Order Number 9018974

Sequence and structure of *mtr*, an amino acid transport gene, of *Neurospora crassa*

Koo, Kenneth, Ph.D.

University of Hawaii, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

SEQUENCE AND STRUCTURE OF *MTR*, AN AMINO ACID TRANSPORT GENE, OF NEUROSPORA CRASSA

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN BIOMEDICAL SCIENCES
(GENETICS)

DECEMBER 1989

By

Kenneth Koo

Dissertation Committee:

W. Dorsey Stuart, Chairman

David S. Haymer

Y. Edward Hsia

John A. Hunt

Bruce G. Adams

ACKNOWLEDGEMENTS

This one is for Karen. She has always provided an atmosphere of love, warmth, and responsible freedom. I am fortunate to be her husband. Kimberly and Kaleigh are my legacy. They have educated me into the role of being their father; a role which I will always cherish. Thank you for knowing where I was at night, girls!

My parents are models of resolve and have always been unrelenting sources of support. They will always find comfort, peace, and a place in our home.

Dorsey has been a profound and major influence in my young adult life. His friendship, honesty and giving nature are gifts which I will forever treasure.

Jo-anne and Daphne, the secs of the department, are friends who always gave me grief. I will truly miss the bitching, razzing and just sharing good times together.

In the lab, Doreen Leong and Elie Kato have kept me from being totally engrossed in my research. I will journey through life with Doreen's music.

I acknowledge and thank the following persons for financial support during my term at the Dept. of Genetics: Dr. G. C. Ashton (Genetics), Dr. S. Saul (Entomology), Dr. F. Greenwood (PBRC), and Dr. J. M. Ivy (HBG).

Drs. Rebecca Cann and Mort Mandel are acknowledged for their support of this work and their friendship.

ABSTRACT

The gene product of the *mtr* locus of Neurospora crassa participates in the transport of neutral aliphatic and aromatic amino acids across the plasma membrane. Mutants of the *mtr* locus have been identified by other workers by selecting for resistance to the poisonous amino acid analog 4-methyltryptophan. Using a transformation and selection scheme, three cosmids were cloned which complement the mtr-6(r) mutant allele. Two of the cloned DNAs were found to be tightly linked to restriction fragment length polymorphisms (RFLPs) which flank the *mtr* locus on linkage group IV.

The *mtr+* gene was subcloned as a 2.9 kbp fragment; nested deletion and restriction site subclones of the fragment were generated and used to determine the DNA sequence. DNA sequence analyses indicated the presence of a 845 bp open reading frame (ORF) with a 59 bp intron. The proposed *mtr* ORF contains sequences which are 100% homologous to N. crassa consensus sequences at the ATG start site, at the exon/intron splice boundaries, and internal intron sequences. The *mtr* ORF is followed downstream by a possible polyadenylation signal.

The 2.9 kbp fragment hybridizes to a poly A+ mRNA transcript of 2,300 nt. S1 nuclease mapping analysis of the transcript confirms the size of the transcript and the presence of the intron. An upstream open reading frame (uORF), found in the same reading frame as the *mtr* ORF, is proposed to be present in the mRNA transcript.

The proposed *mtr* ORF is predicted to translate into a 261 amino acid polypeptide with a molecular weight of 28,613 daltons. Codon usage of the *mtr* polypeptide is biased with 18 of the 61 sense codons being absent. The polypeptide is composed of 164 non-

polar amino acid residues which represent 63% of the amino acids. Hydropathy analysis of the *mtr* polypeptide indicates six possible transmembrane spanning domains with an average length of 23 amino acids which are proposed to span the membrane in alpha helical conformations. The *mtr* polypeptide does not have a signal sequence and the polypeptide is proposed to function as a dimeric protein. Homology to a RNP octamer motif supports the proposal that the *mtr* polypeptide functions as a ribonucleoprotein (RNP) particle.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Cells are units of life which exist within an ever-changing environment. The plasma membrane defines the extent of the cell and differentiates the intracellular contents from that of the environment. This difference is maintained by the activity and chemical nature of the molecules which make up the plasma membrane. The plasma membrane, as well as other biological membranes, is composed primarily of phospholipids, proteins and oligosaccharides. The arrangement and interactions between these molecules have been described in terms of a fluid mosaic model (Singer and Nicolson, 1972). The phospholipids, due to thermodynamic considerations, are oriented in a lipid bilayer arrangement. The hydrophobic fatty acid chains of the phospholipids are in contact with each other and comprise the inner core of the bilayer. The ionic and polar residues of the phosholipids are exposed to the hydrophilic environment and constitute the surfaces of the bilayer.

The arrangement of proteins in the membrane follows the general thermodynamic principles of maximizing hydrophobic and hydrophilic interactions. To maximize these interactions and hence achieve the lowest free energy state, the hydrophobic amino acid residues of a membrane protein are sequestered into the central core of the bilayer and the polar amino acid residues are exposed to the aqueous environment. That the globular proteins are embedded in the phosholipid bilayer matrix is demonstrated by freeze-etching experiments (Bayer and Remsen, 1970). Other non-covalent forces, such as hydrogen bonding and Van der Waals forces, also contribute to stabilize the macromolecular structure of membrane proteins.

The chemical nature of the phospholipids and membrane proteins prevents the direct mixing of the plasm of the cell with the aqueous exterior. However, the plasm and its dynamics are influenced by the environment. The communication between the plasm and the extracellular milieu is primarily mediated through the activity of membrane proteins.

Membrane proteins perform a number of different functions. Some membrane proteins function as ion-exchange pumps, as transporters of various metabolites and ions, and as receptors for chemicals and photons. Transport of metabolites and ions across the hydrophobic membrane, into as well as out of the cytoplasm, is a fundamental phenomena of cells. The energy expended to transport a metabolite is far less than that to synthesize it. Ions, necessary for many catalytic reactions within the cell, can only be obtained from the environment. Biochemical, physiological and genetic approaches have been employed to study the phenomena of the transport process. These approaches have revealed many important aspects. These include kinetic criteria, such as specificity and affinity of the transport process, energy coupling considerations, isolation and identification of proteins which transport the metabolites, and the identification of genes that affect transport.

The focus of this research effort is the transport of neutral aliphatic and aromatic amino acids into cells of Neurospora crassa. Mutants of the *mtr* locus have been selected for resistance to the poisonous amino acid analog, 4-methyltryptophan. These *mtr* mutants are defective in transport of neutral aliphatic and aromatic amino acids across the cell membrane via the N transport system (Perkins et. al., 1982). Much information has been documented about the genetics of the *mtr* locus, and about the physiological characteristics of *mtr* mutants (Stadler, 1966; Lester, 1966; Larimer and DeBusk, 1977). Some biochemical studies of the *mtr* gene product have been undertaken but they have been hindered because of the difficulty in isolating this membrane protein. To increase our understanding of the transport of neutral aliphatic and aromatic amino acids at the molecular level, the present investigation is specifically directed towards the determination of the sequence and structure of the *mtr* locus of Neurospora crassa. Through the DNA sequence, the amino acid sequence is predicted and a possible secondary structure of the *mtr* gene product can be inferred. The study of the secondary and tertiary structure of the *mtr* gene product is critical towards an the understanding of this fundamental process of amino acid transport.

Membrane Proteins and Structure

Membrane Proteins

Determination of protein structure is a common approach used to understand the function of proteins at the molecular level. An understanding of the transport process requires details about (a) the protein components; (b) the structure of the proteins, including the topology with respect to each other and with the membrane; (c) the identification of recognition sites for the substrate and energy coupling; and (d) the conformational changes of the proteins which occur during the transport process (Shuman, 1987).

Membrane proteins can be categorized operationally into two types, peripheral or extrinsic proteins and integral proteins. Peripheral proteins are described as being loosely attached to the membrane. They are dissociated from the membrane with mild treatment, such as an increase in ionic strength of the medium (Singer and Nicolson,

1972) or by cold osmotic shock (Ames, 1986). Anchored membrane proteins, examples of peripheral proteins, have part of the polypeptide inserted into the bilayer (Engelman and Steitz, 1984). The interaction between the polypeptide with other membrane proteins, and acylation of the polypeptide to fatty acids via $-NH_2$, -SH, and -OH groups of the amino acids enable the protein to be associated with the membrane (Sefton and Buss, 1987). Classic examples of peripheral proteins are spectrin and ankyrin of human erythrocytes. Spectrin is associated with the membrane via its interaction with ankyrin. Ankyrin, in turn, is bound to Band 3 - an integral protein - which is the anion transporter in red blood cells.

Integral membrane proteins are released from the membrane with the use of detergents, chaotropic agents, and organic solvents. Channel proteins and globular membrane proteins are examples of integral membrane proteins. Channel proteins, exemplified by the porins in the outer membrane of E. coli, form aqueous channels through the membrane. Globular membrane proteins have segments of the polypeptide spanning the bilayer, resulting in a tertiary structure within the hydrophobic region of the bilayer (Engelman and Steitz, 1984).

Globular membrane proteins may span the membrane once, as seen in the cell adhesion molecules, receptors, and immunoglobulins or traverse the membrane several times as observed for many proteins which transport metabolites (Ferguson and Williams, 1988). The membrane proteins which span the membrane once have either the amino (NH_2) or carboxyl (COOH) ends exposed to either the cytoplasm or the extracellular surroundings. Examples of membrane proteins which traverse the membrane several times are bacteriorhodopsin - the purple membrane protein of

Halobacterium halobium (Engelman and Steitz, 1984), and the lac permease of E. coli (Kaback, 1986).

Structure

The atomic descriptions of protein structure, revealed in the 1950s and 1960s, did not uncover the principles which govern protein folding. However, these experiments did determine fundamental secondary structures of local folding of proteins. These secondary structures are the alpha-helix and beta conformation. The majority of alpha-helices found in polypeptides are right-handed helices. On average, there are 3.4 amino acid residues per turn of the helix. Hydrogen bonding between the carbonyl oxygen and -NH-group of the peptide bond three residues apart, between the R groups contribute to the stability of the helix. The beta conformation is stabilized by hydrogen bonding between beta chains. These chains can be oriented in either a parallel or antiparallel arrangement. For a more detailed introduction of secondary structures, the reader is referred to Lehninger (1975) and King (1989).

Alpha-helices and beta sheets pack close together in proteins (Chothia, 1984) and form major categories of tertiary structure. These categories include antiparallel alpha, parallel alpha/beta and antiparallel beta structures (Richardson, 1984). One of the topological regularities in these structures is that the secondary structures adjacent in sequence are often in close proximity in three dimensions (Chothia, 1984).

Much is known about the secondary and tertiary structures, and about the atomic arrangement of enzyme-substrate complexes for soluble proteins. However, similar knowledge of membrane associated proteins is limited. The determination of the structures of membrane proteins has been hampered by the difficulty of obtaining protein crystals needed for diffraction studies. Only a few membrane proteins have been successfully crystallized. These include bacteriorhodopsin of H. halobium, porin, mitochondrial cytochrome oxidase and the light-harvesting chlorophyll a/b protein (Eisenberg, 1984). However, the crystals are often too small for high resolution X-ray diffraction studies.

The near absence of structural information has not affected the compilation of amino acid sequence information of membrane proteins. The amino acid sequence has been obtained by rapid DNA sequencing techniques. This discrepancy in the amount of structural information versus amino acid sequence data has prompted models (see below) which try to relate the two sources of information. The basis of these models has been the data obtained from a few exemplary membrane proteins.

Bacteriorhodopsin, involved in H^+ translocation, has been studied extensively. It is composed of seven rods which traverse the bilayer in a perpendicular manner. At least four of the seven rods exhibit an alpha helical arrangement as determined by X-ray diffraction studies. The amino acid sequence contains seven relatively hydrophobic regions sufficient to cross the bilayer (30-35 Å) when coiled as alpha helices (Eisenberg, 1984).

Another membrane protein which has been studied extensively is the ß-galactoside or lac permease of E. coli. Unlike bacteriorhodopsin, the lac permease has not been crystallized but can be readily purified. From circular dichroism studies of the purified protein, it has been shown that about 85% of the protein is arranged in alpha helical secondary structures. In light of the circular dichroism data, it is postulated that twelve of the hydrophobic segments of the lac permease are alpha helices. The mean length of the hydrophobic segments are 24 ± 4 amino acids (Kaback, 1986).

Other secondary structures have been implicated to exist in membrane proteins. These structures include beta helices, beta sheets, and bent alpha helices. From stereochemical and free energy considerations, it is proposed that the alpha helix is the preferred secondary structure by which proteins are inserted into the membrane (Engelman and Steitz, 1984).

There are many algorithms which calculate the degree of hydrophobicity of a protein. These algorithms have been based on the free energy values of transferring an amino acid from an aqueous solution to an organic phase. The Eisenberg algorithm (Eisenberg, 1984) is a consensus of various free energy values. The range of this scale is from -1.8 (Arginine) to + 0.73 (Isoleucine). Another combined-scale is used in the algorithm of Kyte and Doolittle (1982). The range of the Kyte and Doolittle scale is -4.5 (Arginine) and +4.5 (Isoleucine). Thus, the primary difference between the two scales is the assignment of values for the various amino acids.

