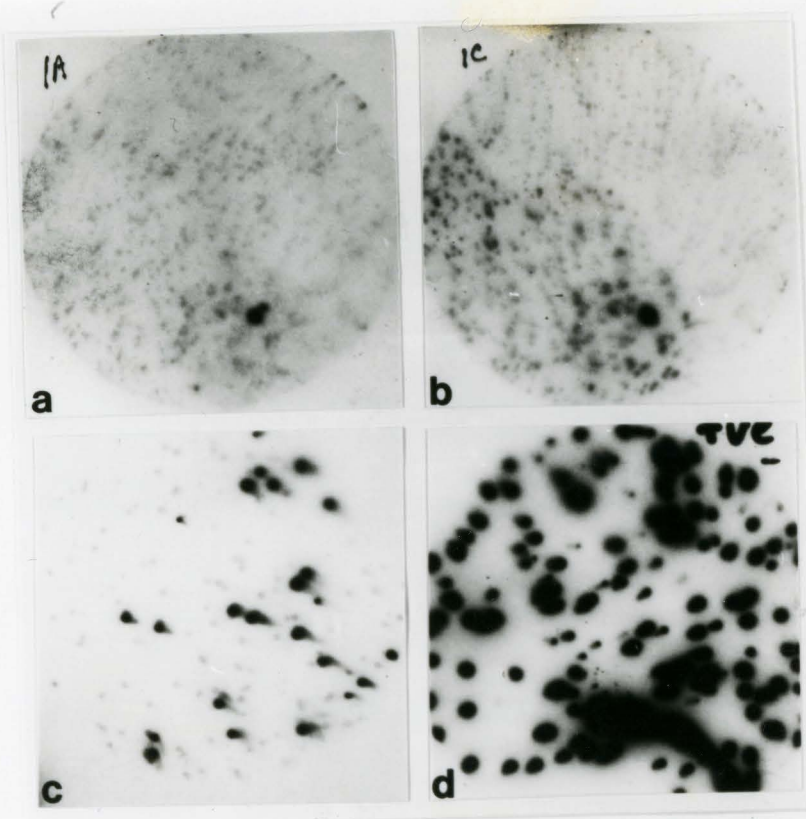


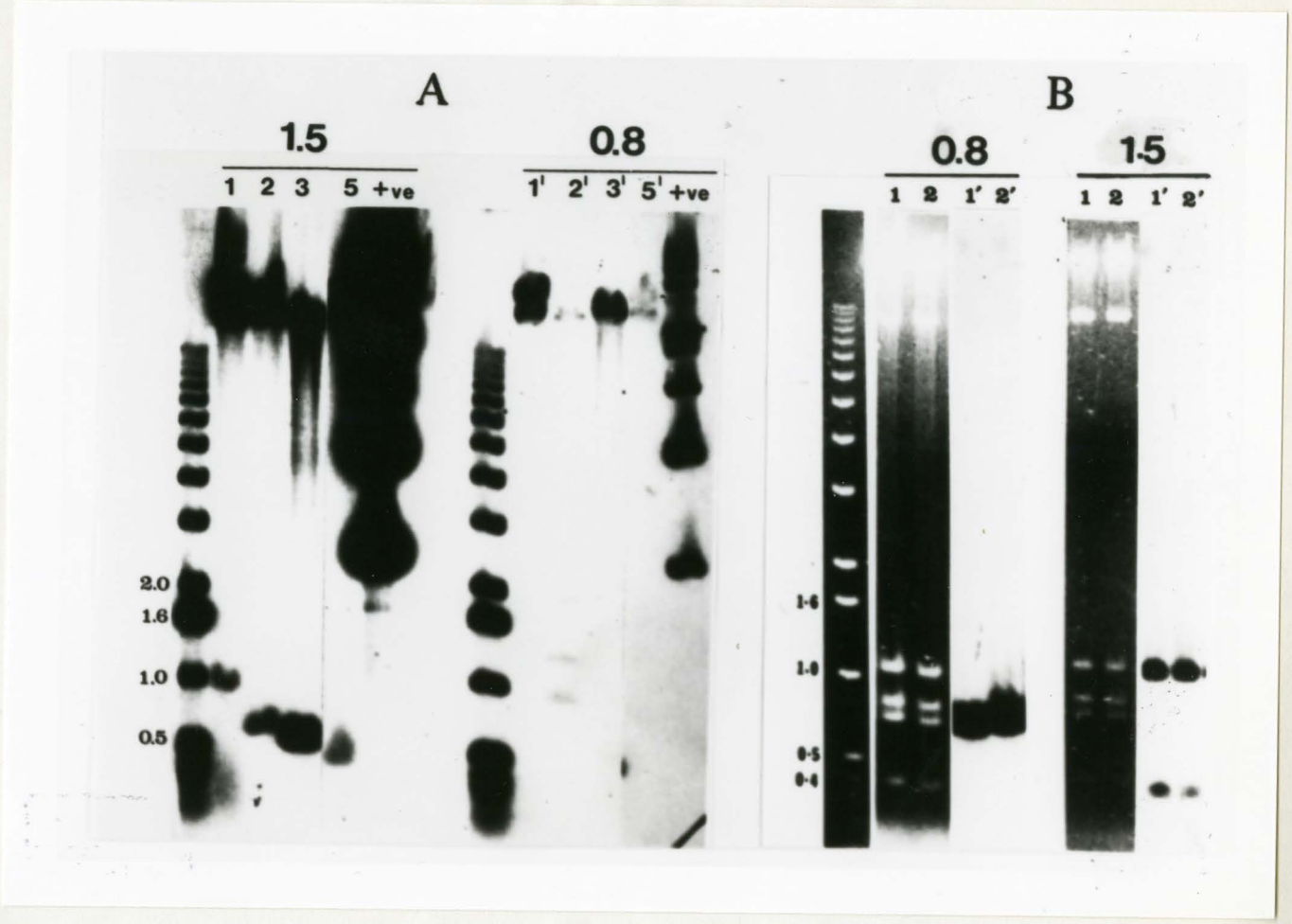
Figure 9 **Detection of a positive clone.** Clone WT-1 was identified as a hybridising signal found on duplicate filters (a, b). The agar plug was excised from the plate and replated at a lower concentration for secondary screening (c). Panel (d) depicts a positive control filter.

WT-2, WT-3, WT-5, WT-6, and WT-8. Clones WT-4, and WT-7 hybridized to the 5'-end fragment. The positive plaques were removed as plugs and plaque-purified as described in Materials and Methods. Clones WT-6, WT-7, and WT-8 did not show any hybridization after the first round of plaque purification and were abandoned. Clone WT-4 was also abandoned.

Screening of the λgt10 library yielded 11 positive clones, two of which hybridized to both the 1.5 and 0.8 kbp fragments. Since four 3'-end clones had already been obtained, only the clones CHP1A and CHP1D, which hybridized to both probes, were plaque purified.

After plaque purifying the positive clones, small scale phage DNA preparations were performed to examine the insert sizes. The DNA was digested with Eco RI and electrophoresed on an agarose gel. The gels were blotted onto nitrocellulose in duplicate such that one blot could be probed with the 3'-end probe and the other with the 5'-end probe. Analysis of the WT-clones showed that the inserts ranged in size from approximately 500 to 1100 base pairs in length. The Southern blots (Figure 10a) confirmed that these were positive clones. There is a large discrepancy in signal which was partially due to incomplete digestion of the samples, specifically WT-1, WT-2, and WT-5. Another reason for the discrepancy was determined when sequencing showed that these clones resided in the 3'-untranslated

Figure 10 **Southern blot analysis of inserts from positive clones.** (A) DNA from clones WT-1 (lanes 1 and 1′), WT-2 (lanes 2 and 2′), WT-3 (lanes 3 and 3′), and WT-5 (lanes 5 and 5′) was digested with Eco RI and electrophoresed in 0.8% agarose gels. The gels were blotted onto nitrocellulose and probed with either ^{32}P-labelled 1.5 kbp (1.5) or 0.8 kbp (0.8) human P1 fragments. Plasmid P1-C5 was used as the positive control. (B) DNA from clones CHP1A (lanes 1 and 1′) and CHP1D (lanes 2 and 2′) was digested with Eco RI and electrophoresed in 0.8% agarose gels. Lanes 1 and 2 show ethidium bromide staining of the gels. Portions of the same gel were blotted on nitrocellulose (lanes 1′ and2′) and probed with either the ^{32}P-labelled 0.8 kbp or 1.5 kbp human P1 fragments. Molecular weight size markers are given in kbp for both blots.


A
1.5
0.8
1 2 3 5 +ve
1' 2' 3' 5' +ve
2.0
1.6
1.0
0.5
B
0.8
1.5
1 2 1' 2'
1 2 1' 2'
1·6
1·0
0·5
0·4

region where there are some sequence differences from the human gene. Although these are minor differences, the hybridization and washes were performed at 68°C, under high stringency which should allow for detectable differences in the two sequences. The positive control for this Southern blot was undigested plasmid P1-C5. This signal was very strong, as expected.

Eco RI digestion of DNA prepared from either CHP1A or CHP1D yielded four fragments on an agarose gel. These fragments were approximately 1.1 kb, 0.85 kb, 0.75 kb and 0.4 kb in length (Figure 10b). Southern blot analysis of clones CHP1A and CHP1D showed that the 1.1 and 0.4 kb fragments hybridized to the 3'-probe (Figure 10b). The other two fragments migrated closely on the agarose gel and thus it was difficult to determine if the 5'-probe hybridized to only the 0.75 kb fragment or to both (Figure 10b). It was later shown (see Northern analysis) that only the 0.75 kb fragment was specific for the 5'-probe. This suggested that the 0.85 kb fragment represented either an unrelated fragment or some additional 5'- or 3'-sequence which is not present in the human cDNA clone.

The combined size of the three hybridizing fragments (2.25 kb) corresponds well with the size of the human cDNA clone (Jindal et al., 1989).

3.3 Subcloning of the P1 cDNA Inserts

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 there are internal Eco RI sites (Figure 8a). Clones WT-1, WT-2, WT-3, WT-4, and WT-5, were digested with Kpn I and Sst I and "shotgun" ligated into pTZ18U at a weight ratio of insert to vector of 4:1. Prior to ligation a small aliquot of digested DNA was examined to ensure that there were no Kpn I or Sst I sites within the insert. All clones lacked sites for these enzymes. In addition, this method inserts approximately 1.0 kb of phage DNA on either side of the insert. This type of directional cloning allows for easy screening since all inserts will be in the same orientation.

The ligation mixtures were plated onto agar plates containing ampicillin, X-gal, and IPTG. Bacteria containing recombinant plasmids appeared as white colonies and represented approximately 95% of all colonies. Six colonies from each ligation were chosen and small scale plasmid DNA was prepared to screen for the correct insert. One clone for each of WT-1, WT-2, WT-3, WT-4, and WT-5 was then chosen for a large scale preparation of plasmid DNA. This CsCl-purified DNA was then used for sequencing. Although these clones contained phage DNA, this DNA contained the sequences for the λgt11 forward and reverse primers, which were used for sequencing both ends of the insert.

Since clones CHP1A and CHP1D appeared to be identical, only CHP1A was used for subcloning. The insert for this clone was subcloned as Eco RI fragments since there were no enzyme sites in λgt10 near the Eco RI site that could be used to clone the entire insert. A "shotgun" ligation was also used for CHP1A, with a weight ratio of insert to vector of 10:1. A high ratio was used to reduce the frequency of vector self-ligation. After transfection, approximately 30% of the colonies appeared to be recombinant. Twenty-four colonies were chosen for small scale plasmid DNA preparations. Of the 24 clones, 3 contained the 1.1 kb fragment, 6 contained the 0.75 kb fragment, and 4 had the 0.4 kb insert. One clone was interesting as it contained both the 0.75 and the 0.85 kb fragments. An additional 7 clones contained the 0.85 kb fragment. These were saved for Northern analysis to determine if this fragment represented part of the P1 gene. Three of the 24 clones examined did not contain any fragment.

These clones were then digested with Rsa I to determine the orientation of the insert. The initial 24 clones examined contained the 1.1 and 0.4 kb inserts in both orientations (Figure 11). The plasmids containing the 0.75 and 0.85 kb inserts had to be digested and then religated to find the second orientation (Figure 11). Also, the 0.75 kb insert did not contain an Rsa I site within the insert so the enzyme Dde I was used to determine the two orientations.

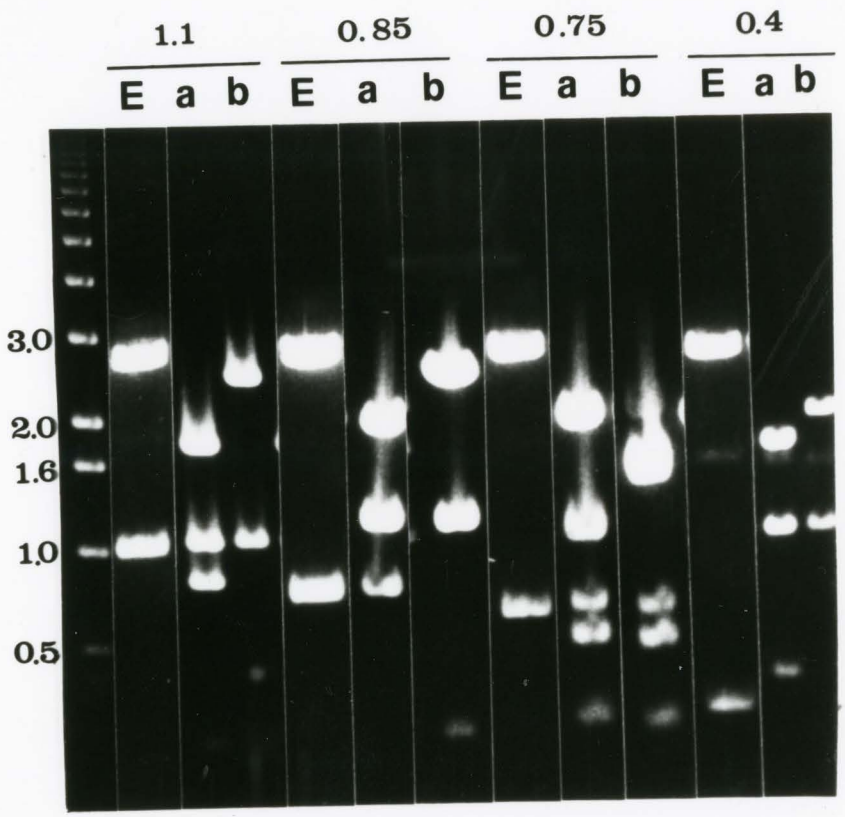


Figure 11 **Agarose gel depicting the orientations of the CHP1A Eco RI subclones.** The Eco RI fragments from CHP1A (1.1, 0.85, 0.75, 0.4) were subcloned into pTZ18U at the Eco RI site. Each clone was digested with Eco RI (E) to examine the insert and with Rsa I (fragments 1.1, 0.85, 0.4; lanes a and b) or Dde I (fragment 0.75, lane a and b) to determine the two orientations of the fragment in the vector.

Both orientations for each clone were then used to prepare large scale DNA samples. These were then used for sequencing and the construction of deletions.

3.4 Northern Analysis

Since CHP1A contained what appeared to be a full length clone it was used for a Northern analysis. Hamster and human total RNA was electrophoresed in an agarose-formaldehyde gel in alternating lanes. Ten micrograms of CHO RNA was loaded in one lane followed by 2 ug of HeLA RNA. The gel was then blotted onto nitrocellulose and cut into strips. To confirm that the fragments were part of the P1 gene, they were isolated from plasmid DNA and labelled to high specific activity to probe the strips of nitrocellulose. As a positive control the 1.5 kb human cDNA clone was used to probe one strip. Figure 12 shows clearly that the 0.85 kb fragment recognizes a separate transcript and is thus an unrelated fragment that was ligated into the vector with the P1 cDNA. The partial nucleotide sequence showed that it did not seem to encode for any functional protein (results not shown).

The transcript recognized by the other three fragments is approximately 2.2 kb in length and similar in size to the transcript recognized by the human 1.5 kb probe (Figure 12, lanes 1 and 2).

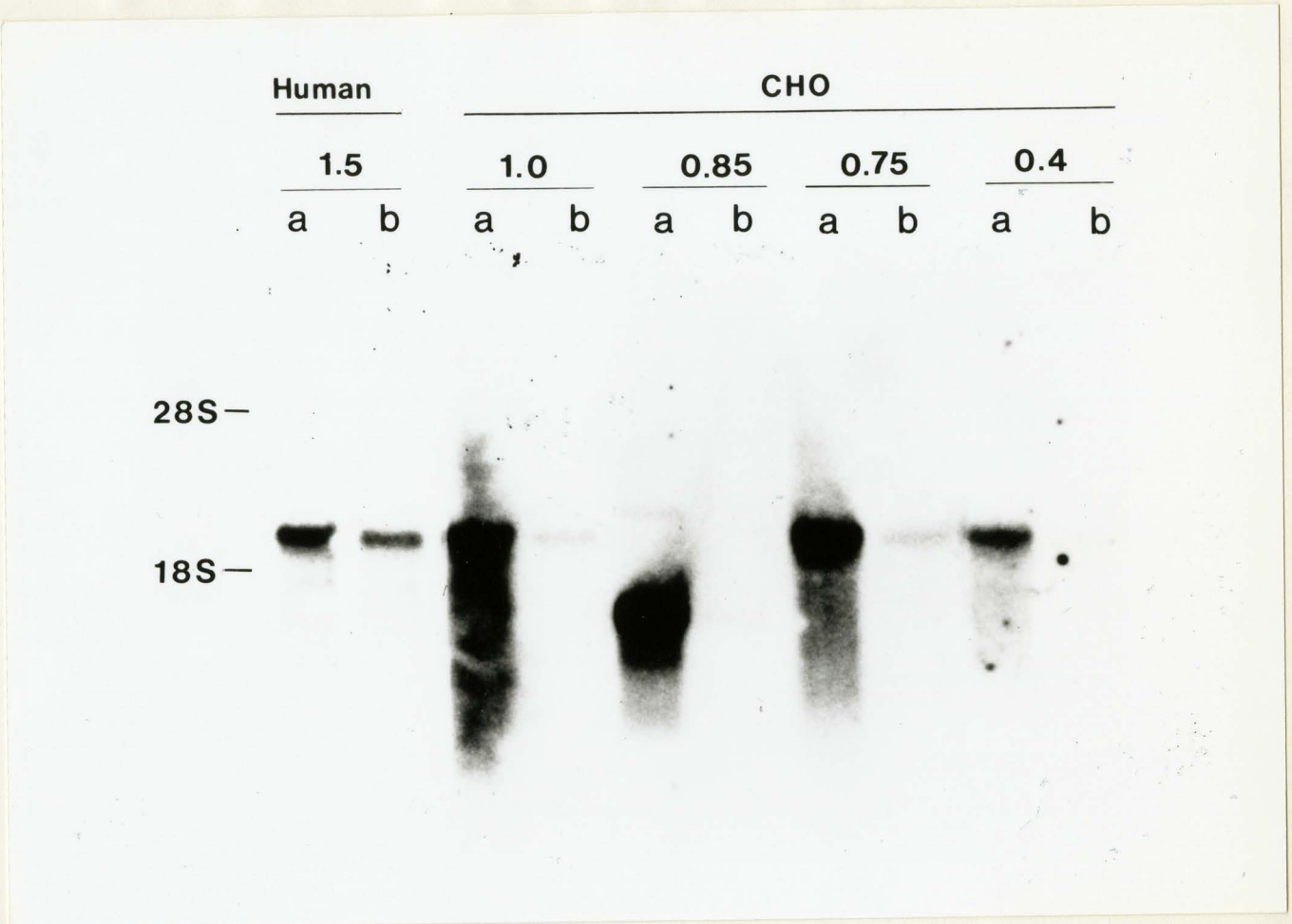


Figure 12 **Northern blot analysis of total CHO and human cells RNA probed with the four Eco RI fragments from CHP1A and the 1.5 kbp human P1 cDNA fragment.** Approximately 10 ug of CHO (a) and 2 ug of human (b) RNA were electrophoresed in adjacent lanes in a 1.2% agarose-formaldehyde gel. After blotting onto nitrocellulose, portions of the blot were hybridized to different ^{32}P-labelled human (1.5 kbp fragment) and CHO (1.0 kbp, 0.85 kbp, 0.75 kbp, and 0.4 kbp fragments) cDNA probes. The positions of 28S and 18S on the gel are indicated.

3.5 Sequencing of the pTZ-18U Subclones

Initially, a sequencing gel of the clones WT-1, WT-2, WT-3, and WT-5 was used to examine the location of these inserts with respect to the human cDNA clone, and to determine if they were overlapping clones. Both the λgt11 forward and reverse sequencing primers were used for this purpose. Figure 13 outlines the location of clones WT-1, WT-2, WT-3, and WT-5 to the 3'-end of the gene. The insert for clone WT-1 was 974 bp in length, its 5'-end located 68 bp downstream from the Eco RI site (position 1109) and its 3'-end located at the first AAT of the polyadenylation signal. Clones WT-2, WT-3, and WT-5 were all located within the 974 bp of the WT-1 insert. Clone WT-2 started 245 bp downstream of WT-1, clone WT-3 started 132 bp downstream of WT-2, and clone WT-5 started 157 bp downstream from WT-3. This allowed for sufficient overlap to sequence the entire 974 bp of one strand. The 1.1 kb Eco RI subclone from CHP1A was used to sequence the other strand and to confirm the sequence data obtained from the WT-clones. Sequencing of this clone was performed using deletions of the 1.1 kb subclone and oligonucleotide primers. All the deletions that were generated contained large deletions which only allowed for sequencing of the 5'-end of this fragment. Primers AB537 and AB538 were used to sequence the remaining part of the insert (Figure 13).

The other Eco RI subclones were used to sequence the

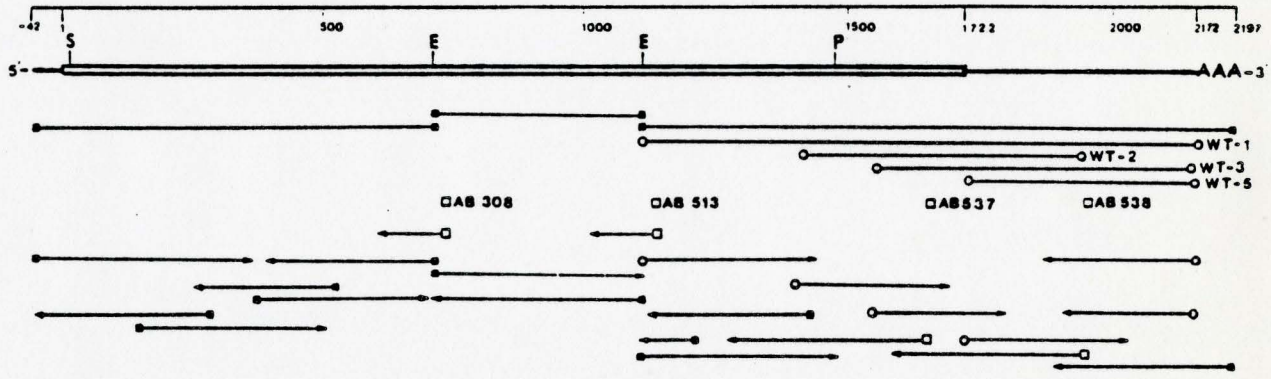


Figure 13 Partial restriction map and the strategy employed for sequencing the P1 cDNA from CHO cells. Clones WT-1, WT-2, WT-3, and WT-5 are from the λgt11 cDNA library prepared from CHO cells. The restriction sites indicated are: E, Eco RI; P, Pst I; and S, Sma I. The subcloned Eco RI fragments from the clone CHP1A are represented by lines terminating on both ends with solid squares. The start points and direction of sequencing are indicated by the arrows; closed squares indicate sequencing of the Eco RI subclones of CHP1A; open circles represent sequence obtained from WT clones; open squares represent sequence obtained using oligonucleotides and either phage DNA (for primers AB308 and AB513) or subclones (for primers AB537 and AB538) of CHP1A. The protein coding region is indicated by the large open rectangle. The map contains numbering for reference and is drawn to scale.

remaining portion of the gene. The two Eco RI subclones of the 0.4 kb fragment were used to sequence this entire insert, in both orientations, utilizing the universal sequencing primer on pTZ18U. For the 0.75 kb fragment, several deletions were sequenced, in addition to the Eco RI clones, to determine the entire sequence of this insert (Figure 13).

Finally, phage DNA was sequenced directly, using primers AB308 and AB513, to sequence across the Eco RI sites within the gene (Figure 13).

3.6 Nucleotide sequence of CHO P1 cDNA

The complete nucleotide sequence of the P1 cDNA is presented in Figure 14. The P1-specific insert of clone CHP1A was 2238 nucleotides in length and contained two internal Eco RI sites. These sites are denoted by arrows in Figure 14. The sequence consists of a single open reading frame of 1761 nts, from the first base after the EcoRI linker (nt -42) to the TAA stop codon (nt 1720 to 1722). The 5'-untranslated region is 42 nucleotides long and the 3'-untranslated region is 450 nucleotides long, followed by a 24 nucleotide poly (A) tail. The poly (A) tail is located 15 nucleotides downstream of the consensus polyadenylation signal of AATAAA at nucleotides 2150-2155 (Proudfoot and Brownlee, 1976).

Although the single open reading frame starts from the

first nucleotide, the first ATG is 42 nucleotides downstream. This was determined to be the initiating codon by two criteria. First, the sequences around the ATG codon are consistent with the consensus sequence (GCCGCCPuCCATG) for initiation of translation in vertebrates (Kozak, 1987). There is a purine at position -3, a G at positions -3 and-6, a C at positions -5, -7, and -8, and an absence of T in the first 17-20 nucleotides upstream of the ATG codon. The leader sequence of 42 nucleotides is consistent for the size range of 20 to 100 nucleotides found for most vertebrate mRNAs (Kozak, 1987).

Second, the human 5'-untranslated sequence is identical to the CHO sequence for the first 17 nucleotides upstream of the ATG and shows 76% homology over the entire 5'-untranslated region. A primer was constructed for the human 5'-untranslated sequence for primer extension studies. It was determined from these studies that the human cDNA clone isolated was a full-length clone (Jindal et al., 1989). This suggests that CHP1A also contains the entire coding sequence of this protein.

The nucleotide sequence shows approximately 91% identity to the human sequence. Many of the nucleotide changes were in the first 50 nucleotides and in the 3'-untranslated region, where there are some small (1-3 nt) deletions in the CHO sequence. Other changes are mainly restricted to the third position of a codon and only 18