The demonstration that the four 24 ± 4 amino acid hydrophobic segments are arranged as alpha helices in bacteriorhodopsin is significant. The number of amino acid residues in the hydrophobic regions is highly indicative of a membrane spanning segment (Doolittle, 1986). However, it has been demonstrated that other secondary structures may be associated with the hydrophobic segments.

Mechanisms of Transport and Energy Coupling

Mechanisms of Transport

There are several different modes by which metabolites and ions cross the lipid bilayer and enter into the cell. In passive transport, molecules cross the membrane at a rate determined by the concentration and electrical gradient - the electrochemical gradient - across the membrane. Small molecules, such as oxygen, carbon dioxide and water enter the cell by simple diffusion. In some instances, simple diffusion depends on membrane channels or pores. Discrimination between solutes is achieved by the size selectivity of the pores (Stein, 1986). Another example of passive transport is facilitated diffusion. This system is characterized by a membrane associated carrier protein which binds to the solute and transfers it across the membrane (Alberts et. al.,1983). The transfer process is due to the movement of the carrier protein's binding site across, or at least part of, the membrane (Stein, 1986). Uridine and leucine are transported into the human red blood cell by facilitated diffusion.

Unlike passive transport, active transport of molecules require the coupling of some form of energy (see below). In active transport, the substrate is transferred against an electrochemical gradient and is accumulated within the cell at a higher concentration than outside the cell. Three different active transport mechanisms have been found in bacteria (Wilson, 1978). A group translocation system, e.g. the phosphotransferase system (PTS), is characterized by chemical modification of the metabolite during the transport process. Vectorial phosphorylation, of the modification reaction, is the chemical energy source for this transport process. This system is exemplified by the transport of hexoses and hexitols. This system has not been found in higher phylogenetic species and is not ubiquitous in bacteria (Kaback, 1986).

The osmotic shock insensitive or membrane bound transport systems are characterized by the firm attachment of one protein component to the cell membrane. This mechanism is typified by the transport of lactose and proline (Furlong, 1987). The osmotic shock sensitive or binding protein transport system has a binding protein which is easily dissociated from the membrane by osmotic shock. This binding protein is present in the periplasmic space in bacteria and binds to the substrate with high affinity.

Other membrane bound proteins are required to transport the metabolite across the cell membrane. This transport mechanism is exemplified by the transport of L-Histidine via the high affinity histidine permease which is composed of proteins coded by the hisJ, Q, M, P genes and the transport of maltose by the maltose permease, composed of the products of the malE, F, G, and K genes, in bacteria (Ames,1986). Neither the membrane bound nor the binding protein transport systems modify the metabolite during the translocation process.

Energy Coupling

The mechanism of energy coupling to the transport process is complex. In some transport systems, hydrolysis of the $\beta - \gamma$ bond of ATP or from some other source e.g. acetylphosphate is coupled to the transport process. This ATP-coupled solute transport system is observed in the proton (H^+) and potassium (K^+) ATPases, and in the osmotic shock sensitive transport systems (Rosen, 1987). The sodium-potassium (Na^+: K^+) ATPase in higher phylogenetic organisms and the calcium (Ca^+) ATPase of the sarcoplasmic reticulum are further examples of this system. These ATP-coupled solute transport systems have been termed primary transport systems.

Another mode of coupling energy to the transport process is the concurrent transport of an ion with the solute. Such ion-coupled solute transport systems have been termed secondary transport systems. The sodium ion (Na^+) is commonly involved in mammalian transport systems, whereas the hydrogen ion is used in bacterial systems. When the ion is transported in the same direction as the solute, the process is termed a symport system. The antiport system is described by the counter directional flow of the ion and the solute. The reader is referred to the article by Maloney (1987) for examples of ion-coupled transport systems.

The widely favored chemiosmotic hypothesis states that the immediate driving force in energy-coupled membrane systems is the proton electrochemical gradient, (Kaback, 1986). The simple chemiosmotic cycle is composed of a primary transport system and a secondary transport system. The primary transport system extrudes a H^+ via a phosphate hydrolysis reaction. The secondary transport system accumulates the solute, together with H^+ or other cation, in response to the existing electrochemical gradient of the hydrogen ion (Maloney,1987).

Transport Systems in Prokaryotic Organisms

The following presentations of transport systems in prokaryotes are a few selected examples. The reader is referred to the review article of Postma (1987) on the phosphotransferase transport system for various sugars. Description of transport for a large number of different sugars, polyols and carboxylates is presented by Lin (1987). A most fascinating aspect of transport of sugars, as well as of amino acids, is that there is no one general mechanism by which a metabolite is taken up by the cell. Compounded on this phenomenon is that any one substrate may be transported into the cell by different and distinct transport systems. Active transport of D-galactose exemplifies this one substrate:multiple transport system aspect. D-galactose is transported into the cell by two inducible systems with differing K_m values. These systems are genetically distinct; the loci responsible for these different transport systems have been mapped and delineated from each other (Lin, 1987).

The lac permease

Lactose and other galactosides are transported into E. coli by the lac permease. The lac permease (K_M = 190 μM) is an example of an ion-driven transport system. Specifically, it is a H^+:lactose symport process; one proton is transported per lactose molecule. The energy source is the proton motive force or the proton electrochemical gradient. It has been suggested that the protein exists in a monomeric form which transports lactose by facilitated diffusion. However, the proton electrochemical gradient results in dimerization of the protein; the dimer is the form by which the protein participates in active transport (Kaback, 1986). The binding of protons to the permease has only been inferred from the coupled transport of protons and galactosides.

The lac permease, or M protein, is coded for by the lacY gene of the lac operon. The lacY gene is 1,254 bp long (Buchel et. al., 1980); the molecular weight of the gene product, predicted from the DNA sequence, is 46.5 kd. The molecular weight of the lac permease, based on SDS-PAGE, is 33.0 kd. The difference in the molecular weights is due to anomalous electrophoretic migration of this highly hydrophobic protein (Cronan et. al., 1987). Such electrophoretic anomalies have been observed for other integral membrane proteins. Excess binding of sodium dodecyl sulfate to the protein results in an electrophoretic migration that is not solely based on molecular weight of the protein.

As previously mentioned, circular dichroism analysis of the purified M protein indicates that 70 - 85% of the polypeptide has an alpha helical secondary structure. Hydrophobic analysis of the predicted amino acid sequence suggests that the 8 or 12 hydrophobic regions may be in $\alpha-$ helix conformation. Each of the hydrophobic regions has about 20 amino acids (Kaback, 1986).

Several lines of evidence indicate that the permease spans the membrane. The permease activity is inactivated or inhibited by N-ethylmaleimide treatment, by protease treatment, and by antibodies in both right-side-out (RSO) and inside-out (ISO) vesicles (Cronan et. al., 1987). These experiments indicate that the "active sites" are symmetrically located on both sides of the membrane. This symmetry is most easily explained by a membrane channel formed within the hydrophobic transmembrane domains of the protein. The substrate crosses the membrane through the hydrophilic core of the channel. Although the "active sites" are symmetrical, the protein is asymmetrically situated in the membrane with both the N- and C- terminals of the protein located on the cytoplasmic side of the membrane (Cronan et. al., 1987).

Histidine Transport

Much of the research on the high affinity histidine permease in S. typhimurium originates from the G. F. L. Ames group. The permease is composed of proteins coded by the hisJ, Q, M, and P genes. The reader is referred to the excellent review by Ames (1986).

The high affinity histidine permease system (K_M = 0.01 μM) is an example of an osmotic-shock sensitive transport system. Through the use of metabolic poisons, e.g. arsenate (affecting ATP levels) and cyanide (proton uncoupler), it has been inferred that the histidine permease system is energized by hydrolysis of a PO_4 bond, probably ATP or acetylphosphate (Ames, 1986).

The hisJ gene product, a histidine binding protein, is found in the periplasmic space and is released by osmotic shock. The hisJ protein contains a N-terminal signal sequence needed for secretion into the periplasm. Most secreted proteins contain a signal sequence which is recognized by a signal recognition particle (Walter and Blobel, 1982). The products of the hisM and hisQ genes are integral membrane proteins and are highly hydrophobic. Both gene products, based on the predicted secondary structure from the amino acid sequence, have three α-helical membrane spanning regions of about 20 amino acids each. Based on sequence homology, it has been proposed that the hisM and hisQ genes arose from gene duplication and subsequent divergent evolution. The hisP gene product is not highly hydrophobic and is believed to be a peripheral membrane protein (Higgins et. al., 1982). The hisM, hisQ, and hisP proteins do not have N-terminal signal sequences.

A current model of histidine transport is the initial binding of histidine to the hisJ protein in the periplasm. This binding induces a conformational change in the hisJ protein which allows the liganded hisJ protein to interact with the hisP protein. ATP hydrolysis is linked, by some unknown mechanism, to the hisP protein (Ames, 1986). Together, the interaction of the ligand/hisJ protein with the hisP protein and the ATP hydrolysis, induces conformational changes in the hisM and hisQ proteins. These conformational changes are such that the histidine binding sites of the hisJ, hisM, and hisQ proteins are juxtaposed to each other. The histidine molecule is released from the hisJ protein and is then bound to the hisM, hisQ proteins. The mechanism by which histidine then crosses the membrane is probably similar to a simple facilitated diffusion model in which the carrier protein's binding site is transferred across, or least part of, the membrane.

Transport Systems in Eukaryotic Organisms

Much information about transport systems in eukaryotes is known. The reader is referred to the following review articles on amino acid transport in eukaryotic microorganisms (Wolfinbarger, 1980; Horak, 1986). Numerous transport systems in higher eukaryotes have been well documented (Oxender et. al., 1985; Yim and Stuart, 1983; Birnbaum et. al., 1986). The following presentations of transport systems in eukaryotes are a few selected examples.

Like prokaryotes, most eukaryotic organisms display the one substrate: multiple transport system phenomenon. Prokaryotes and eukaryotic cells also have transport systems with broad specificities for classes of structurally related amino acids (discussed below). However, an intriguing feature in eukaryotes is the general lack of transport systems with specificities for single amino acids (Wolfinbarger, 1980). By contrast, in S. typhimurium histidine is transported by five different transport systems one of which, the hisJ, M, Q, P permease, transports histidine only.

Glucose Transporter

Glucose is transported across the plasma membrane in virtually all mammalian cells. In epithelial cells, like the brush border membranes of the kidney and small intestines, glucose is transported by a Na^+ dependent active transport carrier. In nonepithelial cells, like erythrocytes, glucose enters the cell via a carrier which functions in a facilitated diffusion mechanism. The two proteins are immunologically distinct (Birnbaum et. al., 1986).

The passive glucose transporters in both human and murine cells are about 55.0 kd from SDS-PAGE analysis. The protein is an integral membrane protein and it is glycosylated probably with a single, heterogeneous N-linked oligosaccharide. The gene coding for the transporter has been cloned for human cells (Mueckler et. al., 1985), and for murine cells (Birnbaum et. al., 1986). The homology at the DNA level between the human and rat sequences is 89% in the coding regions. In both cases, the sequences predict 492 amino acid residues polypeptides which correspond to non-glycosylated proteins of M_r = 54,117 daltons (human) and M_r = 56,133 (rat).

Hydropathy analysis, by the Kyte and Doolittle algorithm (1982), of the human glucose transporter amino acid sequence suggests twelve membrane spanning domains in α–helical conformation. Each domain is composed of about 21 amino acids. The Chou-Fasman method (1978) predicts β turns in the hydrophilic loops which separate the membrane spanning domains. No evident signal sequence was observed at the N-terminal of the polypeptide. Only one of the two potential N-glycosylation sites were found to be amenable to glycosylation. Like the lac permease of E. coli, the N- and C- terminals were depicted to be on the cytoplasmic side of the membrane (Mueckler et. al., 1985).

Yeast Histidine Permease

A general review of amino acid transport in the yeast, Saccharomyces cerevisiae, can be found by Horak (1986). S. cerevisiae has at least twelve (I-XII) constitutive and four "adaptive" amino acid transport systems. One of the four "adaptive" transport systems is the general amino acid permease system (GAP). The GAP system is repressed in glucose-ammonium medium. The GAP system transports all classes of amino acids.

Histidine is transported by two constitutive systems (III & IV), and by the GAP system. The hip1 gene is responsible for the high affinity constitutive system (III). The hip1 gene maps to linkage group VIIR. The hip1 gene has been cloned by transformation and complementation of a hip1⁻ mutant strain (Tanaka and Fink, 1985). The hip1 open reading frame is 1,596 bp. The predicted amino acid sequence indicates five potential N-linked glycosylation sites. No signal sequence was observed. Hydropathy analysis of the amino acid sequence suggests at least four hydrophobic regions sufficient to span the membrane.