Figure 14 **Complete nucleotide and deduced amino acid sequence of CHO P1 cDNA.** The numbers on the left refer to the nucleotide sequence beginning with 1 at the first base of the putative initiation codon ATG. The numbers on the right correspond to the amino acid sequence of the protein. The putative mitochondrial targeting presequence at the N-terminal end (amino acids 1-26) and the polyadenylation signal (nucleotides 2150-2155) are underlined. The positions of the internal Eco RI sites in the sequence is marked with an arrowhead.

```
-62                                                   GGC CTG CTT CGC CTC GTG CCC TCC CGC CGC CGC CCC GCA GAA

   1 ATG CTT CGA CTA CCC ACA GTC CTT CGC CAG ATG AGA CCA GTG TCC CGG GCG CTG GCT CCT CAT CTC ACT CGG GCC TAT GCC AAA GAT GTA
     Met Leu Arg Leu Pro Thr Val Leu Arg Gln Met Arg Pro Val Ser Arg Ala Leu Ala Pro His Leu Thr Arg Ala Tyr Ala Lys Asp Val  30
             +                       +                       +                             +

  91 AAA TTT GGT GCA GAT GCT CGA GCC TTA ATG CTT CAA GGT GTA GAC CTT TTA GCG GAT GCT GTA GCT GTT ACA ATG GGG CCA AAG GGA AGA
     Lys Phe Gly Ala Asp Ala Arg Ala Leu Met Leu Gln Gly Val Asp Leu Leu Ala Asp Ala Val Ala Val Thr Met Gly Pro Lys Gly Arg  60

 181 ACA GTG ATT ATT GAA CAG AGT TGG GGA AGT CCC AAA GTA ACA AAA GAT GGG GTC ACT GTT GCA AAG GCA ATT GAT CTA AAG GAT AAA TAC
     Thr Val Ile Ile Glu Gln Ser Trp Gly Ser Pro Lys Val Thr Lys Asp Gly Val Thr Val Ala Lys Ala Ile Asp Leu Lys Asp Lys Tyr  90

 271 AAA AAT ATC GGA GCT AAA CTT GTT CAA GAT GTT GCC AAT AAC ACA AAT GAA GAG GCT GGG GAT GGC ACT ACC ACT GCT ACT GTT CTG GCA
     Lys Asn Ile Gly Ala Lys Leu Val Gln Asp Val Ala Asn Asn Thr Asn Glu Glu Ala Gly Asp Gly Thr Thr Thr Ala Thr Val Leu Ala 120

 361 CGT TCT ATT GCC AAG GAA GGC TTT GAG AAG ATC AGC AAA GGG GCT AAT CCA GTA GAA ATC CGG AGA GGT GTG ATG TTG GCT GTT GAT GCT
     Arg Ser Ile Ala Lys Glu Gly Phe Glu Lys Ile Ser Lys Gly Ala Asn Pro Val Glu Ile Arg Arg Gly Val Met Leu Ala Val Asp Ala 150

 451 GTA ATT GCT GAA CTG AAA AAA CAA TCT AAA CCT GTG ACA ACC CCT GAA GAA ATT GCT CAG GTT GCT ACA ATT TCT GCG AAT GGA GAC AAA
     Val Ile Ala Glu Leu Lys Lys Gln Ser Lys Pro Val Thr Thr Pro Glu Glu Ile Ala Gln Val Ala Thr Ile Ser Ala Asn Gly Asp Lys 180

 541 GAC ATT GGA AAC ATC ATC TCT GAT GCA ATG AAA AAG GTT GGA AGG AAA GGC GTC ATC ACA GTG AAG GAT GGA AAA ACC CTG AAT GAT GAG
     Asp Ile Gly Asn Ile Ile Ser Asp Ala Met Lys Lys Val Gly Arg Lys Gly Val Ile Thr Val Lys Asp Gly Lys Thr Leu Asn Asp Glu 210
                                                                                                                  ▼
 631 TTA GAA ATT ATT GAA GGC ATG AAG TTT GAC AGA GGA TAT ATT TCC CCG TAT TTT ATT AAC ACA TCA AAA GGT CAA AAA TGT GAA TTC CAA
     Leu Glu Ile Ile Glu Gly Met Lys Phe Asp Arg Gly Tyr Ile Ser Pro Tyr Phe Ile Asn Thr Ser Lys Gly Gln Lys Cys Glu Phe Gln 240

 721 GAT GCC TAT GTT CTG TTG AGT GAA AAG AAA ATT TCT AGT GTC CAG TCC ATT GTA CCT GCT CTT GAA ATT GCT AAT GCT CAT CGT AAG CCC
     Asp Ala Tyr Val Leu Leu Ser Glu Lys Lys Ile Ser Ser Val Gln Ser Ile Val Pro Ala Leu Glu Ile Ala Asn Ala His Arg Lys Pro 270

 811 TTG GTC ATT ATT GCT GAA GAT GTT GAT GGA GAA GCT CTA AGC ACA CTG GTT TTG AAC AGG CTA AAA GTT GGT CTT CAG GTG GTA GCA GTC
     Leu Val Ile Ile Ala Glu Asp Val Asp Gly Glu Ala Leu Ser Thr Leu Val Leu Asn Arg Leu Lys Val Gly Leu Gln Val Val Ala Val 300

 901 AAA GCT CCA GGT TTT GGG GAC AAT AGG AAG AAT CAG CTT AAA GAT ATG GCT ATT GCT ACT GGT GGT GCG GTG TTT GGA GAA GAG GGG CTA
     Lys Ala Pro Gly Phe Gly Asp Asn Arg Lys Asn Gln Leu Lys Asp Met Ala Ile Ala Thr Gly Gly Ala Val Phe Gly Glu Glu Gly Leu 330

 991 AAT CTA AAT CTT GAA GAT GTT CAA GCT CAT GAC TTG GGA AAA GTT GGA GAG GTC ATT GTC ACC AAA GAT GAC GCC ATG CTT TTG AAA GGA
     Asn Leu Asn Leu Glu Asp Val Gln Ala His Asp Leu Gly Lys Val Gly Glu Val Ile Val Thr Lys Asp Asp Ala Met Leu Leu Lys Gly 360
                                             ▼
1081 AAG GGT GAG AAA GCT CAG ATT GAG AAA CGA ATT CAA GAA ATC ACT GAG CAG CTA GAA ATC ACA ACT AGT GAA TAT GAA AAG GAA AAA CTA
     Lys Gly Glu Lys Ala Gln Ile Glu Lys Arg Ile Gln Glu Ile Thr Glu Gln Leu Glu Ile Thr Thr Ser Glu Tyr Glu Lys Glu Lys Leu 390

1171 AAC GAA AGA CTC GCT AAA CTT TCA GAT GGA GTA GCA GTG TTG AAG GTT GGA GGA ACA AGT GAT GTT GAA GTG AAT GAG AAG AAA GAC AGA
     Asn Glu Arg Leu Ala Lys Leu Ser Asp Gly Val Ala Val Leu Lys Val Gly Gly Thr Ser Asp Val Glu Val Asn Glu Lys Lys Asp Arg 420


1261 GTT ACA GAT GCT CTC AAT GCT ACA AGA GCA GCT GTT GAA GAA GGC ATT GTT TTA GGA GGG GGC TGT GCT CTG CTT CGA TGC ATC CCA GCC
     Val Thr Asp Ala Leu Asn Ala Thr Arg Ala Ala Val Glu Glu Gly Ile Val Leu Gly Gly Gly Cys Ala Leu Leu Arg Cys Ile Pro Ala 450

1351 TTG GAT TCC TTA AAG CCT TCT AAT GAA GAT CAG AAA ATA GGT ATA GAA ATT ATT AAA AGA GCA CTC AAA ATT CCT GCA ATG ACA ATT GCT
     Leu Asp Ser Leu Lys Pro Ser Asn Glu Asp Gln Lys Ile Gly Ile Glu Ile Ile Lys Arg Ala Leu Lys Ile Pro Ala Met Thr Ile Ala 480

1441 AAG AAT GCA GGT GTT GAA GGA TCT TTG ATA GTT GAG AAA ATT CTG CAG AGT TCC TCG GAA ATT GGT TAT GAT GCT ATG CTC GGA GAT TTT
     Lys Asn Ala Gly Val Glu Gly Ser Leu Ile Val Glu Lys Ile Leu Gln Ser Ser Ser Glu Ile Gly Tyr Asp Ala Met Leu Gly Asp Phe 510

1531 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 TCC TTG CTA ACT
     Val Asn Met Val Glu Lys Gly Ile Ile Asp Pro Thr Lys Val Val Arg Thr Ala Leu Leu Asp Ala Ala Gly Val Ala Ser Leu Leu Thr 540

1621 ACA GCA GAA GCT GTA GTG ACC GAA ATT CCT AAA GAA GAG AAG GAC CCT GGA ATG GGC GCA ATG GGT GGA ATG GGA GGG GGT ATG GGA GGT
     Thr Ala Glu Ala Val Val Thr Glu Ile Pro Lys Glu Glu Lys Asp Pro Gly Met Gly Ala Met Gly Gly Met Gly Gly Gly Met Gly Gly 570

1711 GGC ATG TTC TAA TTC CTA GAA TAG TGC TTT GCC CTT ATC AAT GAA CTG TGA CAG GAA GCT CAA TGC AGG TTC CTC GGT GAT AAC TTA AGA
     Gly Met Phe ...                                                                                                         573
1801 GAA GTC ACC TGA AGA AAA TGA AGA GAA GGC TGG CTG ATC ACT ATA ACC ATC AGT TAC TGG TTC CCA TTG ACA ATA CAT AAT GGT TTA CTG
1891 CTG TCA TTG TCC ATG CCT ACA GAT AAT TTA TTT TGT ATT TTT GAA TAA AGA CAT TTG TAC ATT CCT GAT ACT GGG TGC AAG AGC CAT ATA
1981 CCA GTG TCC TGC TTT CAA CTT AAA TCA TTG AGG CAG TTC TAC TAT TCT GTT ACA GTC AGG ATT CTA GTG CTT GCG TGA CCA GAT GAG TTC
2071 AGA AGC AGC CTT TCT GTG GAG AGT GAG AAT AAT TGT GTA CAA AGT AGA GAA GTA TCC AAT TAT GTG ACA ACT TTT GTG TAA TAA AAT TTT
2161 GTT TAA AAG TTA AAA AAA AAA AAA AAA AAA AAA AAA
```

nucleotide differences result in amino acid changes.

3.7 Deduced Amino Acid Sequence

Beginning at the first ATG (nt 1 to 3) the single open reading frame encodes a protein of 573 amino acids with a calculated molecular weight of 60,983 daltons. Based on the similarity to the human protein the first twenty-six amino acids appear to fit the requirements for a mitochondrial presequence (see Section 4.2). Provided that the mitochondrial presequence is cleaved after codon 26 (Tyr) then the mature protein is 547 amino acids with a calculated molecular weight of 57,949 daltons. This molecular weight is in close agreement with that predicted for the human P1 sequence, and estimates made from polyacrylamide gels (Gupta et al., 1985; Jindal et al., 1989). The mature protein has a calculated pI of 6.14 which is in agreement with that estimated from 2-D polyacrylamide gels (Gupta et al., 1982). The complete amino acid sequence of the protein is displayed in Figure 14.

3.8 Amino Acid Alignment with Other Chaperonins

The deduced amino acid sequence from the CHO cDNA for P1 was compared at the amino acid level to the human P1 protein, as well as, the 65 kDa antigen of M. leprae (Mehra et al., 1986) and M. bovis BCG (Shinnick, 1987), the E. coli groEL protein (Hemmingsen et al., 1988), the Rubisco large

subunit binding protein from wheat (Triticum aestivum) chloroplasts (Hemmingsen et al., 1988) and the 62 kDa protein of C. burnetii (Vodkin and Williams, 1988). Figure 15 outlines the homology between these proteins.

The deduced amino acid sequence of P1 from CHO cells is 97% identical to the human sequence. They only differ in 18 amino acids, two of which are in the putative presequence. Nine of the other changes are conservative substitutions.

The other proteins showed amino acid identity that ranged from 42%, for Rubisco binding protein, up to 50% for the 62 kDa antigen of C. burnetti. Conservative changes account for an additional 20-25% similarity. Due to differences in the length of these proteins, small insertions were included to maximize the homology to the P1 protein. These insertions are represented by triangles in Figure 15. These values are summarized in Table 4.

Figure 15 **Comparison of the sequence of mammalian P1 protein with the related protein from other species.** Lane 1, CHO P1 sequence; Line 2, human P1 (Jindal et al., 1989); Line 3, 65-kDa antigen from M. leprae (Mehra et al., 1986); Line 4, 65-kDa antigen from M. bovis BCG (Shinnick, 1987); Line 5, E. coli groEL protein (Hemmingsen et al., 1988); Lane 6, Rubisco large subunit binding protein from wheat (Triticum aestiverm) chloroplasts (Hemmingsen et al., 1988); Lane 7, 62-kDa antigen from C. burnetii (Vodkin and Williams, 1988). The amino acid residues that are identical to the CHO P1 sequence are shown by a dash. Any addition and/or deletions in the sequence alignment are indicated by a filled triangle.

```
CHO          MLRLPTVLRQMRPVSRALAPHLTRAYAKDVKFGADARALMLQGVDLLADAVAVTMGPKGRTVIIEQSWGS   70
HUM          -------F--------V-------------------------------------------------------   70
M. leprae                            M--TIAYDEE--RGLER-LNS-----K--L-----N-VL-KK--A   45
BCG                                  M--TIAYDEE--RGLER-LNA-----K--L-----N-VL-KK--A   45
GROEL                                MA-------N---VK--R--NV-----K--L-----N-VLDK-F-A   46
RUBISCOBS                            G-D--EIA-DQKS--ALQA--EK--N--G--L--R--N-VLDEYA-N   46
C. burnetii                          MA--VL--SHEVLNA-SR--EV--N--K--L-----N-VLDK-F-A   46

             PKVTKDGVTVAKAIDLKDKYKNIGAKLVQDVANNTNEEAGDGTTTATVLARSIAKEGFEKISKGANPVEI  140
             ----------S-------------------------------------------------------------  140
             -TI-N---SI--E-E-E-P-EK---E--KE--KK-DDV-----------QALV---LRNVAA----LGL  115
             -TI-N---SI--E-E-E-P-EK---E--KE--KK-DDV-----------QALVR--LRNVAA----LGL  115
             -TI-----S--RE-E-E--FE-N--QN-KE--SKA-DA----NN------QA-IT--LKAVAA-N--NDL  116
             ---VN----I-R--E-ANPNE-A--A-IRE--SK--DS--------C----E-I-L-ILSVTS-----SL  116
             -TI-----S---E-E-E--FE-N--QN-KE--SR-SDD-----------QA-LV--IKAVIA-N--NDL  116

             RRGVMLAVDAVIAELKKQSKPVTTPEEIAQVATISANGDKDIGNIISDAMKKVGRKGVITVKDGKTLNDE  210
             ----------------------------------------E-------------------------------  210
             K--IEK---K-TET-L-DA-E-E-K-Q--AT-A---A--QS--DL-AE--D---NE-----EESN-FGLQ  184
             K--IEK--EK-TET-L-GA-E-E-K-Q--AT-A---A--QS--DL-AE--D---NE-----EESN-FGLQ  184
             K--IDK--T-AVE---AL-V-CSDSKA----G----TS-ETV-KL-AE--D---KE-----E--TG-Q--  186
             KK-IDKT-QGL-E--ERKAR--KGSGD-KA---S---GN-EL--AN-A--ID---PD--LSIESSSSFETT  186
             K--IDK--T-AV-----I---CKDQKA----G-----S--S--D--AE--E---KE-----E--SG-ENA  186

             LEIIEGMKFDRGYISPYFINTSKGQKCEFQDAYVLLSEKKISSVQSIVPALEIANAHRKPLVIIAEDVDG  280
             ----------------------------------------I-------------------------------  280
             --LT---R--K----G--VTDAER-EAVLEEP-I--VSS-V-T-KDLL-L--KVIQAG-S-L------E-  254
             --LT---R--K----G--VTDPER-EAVLE-P-I--VSS-V-T-KDLL-L--KVIGAG---L------E-  254
             -DVV---Q-----L------KPETGAV-LESPFI--AD----NIREML-V--AVAKAG---L------E-  256
             VDVE---EI-------Q-VTNLEKSIV--EN-R--ITDQ--T-IKE-I-L--QTTQL-C--F-V---IT-  256
             --VV---Q-----L------NQQNMSA-LENPFI--VD----NIRELI-L--NVAKSGR--LV----IE-  256

             EALSTLVLNRLKVGLQVVAVKAPGFGDNRKNQLKDMAIATGGAVFGEEGLNLNLEDVQAHDLGKVGEVIV  350
             ----------------------------------------T-------P-----------------------  350
             -------V-KIRGTFKS----------R--AN-Q----L--AQ-IS--AVG-T--NTDLSL---ARK-VN  323
             -------V-KIRGTFKS----------R--AN-Q----L---Q-IS--AVG-T--NADLSL---ARK-V-  323
             ---A-A-V-TIRGIVK-A---------R--AN-Q-I-TL---T-IS--AIGNE--KATLE---QAKR-VI  325
             ---A---V-K-RGIIN-A-I---S--ER--AV-Q-I--V---AEYLAKDA-G-LV-NATVDQ--TARKITI  325
             ---A---V-NIRGVVK-A---------R--AN-Q-I-VL---K-IS--AVG-S--AASLD---SAKR-V-  325

             TKDDAMLLKGKGEKAQIEKRIQEITEQLEITTSEYEKEKLNERLAKLSDGVAVLKVGGTSDVEVNEKKDR  420
             ----------D----------I---DV-------------------------------------------  420
             ---ETTIVE-A-DTDA-AG-VAQ-RTEI-NSD-D-DR---Q------AG----I-A-AATE--LK-R-N-  393
             ---ETTIVE-A-DTDA-AG-VAQ-RQEI-NSD-D-DR---Q------AG----I-A-AATE--LK-R-N-  393
             N--TTTIID-V--E-A-QG-VAQ-RQ-I-EA--D-DR---Q--V---AG----I---AATE--NK---A-  395
             HQTTTT-IADAAS-DE-QA-VAQLKKE-SE-D-I-DS---A--I----G----I---A-TET-LEDRQL-  395
             ----TTIID-S-DAGD-KN-VEQ-RKEI-NSS-D-D----Q------AG---I---AATE--NK---A-  395

             VTDALNATRAAVEEGIVLGGGCALLRCIPALDSLKPSNADEQKIGIEIIKRALKIPAMTIAKNAGVEGS  490
             ----------------------------T-A-AA----------T----------------------  490
             IE--VRNAK--------A---VT--QAA----K--LTGAAA-EAT-AN-V-V--EA-LKQ--F-S-N-PG  462
             IE--VRNAK--------A---VT--QAA-T--E--LEGAAA-EAT-AN-V-V--EA-LKQ--F-S-L-PG  462
             -E---N---------V-A---V--I-VASK-AD-RGQ-AA----NV--KVAL--MEA-LRQ-VL-C-E-P-  465
             IE--K---F--I-----P---A-YVHLSTYVPAI-ETIEDH-ERL-AD--QK--QA--SL--N------E  467
             -E---N---------V-P---V--I-VLKS---VEVE-AA----RV-V--AR--MAY-LSQ-V--T--QAA  465

             LIVEKILQASSSEIGYDAMLGDFVNMVEKGIIDPTKVVRTALLDAAGVASLLTTAEAVVTEIPKEEKDPG  560
             ------N-A----V-----A----------------------------------V-------------  560
             VVA--VRNAL-VGH-LN-AT-EYEDLLKA-VA--V--T-S--QN--SI-G-FL-T----ADK-EKTAA-A  532
             VVA--VRNALPAGH-LN-QT-VYEDLLAA-VA--V--T-S--QN--SI-G-FL-T----ADK-EK--ASV  532
             VVANTVKGAGDGHY--N-ATEEYG--IDN--L-----T-S--QY--S--G-MI-T-CM--DL--NDAADL  535
             VVI---KEA-EW-N--N--TDKYE-LI-S-V---A--T-C--QN--S-SGNVL-TQ-I-V-K--PKPKVA  537
             VVAD-V-NNKDVNY--N-AT-EYGD-I-N--L-----T----QN--SI-G-MI-T-CM---A--KKEESN  536

             NGANGGNGGGNGGGNF                                                        573
             ----------------                                                        573
             SDPT-----NDF                                                            541
             P-AG-D---NDF                                                            540
             GA-G-A---NG-N-GNN                                                       548
             EP-E-QLSV                                                               543
             P-AG-D---NG-N-GNGGNN                                                    552
```

Organism[a]	Human	Hamster	C. burnetii	E. coli	M. leprae	chloroplast
Human (P1)		96.7[c]	50.1	47.7	46.7	42.2
Hamster (P1)	98.4[b]		50.0	47.5	46.8	42.4
C. burnetii	72.2	72.3		75.0	57.8	47.3
E. coli	69.2	69.5	84.6		57.4	46.5
M. leprae	67.0	67.6	76.1	75.0		48.4
Plant chloroplast	65.4	66.0	70.1	70.5	69.7	

a The P1-related protein from C. burnetii was identified as an antigenic 62 kDa heat-shock protein (Vodkin and Williams, 1988), from E. coli as the product of the GroEL gene (Hemmingsen et al., 1988), from M. leprae as the 65 kDa antigen (Mehra et al., 1986) and from plant chloroplasts as the rubisco binding protein (Hemmingsen et al., 1988).

b The lower triangle indicates percent similarity of the sequences as determined by the BESTFIT program of the University of Wisconsin, GCG software package, using the default settings.

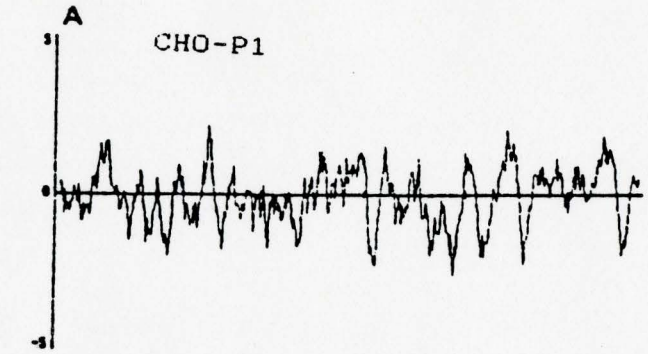
c The upper triangle indicates amino acid sequence homology over the entire length.

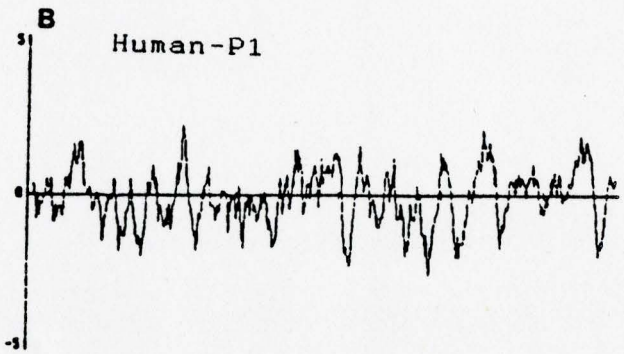
Table 4 Matrix of similarity based on the sequence of P1-related proteins from different species (taken from Gupta et al., submitted).

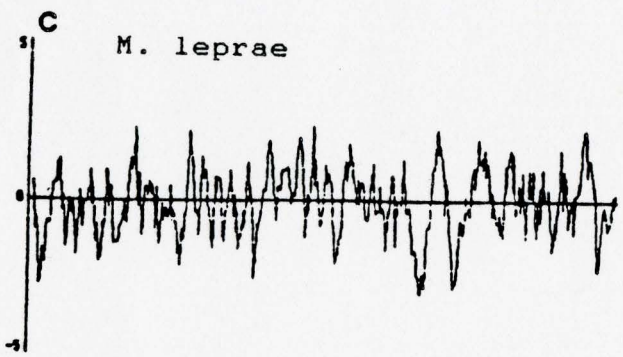
3.9 Hydrophilicity Plots

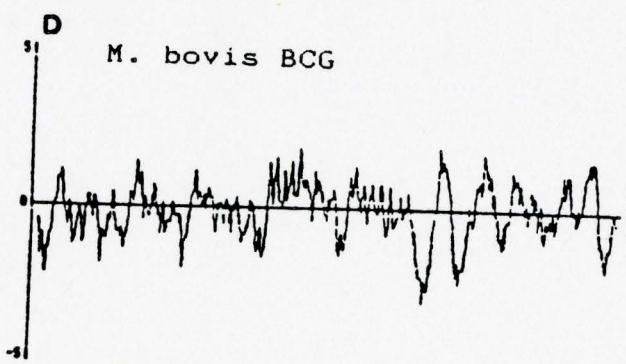
The amino acid hydrophilicity plots of the seven proteins were examined for any significant differences (Figure 16). The plots were calculated using the method of Kyte and Doolittle (1982), with a window of seven amino acids. The plots are all very similar and show no long hydrophobic stretches. This is consistent with the solubility of these proteins.

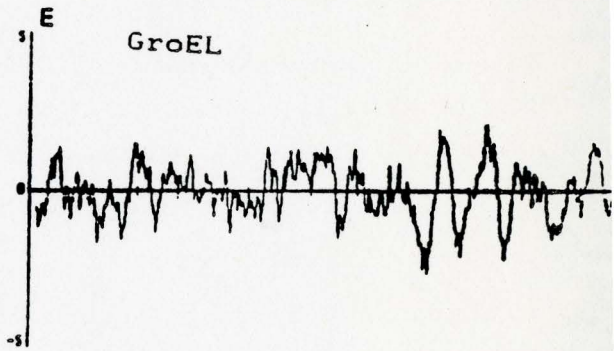
Figure 16 **Hydrophilicity plots for the P1 protein and related proteins.** (A) CHO P1 protein; (B) human P1 protein; (C) 65-kDa protein antigen from M. leprae; (D) 65-kDa antigen from M. bovis BCG; (E) groEL protein from E. coli; (F) Rubisco binding protein from wheat (Triticum aestiverm) chloroplasts; (G) 62-kDa antigen from C. burnetii. An amino acid window of seven was used to calculate the values by the method of Kyte and Doolittle, 1982.

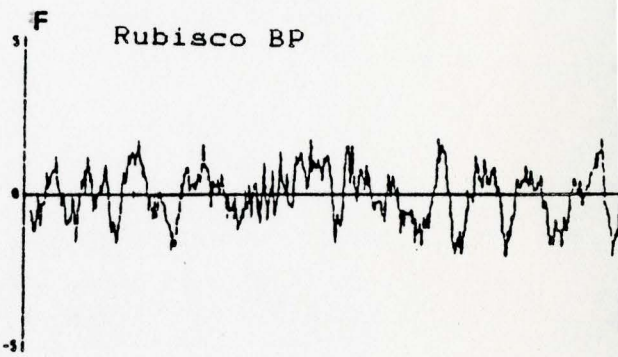

A
CHO-P1

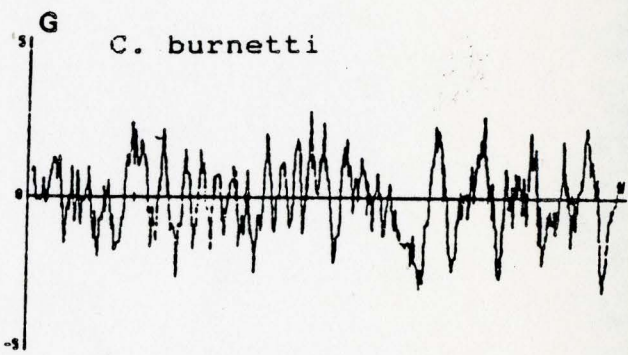

B
Human-P1


C
M. leprae


D
M. bovis BCG


E
GroEL


F
Rubisco BP


G
C. burnetti

4.0 DISCUSSION

4.1 cDNA Library Construction

The main focus of this project was to construct a cDNA library from CHO cells to isolate the cDNA clone for the P1 protein. The cDNA synthesized from CHO RNA resulted in cDNA fragments which ranged in size from 300 bp to greater than 3 kbp in length (Figure 5). This is in agreement with Sartoris and coworkers, who found that 60-80% of their cDNA yield was 0.8-2.0 kb in length (Sartoris et al., 1987). Since the transcript of interest was approximately 2.3 kb and relatively abundant, the cDNA synthesized likely contained a full-length copy of this transcript. Similarly, the completed library was 1.5 x 10^5 pfu/ug of cDNA which is only slightly smaller than the average library size of 5 x 10^5 pfu/ug of cDNA produced by similar techniques reported in the literature (Gubler and Hoffman, 1983; Okayama and Berg, 1982; Pape and Kim, 1987; Sartoris et al., 1987). However, the library, which yielded several P1 clones, contained only partial clones, the majority of them 3'-end fragments. From these results improvements can be suggested for future cDNA library constructions.

Since the largest clone obtained from this library was under 1 kb in length it suggests that either the RNA was slightly degraded or the secondary RNA structures were

not denatured completely. To eliminate secondary structures the samples should be treated at 70°C for 15 or 20 minutes, then cooled quickly on ice. Any secondary structures that remain could cause the reverse transcriptase to fall off the RNA template before reaching the 5'-end of the transcript.

One other point that should be noted with respect to the construction of this library concerns the Eco RI methylation procedure. The largest P1-related fragment cloned was from a library constructed using $Col^{r}22a$ cells (results not shown). This clone was 1.1 kb in length with its 5'-end located at an Eco RI site (nucleotide 1108). This suggests that Eco RI methylation was incomplete. Since this procedure was also used for the construction of the CHO library it may also explain why the clones obtained were not full-length. One way to avoid this would be to run a test reaction using pTZ18U as the sample. If the pTZ18U plasmid is digested by Eco RI after Eco RI methylation then the reaction conditions could be altered until effective concentrations of Eco RI methylase and S-adenosyl methionine were determined.

Once a partial clone was obtained, an alternative to constructing a new library or obtaining the well characterized λgt10 library, would have been to use primer extension studies. Synthesizing an oligonucleotide near the 5'-end of the clone would provide a primer for the

construction of a new cDNA library that would be specific for the P1 gene, and yield clones containing information upstream from the primer. Since the human cDNA sequence was obtained before the λgt10 library was screened, another approach could have been to construct a cDNA library using the polymerase chain reaction (Frohman et al., 1988). Instead of using oligo d(T) as the only primer this procedure would utilize an additional oligonucleotide (for the 5'-end) which would have to be designed from the human sequence.

4.2 The P1 Sequence Contains a Mitochondrial Presequence

The majority of mitochondrial proteins are nuclear coded, synthesized as precursors with amino-terminal presequences, and imported into the mitochondria (Hurt and van Loon, 1986). Although there is no consensus presequence conserved amongst these proteins, there are common characteristics. Matrix and inner membrane proteins usually have several positively charged amino acids separated by three to five uncharged residues over the entire presequence. They are also rich in hydroxylated amino acids and devoid of acidic residues. Proteins destined for the intermembrane space have, in addition to the amino-terminal sequence, a long stretch of hydrophobic amino acids adjacent to the amino-terminal sequence. These two types of presequences are cleaved when the protein

enters the mitochondria. The outer membrane proteins have a similar presequence as the intermembrane space proteins except that they lack a cleavage site (Hurt and van Loon, 1986).

Earlier work on the P1 protein localized it to the mitochondrial matrix by cell fractionation studies (Gupta and Austin, 1987). From this work the cDNA for P1 was expected to encode a mitochondrial targeting sequence at the amino-terminal end which would direct the P1 protein to the matrix.

The first 26 codons downstream of the initiating ATG codon for the P1 cDNA do encode amino acids that fit the requirements for a mitochondrial presequence (Hurt and van Loon, 1986). This sequence lacks acidic amino acids and contains several basic and hydroxyl group containing residues in a predominantly hydrophobic region.

One other feature of mitochondrial presequences is that most of them form an amphiphilic alpha-helix with positively charged residues on one side of the helix and hydrophobic residues on the opposite face (Roise and Schatz, 1988). To determine if amino acids 1-26 form an amphiphilic alpha-helix, a helical wheel plot was drawn (Figure 17a). This resulted in an amphiphilic alpha-helix with the arginine residues at positions 9, 12, 16, and 24 located on the same face, across from the hydrophobic residues. This plot is nearly identical to the human

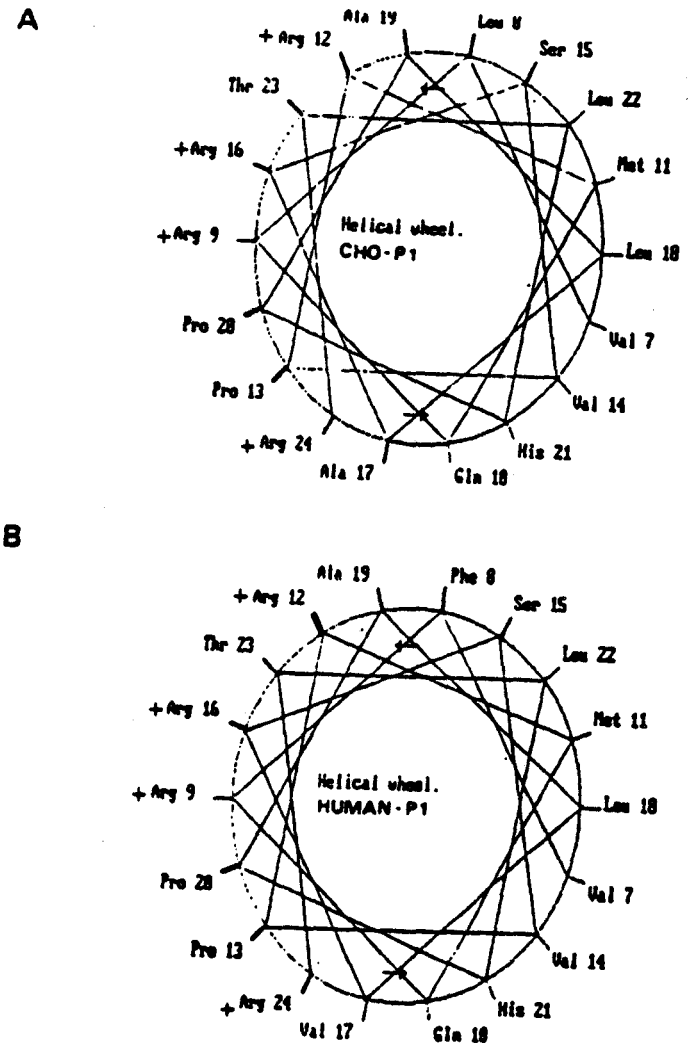


Figure 17 **Helical wheel diagram of the putative mitochondrial presequence.** Comparison of CHO (A) and human (B) mitochondrial presequences of the P1 protein. +, positively charged amino acids.

helical wheel plot for the P1 protein which is presented in Figure 17b.

Further evidence that the first 26 amino acids represent a targeting sequence comes from closer examination of the amino acid sequence alignment (Figure 15). The other chaperonin proteins show the greatest sequence homology to the human and hamster P1 sequences when the alignment begins from codon 27. Since most matrix proteins have amino-terminal presequences this result also provides strong supporting evidence that clone CHP1A contains the entire coding sequence.

4.3 The P1 Protein Shows Homology to Molecular Chaperones

The amino acid sequence alignment (Figure 15) clearly demonstrates that there is sequence similarity to the chaperonin class of molecular chaperones. The P1 protein shows 40-50% identity with these proteins and an additional 20-25% similarity based on conservative amino acid substitutions.