Amino Acid Transport Systems in Neurospora crassa

It is generally accepted that the filamentous fungus N. crassa (a member of the class Ascomycetes) possesses five amino acid transport systems (Wolfinbarger, 1980). These systems have been categorized on the basis of the polarity of the amino acid transported. These systems are: the neutral (N) system - transport of neutral aliphatic and aromatic amino acids; the basic (B) system - transport of cationic amino acids; the general (G) system - transport of all classes of amino acids; the acidic (A) system - transport of anionic amino acids and a specific system for methionine. The A - and methionine - specific systems are activated only under special conditions of nitrogen and sulfur starvation, respectively; whereas the N, B, and G transport systems are expressed constitutively when the cells are grown on minimal media.

The transport activity of the N system has been described as a low affinity (K_M = 48-57 μM) and high velocity (V_{max} = 3.9-4.6 nmol min^{-1} mg^{-1}) process (Larimer and DeBusk, 1977; DeBusk and DeBusk, 1980). The B system's activity is described as high affinity (K_M = 1-3 μM) and intermediate velocity (V_{max} = 0.4 - 0.6 nmol min^{-1} mg^{-1}). The activity of the G system is characterized as a high affinity (K_M = 1-8 μM) and low velocity (V_{max} = 0.017-0.237 nmol min^{-1} mg^{-1}) process (DeBusk and DeBusk, 1980). All three systems are active transport systems. The concentration gradient of amino acids (internal to external concentration) established by the N and B systems is about 700- and 400- fold, respectively (DeBusk and DeBusk, 1965; Roess and DeBusk, 1968). Although the G system functions constitutively in an ammonium ion containing minimal medium, its transport activity is about 3 - 5 times lower than the B and N systems, respectively (Wolfinbarger and DeBusk, 1971). The activity of the G system is stimulated when the cells are cultured on ammonium free minimal medium.

Earlier studies on amino acid auxotrophic mutants reported the inhibition of growth when the mutants were grown in the presence of other amino acids. Thus, a histidine-requiring auxotroph fails to grow in the combined presence of a neutral and basic amino acid (Wolfinbarger, 1980). Histidine, being a zwitterion at the particular pH of the media, is transported by the N, B, and G systems (see Fig. 1.1). Transport studies, using radioactively labelled amino acids, confirmed the growth inhibition studies. The same amino acids which inhibited the growth of an amino acid auxotroph also competed in the uptake for that particular required amino acid. The delineation of the transport systems were ascertained by using transport mutant strains, and by genetic analyses which demonstrate independent segregation of the loci responsible for these transport systems (Rao et. al., 1975). The loci coding for gene products of the N, B and G transport systems are *mtr*, pmb, and pmg, respectively (see Fig. 1.1).

Biochemical investigations have been designed to isolate proteins from the plasma membrane without the disruption of the cell (Stuart and DeBusk, 1971; Travis et. al., 1972). Following extraction of conidia with a solution of saturated potassium chloride, three peaks of material were obtained from a L-arginine affinity column. The material in fraction A had little or no affinity to the L-arginine ligand. The material in peaks B and C exhibited affinity to L-arginine. Similar analyses of mutants for the N and B

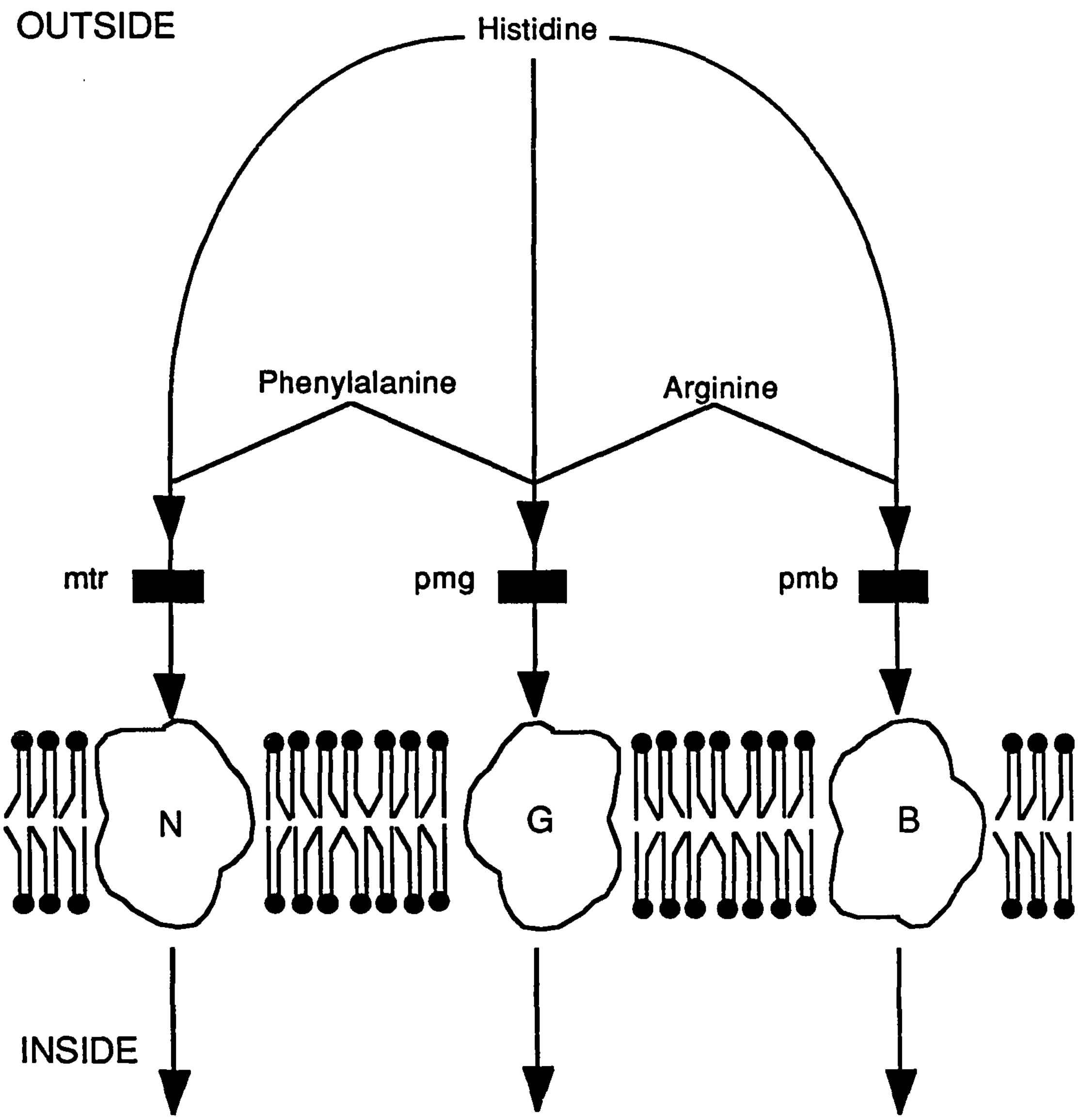


Fig. 1.1 A schematic representation of the three major amino acid transport systems of Neurospora crassa. Histidine is transported by all three systems. Phenylalanine is transported by the N and G systems while arginine is transported by the B and G systems. Mutations that reduce or eliminate these transport activities are represented by the heavy black lines which intersect the arrows and are denoted by mtr, pmb, and pmg. It is not known if the systems consist of more than one gene product (Stuart et. al., 1988).

transport systems resulted in the proposal that the material in fraction B was the product of the pmb locus (Stuart and DeBusk, 1971). Disc electrophoresis of fraction B material resulted in a single band. Biochemical analyses of peak B material indicated that it was a glycoprotein composed of a minimum 113 amino acid residues and eight types of sugars. No amino sugars were detected; this suggested that N-glycosidic linkages to asparagine residues were not present in this glycoprotein (Stuart and DeBusk, 1975).

UV absorption spectra of peak B material showed a significant amount of 260 nm absorbing material. The amount of 260 nm absorbing material was reduced when peak B was treated with RNase. No loss of protein was detected with this RNase treatment. Peak B material was also labelled *in vivo* with ^{14}C-uridine; the material, which when digested with RNase, co-chromatographed with 3'-uridine monophosphate. In addition, preincubation of conidia with RNase resulted in a decrease in transport activity of L-phenylalanine (Stuart and DeBusk, 1975). The effect was also observed when the RNase was covalently bound to polyacrylamide beads, indicating that the site of enzymatic activity was at the cell surface (Stuart and Woodward, 1975). These experiments eventually led to the isolation and characterization of extrinsic RNA (ext RNA) molecules (Stuart and Woodward, 1977). The RNAs are heterogeneous in size and range from 150-380 nts. Biphasic melting profiles of the RNAs indicate some degree of secondary structure.

At least two other cases have been reported that demonstrate RNA molecules being associated with proteins involved in transport. One report is the RNA molecule associated with the signal recognition particle (SRP) (Walter and Blobel, 1982). This RNA is 260 nts long, homologous to 7SL RNA (Ullu et. al., 1982), and is complexed with six proteins to make up the SRP. The SRP is involved with the translation and translocation of secreted proteins. The other report is of a RNA moiety necessary for the post-translational uptake of nuclear encoded proteins into the mitochondrial matrix (Firgaira et. al., 1984). Like the SRP-RNA, and the ext RNA, large amounts of RNase were needed to inhibit the transport activity. This suggests, as it was demonstrated for the SRP-RNA (Walter and Blobel, 1983), that the RNA is highly protected by the proteins to which the RNA is complexed.

N Transport System of Neurospora crassa

The *mtr* gene is involved in the activity of the N system. It was first identified in mutants that were resistant to 4-methyl-tryptophan (4-MT), a poisonous methyl derivative of tryptophan (Stadler, 1966; Lester, 1966). The mutants were also found to be resistant to para-fluoro-phenylalanine. The phenotype of resistance to 4-MT was not attributed to de-repression of tryptophan synthetase activity nor to the absence of early feedback inhibition in the tryptophan biosynthetic pathway. The mutants, however, exhibited a decrease in uptake of 4-MT, tryptophan and other amino acids of similar polarity. Uptake of sugars and amino acids of differing polarity was not altered in these mutants when compared to that of wild type cells. Transport of 4-MT across the plasma membrane however can occur via the G system. But the activity of the G system, in ammonium ion containing media, is sufficiently low so that the inhibitory effects of low concentration of 4-MT on growth are negligible when analyzed for a *mtr* mutant with a wild type G system.

Several lines of evidence, both biochemical and genetic, demonstrate that the *mtr* gene codes for a structural protein which is directly involved in the transport of neutral aliphatic/aromatic amino acids. The evidence is the result of numerous independent investigations:

I. Mutations of the *mtr* locus show reduced transport activity of neutral aliphatic/ aromatic amino acids. This reduction was determined by transport studies of radioactively labelled amino acids (Stadler, 1966; Lester, 1966). Futhermore, various *mtr* mutants exhibit reduced but differing V_{max} and K_M values (Larimer and DeBusk, 1977). The differing V_{max} and K_M values indicate that these mutants are not the results of mutations of a regulatory gene. Mutations at a regulatory gene would yield phenotypes that are either null, inducible, repressible or constitutive. Whereas, mutations of a structural gene would yield such differing biochemical values. On a molecular basis, the differing and reduced transport activity is attributed to an alteration of the *mtr* gene product. It is probable that the various mutant *mtr* gene products exhibit post-translational modification differences. But such post-translational modifications are probably the result of the mutated *mtr* gene product and not due to alterations in an unidentified modification system.

II. Wild type cells exhibit decreased amounts of transport activity after they have been subjected to methods which release macromolecules associated with the plasma membrane where one would expect the *mtr* gene product to be located (Stuart and DeBusk, 1971). The methods of release do not significantly affect cell viability nor disrupt the cell's integrity as visualized by electron microscopy (Travis et. al., 1972).

III. The released membrane-associated molecules bind to amino acids (Stuart and DeBusk, 1971). In a re-construction experiment, the released material was also capable of significantly increasing the permeation rate of amino acids when compared to lipid films that were depleted of such constituents (Stuart and DeBusk, 1973).

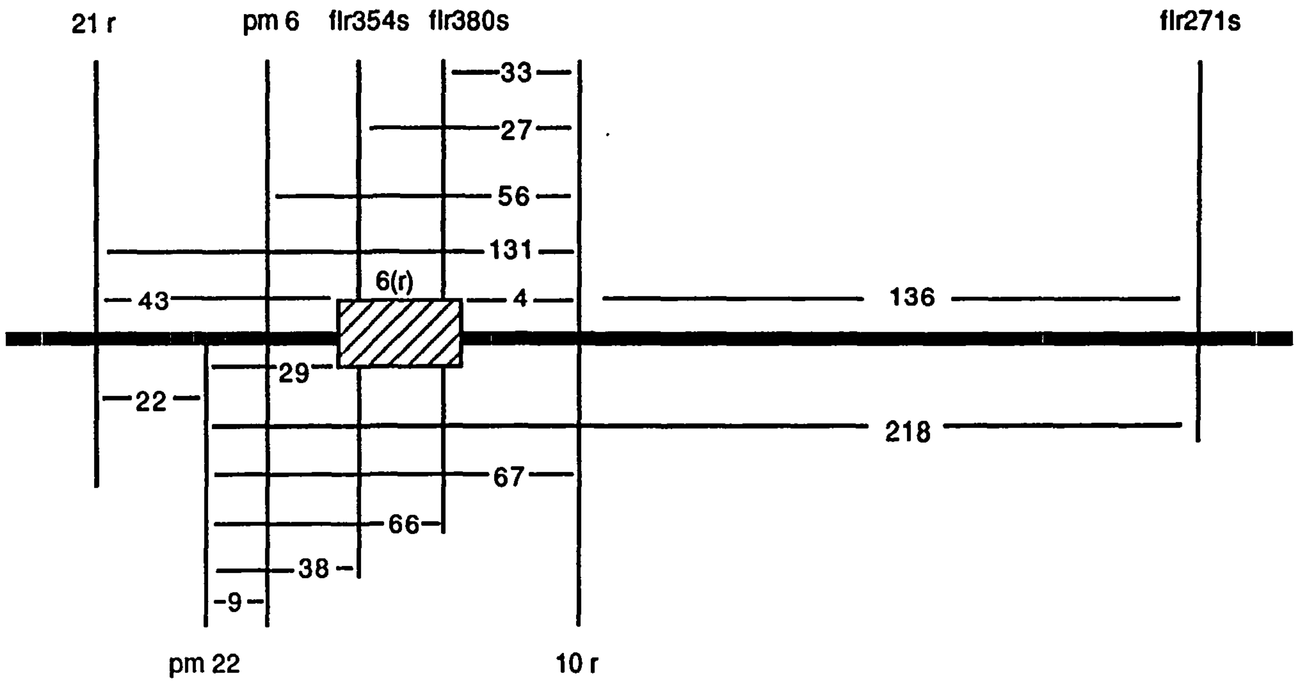


Fig.1.2 Recombination of the *mtr* locus (redrawn from Larimer and DeBusk, 1977). The numbers connecting the alleles represent the number of recombinants per 10^6 ascospores. The vertical lines denote the relative positions to which the different mtr alleles map.

IV. It has been reported that the released material from an *mtr* mutant strain exhibited a decrease in binding activity to that from a wild type strain (Wiley, 1970). However, it is not clear from this report whether the *mtr* gene product is acting only in a binding protein capacity or is involved in the actual translocating mechanism.

The *mtr* locus has been mapped to the right arm of linkage group IV (Stadler, 1966). Allelic mutants have been shown to undergo intralocus recombination, thus providing a fine structure map of the *mtr* locus (Larimer and DeBusk, 1977; Stadler, 1966). The recombination map is shown in Fig. 1.2 (Larimer and DeBusk, 1977). The failure of the 6(r) allele to recombine with alleles flr354s and flr380s suggests that the 6(r) allele may be a result of either a short intragenic deletion or an inversion.

Allelic complementation can be readily assayed by establishing heterokaryons that are homokaryotic for the *his*-2 gene (mutants of the his-2 require histidine) and

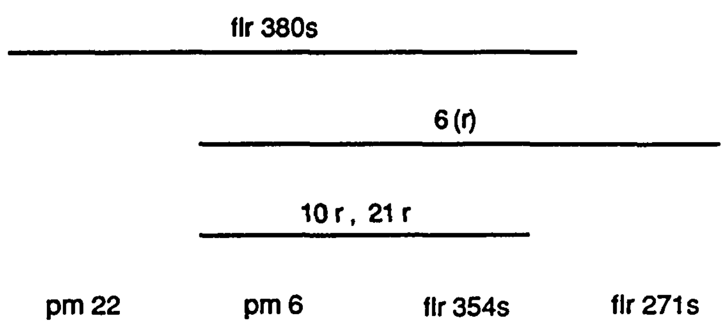


Fig. 1.3 Complementation map of the *mtr* locus (redrawn from Larimer and DeBusk, 1977). Alleles on a single line or on overlapping lines fail to complement with each other. The 6(r) allele complements the pm 22 allele and restores wild type transport activity.

heterokaryotic for the two *mtr* alleles (Larimer and DeBusk, 1977; Stadler and Kariya, 1969). Complementaion is indicated by growth of the heterokaryon on histidine - 10X arginine media. Uptake of histidine, to fulfill the *his*-2 requirement, via the B and G systems is competively blocked by the excess arginine. Larimer and DeBusk (1977) reported the occurrence of allelic complementation at the *mtr* locus using the above experimental design. The complementation map describes a single cistron indicating that the *mtr* gene codes for a product that functions as a homomultimeric protein. The complementation map is shown in Fig. 1.3.

The present research effort is a natural extension of the vast amount of genetic, physiological, and biochemical data gathered from the last twenty years of research. This dissertation describes the cloning and sequencing of the *mtr* locus in Neurospora crassa. Through analysis of the predicted amino acid sequence, a secondary structure of the *mtr* polypeptide is proposed. Determination of the structure may assist future research on the mechanism by which neutral aliphatic and aromatic amino acids are transported across the plasma membrane.

CHAPTER TWO

MATERIALS AND METHODS

Strains, Maintenance and Media

Strains of *Neurospora crassa*

The majority of *N. crassa* strains used in this study were obtained from the Fungal Genetics Stock Center (FGSC) or from Dr. D. D. Perkins (Stanford University). New strains were generated by mating crosses and selection; the crosses were performed by Dr. W. D. Stuart and L. Pineda. A summary of the strains and their origin is shown in Table 2.1. All strains were grown on 1X Vogel's Minimal Medium (N medium) in cotton-plugged test tubes; supplements were added depending on the strain's phenotype.

Strains of *Escherichia coli*

E. coli strain NM522, genotype supE, thi, Δ(lac-proAB), Δhsd5(r^-,m^-), {F', proAB, laclqZΔM15, was purchased from Pharmacia Lab. Streaks of NM522 cells on M9/thiamine plates were used to select for cells which contain the F (fertility) factor. Individual colonies on the M9/thiamine plates were used to inoculate 2X YT medium to establish glycerol stocks and for preparation of competant cells.

E. coli strain LE392, genotype F^-, hsdR514(rk^-,mk^+), supE44, supF58, lacY1 or Δ(laclZY)6, galK2, galT22, metB1, trpR55, λ^-, was a gift from Dr. S. J. Vollmer.

Table 2.1 <u>Neurospora crassa</u> strains used

Loci	Alleles	FGSC No.	Supplements	Source
74-OR23-1A	wild type Mating type A	987	-	FGSC
mtr	6(r) Mating type A	3043	-	DDP
his-2	Y152m14 Mating type a	21	L-Histidine	DDP
his-2; mtr	Y152m14 6(r) Mating type a	_	L-Histidine	3043 X 21
his-2	T51M152(t) Mating type a	4625	L-Histidine	DDP
mtr; pmb; pmg	Pm-22(r); can-37(r); pmg Mating type A	4607	-	FGSC
mtr; pmb; pmg; his-2 quad-2	Pm-22(r); can-37(r); pmg; T51M152(t) Mating type a		L-Histidine	4607 X 4625
mtr; pmb; pmg; his-2; X-40	6(r); can-37(r); pmg; T51M152(t) Mating type?		L-Histidine	quad-2 X 3043
mtr; su (mtr)	21 (r); 21-2 Mating type A	1718	-	FGSC
mtr	flr271 (s) Mating type A	1317	-	FGSC
mtr	flr354 (s) Mating type A	1318	-	FGSC
al-2; nuc-2; arg-12; inl cot-1		4411- 4430	L-Arginine Inositol	FGSC

Individual colonies on 1X YT plates were used to establish glycerol stocks and preparation of competant cells.

E. coli strains transformed with plasmids carrying the ampicillin resistance marker were maintained and grown in an identical manner as the host bacterium except for the constant selection of ampicillin resistance.

Media

All media were autoclaved. Heat-labile supplements and antibiotics were added after the media had cooled. The components of N medium and synthetic crossing medium can be found in the review by Davis and DeSerres (1970). The supplements for the various *N. crassa* strains are listed in Table 2.1

The components of M9 salt solution are described by Maniatis et. al., (1982). Thiamine was added to a final concentration of 1 μg/ml for M9/thiamine plates. 2X YT medium contains 16.0 g Bacto-Tryptone, 10.0 g Bacto-Yeast Extract, 5.0 g sodium chloride per liter and adjusted to pH 7.4 with sodium hydroxide. 1X YT contains half the amount of Bacto-Tryptone and Yeast Extract as 2X YT. Ampicillin was added to media at a final concentration of 50-100 μg/ml.

General Materials and Methods of Nucleic Acid Analysis

The methods presented in this chapter, though written in a "manual-like" style, are not intended to be authoritative. Numerous variations of the described protocols exist in the literature and readers are encouraged to adopt protocols which satisfy the intricacies of experimental research. The methods of this particular research effort were generally performed in the manner described to achieve some degree of reproducibility.

Materials

All reagents were analytical grade or better. All aqueous solutions were prepared with double-distilled water and autoclaved unless stated otherwise. Some solutions were filter sterilized with a 0.45 µm filter.

Phenol (redistilled or molecular biology grade) was prepared by adding 8-hydroxyquinoline, an anti-oxidant, to a final concentration of 0.1%. Phenol was saturated with water and stored at -20°C. Prior to usage, phenol is thawed and equilibrated with TE, pH 8.0.

Chloroform: Chloroform:Iso-amyl alcohol (24:1)

TE, pH 8.0: 10 mM Tris.Cl, pH 8.0; 1.0 mM EDTA, pH 8.0

0.5 M EDTA: 186.1 g Disodium ethylene diamine tetraacetate. $2H_2O$ in water. Adjusted to pH 8.0 with sodium hydroxide. Dissolved and adjusted to 1 liter with water.

7.5 M NH_4OAc: 578.1 g Ammonium acetate dissolved and adjusted to 1 liter with water. Filter sterilized.

5X TBE: 54.0 g Tris base; 27.5 g Boric acid; 20.0 ml 0.5 M EDTA, pH 8.0 dissolved and adjusted to 1 liter. This solution is usually not sterilized.

6X Loading Buffer: 0.25% Bromophenol blue; 30% Glycerol

10 mg/ml EtBr: 0.5 g Ethidium bromide dissolved and adjusted to 50 ml. This mutagen is not sterilized.

95% EtOH

Methods

(i) Purification of Nucleic Acids

Analysis and enzymatic manipulation of nucleic acids require that the nucleic acid preparation is relatively free of contaminating proteins. Proteins in nucleic acid solutions are generally removed with the organic solvents, phenol and chloroform - extracting once with phenol, once with phenol/chloroform and once with chloroform.

The three extraction methods are similar. The nucleic acid, dissolved in a low ionic strength buffer e.g. TE, is mixed with an equal volume of the organic solvent. Rapid separation of the two phases is achieved by centrifugation and the upper aqueous phase is recovered. Phenol/chloroform extraction entails the mixing of the nucleic acid solution with 1/2 volume of phenol, followed by the addition of 1/2 volume of chloroform and subsequent mixing, centrifugation and recovery of the aqueous phase. Although chloroform is a poor de-proteinizing agent, chloroform extraction essentially removes traces of phenol that would interfere with the spectrophotometric quantitation of the nucleic acid sample.

(ii) Concentration of Nucleic Acids

Traces of phenol and chloroform can be removed by precipitating the nucleic acid with ethanol. The presence of a moderate concentration of monovalent cations in the nucleic acid solution is required for effective precipitation. The final concentration of some monovalent cations can be found in Maniatis et. al., (1982).

Precipitation of DNA is usually carried out by adding 1/2 volume 7.5 M NH_4OAc, and then 3 volumes of cold 95% EtOH. Storage of the mixture at low temperatures

(-20°C) allows for the precipitation of nucleic acid. Storage period may vary from 10 min to overnight depending on the experimental conditions.

The nucleic acid is recovered by centrifuging at 10,000 to 12,000 g for 10 min. The supernatant is decanted. Trace amounts of cations are removed by rinsing the nucleic acid pellet with cold 70% EtOH, re-centrifuging and decanting of the supernatant. The centrifuge tube is placed in an inverted position to drain excess moisture. Air drying or vacuum dessicating removes traces of the supernatant from the pellet.

The nucleic acid pellet can now be dissolved with a buffer at a desired concentration. Heating at 42°C and mixing of the sample aids in dissolving the pellet. The concentration and purity of the nucleic acid sample can be determined either by the UV - light absorbance values at 260 and 280 nm wavelength or by agarose mini-gel analysis.

(iii) Agarose Mini-gel

The quality and concentration of a DNA sample can be determined by electrophoresing an aliquot in a mini-gel. This procedure is routinely used to assess enzymatic manipulations of a DNA sample. Usually 1% Agarose (BRL) in 1X TBE is heated in a microwave oven until the agarose is completely dissolved. Ethidium bromide (EtBr) is then added to a concentration of 0.5 μg/ml. The molten gel is poured onto 2-inch x 3-inch glass slides. A gel comb is mounted above the glass slide so as to form a slot in the gel when the molten gel polymerizes. The hardened gel is then placed in an electrophoretic chamber containing 1X TBE and EtBr.

The DNA sample, in 1X loading buffer, is pipetted into the slot. Electrophoretic conditions vary from 20-200 V. The great flexibility of voltage applied reflects the intent of the analysis which is the monitoring of enzymatic steps. The nucleic acid sample is visualized by using a long-wave U.V. transilluminator.