Closer examination of the seven sequences showed that there were short regions where there was very high identity in comparison to the overall similarity. These regions of high conservation are located at amino acids 109-120, 215-230, 271-289, 297-310, 383-397, and 423-437 in the hamster sequence. The amino acid changes in these regions are very conservative changes such as leucine for isoleucine,

aspartic acid for glutamic acid, and asparagine for glutamine. Several of these changes result in the hamster and human sequences having one amino acid while the other sequences all contain the alternate residue. These regions most likely represent functional domains of the protein.

Another interesting feature of the primary sequence was the carboxy-terminal region. With the exception of the Rubisco binding protein all the chaperonin proteins contained a Gly-Gly-Met repeating motif at the carboxy-terminus. It has not been determined what significance, if any, this repeating motif has on the functions of these proteins. However, intracellular localization studies of the 65 kDa protein in M. leprae have suggested that the protein may be membrane associated (Gillis et al., 1985). It has been postulated from this data that the repeating motif may serve as an anchor, to the membrane or peptidoglycan fraction of the 65 kDa antigen and for the C. burnetii protein (Vodkin and Williams, 1988). However, the hydropathy profiles of these proteins do not appear to contain any hydrophobic membrane-spanning stretches (Figure 16).

4.4 The P1 Protein and Chaperonins Have Similar Properties

A comparison of the properties and amino acid composition of the P1 protein with the other chaperonin proteins is presented in Table 5. The mature proteins all

have similar relative molecular weights. They are all slightly acidic and show conservation in the number of aromatic and hydrophobic residues they contain. The hamster P1 protein has one more basic residue than the human P1 protein which was a result of a single nucleotide change. A lysine codon, AAG, in the hamster protein has been changed to a threonine codon, ACG, in the human sequence at residue 455. The two amino acid changes in the presequence of the hamster P1 protein decreased the number of aromatic residues and hydrophobic residues by one. An additional substitution of a valine for an alanine in the hamster P1 protein (amino acid 557) accounted for the second extra hydrophobic residue present in the human protein.

Recent work in our laboratory has determined that other properties of the P1 protein are also similar to the chaperonins. The P1 protein from both rat liver mitochondria and CHO cells was eluted from a precalibrated gel filtration column. From this experiment and from sucrose density sedimentation experiments the P1 complex was estimated to be approximately 440 kDa (C.S.K. Mayanil, personal communication). Examination of the subunit molecular weight of the purified protein fraction by SDS-polyacrylamide gel electrophoresis showed that the P1 complex consisted of identical subunits (C.S.K. Mayanil, personal communication). This suggested that the P1

Property	CHO	Human	Yeast	GroEL	Rubisco	M. leprae	C. burnetii
Residues	573	573	572	548	543	541	552
M_r P.	60983	61049	60656	-	-	-	-
M_r M.	57949	57939	58088	57203	57454	56827	58300
Basic	76	75	72	62	59	66	67
Acidic	80	80	83	81	81	86	81
Net Negative	-4	-5	-11	-19	-22	-20	-14
Aromatic	17	18	24	14	16	16	14
Hydrophobic	182	184	171	168	165	160	176

Table 5 A comparison of the properties and amino acid composition of the P1 and P1-related proteins. M_r P., relative molecular mass of precursor protein; M_r M., relative molecular mass of mature protein; basic, number of lysine and arginine residues in protein; acidic, number of glutamic and aspartic acid residues in protein; net negative, charge obtained by adding the basic and acidic residues; aromatic, number of phenylalanine, tryptophan, and tyrosine residues in protein; hydrophobic, number of aromatic, isoleucine, leucine, methionine, and valine residues in protein.

complex existed as a homooligomer of seven subunits. This is in agreement with the other chaperonins although they are homooligomeric proteins of either 14 or 6 subunits. The GroEL protein is a homooligomeric complex of 14 subunits arranged into two rings of seven subunits each (Hendrix, 1979; Ellis et al., 1989). The Rubisco large subunit binding protein of plant chloroplasts consists of six alpha and six beta subunits (Ellis and Van Der Vies, 1988).

The chaperonin proteins also show weak ATPase activity and slight induction upon heat shock (Hemmingsen et al., 1988). Preliminary experiments in our laboratory suggested that the P1 protein is induced two- to three-fold upon heat shock (R.S. Gupta, personal communication). In addition, we have recently demonstrated that the P1 protein showed a time- and concentration-dependent hydrolysis of [γ -^{32}P]ATP (R.S. Gupta, personal communication).

Recently a minor heat shock protein from Tetrahymena thermophila was purified (McMullin and Hallberg, 1987). This protein was found to be a nuclear encoded mitochochondrial protein that was induced approximately two- to three-fold during heat shock. This group also observed that the protein sedimented as part of a high molecular weight complex (McMullin and Hallberg, 1987). This protein is likely a homologue of the P1 protein.

4.5 Comparison of the P1 Protein to the Yeast HSP60 Protein

Reading and coworkers have recently reported the cloning and sequencing of the yeast HSP60 gene (Reading et al., 1989). The HSP60 gene codes for a protein of 572 amino acids with a molecular weight of 60,830 daltons. This protein hybridizes to a transcript estimated to be 1.9 kb in size upon Northern hybridization, and is induced 2-3 fold above basal levels when cells are heat-stressed. This gene was found to be the genetic locus of a temperature sensitive lethal mutant, mif4, which is deficient in mitochondrial assembly factor (Cheng et al., 1989; Reading et al., 1989).

The sequence homology between the yeast hsp60 and the P1 protein was examined. The amino acid alignment is presented in Figure 18. The yeast hsp60 protein contains a mitochondrial presequence that is cleaved after amino acid 22. The alignment of the mature proteins showed that 55% of the yeast hsp60 amino acids were identical to the hamster P1 protein. Of the remaining amino acids, an additional 20% were conserved changes. This protein also showed similar identity to the other chaperonin proteins (Picketts et al., 1989; Reading et al., 1989). Other properties of this protein also appeared to be similar to the P1 protein (see Table 5).

```
COMPARING:  CHOP1-AASEQ
            YEAST HSP60 AASEQ

MLRLPTVLRQMRPVSRALAPHLTRAYAKDVKFGADARALMLQGVDLLADAVAVTMGPKGRTVIIEQSWGS  70
---SSV-RSRATLRPLLRRAY**SSH-EL---VEG--SL-K--ET--E---A-L-----N-L---PF-P  70

PKVTKDGVTVAKAIDLKDKYKNIGAKLVQDVANNTNEEAGDGTTTATVLARSIAKEGFEKISKGANPVEI 140
--I---------S-V----FE-M----L-E--SK---A------S----G-A-FT-SVKNVAA-C--MDL 140

RRGVMLAVDAVIAELKKQSKPVTTPEEIAQVATISANGDKDIGNIISDAMKKVGRKGVITVKDGKTLNDE 210
---SQV--EK--EF-SANK-EI--S--------S-----SHV-KLLAS--E---KE----IRE-R--E-- 210

LEIIEGMKFDRGYISPYFINTSKGQKCEFQDAYVLLSEKKISSVQSIVPALEIANAHRKPLVIIAEDVDG 280
--VT---R----F------TDP-SS-V--EKPLL---------I-D-L-----S-QS-R--L-------- 280

EALSTLVLNRLKVGLQVVAVKAPGFGDNRKNQLKDMAIATGGAVFGEEGLNLNLEDVQAHDLGKVGEVIV 350
---AACI--K-RGQVK-C-------------TIG-I-VL---T--T--*-D-KP-QCTIEN--SCDSIT- 350

TKDDAMLLKGKGEKAQIEKRIQEITEQLEITTSE*YEKEKLNERLAKLSDGVAVLKVGGTSDVEVNEKKD 420
--E-TVI-N-S-P-EA-QE--EQ-KGSID---TNS------Q-------G----IR---A-E---G---- 420

RVTDALNATRAAVEEGIVLGGGCALLRCIPALDSLKPSNEDQKIGIEIIKRALKIPAMTIAKNAGVEGSL 490
-YD--------------LP---T--VKASRV--EVVVD-F---L-VD--RK-ITR--KQ-IE---E---V 490

IVEKILQSSSEI**GYDAMLGDFVNMVEKGIIDPTKVVRTALLDAAGVASLLTTAEAVVTEIPKEEKDP 560
-IG-LIDEYGDDFAK----SKSEYTD-LAT-----F----SG-V--S------A-T-VAIVDA-EPAAAA 560

GMGAMGGMGGGMGGGMF                                                     630
-P-G-P-GMP--P-M-*                                                     630
```

Figure 18 Amino acid alignment of the P1 protein from CHO cells and the Yeast hsp60 protein. Line 1, CHO amino acid sequence of the P1 protein; Line 2, amino acid sequence of the yeast hsp60 protein. Identical residues are indicated by a dash. Any additions and/or deletions are indicated by a filled square.

4.6 Proposed Evolution of the P1 Protein

Since the hamster and human P1 proteins are mitochondrial proteins, and they show homology to proteins from diverse organisms it was interesting to co-relate the homology comparisons with the endosymbiotic theory of mitochondrial origin. Prokaryotes are divided into two major classes, eubacteria and archaebacteria. Sequence comparisons of rRNA from many different organisms further subdivided eubacteria into 8 to 10 subgroups that have minimal intragroup variation (Woese et al., 1983; Yang et al., 1985). Two of these subgroups, cyanobacteria and purple bacteria, were proposed to be involved in fusion with a progenitor eukaryotic cell. These fusions gave rise to chloroplasts and mitochondria, respectively. More recently, studies involving the 16S rRNA sequences suggested that the endosymbiont that gave rise to mitochondria was from the alpha-subdivision of purple bacteria (Woese et al., 1983; Yang et al., 1985). The sequence homologies between each pair of the chaperonin proteins was examined in order to construct a dendrogram for these organisms.

The comparison showed that the mammalian proteins were most identical to the hsp 60 protein from yeast and the 62 kDa protein from C. burnetii, and least similar to the plant chloroplast chaperonin (Table 4). A dendrogram was plotted for this data based on the branching order

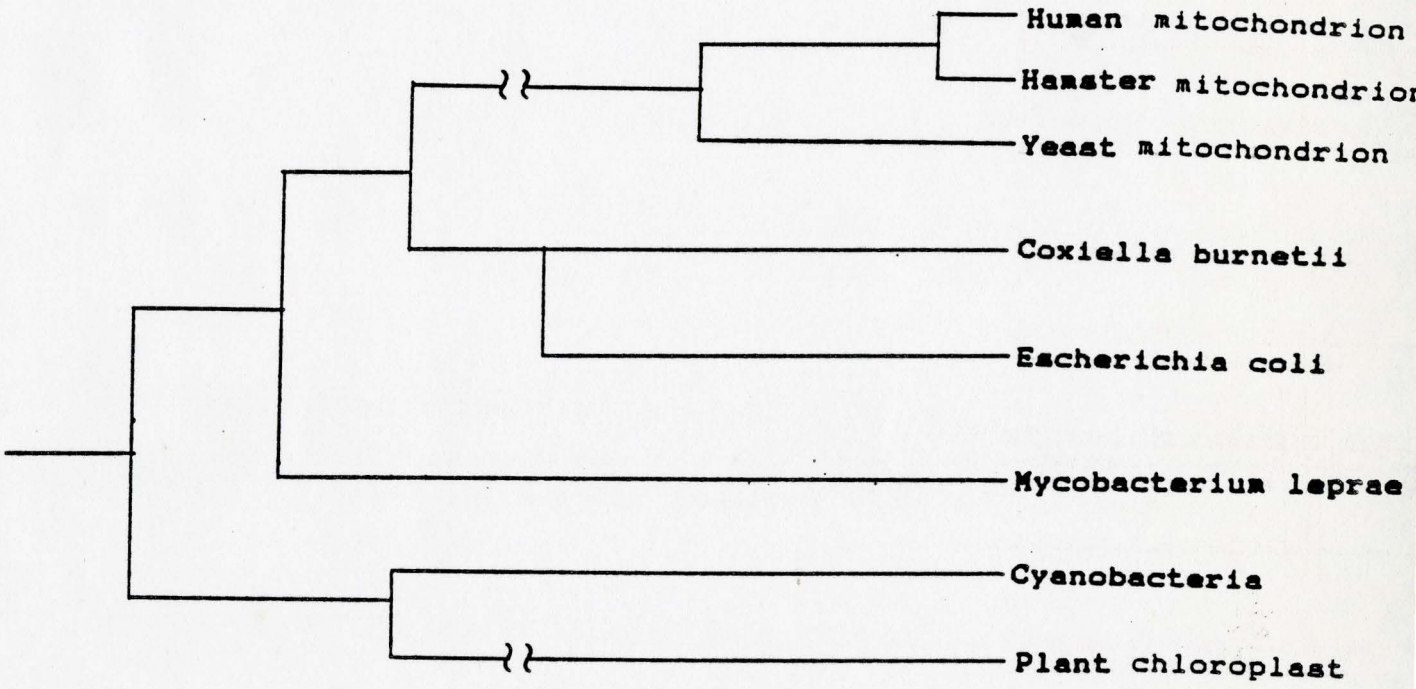


Figure 19 Dendrogram showing the phylogenetic relationship of P1 protein deduced from the sequence analysis. The dendrogram constructed is based on the branching order pattern deduced from small subunit rRNA sequences and the sequence homologies from Table 4 (Woese et al., 1983; Yang et al., 1985).

pattern deduced from small subunit rRNA sequences (Figure 19). The data is in agreement with the endosymbiotic theory for mitochondrial origin as maximum identity of P1 was seen with the corresponding protein from E. coli and C. burnetii, members of the purple bacteria subgroup. Since plant chloroplasts were proposed to evolve from cyanobacteria it is expected that they should show the least identity to the hamster and human P1-protein. The plant chloroplast sequence did show the least identity to the P1 protein so this data is also supportive of the endosymbiotic theory for plant chloroplasts. However, it is also expected that plants should contain a mitochondrial homologue of the P1-protein which should show higher identity to the mammalian P1-protein. This homologue will also probably contain the Gly-Gly-Met sequence at the carboxy-terminus that is absent in the chloroplast protein.

4.7 Proposed Functions of the P1 Protein

At present the exact biological role of the P1 protein in mammalian systems is unclear. The genetic and biochemical studies point to its possible involvement in the in vivo assembly of MTs. However, the molecular cloning of P1 has shown that it is a mitochondrial matrix protein with high identity to chaperonins. This suggests a role in the assembly of such mitochondrial enzyme complexes as pyruvate dehydrogenase, F_1-ATPase, and alpha-

ketoglutarate dehydrogenase. Both possibilities will be addressed.

There are a number of observations that, collectively, suggest a role for P1 in the in vivo assembly of MTs. First, a large number of mutants of CHO cells resistant to podophyllotoxin have been obtained (Gupta et al., 1982; Gupta and Gupta, 1984). These mutants exhibit highly specific cross-rsistance patterns toward other MT inhibitors, and a high frequency (23/26) of Pod^{RII} mutants show an electrophoretic alteration in the P1 protein (Gupta et al., 1982; Gupta and Gupta, 1984; Gupta et al., 1985). This suggests that the genetic lesion in these mutants is MT-specific.

Second, the P1 protein was found to be co-released with tubulin from crude MT preparations, under a number of conditions that cause MT depolymerization. From this data it was suggested that P1 was a MAP (Gupta and Gupta, 1984). An independent study identified the P1 protein as an intermediate filament-associated protein (Mose-Larsen et al., 1982). These studies suggest that P1 somehow interacts with the cytoskeleton.

Third, binding studies with fluorescently labelled colchicine and podophyllotoxin derivatives, which bind free tubulin, showed that in interphase cells the drug binds mainly to mitochondria (Gupta and Dudani, 1989; Moll et al., 1982). Other reports have shown that tubulin shows

specific associations with mitochondrial membranes (Bernier-Valentia et al., 1983; Hargreaves and Avila, 1985). This suggests that the mitochondrial membrane may be the main cellular receptor for free tubulin.