At times when a critical determination of the molecular weight of a DNA sample is required, the length of the agarose gel is increased and the electrophoresis is carried out at 2 to 3 volts/cm of gel. Varying the agarose concentration will improve the accuracy of the molecular weight estimate. Lower concentrations of agarose should be used when resolution of high molecular weight fragments is needed; likewise, low molecular weight fragments are resolved better at higher agarose concentrations. A table of agarose concentrations and their resolving range can be found in Maniatis et. al., (1982).

The concentration of a DNA sample is estimated by comparing the intensity of the UV fluorescence emitted by the sample with that emitted by a DNA sample at a known concentration. Since the intensity of fluorescence is proportional to the amount of DNA present, the molecular weights of the DNA samples have to taken into account so as to obtain a reliable estimate.

Nucleic Acid Preparation

Isolation of *Neurospora crassa* Genomic DNA

This method of isolating genomic DNA from *Neurospora* is a modification of that used by Akins and Lambowitz (1985).

Materials

Lysis Buffer: 50 mM Tris.Cl, pH 8.0; 2.0% SDS; 62.5 mM EDTA

5.0 M KOAc (~ pH 4.8): 29.45 g Potassium acetate; 11.5 ml glacial acetic acid dissolved and adjusted to 100 ml.

Iso-propanol

Methods

In a cotton plugged 250 ml Erlenmeyer flask, 100 ml sterile 1X Vogel's medium is inoculated with a suspension of conidia. Supplements are added to N medium depending on the growth requirements. Cell growth is allowed to reach late log or stationary phase. Cells are harvested by filtration through a glass sintered funnel. The cells are rinsed 2-3 times with water. The mycelial mat is collected and frozen at -80°C.

The frozen mycelial mat is pulverized with acid washed sand and mortar/pestle. The sand, mortar and pestle are pre-cooled at -80°C. The cell/sand mixture is placed into a 30 ml Corex tube, resuspended with 5 ml lysis buffer and incubated at room temperature for 60 min with occasional gentle mixing. Sand and cell debris are pelleted at 3,000 rpm at 4°C for 10 min.

The supernatant is decanted into a 15 ml Corex tube and adjusted to 2.0 M with 3.3 ml 5.0 M KOAc. The solution is inverted 2-3 times and placed in an ice-bath (30 min). The suspension is centrifuged at 10,000 g at 4°C (10 min). The supernatant is decanted into a 30 ml Corex tube containing 5.5 ml 5.0 M NH_4OAc. An equal volume of iso-propanol is added and the solution is allowed to sit overnight at -20°C.

Nucleic acid is pelleted by centrifuging at 10,000 g at 4°C for 10 min. The pellet is dissolved with 1.5 ml TE and placed into a 15 ml Corex tube. To obtain high molecular weight genomic DNA, all manipulations are such so as to minimize shearing of the DNA. Mixing and pipetting of the DNA are performed gently. The solution is extracted once with phenol/ chloroform and precipitated with NH_4OAc/EtOH. The precipitate is recovered by centrifuging as before and rinsed once with 70% EtOH. The pellet is dissolved with 1.0 ml TE. RNA is digested by adding 20 µl 500 µg/ml RNase A and incubating at 37°C for 60 min.

The ribonuclease and any remaining proteins are removed by a phenol/ chloroform extraction, followed by two chloroform extractions of the aqueous solution. DNA is precipitated with NH_4OAc/EtOH and rinsed twice with 70% EtOH. The pellet is dissolved with 1.0 ml TE and stored at 4°C.

Isolation of mRNA from N. crassa

This protocol is a minor modification of that by Reinert et. al., (1981). The contaminating DNA molecules were digested with RNase-free DNase I at a final concentration of 10 µg/ml as described by Ausubel et. al., (1987).

Materials

3.0 M NaOAc, pH 5.5: 40.8 g sodium acetate.$3H_2O$ dissolved in water and adjusted to pH 5.5 with glacial acetic acid. The volume is adjusted to 100 ml.

Extraction Buffer: 100 mM sodium acetate, pH 5.5; 1.0 mM EDTA; 2% SDS

RNase-free DNase I: BRL Cat. No. 8047SA

Methods

The growth, harvest and storage of the mycelial mat is the same as previously described in the extraction procedure of genomic DNA. The frozen mycelial mat is pulverized with acid washed sand and mortar/pestle. The sand and mortar/pestle are pre-cooled at -80°C. The powder is placed into a sterile 35 ml Oakridge tube and resuspended in 5.0 ml Extraction buffer and 2.5 ml phenol. The suspension is mixed vigorously at high speed (Vortex mixer) for 2 min. 2.5 ml chloroform is added and the suspension is mixed as before for 2 min. The suspension is heated at 50°C for 20 min.

The suspension is centrifuged at 5,000 rpm for 5 min and the upper aqueous phase is recovered. 2.5 volumes of 95% EtOH is added and the solution is stored at -70°C (30 min). The nucleic acids are pelleted at 10,000 rpm (10 min) and rinsed once with 70% EtOH. The pellet is dissolved with 2.0 ml Extraction buffer and is extracted 2 times with phenol/chloroform and once with chloroform. The aqueous solution is aliquotted into four 1.5 ml microfuge tubes. The nucleic acids are again precipitated with 2.5 volumes of 95% EtOH at -70°C (30 min), rinsed once with 70% EtOH and air-dried.

The pellet of each tube is dissolved with 200 µl H_2O. An aliquot of 1.0 µl is saved for electrophoresis. The solution is adjusted to 10 mM Mg_2SO_4. Contaminating DNA is digested by adding RNase-free DNase I to a final concentration of 5 µg/ml and incubating at 37°C (15 min). The digestion is stopped by adjusting to 20 mM EDTA. The electrophoresis of an aliquot, together with the previously saved aliquot, is used to determine the extent of the DNase I digestion. The solution is adjusted to 0.3 M NaOAc, pH 5.5. DNase is removed by extracting once with phenol/chloroform and once with chloroform. RNA is precipitated by adding 2.5 volumes 95% EtOH at -70°C (30 min), rinsed once with 70% EtOH, and the pellet is dissolved with 200 µl H_20.

Success in using the above isolation procedure is strongly dependent on the rapid execution of the various steps. Aqueous solutions of RNA is stored at -70°C and not at -20°C. Ethanol precipitated RNA pellets may be stored at -20°C.

Rapid Plasmid Isolation from E. coli

Alkaline Lysis Method

This protocol is a modification of that used by Birnboim and Doly (1979). The removal of rRNA with 5.0 M Lithium chloride is suggested by Pelham (1985). The quality of the plasmid preparation is good for restriction enzyme digestion. The yield of plasmid will vary depending on the plasmid, but 10 µg per 1.5 ml of culture is usually attained.

Materials

GTE: 50 mM Glucose; 25 mM Tris.Cl, pH 8.0; 10 mM EDTA

10 mg/ml Lysozyme: 0.1 g lysozyme dissolved and adjusted to 10 ml with sterile GTE. Store as 1.0 ml aliquots at -20°C.

5.0 M NaOH: 20 g Sodium hydroxide dissolved and adjusted to 100 ml. This solution is not sterilized.

10% SDS: 100 g Sodium dodecyl sulfate in water. Adjust to pH 8.0 with HCl. Adjusted to 1 liter with water. This solution is not sterilized.

Lysis Buffer: 0.2 M NaOH; 1% SDS. This solution is prepared fresh for each isolation.

Methods

Single colonies are used to inoculate 2.5 ml 2X YT medium containing 50-100 µg/ml ampicillin. Cultures are shaken (200-300 rpm) and grown overnight at 37°C.

Bacteria from 1.5 ml of culture are centrifuged at 12,000 g for 15-30 sec in micro-tubes. The pellet is resuspended and mixed in 100 µl GTE. 40 µl 10 mg/ml lysozyme is added and the mixture is incubated at room temperature for 5 min. The spheroplasted bacteria are lysed by adding 200 µl lysis buffer. The tubes are gently inverted 2-3X and stored in ice for 5 min. The mixture is neutralized with 150 µl cold 5.0 M KOAc, pH 4.8. The tubes are moderately mixed (5-10 sec) in an inverted position and stored in ice for 5 min.

Cell debris and the trapped chromosomal DNA is removed by centrifuging at 12,000 g for 10 min. The supernatant is recovered and mixed with 1.0 ml 95% EtOH. The mixture is stored at -20°C. The precipitate is pelleted as before and rinsed once with 70% EtOH. The air-dried pellet is dissolved with 50 µl TE. Most of the rRNA are precipitated out by adding an equal volume of 5.0 M LiCl and storing the mixture in ice for 5-10 min. The precipitate is pelleted as before. The supernatant is recovered and mixed with 200 µl 95% EtOH. The mixture is stored at -20°C for 10 min; centrifuged and rinsed once with 70% EtOH. The air-dried pellet is dissolved with 30 µl TE, precipitated with NH_4OAc/EtOH, and rinsed twice with 70% EtOH. The final pellet is dissolved with 20 µl TE.

Rapid Boiling Method

This protocol is a modification of that of Holmes and Quigley (1981). It incorporates the previously described precipitation with 5.0 M LiCl. The quality of the plasmid preparation is good for restriction enzyme digestion and double-stranded DNA sequencing. The yield is similar to that of the alkaline lysis method.

Materials

STET: 8% Sucrose; 0.5% Triton X-100; 50 mM EDTA, pH 8.0;10 mM Tris.Cl, pH 8.0

Methods

Growth and harvesting of bacteria is the same as the alkaline lysis method. The bacterial pellet is resuspended and mixed with 350 µl STET buffer. The bacteria are spheroplasted with 25 µl 10 mg/ml lysozyme and mixed for 3 sec. with a Vortex at full speed. The spheroplasts are lysed by boiling for 40 sec. The suspension is centrifuged at 12,000 g, 4°C for 10 min.

The pellet of cell debris and trapped chromosomal DNA is removed with a toothpick. The supernatant is mixed with 40 µl 3.0 M NaOAc, pH 5.2. The solution is then mixed with 500 µl iso-propanol. The precipitate is formed at -20°C for 10 min and pelleted as previously described. The pellet is rinsed once with 70% EtOH. Further removal of rRNA with 5.0 M LiCl, precipitation with NH_4OAc/EtOH and preparation is identical to that of the alkaline lysis method.

Transformation

Transformation occurs when naked DNA from the medium enters a cell and results in a genetic alteration of the cell. The genetic alteration is reflected by the stable passage of the DNA molecule from the parental to daughter cell. The DNA molecule may either confer a new phenotype to the host cell e.g. antibiotic resistance, or complement a phenotypic marker of the host cell. Complementation of a phenotypic marker is sometimes referred to as marker rescue. In both cases the exogenous DNA molecule is expressed and the transformed cell is selected for a novel phenotype.

Transformation of N. crassa

One protocol of N. crassa transformation requires the formation of conidial spheroplasts. The spheroplasts are prepared by the enzymatic removal of the tough cell wall with Novozyme under isotonic conditions. An isotonic solution is needed to prevent the "bursting" of the spheroplasts. This protocol is adapted from the method of Vollmer and Yanofsky (1986).

Materials

Filtering units: cheesecloth formed into a bag in the mouth of an Erlenmeyer flask and secured with rubber bands. Autoclave.

Hemocytometer

Novozyme 234: purchased from Novo Laboratories.
1.0 M Sorbitol: 182 g Sorbitol dissolved and adjusted to 1 liter.

STC: 91 g Sorbitol; 50 mM Tris.Cl, pH 8.0; 50 mM $CaCl_2$. Adjusted to pH 8.0 with NaOH. Dissolved and adjusted to 500 ml with water.

PTC: 200 g Polyethylene glycol (PEG) "4000" (Sigma); 50 mM Tris.Cl, pH 8.0; 50 mM $CaCl_2$. Adjusted to pH 8.0 with NaOH. Dissolved and adjusted to 500 ml with water.

50 mM Spermidine: 12.73 mg Spermidine dissolved in 1.0 ml TE, pH 8.0 and stored at -20°C.

5 mg/ml Heparin: 50 mg Heparin, sodium salt dissolved in10 ml STC and stored in frozen aliquots.

10X FIGS: 5 g Fructose; 2 g Inositol; 2 g Glucose; 200 g Sorbose. Dissolved and adjusted to 1 liter with water.

0.25 mg/ml Benomyl: Dissolved with 95% EtOH and stored at -20°C.

Regeneration "Top" Agar: 20 ml 50X Vogel's Minimal Medium; 825 ml H_2O; 182 g Sorbitol and 28 g Bacto-Difco Agar. Autoclaved and 100 ml 10X FIGS is added.

Plating "Bottom" Agar: 0.5 mg/ml L-Histidine.HCl; 5 mg/ml L-Arginine.HCl in 1x N medium. 10X FIGS and benomyl is added after autoclaving to final concentrations of 1X and 0.5 μg/ml, respectively. It is critical that the volume of "bottom" agar poured into each petri plate is 25 ml. This parameter is as such because growth of benomyl resistant transformants will cease if the amount of benomyl is too great.

Methods

a. Preparation of Spheroplasts

The host strain is inoculated onto 25 ml solid N medium in cotton-plugged 125 ml Erlenmeyer flasks. Four to five flasks should provide sufficient conidia for the protocol. The N medium is supplemented with 0.5 mg/ml L-Histidine. The host strain, X-40 - genotype described in Table 2.1, appears to grow best at this concentration of L-Histidine. The cultures are grown at room temperature for 5-7 days.

Conidia are harvested by adding 10 ml N medium per flask, replacing the cotton plug and swirling the flask. The conidia are allowed to settle for a few minutes. The conidial suspension is poured into the cheesecloth bag of the filtering units and the filtrate is recovered. The concentration of conidia is determined by a hemocytometer count; chains are counted as one.

A volume of 2×10^9 conidia is added to 150 ml liquid N medium containing 1.5% sucrose and 0.5 mg/ml L-Histidine. The conidia is germinated while shaking the cotton-plugged flask (150-200 rpm) for 5-6 hr at room temperature. The cells are harvested when more than 75% are germinated and the germ tubes are 1-4 conidial diameters in length. The cells are harvested by centrifuging in a swinging bucket centrifuge at 1500-2000 rpm for 10 min. The cell pellet is rinsed three times with water.

The pellet is resuspended in 10 ml 1.0 M sorbitol. In a sterile 250 ml Erlenmeyer flask, the conidial suspension is added to 50 mg solid Novozyme 234. Digestion of the cell wall is carried out with shaking (100 rpm) at 30°C for approximately 1 hr ± 10 min. The spheroplasting is monitored by taking a small aliquot of the digested cells and examining the cells microscopically under a cover slip. Spheroplasts are osmotically lysed when water is applied to one edge of the cover slip. Monitoring of the spheroplasts should be done frequently at the later stages of the incubation period. The spheroplasting procedure is perhaps the most critical step for a successful transformation.

The spheroplasts are decanted into a sterile 15 ml conical centrifuge tube and recovered by centrifuging at 500 rpm (10 min) in a swinging bucket table top centrifuge. The pellet is rinsed twice with 10 ml 1.0 M sorbitol and then once with STC. The final spheroplast pellet is suspended in 16.0 ml STC, 200 μl DMSO and 4 ml PTC. The cells can then be used directly for transformation or stored frozen in 1.0 ml aliquots at -80°C.

b. Transformation of conidial spheroplasts

2.0 μl 50 mM Spermidine, 5.0 μl DNA (usually at 1.0 mg/ml) and 5.0 μl 5 mg/ml heparin are aliquotted into a sterile 15 ml screw-cap tube. The solutions are mixed by flicking the tube. The contents are briefly spun (pulsed) in a table top centrifuge and then placed into an ice-bath. 50-100 μl of thawed spheroplasts are added to the DNA solution and incubated on ice (30 min). Incubation periods of 20 min on ice have been successful. 1 ml PTC is added and mixed well by flicking the tube. The mixture is incubated further at room temperature (20 min).

15 ml Regeneration "Top" agar, incubated at 50-55°C, is poured into the tube. The contents are quickly mixed by flicking and inverting the tube 2-3 times and then

uniformly poured onto the plating "bottom" agar. Bubbles are removed by flaming. Prewarming the bottom agar plates prevents the premature hardening of the top agar. The plates are kept in an upright position until the top agar has solidified (about 5 min). The plates are incubated in an inverted position at 34°C.

Selection for N. crassa transformants

Transformants of the host X-40 strain were selected for two phenotypic markers, the mtr phenotype and benomyl resistance. The marker for benomyl resistance is semi-dominant. It is contained in the cosmid pSV50 which was used in the construction of the ordered genomic library of N. crassa. Selection for the *mtr* phenotype is similar to that used by Stadler and Kariya (1969). The host X-40 strain requires L- Histidine at the conditional temperature of 34°C. Three of the major transport systems are defective in the X-40 strain, thus inhibiting the uptake of L-Histidine. Transport of L-Histidine via the B system (basic amino acids) and the G system (general amino acid permease) is inhibited further by a 10-fold excess of L-Arginine. Transformation with DNA which contain the *mtr* gene would fulfill the L-Histidine requirement by restoring the N transport system.

Methods

The transformed cells are incubated at 34°C. An indication of a successful transformation can be seen about 24-36 hr after plating. Stable transformants are scored after 3 days of growth. In some cases, the plates are subdivided into sectors to facilitate the count. The incubation period to detect transformants will vary depending on the host strain and the phenotypic marker that is being investigated. The above observations are those for the *mtr* locus under these conditions.

Transformation and Selection of E. coli

The procedure of preparing E. coli competant cells is that of Mandel and Higa (1970) as described by Maniatis et. al., (1982). Strains NM522 and LE392 were grown to logarithmic phase (A_{600} = 0.2-0.4) and harvested. Competant E. coli cells are usually used for transformation 12-24 hr after being treated with the Tris.Cl/$CaCl_2$ solution. Preparation of frozen competant NM522 cells by this method does not yield very high transformation frequencies.

Materials

Tris.Cl/$CaCl_2$ solution: 10 mM Tris.Cl, pH 8.0; 50 mM $CaCl_2$. Autoclaved and stored at 4°C.

100 mM IPTG: 23.8 mg Isopropyl-β-D-thiogalactopyranoside dissolved in 1.0 ml water and stored at -20°C.

2% Bluo-Gal: Bethesda Research Laboratories. Dissolved in dimethylformamide (DMF) and stored at -20°C.

Methods

The cells are grown to logarithmic phase (A_{600} = 0.2-0.4) and pelleted by centrifuging at 5000g, 4°C for 10 min. The cell pellet is resuspended in a volume of Tris.Cl/$CaCl_2$ (pre-chilled in an ice-bath) which is half of the culture volume. The cells are maintained in the ice-bath for 15 min and then centrifuged again using the above conditions. The pellet is resuspended in a pre-chilled volume of Tris.Cl/$CaCl_2$

solution that is 1/15 of the culture volume. The cells are dispensed as 200 µl aliquots into pre-chilled microfuge tubes and stored at 4°C until used for transformation. It is important to maintain the temperature of the cells and solutions at 4°C.

40-100 ng of vector DNA in 100 µl Tris.Cl/$CaCl_2$ solution is usually added to 200 µl of E.coli competant cells. The DNA/ cell mixture is incubated an ice-bath for 30 min. After the incubation period, the cells are heat-shocked at 42°C (2 min) or at 37°C (5 min) and 1 ml 2X YT medium is added. The cells are then incubated for 45-60 min at 37°C, to allow for β-lactamase expression. 200 µl of the cell/media mixture is plated onto 1X YT & Amp plates with a sterile spreading bar. The plates are kept in an upright position (at least for 30 min) to allow the moisture to diffuse into the agar. The plates are then incubated overnight at 37°C in an inverted position.

Use of the "blue/white" screening method to detect recombinant E. coli cells is based on the enzymatic activity if the β-galactosidase (lac Z) gene of the lac operon. β-galactosidase exists as a tetrameric protein. A substrate of β-galactosidase is the colorless X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) or Bluo-Gal. The substrate is converted by the lac Z protein to a blue end product. E. coli strains, like NM522, carry mutations in the lac Z gene which prevent the formation of an active lac Z protein and hence are unable to perform the color reaction. However, the introduction of a plasmid which carries an intact portion of the lac Z gene will allow for the association of the subunits; thus the enzymatic activity of lac Z is restored and the blue end product is formed. This restoration is termed α-complementation. Recombinant plasmids, whose inserts interrupt the N terminal region of the lac Z gene, are detected in the E. coli cells by the failure of producing the blue end product.

To use the "blue/white" screening method with E. coli strain NM522 cells, 10 µl 100 mM IPTG (isopropylthio-β-D-galactoside) and 50 µl 2% Bluo-Gal are spread together onto a 1X YT & Amp agar plate. It is important to spread the mixture uniformly over the plate to avoid "edge" effects; often the edges of the plate do not have sufficient substrate and hence there is no color reaction. The cell/media mixture is then spread evenly onto the plate. Transformed cells and their (in)ability to form the blue end product can be seen after an overnight incubation at 37°C.

Restriction Fragment Length Polymorphism Mapping

Mapping of genes by conventional methods is not always possible. The difficulty may be due to the lack of mutants, the unknown function of the gene, or the presence of several genes of similar function. A method of mapping genes that does not rely on gene function and mutants is restriction fragment length polymorphism (RFLP) mapping. This method of mapping genes or DNA fragments using RFLPs that exist between individuals was first proposed by Botstein et. al., (1980). The thousands of nucleotide differences scattered throughout the genome provide an unprecented number of markers which can be used to map the genome.

A RFLP mapping kit has been developed to map DNA fragments onto the seven linkage groups of N. crassa (Metzenberg et. al., 1984). The fundamental technique employed is the crossing of an inbred parental strain, carrying a large number of morphological and biochemical markers for all seven linkage groups, to an exotic wild type strain. The strains exhibit an extremely high degree of restriction length polymorphisms between them. The kit is composed of the two parental strains and 18 selected progeny. The progeny of the cross were selected by the criteria of maximizing the segregation of the conventional genetic markers and of dispersed, repetitive 5S RNA markers.

A RFLP at a specific location in the genome will tend to co-segregate with, or be linked to, markers which flank that RFLP. Each progeny from the cross will exhibit a RFLP from either parent. Hence, a RFLP which is tightly linked to a conventional genetic marker will tend to segregate among the 18 progeny in a pattern identical to the segregation pattern of the genetic marker. It is important to note that different but linked RFLPs, generated by various restriction enzymes, will also exhibit the same segregation pattern.

The segregation pattern in one linkage group will differ from that of another linkage group. The pattern will also differ within a linkage group depending on the number of map distance which separate the markers. As more N. crassa genes are mapped, the segregation pattern of RFLPs within linkage groups will continue to be refined. A preliminary segregation pattern of the dispersed, repetitive 5S RNA markers is provided in the Fungal Genetics Newsletter (FGNL)(Metzenberg et. al., 1984); the annual FGNL provides an updated version of the segregation patterns for all seven linkage groups.

Southern Blot

The protocol used to blot DNA samples onto nitrocellulose (NC) paper or nylon membrane (Nytran) was a modification of that used by Southern (1975). This modification was described by Kenneth Hardy (Appendix in a Hoeffer Catalog).

Materials

20X SSPE: 174 g sodium chloride, 27.6 g monobasic sodium phosphate. H_2O, 7.4 g EDTA, disodium salt. Adjust to pH 7.4 with sodium hydroxide and to 1 liter with water.

Depurination Solution: 0.1 M HCl; 8.3 ml conc. HCl adjusted to 1 liter with water.

Denaturation Solution: 1.5 M sodium chloride, 0.5 M sodium hydroxide.

Neutralizing Buffer: 2.5 M Ammonium acetate

Transfer Buffer: 1.0 M Ammonium acetate

Methods

Genomic DNA samples from all twenty strains were prepared by the method described in Section 3A. In a preliminary experiment, 4.0 µg DNA from each of the parental strains were digested with the following restriction enzymes: Bam HI, Eco RI, Hind III and Xho I. Similar procedures (described below) were used in the Southern blotting of Hind III digested DNA samples of all twenty strains. To determine whether the digestion is complete, an aliquot (equivalent to 0.4 µg) is electrophoresed in an agarose mini-gel and the digests were compared to undigested DNA samples. The digested samples should be seen as a faint smear, whereas the undigested samples should appear as a single band of about 20-30 kb with a small degree of degradation below the band.

The remains of the digested samples are electrophoresed overnight in a 0.7 % agarose gel (dimensions 15 X 20 cm) at 1.5 volts/cm in 1X TBE with EtBr. The gel is visualized the next day with an UV-transilluminator to determine that the electrophoresis was satisfactory. The gel is then immersed into 500 ml depurination solution and gently rocked for 45-60 min at room temperature. The HCl solution is decanted and the gel is briefly rinsed with H_20. The gel is gently agitated in 500 ml denaturation solution for 30-45 min at room temperature. The denaturation solution is removed and the gel is again briefly rinsed with H_2O. The gel is placed in 500 ml neutralizing solution and gently rocked for 30-60 min at 4°C.

During the manipulations and soaking of the gel in the various solutions, 3 sheets of Whatmann 3MM paper and 1 sheet of Nytran membrane are pre-wetted in H_2O and then soaked in transfer buffer. The Nytran membrane should be soaked in transfer buffer for at least 1 hr at room temperature. The membrane and two of the Whatmann paper sheets are cut to about the same dimensions as the agarose gel. The third sheet of Whatmann paper is cut about 6 inches longer than the length of the gel.

The longest sheet of Whatmann paper is placed onto a plexiglass support with about 3 inches of paper hanging over the edges of the plexiglass. The gel is removed from the neutralizing solution and positioned on the paper with the top surface of the gel now facing the paper and plexiglass. Trapped air bubbles between the gel and the paper is squeezed out from the edges. The Nytran membrane is then fixed onto the gel while avoiding to trap any air bubbles. The two shorter Whatmann sheets are then layered onto the membrane.