Fourth, mitochondria show a close association with both MTs and intermediate filaments in interphase cells (Ball and Singer, 1982; Gupta et al., 1985; Gupta and Dudani, 1987; Schnapp et al., 1985). It has been suggested that MTs help determine mitochondrial orientation and distribution. However, when MTs are depolymerized by treatment with colchicine or podophyllotoxin the mitochondrial distribution is not altered. When the drug is removed MT reassembly appears to follow the mitochondrial distribution (R.S. Gupta, unpublished results).

These observations,in conjunction with the chaperonin function, suggests that the P1 protein may be involved in MT assembly. Although the cellular localization of P1 makes this difficult to reconcile, it may be speculated that the P1 protein plays a role in MT assembly before entering the mitochondria. This role may be to transport the tubulin monomers to the mitochondrial membrane. At the membrane, tubulin dimers could be formed and growth of the MT would begin from this point.

Although the P1 protein may have some indirect role in MT assembly, there is also supporting evidence for a

role of the P1 protein in oligomer assembly within the mitochondria. This evidence is based on studies with yeast mif mutants and some observations made in this laboratory.

Yeast mif mutants are defective in mitochondrial import functions, hence their name. They were selected using a yeast strain that was defective at the cytosolic ornithine transcarbamylase (OTC) locus, but had a human mitochondrial OTC gene joined to an inducible Gal1 promoter (Cheng et al., 1989). Temperature-sensitive mutants of this strain were derived by EMS mutagenesis. They looked for mutants that did not contain OTC activity and accumulatd the OTC precursor and precursors of yeast mitochondrial proteins.

Examination of these mutants resulted in several complementation groups. Mutants of the processing enhancing protein (PEP) and the mitochondrial processing peptidase (MPP) constituted groups mif1 and mif2, respectively (Cheng et al., 1989). The mif4 group was a result of mutations in the hsp60 gene, which has been shown to have high identity with the P1 protein (Figure 18).

One mif4 mutant, ts143, accumulated mature-sized OTC subunits. It did not have any OTC activity which suggested that the OTC subunits did not assemble into the functional homotrimeric enzyme. These mutants were then tested for their ability to assemble yeast mitochondrial proteins. They examined the assembly of the beta-subunit of F_1-ATPase

into the F_1-ATPase complex, a matrix protein; cytochrome b_2, an intermembrane protein; and the Rieske Fe/S protein of complex III, another intermembrane protein. They found that the mature-sized beta-subunit of F_1-ATPase accumulated in the matrix, but it was not assembled into a functional enzyme. This was determined by the treatment of mitochondrial suspensions with chloroform (Cheng et al., 1989). Assembled enzyme complexes partitioned into the aqueous phase whereas non-assembled complexes remain in the organic phase.

Both cytochrome b_2 and the Fe/S precursor were transported into the matrix of the mitochondria and accumulated in intermediate sized form. They concluded from this that hsp60 must interact with imported precursors such that they obtain a conformation required for further transport. This might include covalent modifications, such as iron-sulphur cluster formation in the Fe/S complexes, proteolytic processing, or assembly into complexes.

This recent work, in conjunction with the sequence homology, strongly suggests that assembly of mitochondrial enzyme complexes may be a major function of the P1 protein. Experimental observations made in this laboratory support this hypothesis.

First, several attempts have been made to isolate second step mutants of P1 that change the second gene copy to the mutated form, M1. These experiments have been

unsuccessful. It can be suggested, that by mutating the second copy of this gene, the mitochondrial enzyme complexes likely remain unassembled, which would result in cell death.

Second, the Pod^{RII} cells grow more slowly than the wild type cells in the absence of the drug. Although the slow growth may be attributable to other factors it may be the result of inefficient assembly of the mitochondrial enzyme complexes. This would result in a lower rate of energy production and thus slower growth.

4.8 Future Direction and Experiments

The molecular cloning of the P1 gene has confirmed the results which localized the protein to the mitochondria (Gupta and Austin, 1987). It has also led to the suggestion that one of the possible functions of the P1 protein is to assemble oligomeric enzyme complexes in the mitochondria. Despite this, there are many questions which need addressing in the future.

The main questions that arise concern the role of this protein in MT structure and function. The biochemical and genetic data suggest that P1 is involved. The high frequency of mutants obtained, with the electrophoretic alteration in the P1 protein, is so great that the mutation in this protein cannot be attributed to chance and must be related to the action of the drug. The antimitotic drugs

are known to interact with MTs and inhibit their growth. Thus, the question remains, what is the role of P1 in MT assembly?

The molecular cloning of P1 and the purification of the P1 protein allows for additional experiments to be performed. If P1 interacts with MTs does P1 bind to tubulin? This experiment can be performed using purified P1 protein or setting up an _in vitro_ expression system with the cloned cDNA, to obtain the P1 protein, to determine if it binds tubulin in a binding assay. If the P1 protein does bind tubulin then the cDNA for P1 can be used to determine the tubulin binding domain. A binding domain for tau protein has been proposed from its sequence. The binding domain is a Pro-Gly-Gly-Gly motif that is part of a 31 amino acid repeat found in the tau gene (Goedert _et al._, 1989; Lee _et al._, 1988). Although the P1 protein does not contain any large repeats or the Pro-Gly-Gly-Gly motif this sequence is not likely a consensus sequence since other MT-associated and MT-related proteins likely bind to different sites on the tubulin dimer. It can be speculated that the Gly-Gly-Gly-Met sequence at the carboxy-terminus might be a possible binding sequence.

There are also other ways to determine if the P1 protein is involved with MTs. If the mutation in P1 is causing drug resistance then overexpression of M1 should also provide some drug resistance. The hamster cDNA for M1

should be subcloned into a mammalian expression vector, which can then be transformed into both wild type and Pod^{RI} cells. Drug resistance studies on these cells should show some resistance to MT inhibitors, if P1 does interact with tubulin. Alternatively, the P1 protein could be linked to a sepharose column. Passing a cell extract over the column should allow the proteins with which P1 interacts to bind to the column. Analysis of these proteins should show whether tubulin binds to the P1 protein.

Further characterization of both the first and second step podophyllotoxin mutants should also be examined. Since most MT mutants show alterations in tubulin, these genes should be examined more closely to ensure there are no mutations in these genes. Three alpha-tubulin genes are known to be differentially expressed in CHO cells. The exons for these genes could be examined by PCR amplification for a mutation. Likewise, the beta-tubulin genes could be examined.

The subcellular localization of the P1 protein to the mitochondria makes it difficult to reconcile its direct role in MT assembly. However, the P1 protein appears to be coreleased with tubulin in MT preparations. This leads to the question - Is there a cytoplasmic form of the P1 protein? The cDNA for P1 can be used to rescreen the library for additional clones which can be examined for their lack of a mitochondrial targeting sequence. The

likelihood of a cytoplasmic form of P1 is strong since other stress proteins, such as hsp70, have several homologues. It is also interesting that hsc70 is altered in the $Col^{R}22a$ mutants, in addition to the P1 protein. Perhaps there is some interaction of the two proteins with tubulin.

With the cloning of the wild type P1 cDNA it would be interesting to find the mutation that is causing the charge alteration as seen by 2D-gel analysis. There are several ways that this could be studied. One method is to use an RNase A cleavage assay (Myers et al., 1985). This method involves the enzymatic cleavage of RNA at a single base mismatch in an RNA:DNA hybrid. After hybridizing the cDNA for P1 to radiolabelled RNA from Pod^{RII} cells and digesting the mismatch with RNase A the RNA could be separated by electrophoresis. By estimating the size of the fragments on the blot the mutation could be sublocalized to a fragment which could then be used to further pinpoint the mutation. This type of approach has been used in the localization of mutations at the Lesch-Nyhan locus (Gibbs and Caskey, 1987). A similar approach that could be used would be denaturing gel electrophoresis (Noll and Collins, 1987).

An alternative approach would be to isolate the peptide fragments of P1 and M1 that are responsible for the loss of one protease sensitive site. Amino acid sequencing

of the amino terminal ends of these three peptides should localize the mutation. Two oligonucleotides could then be used to amplify the genomic sequence between these primers, which should contain the mutation. Direct sequencing of the amplified material from Pod^{RII} cells should yield the mutation.

Once the mutation is localized it would be interesting to examine this region of the gene. It appears that the mutation is responsible for the increase in resistance to the drug podophyllotoxin. Studying this region should help to obtain a better understanding of how the drug interacts with this protein.

The extensive homology of the P1 protein to the 65 kDa antigen of mycobacteria suggests that this is the mammalian homologue of this protein. Since this protein is found in a wide variety of pathogenic bacteria much work has been pursued toward the development of better vaccines and immunodiagnostic tests for these diseases. Examination of the sequence of the 65-kDa antigen and the P1 protein could lead to the development of a peptide that is specific for the mycobacterial protein. The 65 kDa antigen has also been shown to play an important role in the development of the chronic autoimmune disease adjuvant arthritis in rats (Van Eden et al, 1988). Since antibodies and T-cell clones reactive with the 65-kDa antigen have also been observed in humans suffering from the related autoimmune disease

rheumatoid arthritis the human P1 protein may be a likely cellular target for the development of autoimmune response. Further investigations on the cellular function of P1 and its possible role in autoimmune diseases must also be examined.

The sequence of the P1 protein is also a very useful tool for future research on its role in the mitochondria. The cDNA can be inserted into an expression vector which can be used to isolate functional P1 protein in an *in vitro* expression system. The protein can then be studied in greater detail to answer questions such as: How does it interact with mitochondrial proteins? What mitochondrial proteins bind to this complex? Once these questions are answered, *in vitro* mutagenesis can be used to analyse which amino acids are important in the interaction of the protein with the various mitochondrial proteins.

4.9 Conclusions

The complete cDNA sequence of the P1 protein from Chinese hamster ovary cells has been described. The deduced protein sequence contained a mitochondrial presequence that localizes the protein to the matrix. This confirmed previous work from this laboratory that suggested the protein was mitochondria-related (Gupta and Dudani, 1987; Gupta and Austin, 1987). The amino acid sequence showed extensive homology to a class of molecular

chaperones called chaperonins. This class of proteins included the GroEL protein of E. coli, the Rubisco large subunit binding protein of higher plants, and the yeast hsp60 protein. The degree of similarity between these various proteins and the mammalian P1 protein is in agreement with the endosymbiotic theory of the evolution of mitochondria and chloroplasts (Woese et al., 1983; Yang et al., 1985).

The hamster cDNA for the P1 protein showed very high identity with the human sequence throughout the coding region, as well as the 3'-untranslated region. Such high conservation of the untranslated region is an interesting finding that may be important for the stability of the mRNA. The high identity with the human protein sequence and sequences from a wide range of organisms suggests this protein plays an important role within the cell.

The results presented here suggest that P1 likely has a role in assembling mitochondrial enzyme complexes. However, its role in MT assembly that has been suggested from earlier work, could not be proven. Future work should be directed to determining the role this protein has in microtubule assembly/function. It should also be directed at uncovering the mutation that changes P1 to M1 in the Pod^{RII} cells. These experiments should help increase the understanding of the exact cellular function(s) of the P1 protein.

5.0 APPENDIX

Analysis of the cDNA Products

First Strand Yield. The overall yield of the first strand reaction is calculated from the amount of acid-precipitable counts determined as described in *First Strand Synthesis*. In order to perform the calculation, you must first determine the specific activity (SA) of the radioisotope in the reaction mixture. The specific activity is defined as the counts per minute (cpm) of an aliquot of the reaction mixture divided by the quantity (pmol) of the same nucleotide in that aliquot. The amount of dCTP in the 5-µl aliquot is calculated by multiplying the amount of unlabeled dCTP in the reaction mixture (25,000 pmol) by the actual amount sampled:

$$\text{pmol dCTP/5 µl} = (25{,}000 \text{ pmol dCTP/50 µl}) \times (10 \text{ µl/100 µl}) \times 5 \text{ µl} \quad [1]$$
$$= 250 \text{ pmol dCTP/5 µl}$$

For [α^{32}P] dCTP, the specific activity is given by the relationship:

$$\text{SA (cpm/pmol dCTP)} = \frac{\text{cpm/5 µl}}{250 \text{ pmol/5 µl}} \quad [2]$$

The amount of dCTP contributed by the radiolabeled material is insignificant relative to the unlabeled nucleotide and is ignored in equation 2.

Now that the specific activity is known, the yield of first strand product in the 10-µl reaction can be calculated from the amount of acid-precipitable counts determined from the washed filter:

$$\text{Yield (µg DNA)} = \frac{\text{(cpm)} \times (100 \text{ µl/5 µl}) \times (4 \text{ pmol dNTP/pmol dCTP})}{\text{(cpm/pmol dCTP)} \times (3030 \text{ pmol dNTP/µg DNA})} \quad [3]$$

The correction in the numerator takes into account that, on the average, 4 nucleotides will be incorporated into the cDNA for every dCTP which is scored by this assay. The factor in the denominator is the amount of nucleotide which corresponds to 1 µg of single-stranded DNA.

Second Strand Yield. The yield of the second strand reaction is determined in an identical manner, except that the composition of the second strand reaction dictates the specific activity of the isotope. The amount of dCTP in the aliquot spotted for specific activity determination is given by equation 4:

$$\text{pmol dCTP/5 µl} = (80{,}000 \text{ pmol dCTP/345 µl}) \times (10 \text{ µl/100 µl}) \times 5 \text{ µl} \quad [4]$$
$$= 116 \text{ pmol dCTP/5 µl}$$

For [α^{32}P] dCTP, the specific activity is given by the relationship:

$$\text{SA (cpm/pmol dCTP)} = \frac{\text{cpm/5 µl}}{116 \text{ pmol/5 µl}} \quad [5]$$

Again, the amount of dCTP contributed by the radiolabeled material is insignificant relative to the unlabeled nucleotide and is ignored in equation 5.