The plexiglass, gel, membrane and paper are then placed on the edges of a chamber containing transfer buffer, and the entire structure is kept at 4°C. Pre-cut paper towels are stacked onto the two Whatmann sheets and a heavy weight is placed on the top of the stack. Transfer buffer is added to the reservoir chamber so that the overhanging Whatmann paper is immersed in the buffer. The DNA samples were blotted onto the membrane overnight at 4°C.

The gel was inspected with an UV transilluminator to determine the extent of the transfer. There should not be any indication of fluorescence in the gel. The membrane is soaked in 6X SSPE for 10-15 min, air-dried at room temperature and baked for 2 hr at 80°C under vacuum sandwiched between two sheets of Whatmann paper. The membrane is stored at room temperature until hybridization.

Nick Translation

The protocol used to nick translate DNA is that described in the procedure which accompanied the Nick Translation System from Bethesda Research Laboratories (BRL). The procedure is a modification of that originally developed by Rigby et. al. (1977).

Materials

Nick Translation System: Bethesda Research Laboratories Cat. No. 8160SB

^{32}P-dCTP: NEN. (Sp. Act. 800 Ci/mmol)

Methods

About 1.0-2.0 µg of cosmid DNA was mixed with 0.8 µl A2 solution (contains 0.2 mM dATP, dGTP, and dTTP in reaction buffer), 5.0 µl ^{32}P-dCTP, 1.0 µl Pol C (contains 0.4 U/µl *E. coli* polymerase I and 40 pg/µl DNase I) and H_2O to a final volume of 10 µl. The mixture is gently mixed and then pulsed in a microcentrifuge. The reaction is incubated at 15°C for 90 min. The reaction is stopped by adding 0.1 volume of stop buffer (300 mM EDTA).

Several different methods can be used to separate the unincorporated nucleotides from the labelled DNA. Exclusion chromatography with Sephadex G-25 in a spun column format (Maniatis et. al., 1982) is an effective and rapid procedure. Precipitation with ethanol is another reliable method. After the reaction is stopped, about 10 µg of yeast tRNA is added to the reaction to serve as a carrier. An 1.0 µl aliquot is removed and mixed with a scintillation cocktail (BioFluor). This aliquot serves as a reference for the total amount radioactivity present in the reaction. The labelled DNA is precipitated with

NH_4OAc/EtOH and the pellet is rinsed once with 70 % EtOH. After the pellet is air-dried, the pellet is dissolved in TE. The volume of TE is equal to that when the reference 1.0 µl aliquot was removed. An additional 1.0 µl aliquot is removed, mixed with Biofluor and counted at the same time as the reference aliquot in a scintillation counter. The percent incorporation is calculated from the cpm's of both samples and the total amount of radioactivity in the labelled DNA is calculated. The calculated total amount of radioactivity may be an overestimate because of precipitated unincorporated nucleotides. However, this overestimate has not been observed to affect the hybridization results.

Hybridization

Materials

Formamide: deionized by stirring in mixed bed resin (Amberlite MB-1 or Bio-Rad AG 501-X8). The formamide is collecting as the filtrate with a Buchner funnel under vacuum and stored at -20°C.

1.0 M Tris.Cl, pH 7.6: 12.1 g Tris base dissolved in water. Adjusted to pH 7.6 with HCl and brought to a final volume of 100 ml.

5 mg/ml herring sperm DNA: 50 mg herring sperm DNA dissolved in 10 ml TE. The DNA is sheared by passing the solution several times through a 18G needle fitted to a syringe. The solution is dispensed into aliquots and boiled for 10 min. The aliquots are stored at -20°C.

100X Denhardt's solution: 1.0 g Ficoll-400; 1.0 g polyvinylpyrrolidone; 1.0 g bovine serum albumin (Pentax Type V) dissolved and adjusted to a final volume of 50 ml with H_2O. Filter the solution with a 0.45 µm filter and store as aliquots at -20°C.

20X SSC: 175.3 g sodium chloride; 88.2 g sodium citrate dissolved in H_2O. The solution is adjusted to pH 7.0 with 1.0 M NaOH and brought up to a final volume of one liter.

Dextran Sulfate

Non-fat Dry Milk: Your local super-market

Methods

The hybridization solution is prepared by mixing 10.0 ml formamide, 5.0 ml 20X SSC, 0.5 ml 1.0 M Tris.Cl (pH 7.6), 0.25 ml 100X Denhardt's solution, 0.25 ml 10% SDS, 2.5 g dextran sulfate, 0.07 g non-fat dry milk (NFDM), and 1.26 ml H_2O. The solution is stirred until the dextran sulfate is completely dissolved. Herring sperm DNA is boiled for 5 min. and quick cooled in an ice bath while the hybridization solution is being mixed. 100 µl 5 mg/ml herring sperm DNA is added to the hybridization solution. Pieces of nitrocellulose paper or Nytran membrane is mixed into the hybridization solution. 1.76 ml of the hybridization solution is removed and saved in a test tube. This aliquot is replaced by adding 2.4 ml H_2O to the main hybridization solution.

The membrane with the blotted DNA samples is pre-soaked in 4X SSC and then placed into a Seal-a-Meal bag. Excess moisture is squeezed out of the bag and the main hybridization solution is added. After careful removal of air bubbles, the bag is sealed with a Seal-a-Meal device. The membrane is prehybridized at 42°C for 3-16 hr in a water bath.

1 X 10^7 cpm of labelled DNA is boiled for 2-3 min. and quick cooled in an ice bath. The boiled probe is mixed with the 1.76 ml hybridization solution which was saved. The bag is opened at one corner and the probe, in hybridization solution, is added into the bag. The radioactive probe is worked into the main hybridization solution. After the air bubbles is squeezed out of the bag, the corner of the bag is sealed. The bag is sealed twice with the Seal-a-Meal device to prevent any leakage. The probe is hybridized to the membrane at 42°C overnight.

The bag is opened at one corner and the radioactive hybridization solution is collected. The bag is then opened on three sides and the membrane is removed. The membrane is washed with 250 ml 2X SSC, 0.1% SDS, and 0.25% NFDM at room temperature for 15-30 min in a tray. This wash is performed twice. The membrane is then washed with 0.1X SSC, 0.1% SDS, and 0.25% NFDM at 65°C for 30 min. This wash at high stringency conditions is performed twice.

After the wash, the membrane is placed on a sheet of Whatmann paper and air dried. The membrane is then taped onto a sheet of paper and covered with Saran wrap. The membrane is then placed between two intensifying screens in a X-ray film holder. The X-ray film holder is clamped between two pieces of cardboard and kept at -70°C. The length of exposure time will vary from 4-72 hr or even longer.

The X-ray film is developed in a dark room. The film is immersed in developer solution for 5 min. The film is then rinsed with water or Stop solution for 2 min. Final development of the film occurs with the immersion of the film in fixative solution for 5 min. The film is then rinsed with water for 5-10 min and air dried.

The membrane can be re-probed with another radioactive probe after it has been stripped. A solution of 0.1X SSC, and 0.25% NFDM is boiled and then poured onto the membrane. The solution is allowed to cool to room temperature. This process is repeated 3-4 times until the radioactivity has reached background level as judged with a Geiger counter.

LMT-Agarose, DNA purification, and Ligation

Many techniques in molecular biology require the isolation of a DNA fragment of interest. These techniques include oligo-labelling, end-labelling, restriction enzyme digestion and cloning of the DNA fragment. Three current methods of recovering DNA fragments from agarose gels are: (a) use of DEAE membrane paper e.g. NA 45 from Schleicher & Schuell, (b) use of glass beads e.g. Geneclean from Bio 101, and (c) use of low melting agarose. The first two methods of retrieval are primarily based on ionic character of nucleic acids. The use of low melting temperature agarose gels rely on the quality of the agarose.

The DNA sample is electrophoresed in the standard manner as described in Section 2B(iii). Unlike the commonly used grades of agarose, low melting agarose (Sea-Plaque) is extremely fragile and can be easily be melted if the applied voltage is too high. The electrophoretic buffers can be 1X TBE or 1X TAE. After the electrophoresis is finished, the gel is placed on a UV transparent plexiglass and the EtBr stained DNA fragment is visualized with a long wavelength UV transilluminator to avoid nicking of the fragment. The agarose plug, containing the DNA fragment, is excised from the gel and placed into a microfuge tube.

The agarose plug is heated at 65-70°C (5 to 10 min) and its volume is determined while the agarose is liquefied. After the agarose is melted, it remains in a liquid state provided that its temperature does not go below 37°C. Nucleic acid is extracted from the agarose by adding an equal volume of TE to the molten gel and extracting the solution once with TE buffered phenol, once with phenol: chloroform and once with chloroform. The nucleic acid is precipitated by NH_4OAc/EtOH and the pellet is rinsed once with 70% EtOH.

DNA fragments are reliably recovered using the above procedure. However, the procedure is tedious and the recovery yield is low. Usually, 10-25% of the DNA fragment is recovered. The low recovery yield can be avoided by performing subsequent enzymatic manipulations of the DNA fragment in the presence of the low melting temperature agarose (Struhl, 1985). Ligation reactions are routinely performed in this manner to generate desired clones. This technique has also been applied successfully to oligo-labelling reactions of DNA fragments.

Typical ligation reactions and transformations with LMT-Agarose are as follows. An aliquot, equivalent to 100-300 ng DNA, of the insert DNA in LMT-Agarose is liquefied as before. 10-30 ng of vector DNA and a predetermined volume of water is mixed into the agarose solution and the solution is further incubated at 37°C (15 min). 10 X ligation reaction buffer and 1-3 μl 1 U/μl T4 DNA ligase is then mixed into the solution. The final volume should be kept to a minimum and should not exceed 50 μl. The reaction mixture is incubated at 15°C or at room temperature depending on the type of ends are being ligated. Ligations of sticky or protruding ends are performed at 15°C. Blunt end ligations are carried at room temperature. At 15°C, the agarose will solidify but it does not appear to affect the reaction (Struhl, 1985). The length of incubation period is usually overnight but ligation reactions of 4 hrs have been successful.

At the end of the incubation period, the volume of the reaction mixture is adjusted to 100 μl with the Tris.Cl/$CaCl_2$ solution (see Section 4C) and is heated at 65°C as before. The condensation is brought down with a brief pulse in the centrifuge and the tube is immersed into an ice-bath. As soon as the temperature of the tube appears to have sufficiently cooled without the contents having solidified, the contents are gently mixed with 200 μl of competant E. coli cells. Further manipulations of transformation, heat-shock, incubation in media and plating is identical to that described in Section 4C.

Restriction Enzyme Mapping

All of the restriction enzyme mapping used in this project involved single and double digestions. For most purposes, 1-2 μg DNA is sufficient for restriction enzyme mapping with two enzymes and takes into account of the preliminary analyses. Three to five units of enzyme are used per 1.0 μg DNA to ensure that most of the DNA is digested. These amounts of enzyme will vary depending on the particular enzyme being used. Restriction enzyme digests were performed in the supplier's suggested buffer. Most digestions were performed at 37°C for one hour but these parameters are not cast in stone.

In double digestions involving enzymes with different preferred buffers, the DNA is first digested with the enzyme which functions best in the lower ionic strength buffer. An aliquot of the reaction is removed and electrophoresed in a minigel to determine the completeness of the digestion. If the digestion is judged to be complete, another aliquot is saved and the remains are digested with the second restriction enzyme. The buffer conditions of the second digestion should be modified as needed. After the second digestion is finished, another aliquot is electrophoresed in a minigel to determine the completeness of the digestion.

Knowing that the digestions are complete, electrophoresis of the DNA fragments can now be performed in a larger gel at an applied voltage of 2-3 volts/cm. The increased length of the gel, versus a minigel, provides for better resolution of the fragments. After electrophoresis and staining of the gel with EtBr (see Section 2B(iii)), the gel is photographed while being illuminated with UV light. The R_f values of the bands are estimated; this is done for the DNA being investigated and for molecular weight DNA markers which are also electrophoresed in the gel. A graph, of R_f values versus the

log_{10} base pairs, is plotted for the molecular weight DNA markers. The molecular weight or the number of base pairs of the DNA fragments being investigated is extrapolated from the graph.

The sizes of the fragments from the single digest are compared to those from the double digestion. If the number of sites for the two enzymes are few, the assignment of relative positions of restriction sites is a worthy exercise in logic. However, this task can easily be complicated with restriction enzymes which "cut" frequently.

Subcloning

Materials

The sources of enzymes include: Boehringer Mannheim Biochemicals (restriction enzymes), New England Biolabs (restriction enzymes, PolK, and T4 polymerase), and Bethesda Research Laboratories (T4 DNA ligase and BAP). Vector pTZ18R and nucleotides (dNTPs) were purchased from Pharmacia Labs. LMT-agarose was obtained from FMC.

Generation of pBN3

The rationale of creating pBN3 is presented in Fig 3. 5. 11.3 µg of Bgl II digested 5:4H DNA was electrophoresed in a preparative 1% LMT-agarose gel. 15.0 µl (70 ng of DNA) of the liquefied agarose gel, containing the 15.0 kbp fragment was mixed with 1.5 µl H_2O, 2.0 µl 10X T4 ligase buffer at 37°C and maintained at this temperature for 15 min. 1.5 µl 1.4 U/µl T4 DNA ligase was then added. The contents were briefly mixed by flicking the tube and brought down with a quick pulse in a microcentrifuge. The reaction was incubated at 15°C overnight.