Now that the specific activity is known, the yield of second strand product can be calculated from the amount of acid-precipitable counts determined from the washed filter:

$$\text{Yield (µg DNA)} = \frac{\text{(cpm)} \times (345 \text{ µl/10 µl}) \times (4 \text{ pmol dNTP/pmol dCTP})}{\text{(cpm/pmol dCTP)} \times (3030 \text{ pmol dNTP/µg DNA})} \quad [6]$$

(taken from BRL cDNA synthesis manual)

6.0 References

Anderson, D.C., Wible, L.J., Hughes, B.J., Smith, C.W., and B.R. Brinkley. (1982) Cytoplasmic MTs in polymorphonuclear leukocytes: Effects of chemotactic stimulation and colchicine. Cell **31**, 719-729

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. Science **181**, 223-230

Aviv, H. and Leder, P. (1972) Purificcation of biologically active globin mRNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**, 1408-1412

Ball, E.H., and S.J. Singer. (1982) Mitochondria are associated with microtubules and not with intermediate filaments in cultured fibroblasts. Proc. Natl. Acad. Sci. USA **79**, 123-126

Bardwell, J.C.A., and E.A. Craig. (1984) Major heat shock gene of Drosophila and the E.coli heat inducible dnaK gene are homologous. Proc. Natl. Acad. Sci. USA **81**, 848-852

Bernier-Valentin, F., Aunis, D., and B. Rousset. (1983) Evidence for tubulin-binding sites on cellular membranes: Plasma membranes, mitochondrial membranes, and secretory granule membranes. J. Cell Biol. 97, 209-216

Bibring, T., and J. Baxandall. (1977) Tubulin synthesis in sea urchin embryos: Almost all tubulin of the first cleavage mitotic apparatus derives from the unfertilized egg. Dev. Biol. 55, 191-195

Birnboim, H.C., and J. Doly. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**, 1513-1523

Bochkareva, E.S., Lissin, N.M., and A.S. Girshovich. (1988) Transient association of newly synthesized unfolded proteins with the heat-shock GroEL protein. Nature **336**, 254-257

Bole, D.G., Hendershot, L.M., and J.F. Kearney. (1986) Posttranslational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting and secreting hybridomas. J. Cell Biol. **102**, 1558-1566

Boom, W.H., Husson, R.N., Young, R.A., David, J.R., and W.F. Piessens. (1987) In vivo and in vitro characterization of murine T-cell clones reactive to Mycobacterium tuberculosis. Infect. Immun. 55, 2223-2229

Borisy, G.G., and E.W. Taylor. (1967) The mechanism of action of colchicine: Binding of ^{3}-H colchicine to cellular proteins. J.Cell Biol. **34**, 525-534

Brinkley, B.R. (1985) Microtubule organizing centers. Ann. Rev. Cell Biol. **1**, 145-172

Britton, W.J., Hellqvist, L., Basten, A., and R.L. Raison. (1985) Mycobacterium leprae antigens involved in human immune responses: Identification of four antigens by monoclonal antibodies. J. Immunol. **135**, 4171-4177

Bryan, J., and L. Wilson. (1971) Are cytoplasmic MTs heteropolymers? Proc. Natl. Acad. Sci. USA **68**, 1762-1766

Buchanan, T.M., Nomaguchi, H., Anderson, D.C., Young, R.A., Gillis, T.P., Britton, W.J., Ivanyi, J., Kolk, A.H.J., Closs, O., Bloom, B.R., and V.J. Mehra. (1987) Characterization of antibody-reactive epitopes on the 65-kilodalton protein of Mycobacterium leprae. Mycobacterial species specificity. Infect. Immun. **55**, 1000-1003

Bukau, B., and G.C. Walker. (1989) Cellular defects caused by deletion of the E.coli dnaK gene indicates roles for heat shock protein in normal metabolism. J. Bacteriol. **171**, 2337-2346

Burton, P.R., Hinclley, R.E., and G.B. Pierson. (1975) Tannic acid stained MTs with 12, 13, and 15 protofilaments. J.Cell Biol. **65**, 227-233

Cabral, F., Sobel, M., and M.M. Gottesman. (1980) CHO mutants resistant to colchicine, colcemid, or griseofulvin have an altered ß-tubulin. Cell **20**, 29-36

Carlier, M-F., and D. Pantaloni. (1981) Kinetic analysis of GTP hydrolysis associated with tubulin polymerization. Biochemistry **20**, 1918-1924

Chandrasekhar, G.N., Tilly, K., Woolford, C., Hendrix, R.W., and C.P. Georgopolous. (1986) Purification and properties of the GroES morphogenetic protein of E.coli. J. Biol. Chem. **261**, 12414-12419

Cheng, M.Y., Hartl, F.U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Hallberg, E.M., Hallberg, R.L., and A.L. Horwich. (1989) Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. Nature, **337**, 620-625

Chirgwin, J.J, Przbyla, A.E., MacDonald, R.J., and Rutter,

W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18, 5294-5299

Chirico, W.J., Waters, M.G., and G. Blobel (1988) 70K heat shock related proteins stimulate protein translocation into microsomes. Nature 332, 805-810

Ciesielski-Treska, J., Bader, M.F., and D. Aunis. (1982) Microtubular organization in flat epitheloid and stellate process-bearing astrocytes in culture. Neurochem. Res. 7, 275-286

Clarkson, T.W., Sager, P.R., and T.L.M. Syverson. (1986) The Cytoskeleton: A target for toxic agents. Plenum Press, New York

Cleveland, D.W., Ilwo, S-Y., and M.W. Kirschner. (1977) Purification of tau, a MAP that induces assembly of MTs from purified tubulin. J. Mol. Biol. 16, 207-225

Cleveland, D.W., Lopata, M.A., MacDonald, R.J., Cowan, N.J., Rutter, W.J., and M.W. Kirschner. (1980) Number and evolutionary conservation of ã- and ß-tubulin and cytoplasmic ß- and gamma-actin genes using specific cloned cDNA probes. Cell 20, 95-105

Cleveland, D.W., and K.F. Sullivan. (1985) Molecular biology and genetics of tubulin. Ann. Rev. Biochem. 54, 331-365

Cleveland, D.W. (1988) Autoregulated instability of tubulin mRNAs: A novel eukaryotic regulatory mechanism. Trends Biochem. Sci. 13, 341-345

Cortese, F., Bhattacharyya, B., and J. Wolff. (1977) Podophyllotoxin as a probe for the colchicine binding site of tubulin. J. Biol. Chem. 252, 1134-1140

Cowan, N.J., Dobner, P.R., Ruchs, E.V., and D.W. Cleveland. (1983) Expression of human ã-tubulin genes: Interspecies conservation of 3′ untranslated regions. Mol. Cell. Biol. 3, 1738-1745

Cowan, N.J., and L. Dudley. (1983) Tubulin isotypes and the multigene tubulin families. Int. Rev. Cytol. 85, 147-173

Darnell, J., Lodish, H., and D. Baltimore. (1986) Molecular Biology of the Cell. W.H. Freeman and Company, New York.

David-Pfeufy, T., Erickson, H.P., and D. Pentaloric. (1977)

GTPase activity of tubulin associated with MT assembly. Proc. Natl. Acad. Sci. USA 74, 5372-5376

David-Pfeufy, T., LaPorte, J., and D. Pantaloni. (1978) GTPase activity at ends of microtubules. Nature **272**, 282-284

DeBrabander, M.J., Van De Veire, R.M., Aerts, F., Burgess, M., and P.A. Jansen. (1976) The effects of methyl-5-(2-thienyl-carbonyl)-1H-benzimidazol-2-yl carbamate (R17934), a new synthetic antitumoral drug interfering with MTs, in mammalian cultured cells in vitro. Cancer Res. 36, 905-916

DeBrabander, M.J., Demey, J., Joniau, M., and S. Geuens. (1977) Immunocytochemical visualization of MTs and tubulin at the light- and electron-microscopy level. J. Cell Sci. 28, 283-302

DeBrabander, M.J., Geuens, G., Nuydens, R., Willebrords, R., and J. DeMay. (1981) Taxol induces the assembly of ree MTs in living cells and blocks the organizing capacity of the centrosomes and the kinetochores. Proc. Natl. Acad. Sci. USA 78, 5608-5612

Deshaies, R.J., Koch, B.D., and R. Schekman. (1988a) The role of stress proteins in membrane biogenesis. Trends Biochem. Sci. 13, 384-388

Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A., and R. Schekman. (1988b) A subfamily of stress proteins facilitates translocation of secretor and mitochondrial precursor polypeptides. Nature 332, 800-805

Drubin, D.G., Caput, C., and M.W. Kirschner. (1984) Studies on the expression of the MAP, tau, during mouse brain development, with newly isolated cDNA probes. J. Cell Biol. 98, 1090-1097

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

Earnshaw, W.C., Honda, B.M., and R.A. Laskey. (1980) Assembly of nucleosomes: The reaction involving X.laevis nucleoplasmin. Cell 46, 291-300

Elliott, E.M., Okayama, H., Sarangi, F., Henderso, G., and V. Ling. (1985) Differential expression of three ã-tubulin genes in CHO cells. Mol. Cell. Biol. 5, 236-241

Ellis, R.J. (1987) Proteins as molecular chaperones. Nature 328, 378-379

Ellis, R.J., and S.M. Van Der Vies. (1988) The Rubisco subunit binding protein. Photosynthesis Research **16**, 101-115

Ellis, R.J., Van Der Vies, S.M., and S.M. Hemmingsen. (1989) The molecular chaperone concept. In: Symposium 55 of the Biochemical Society (ed. J. Ray)

Frohman, M.A., Dush, M.K., and G.R. Martin. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85**, 8998-9002

Fourney, R.M., Miyakoshi, J., Day III, R.S., and M.C. Paterson. (1988) Northern blotting: Efficient RNA staining and transfer. Focus **10**, 5-7

Georgopolous, C.P., Hendrix, R.W., Casjens, S.R., and A.D. Kaiser. (1973) Host participation in bacteriophage lambda head assembly. J. Mol. Biol. **76**, 45-60

Gibbs, R.A., and C.T. Caskey. (1987) Identification and localization of mutations at the Lesch-Nyhan locus by Ribonuclease A cleavage. Science **236**, 303-305

Gillis, T.P., Miller, R.A., Young, D.B., Khanolkor, S.R., and T.M. Buchanan. (1985) Immunochemical characterization of a protein associated with Mycobacterium leprae cell wall. Infect. Immun. **49**, 371-377

Goedert, M., Spillantini, M.G., Potier, M.C., Ulrich, J., and R.A. Crowther. (1989) Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: Diferential expression of tau protein mRNAs in human brain. EMBO J. **8**, 393-399

Goloubinoff, P., Gatenby, A.A., and G.H. Lorimer. (1989) GroE heat-shock proteins promoteassembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in E.coli. Nature 337, 44-47

Gubler, U. and Hoffman, B.J. (1983) A simple and very effective method for generating cDNA libraries. Gene **25**, 263-269

Gupta, R.S. (1981) Resistance to the MT inhibitor podophyllotoxin: Selection and partial chaacterization of mutants in CHO cells. Somatic Cell Genet. 7, 59-71

Gupta, R.S., Ho, T.K.W., Moffat, M.R.K., and R. Gupta. (1982) Podophyllotoxin-resistant mutants of CHO cells: Alteration in a microtubule-associated protein. J. Biol.

Chem. **257**, 1071-1078

Gupta, R.S. (1983) Podophyllotoxin-resistant mutants of CHO cells: Cross-resistance studies with various MT inhibitors and podophyllotoxin analogues. Cancer Res. **43**, 505-512

Gupta, R.S. (1984) Griseofulvin resistance mutation of Chinese hamster ovary cells that affects the apparent molecular weight of a 200,000 Dalton protein. Mol. Cell. Biol. **4**, 1761-1768

Gupta, R.S. (1985) Cross-resistance of vinblastine- and taxol-resistant mutants of Chinese hamster ovary cells to other anticancer drugs. Cancer Treatment Reports **69**, 515-521

Gupta, R.S., Venner, T.J., and A. Chopra. (1985) Genetic and biochemical studies with mutants of mammalian cells afected in MT-related proteins other than tubulin: Mitochondrial localization of a MT-related protein. Can. J. Biochem. Cell Biol. **53**, 489-502

Gupta, R.S. (1986) Cross-resistance of nocodazole-resistant mutants of CHO cells toward other microtubule inhibitors: Similar mode of action of benzimidazole carbamate derivatives and NSC 181928 and TN-16. Mol. Pharmacol. **30**, 142-148

Gupta, R.S., and T.J. Venner. (1986) Mitochondrial localization of a MT-related protein. Ann. NY Acad. Sci. **466**, 854-857

Gupta, R.S., and A.K. Dudani. (1987) Mitochondrial binding of a protein affected in mutants resistant to the MT inhibitor podophyllotoxin. Eur. J. Cell Biol. **44**, 278-285

Gupta, R.S., and R.C. Austin. (1987) Mitochondrial matrix localization of a protein altered in mutants resistant to the MT-inhibitor podophyllotoxin. Eur. J. Cell Biol. **45**, 170-176

Gupta, R.S., and A.K. Dudani. (1989) Mechanism of action of antimitotic drugs: A new hypothesis based on the role of cellular calcium. Medical Hypothesis **28**, 57-69

Haas, I.G., and M. Wabl. (1983) Immunoglobulin heavy chain binding protein. Nature **306**, 387-389

Hargreaves, A.J., and J. Avila. (1985) Localization and characterization of tubulin-like proteins associated with brain mitochondria: The presence of a membrane-specific isoform. J. Neurochem. **45**, 490-496

Hayden, J.H., Allen, R.D., and R.D. Goldman. (1983) Cytoplasmic transport in keratocytes: Direct visualization o particle translocation along microtubules. Cell Motil. 3, 1-19

Hemmingsen, S.M., Woolford, C., Van Der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C.P., Hendrix, R.W., and R.J. Ellis. (1988) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333, 330-334

Hendrix, R.W. (1979) Purification and properties of GroE, a host protein involved in bacteriophage assembly. J. Mol. Biol. 129, 375-392

Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28, 351-359

Herrmann, H., Pytella, R., Dalton, J., and G. Wiche. (1984) Structural homology of MAP1 and MAP2 demonstrated by peptide mapping and immunorstivity. J. Biol. Chem. 259, 612-617

Herrmann, H., Dalton, J.M., and G. Wiche. (1985) Microheterogeneity of MAPS, MAP1 and MAP2, and differential phosphorylation of individual subcomponents. J. Biol. Chem. 260, 5797-5803

Holoshitz, J., Matitiau, A., and I.R. Cohen. (1984) Arthritis induced in rats by cloned T lymphocytes responsive to mycobacteia but not to collagen type II. J. Clin. Invest. 73, 211-215

Hurt, E.C., and A.P.G.M. van Loon. (1986) How proteins find mitochondria and intamitochondrial compartments. Trends Biochem. Sci. 11, 204-207

Huynh, T.V., Young, R.A., and R.W. Davis. (1985) In: DNA Cloning: A practical Approach, Volume 1. Edited by D.M. Glover. IRL Press Ltd. pp 49-88.

Jindal, S., Dudani, A.K., Singh, B., Harley, C.B., and R.S. Gupta. (1989) Primary structure of a human mitochondrial protein homologous to the bacterial and plant chaperonins and to the 65 kDa mycobacterial antigen. Mol. Cell. Biol. 9, 2279-2283

Kim, H., Binder, L.I., and J.L. Rosenbaum. (1979) The periodic association of MAPs with brain MTs in vitro. J. Cell Biol. 80, 266-276

Kirschner, M., and T. Mitchison. (1986) Beyond self-assembly: From microtubules to morphogenesis. Cell **45**, 329-342

Kirschner, M.W. (1989) Biological implications of microtubule dynamics. The Harvey Lectures, Series 83, p 1-20. Alan R. Liss Inc.

Kosik, K.S., Joachim, C.L., and D.J. Selkoe. (1986) Microtubule- associated protein tau is a major antigenic component of paired helical ilaments in Alzheimer's disease. Proc. Natl. Acad. Sci. USA **83**, 4044-4048

Krauhs, E., Little, M., Kempf, T., Hofer-Warbinek, R., Ade, W., and H. Ponstingl. (1981) Complete amino acid sequence of ß-tubulin from porcine brain. Proc.Natl.Acad.Sci.USA **78**, 4156-4160

Kushner, S.R. (1978) In: Genetic Engineering. Editors H.B. Boyer and S. Nicosia. Elsevier/North-Holland Amsterdam. p17.

Kyte, J., and R.F. Doolittle. (1982) A simple method for displaying the hydrophobic character of a protein. J. Mol. Biol. **157**, 105-132

Laskey, R.A., Honda, B.M., Mills, A.D., and J.T. Finch. (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature **275**, 416-420

Lee, M.G., Lewis, S.A., Wilde, C.D., and N.J. Cowan. (1983) Evolutionary history of a multigene family: An expressed human ß-tubulin gene and three processed pseudogenes. Cell **33**, 477-487

Lee, M.G., Cowan, N.J., and M.W. Kirschner. (1988) The primary structure and heterogeneity of tau protein from mouse brain. Science **239**, 285-288

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

Lewis, M.J., and H.R.B. Pelham. (1985) Involvement of ATP in the nuclear and nucleolar fractions of the 70 kDa heat shock protein. EMBO J. **4**, 3137-3143

Lewis, S.A., Sherline, P., and N.J. Cowan. (1986a) A cloned cDNA encoding MAP1 detects a single copy gene in mouse, and a brain abundant RNA whose level decreases during development. J. Cell Biol. **102**, 2106-2114

Lewis, S.A., Villasante, A., Sherline, P., and N.J. Cowan. (1986b) Brain specific expression of MAP2 detected using a cloned cDNA probe. J. Cell Biol. 102, 2098-2105

Lim, L., Hall, C., Leung, T., and S. Whatley. (1984) The relationship of the rat brain 68 kDa MAP with synaptosomal plasma membranes and with the Drosophila 70 kDa heat-shock protein. Biochem. J. 224, 677-680

Lindquist, S. (1986) The heat-shock response. Ann. Rev. Biochem. 55, 1151-1191

Lindquist, S., and E.A. Craig. (1988) The heat-shock proteins. Ann. Rev. Genet. 22, 631-677

Lindwall, G., and R.D. Cole. (1984) Phosphorylation affects the ability of tau protein to promote MT assembly. J. Biol. Chem. 259, 5301-5305

Ling, V., and L.H. Thompson. (1974) Reduced permeability in CHO cells as a mechanism of resistance to colchicine. J. Cell. Physiol. 83, 103-116

Liu, X-Y., Maeda, S., Collier, K., and S. Pestka. (1985) Rapid identification of genomic interferon genes by sequencing on phage Lambda DNA templates. Gene Anal. Tchn. 2, 83-88

Lu, M.C., Lien, M.H., Becker, R.E., Heine, H.C., Buggs, A.M., Lipovsek, D., Gupta, R., Robbins, P.W., Grosskinsky, C.M., Hubbard, S.C., and R.A. Young. (1987) Genes for immunodominant protein antigens are highly homologous in M.tuberculosis, M.africanum, and the vaccine strain M.bovis BCG. Infect. Immun. 55, 2378-2382

Mandel, M., and A. Higa. (1970) Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53, 154

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

Mareel, M.M., and M. DeMets. (1984) Effect of MT inhibitors on invasion and related activities of tumour cells. Intl. Rev. Cytol. 90, 125-168

Margolis, R.L., and L. Wilson. (1977) Addition of colchicine-tubulin complex to MT ends: The mechanism of substoichiometric colchicine poisoning. Proc. Natl. Acad. Sci. USA 74, 3466-3470

Margolis, R.L., and L. Wilson. (1978) Opposite end assembly

and disassembly of MTs at steady state in vitro. Cell 13, 1-8

McDonnell, M.W., Simon, M.N., and F.W. Studier. (1977) Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. 110, 119

McMullin, T.W., and R.L. Hallberg. (1987) A normal mitochondrial protein is selectively synthesized and accumulated during heat shock in Tetrahymena thermophila. Mol. Cell. Biol. 7, 4414-4423

Mehra, V., Sweetster, D., and R.A. Young. (1986) Efficient mapping of protein antigenic determinants. Proc. Natl. Acad. Sci. USA 83, 7013-7017

Moll, E., Manz, B., Mocikat, S., and H-P Zimmermann. (1982) Fluorescent deacetylcolchicine: New aspects of its activity and localization in PtK-1 cells. Exp. Cell Res. 141, 211-220

Mose-Larsen, P., Bravo, R., Fey, S.J., Small, J.V., and J.E. Celis. (1982) Putative association of mitochondria with a subpopulation of intermediate-sized filaments in cultured human skin fibroblasts. Cell 31, 681-692

Munro, S., and H.R.B. Pelham. (1986) An Hsp-70 like protein in the ER: Identity with the 78 kDa glucose-regulated protein and immunoglobulin heavy chain binding protein. Cell 46, 291-300

Myers, R.M., Larin, Z., and T. Maniatis. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. Science 230, 1242-1246

Noll, W.W., and M. Collins. (1987) Detection of human DNA polymorphisms with a simplified denaturing gradient gel electrophoresis technique. Proc. Natl. Acad. Sci. USA 84, 3339-3343

Ochs, R.L. (1982) The role of microtubules in cell shape and pigment distribution in spreading erythrophores. Eur. J. Cell Biol. 28, 226-232

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

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

Pape, M.E., and Kim, K-H. (1987) Improvements in production

of cDNA libraries in gt11: Detection of ligation inhibitor and separation of cDNA from free linkers. Gene Anal. Tech. 4, 14-16

Pelham, H.R.B. (1986) Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell **46**, 959-961

Picketts, D.J., Mayanil, C.S.K., and R.S. Gupta. (1989) Molecular cloning of a Chinese Hamster mitochondrial protein related to the "Chaperonin" family of bacterial and plant proteins. J. Biol. Chem. **264**, 12001-12008

Pierson, G.B., Burton, P.R., and R.H. Himes. (1978) Alterations in number of protofilaments in MTs assembled in vitro. J.Cell Biol. **76**, 223-228

Ponstingl, H., Krauhs, E., Little, M., and T. Kempf. (1981) Complete amino acid sequence of ã-tubulin from porcine brain. Proc.Natl.Acad.Sci.USA **78**, 2757-2761

Posner, B.I., Verma, A.K., Patel, B.A., and J.J. Bergeron. (1982) Effect of colchicine on the uptake of prolactin and insulin into Golgi fractions of Rat liver. J. Cell Biol. **93**, 560-567

Pratt, L.F., Okamura, S., and D.W. Cleveland. (1987) A divergent testes-specific ã-tubulin isotype that does not contain a coded C-terminal tyrosine. Mol. Cell. Biol. **7**, 552-555

Proudfoot, N.J., and G.G. Brownlee. (1976) 3'-non-coding region sequences in eukaryotic mRNA. Nature **263**, 211-214

Quertermous, T. (1987) In: Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience. pp 6.1.1-6.1.4

Reading, D.S., Hallberg, R.L., and A.M. Myers. (1989) Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. Nature 337, 655-659

Roise, D., and G. Schatz (1988) Mitochondrial presequences. J. Biol. Chem. **263**, 4509-4511

Roy, H., and S. Cannon. (1988) Ribulose bisphosphate carboxylase assembly: What is the role of the large subunit binding protein? Trends Biochem. Sci. **13**, 163-165

Ruben, S.M., Van DenBrink-Webb, S.E., Rein, D.C., and R.R. Meyer. (1988) Suppression of the E.coli ssb-1 mutation by an allele of GoEL. Proc. Natl. Acad. Sci. USA **85**, 3767-

3771

Saito, K., and K. Hama. (1982) Structural diversity of MTs in the supporting cells of the sensory epithilium of guinea pig organ of Corti. J.Electron Microsc. **31**, 278-281

Sanger, F., Nicklen, S., and A.R. Coulson. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**, 5463-5467

Sartoris, S., Cohen, E.B., and J.S. Lee. (1987) A rapid and improved method for generating cDNA libraries in plasmid and phage lambda vectors. Gene **56**, 301-307

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

Schiff, P.B., Fant, J., and S.B. Horwitz. (1979) Promotion of MT assembly *in vitro* by taxol. Nature **277**, 665-667

Schiff, P.B., and S.B. Horwitz. (1980) Taxol stabilizes MTs in mouse fibroblast cells. Proc. Natl. Acad. Sci. USA **77**, 1561-1565

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

Schnapp, B.J., Vale, R.D., Sheetz, M.P., and T.S. Reese. (1985) Single microtubules from squid axoplasm support bidirectional movement of organelles. Cell **40**, 455-462

Schnapp, B.J., Vale, R.D., Sheetz, M.P., and T.S. Reese. (1986) Microtubules and the mechanism of directed organelle movement. Ann. N.Y. Acad. Sci. **466**, 909-918

Shelanski, M.L., Gaskin, F., and C.R. Cantor. (1973) Microtubule assembly in the absence of added nucleotides. Proc. Natl. Acad. Sci. USA **70**, 765-768

Shinnick, T.M. (1987) The 65-kDa antigen of *Mycobacterium tuberculosis*. J. Bacteriol. **169**, 1080-1088

Shinnick, T.M., Sweetser, D., Thole, J.E.R., van Embden, J., and R.A. Young. (1987) The etiologic agents of leprosy and tuberculosis share an immunoreactive protein antigen with the vaccine strain *Mycobacterium bovis* BCG. Infect. Immun. **55**, 1932-1935

Shinnick, T.M., Vodkin, M.H., and J.C. Williams. (1988) The *M.tuberculosis* 65-kilodalton antigen is a heat shock protein which corresponds to common antigen and to the *E.coli* GroEL protein. Infect. Immun. **56**, 446-451

Shiu, R.P.C., Pouyssegur, J., and I. Pastan (1977) Glucose depletion accounts for the induction of two transformation-sensitive membrane proteins in Rous sarcoma virus-transformed chick embryo fibroblasts. Proc. Natl. Acad. Sci. USA 74, 3840-3844

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

Spiegelman, B.M., Pennigroth, S.M., and M.W. Kirschner. (1977) Turnover of tubulin and the N-site in CHO cells. Cell 12, 587-600

Straus, D.B., Walter, W.A., and C.A. Gross. (1988) E.coli heat shock gene mutants are defective in proteolysis. Genes and Development 2, 1851-1858

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

Stryer, L. (1988) Biochemistry, Third Edition. W.H. Freeman and Co., New York.

Sullivan, K.F., Havercroft, J.C., Machlin, P.S., and D.W. Cleveland. (1986) Sequence and expression of the chicken ß5- and ß4-tubulin genes define a pair of divergent ß-tubulins with complementary patterns of expression. Mol. Cell Biol. 6, 4409-4418

Tabor, S., and C.C. Richardson. (1987) DNA sequence analysis with a modified bacteriophage T_7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84, 4767-4771

Thole, J.E.R., Keulen, W.J., Kolk, A.H.J., Groothuis, D.G., Berwald, L.G., Tiesjema, R.H., and J.D.A. Van Embden. (1987) Characterization, sequence determination, and immunogenicity of a 64-kilodalton protein of M.bovis BCG expressed in E.coli K-12. Infect. Immun. 55, 1466-1475

Thole, J.E.R., Hindersson, P., de Bruyn, J., Cremers, F., van der Zee, J., de Cock, H., Tommassen, J., van Eden, W., and J.D.A. van Embden. (1988a) Antigenic relatedness of a strongly immunogenic 65 kDa mycobacterial protein antigen with a similarly sized ubiquitous bacterial common antigen. Microbial Pathogenesis 4, 71-83

Thole, J.E.R., Van Schooten, W.C.A., Keulen, W.J., Hermans, P.W.M., Janson, A.A.M., De Vries, R.R.P., Kolk, A.H.J., and J.D.A. Van Embden. (1988b) Use of recombinant antigens

expressed in E.coli K-12 to map B-cell and T-cell epitopes on the immunodominant 65-kilodalton protein of M.bovis BCG. Infect. Immun. **56**, 1633-1640

Timasheff, S.N., and L.M. Grisham. (1980) In vitro assembly of cytoplasmic microtubules. Ann. Rev. Biochem. **49**, 565-591

Vale, R.D., Reese, T.S., and M.P. Sheetz (1985a) Identification of a novel translocator, kinesin, induced in MT-based motility. Cell **40**, 39-50

Vale, R.D., Schnapp, B.J., Reese, T.S., and M.P. Sheetz (1985b) Movement of organelles along filaments dissociated from the axoplasm of the squid giant axon. Cell **40**, 449-454

Vale, R.D., Schnapp, B.J., Reese, T.S., and M.P. Sheetz (1985c) Organelle, bead, and MT translocations promoted by soluble factors from the squid giant axon. Cell **40**, 559-569

van Eden, W., Holoshitz, J., Nevo, Z., Frenkel, A., Klajman, A., and I.R. Cohen. (1985) Arthritis induced by a T-lymphocyte clone that responds to Mycobacterium tuberculosis and to cartilage proteoglycans. Proc. Natl. Acad. Sci. USA **82**, 5117-5120

van Eden, W., Thole, J.E.R., van der Zee, R., Noordzij, A., van Embden, J.D.A., Hensen, E.J., and I.R. Cohen. (1988) Cloning of the mycobacterial epitope recognized by T lymphocytes in adjuvant arthritis. Nature **331**, 171-173

Valenzuela, P., Quiroga, M., Zaldivar, J., Rutter, W.J., Kirschner, M.W., and D.W. Cleveland. (1981) Nucleotide and corresponding amino acid sequence encoded by ã- and ß-tubulin mRNAs. Nature **289**, 650-655

Vodkin, M.H., and J.C. Williams. (1988) A heat shock operon in Coxiella burnetii produces a major antigen homologous to a protein in both Mycobacteria and E.coli. J. Bacteriol. **170**, 1227-1234

Wada, M., and H. Itikawa. (1984) Participation of E.coli K-12 GroE gene products in the synthesis of cellular DNA and RNA. J. Bacteriol. **157**, 694-696

Wada, M., Fujita, H., and H. Itikawa. (1987) Genetic suppression of a temperature-sensitive GroES mutation by an altered subunit of RNA polymerase of E.coli K-12. J. Bacteriol. **169**, 1102-1106

Weber, K., Bibring, T.H., and M. Osborn. (1975) Specific visualization of tubulin-containing structures in tissue culture cells by immunofluorescence. Cytoplasmic MTs, vinblastine-induced paracrystals, and mitotic figures. Exp. Cell Res. 95, 111-120

Weisenberg, R.C., Borisy, G.G., and E.W. Taylor. (1968) The colchicine binding protein of mammalian brain and its relationship to MTs. Biochemistry 7, 4466-4479

Weisenberg, R.C. (1972) Microtubule ormation in vitro in solutions containing low calcium concentrations. Science 177, 1104-1105

Whatley, S.A., Leung, T., Hall, C., and L. Lim. (1986) The brain 68 kDa MAP is a cognate form of the 70 kDa mammalian heat-shock protein and is present as a specific isoform in synaptosomal membranes. J. Neurochem. 47, 1576-1583

Wilson, L., Bamburg, J.R., Mizel, S.B., Grisham, L.M., and K.M. Creswell. (1974) Interaction of drugs with MT proteins. Fed. Proc. 33, 158-166

Wilson, L., and J. Bryan. (1974) Biochemical and pharmacological properties of microtubules. Adv. Cell. Mol. Biol. 3, 21-72

Wilson, L. (1986) The cytoskeleton: A target for toxic agents. Edited by Clarkson, T.W., Sanger, P.R., and T.L.M. Syverson. Plenum press, New York.

Woese, C.R., Gutell, R.R., Gupta, R., and H.F. Noller. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol. Rev. 47, 621-669

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, J., and C.R. Woese. (1985) Mitochondrial origins. Proc. Natl. Acad. Sci. USA 82, 4443-4447

Young, D.B., Fohn, M.J., Khanolkar, S.R., and T.M. Buchanan. (1985a) Monoclonal antibodies to a 28,000-molecular-weight protein antigen of Mycobacterium leprae. Clin. Exp. Immunol. 60, 546-552

Young, D.B., Lathigra, R., Hendrix, R., Sweetster, D., and R.A. Young. (1988a) Stress proteins are immune targets in leprosy and tuberculosis. Proc. Natl. Acad. Sci. USA 85, 4267-4270

Young, R.A., Mehra, V., Sweetster, D., Buchanan, T., Clark-Curtis, J., Davis, R.W., and B.R. Bloom. (1985b) Genes for

the major protein antigens of the leprosy parasite M.leprae. Nature 316, 450-452

Zhang, H., Scholl, R., Browse, J., and C. Somerville. (1988) Double stranded DNA sequencing as a choice for DNA sequencing. Nucl. Acid Res. 16, 1220

Ziai, M.R., Giordano, A., Armandola, E., and S. Ferrone. (1988) Purification by ammonium sulfate precipitation of bacteriophage lambda gt11 DNA for restriction analysis of cloned cDNA inserts. Anal. Biochem. 171, 192-196