At the end of the incubation, 150 µl Tris.Cl/$CaCl_2$ was added to the ligation reaction and was heated to 65-70°C for 5 min to liquefy the agarose. The condensation was brought down with a rapid pulse, and the tube was immersed into an ice-bath. The contents, after sufficient cooling, were dispensed into a microfuge tube containing 200 µl competant E. coli LE392 cells. The transformation and plating of the cell/media mix was the same as described in Section 4C.

Generation of pX18R2 and pCVN2.9

Vector preparation

1.1 µg pTZ18R DNA was digested with 5 U Bam HI at 37°C for 1 hr. An aliquot was electrophoresed to determine whether the digestion was complete. The Bam HI digested DNA was digested with 5 U Eco RI. The DNA was precipitated with NH_4OAc/EtOH and the pellet was dissolved in TE, pH 8.0 to a final concentration of 200 ng/µl.

Insert preparation

24 µg pBN3 DNA was digested to completion with 24 U Bgl II. The Bgl II digested DNA was further digested with 25 U Eco RI and precipitated with NH_4OAC/EtOH. The DNA pellet was dissolved with 10 µl 50 mM Tris.Cl, pH 8.0 and mixed with 1.0 µl 200 U/µl bacterial alkaline phosphatase (BAP). The solution was incubated at 37°C for 30 min. The BAP enzymatic activity was stopped by heating at 65°C (30 min.), by extracting once with phenol/chloroform, and once with a chloroform extraction. The DNA was

precipitated with NH_4OAc/EtOH and the pellet was dissolved to a final concentration of 250 ng DNA/μl with 10 μl TE.

Ligation

100 ng of dephosphorylated Bgl II/Eco RI pBN3 DNA was mixed with 20 ng Bam HI/Eco RI pTZ18R DNA, 0.7 U T4 DNA ligase and 10X T4 ligase buffer to a final volume of 5 μl. The reaction was incubated at 15°C overnight. The transformation procedure employed competant E. coli NM522 cells and the "blue/white" screening method to detect recombinant DNA molecules (see Section 4C).

The usual procedure of dephosphorylating the vector was not deemed to be necessary because of the incompatibility of the Bam HI and Eco RI generated ends. However, dephosphorylation of Bgl II/Eco RI digested pBN3 DNA was necessary to prevent the recircularization of an Eco RI fragment from pSV50 which contains the origin of replication (ori) and the Ampicillin (Amp) marker.

Generation of pN807 and pN846

Vector preparation

6 μg pTZ18R DNA was digested to completion with 17 U Bam HI. The final concentration was adjusted to 60 ng/μl.

Insert preparation

12.0 μg pBN3 DNA was double digested with 36 U Bgl II and 34 U Bam HI. The sample was electrophoresed in a 1% LMT-agarose gel and the 2.9 kbp fragment was excised out. The agarose plug was immersed in H_2O at room temperature (60 min.) to remove excess EtBr and 1X TBE buffer.

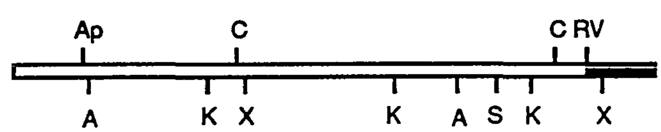
Ligation

The agarose plug, containing 600 ng of the 2.9 kbp fragment, was liquefied by heating and mixed with 120 ng of Bam HI digested pTZ18R vector. 5X T4 ligase buffer and 2 U T4 DNA ligase was mixed into the solution at 37°C for a final volume of 90 µl. The reaction was incubated at room temperature (overnight) instead of the usual 15°C because the great amount of agarose was thought to would have reduced the kinetic energy of the molecules; thus reducing the probability of interstrand ligation.

After the overnight incubation, a half volume aliquot of the reaction mix was liquefied and mixed with an equal volume of Tris.Cl/$CaCl_2$ solution. After sufficient cooling in an ice-bath, the solution was used to transform 200 µl competant E. coli NM522 cells. The transformed cells were spread on X-gal indicator plates to detect recombinant DNA molecules.

Generation of pNE/E4 846

The rationale used to generate pNE/E4 846 is presented in the diagram below. The 0.3 kbp fragment is removed from pN846 with a Eco RV/ Eco RI double digest; this fragment is darkened in the diagram below. The remaining DNA fragment is then readily manipulated and ligated to create a plasmid.



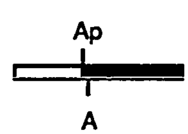
5.0 µg pN846 was double digested with 18 U Eco RV and 15 U Eco RI. The completion of the digest was determined by electrophoresis of an aliquot. The DNA was precipitated with NH_4OAC/EtOH and dissolved to a final concentration of 500 ng/µl.

To fill in the 5' ends generated by the Eco RI digestion, a half volume aliquot was mixed with 2.0 µl 1.0 mM dNTPs, 1.0 µl 10X Nick Translation buffer and 2.5 U Klenow fragment of Pol I (Pol Ik) in a final volume of 10 µl. The reaction was incubated at room temperature for 30 min. The reaction was heat-inactivated at 65°C (5 min).

The reaction sample was electrophoresed in a 1% LMT-agarose gel and the desired band, about 5.6 kbp, was excised out from the gel. After immersion in excess H_20, the agarose plug was melted. About 0.25 volume of the liquefied agarose was mixed 5 U T4 DNA ligase, adjusted to 1X T4 DNA ligase buffer for a final volume of 50 µl and was incubated overnight at 15°C. 25 µl of the ligation mixture was used to transform 200 µl competant E. coli NM522 cells. In this and following constructs, it was not necessary to spread the transformed cells on X-gal indicator plates since the "parental" clones already have inserts which interrupt the lacZ gene.

Generation of pNA/E12 3-1

The rationale used to generate pNA/E12 3-1 is presented in the diagram below. The 0.5 kbp fragment is removed from pNxB 3-1 with an Apa I/Eco RI double digest; the equivalent fragment is darkened in the diagram below. The remaining DNA fragment is then readily manipulated and ligated to create a plasmid.



5.0 μg of pNxB 3-1 DNA was digested to completion with 60 U Apa I. The excess amount of Apa I was required to ensure the completeness of the digestion. The Apa I digested DNA was digested with 25 U Eco RI. Aliquots of both digestions were electrophoresed to determine the extent of the digestion. The DNA was precipitated with NH_4OAc/EtOH and the DNA pellet was rinsed twice with 70% EtOH. The pellet was dissolved with 10 μl TE.

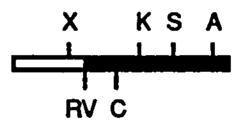
The 5' overhang ends created by the Eco RI digest were made blunt by the polymerase activity of E. coli polymerase I, Klenow fragment (Pol Ik). 5.0 μl of Apa I/Eco RI digested DNA was mixed with 2.0 μl 1.0 mM dNTPs, 1.0 μl 10X Nick Translation buffer, and 5 U Pol Ik in a final volume of 10 μl. The reaction was incubated at room temperature for 30 min. The reaction was heat inactivated by heating at 65°C for 5 min.

The 3' overhang ends created by the Apa I digestion were made blunt by the exonuclease activity of T4 DNA polymerase. After the Pol Ik reaction was heat inactivated, the contents were mixed with 2.0 μl 1.0 mM dNTPs, 2.0 μl 10X T4 DNA polymerase reaction buffer, 2.0 μl 100 mM DTT, 1.0 μl 10 mg/ml BSA and 5 U T4 DNA polymerase in a final of 20 μl. The reaction was incubated at 37°C for 5 min. and then heat inactivated as before.

The reaction sample was electrophoresed in a 1% LMT-agarose gel and the desired band, about 3.2 kbp, was excised out from the gel. About 0.5 vol of the liquefied agarose plug was mixed with 5.0 U T4 DNA ligase and adjusted to 1X T4 DNA ligase buffer for a final volume of 50 μl. The reaction was incubated at room temperature overnight. About 0.5 volume of the ligation reaction was used to transform 200 μl competant E. coli NM522 cells.

Generation of pNE/E2 10-2

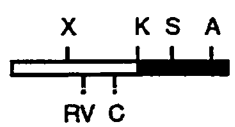
The rationale used to generate pNE/E2 10-2 is presented in the diagram below. The 0.7 kbp fragment is removed from pNxB/B 10-2 with an Eco RV/ Eco RI double digest; the equivalent fragment is darkened in the diagram below. The remaining DNA fragment is then readily manipulated and ligated to create a plasmid.



About 5.0 µg of pNxB/B 10-2 DNA was digested to completion with 18 U Eco RV and 15 U Eco RI. The protocol used to generate pNE/E2 10-2 is similar to that employed to create pNE/E4 846. The 5' overhang ends created by the Eco RI restriction digest were made blunt by the polymerase activity of Pol Ik.

Generation of pNE/K6 10-2

The rationale used to generate pNE/K6 10-2 is presented in the diagram below. The 0.4 kbp fragment is removed from pNxB/B 10-2 with a Kpn I/ Eco RI double digest; the equivalent fragment is darkened in the diagram below. The remaining DNA fragment is then readily manipulated and ligated to form a plasmid.



About 5.0 µg of pNxB/B 10-2 DNA was digested to completion with 15 U Kpn I and 15 U Eco RI. The protocol used to generate pNE/K6 10-2 is similar to that employed to create pNA/E12 3-1. The Eco RI generated 5' overhang ends were made blunt with Pol Ik. The 3' overhang ends created by the Kpn I restriction digest were made blunt by exonuclease activity of T4 DNA polymerase.

Nested Deletion Subcloning

This protocol of generating nested deletions of a DNA fragment is one of a few methods that are presently being used by investigators. It was originally developed by S. Henikoff (1984). A modified version of the protocol was provided by J. Gabe (Codon) and was used in this study. It entails the sequential exonuclease activity of Exonuclease III (Exo III) and single-stranded nuclease activity of S1 nuclease. Effective use of this method requires restriction site information of the target DNA fragment.

Materials

Exonuclease III: Pharmacia Labs Cat. No. 27-0874-01

S1 Nuclease: Pharmacia Labs Cat. No. 27-0920-01

Exo III Reaction Buffer: 66.0 mM Tris.HCl, pH 8.0; 0.66 mM $MgCl_2$

Exo III Stop Buffer: 0.3 M NaOAc in TE, pH 8.0 mixed with an equal volume of phenol/chloroform

S1 Reaction Buffer: 0.25 M NaCl; 30 mM KOAc, pH 4.6; 1.0 mM $ZnSO_4$; 5% Glycerol

Methods

In both cases of pN807 and pN846, 40 μg of DNA was digested to completion with Sac I . The extent of the reaction was monitored by electrophoresing a small aliquot in an agarose mini-gel. Sac I generates 3' overhangs which are not substrates of the Exo III enzyme. The samples were heated at 65°C (5 min) to inactivate any remaining enzyme activity. 10 μg of the linearized DNA fragment was aliquotted out to serve as a control for the Exo III reaction. The other 30 μg of DNA was digested with Sma I which generates blunt ends. Blunt ends, like 5' protruding ends, are suitable substrates of Exo III. Excess Sma I was used to ensure complete digestion. The order of the restriction enzyme digests was performed in this manner so as to detect the complete digestion of the first enzyme which would generate the type of ends not susceptible to Exo III digestion.

The volume of both the control and experimental samples was adjusted to 200 μl with TE; phenol/chloroform extracted once, chloroform extracted once and precipitated with NH_4OAc/EtOH overnight. The DNA is pelleted and rinsed once with 70% EtOH. The experimental and control DNA samples were dissolved with 300 and 100 μl Exo III reaction buffer. A 25 μl aliquot is taken from both DNA samples to serve as a reference for untreated Exo III DNA . The Sma I blunt ends of the experimental and control samples were digested with Exo III by adding 2 and 1 μl ~73 U/ μl Exo III to each sample, respectively. The reaction tubes are incubated at 37°C. At suitable time points, 25 μl aliquots are removed from the reaction tubes and mixed with 375 μl Exo III stop buffer (175 μl 0.3 M NaOAc and 200 μl phenol /chloroform). An aliquot of the control DNA sample may simply be taken after the last time point. The samples are extracted once with phenol/chloroform and then extracted once with chloroform. A 2.0 μl aliquot from all the samples is electrophoresed. The DNA fragment should not exhibit any change in mobility at this time, and there should not be any significant amount of degradation. Two volumes of 95% EtOH is added to the main DNA solution and precipitated at -20°C overnight.

The DNA is pelleted and rinsed once with 70% EtOH. All control and experimental samples are dissolved with 50 µl 0.11 U/ µl S1 nuclease in S1 reaction buffer. The samples are incubated at 37°C for 30 min. A 5-10 µl aliquot of all samples is electrophoresed in a mini-gel. The DNA fragment should now exhibit a decrease in molecular weight. There might be some degradation of the sample at this time but the amount of the fragment should represent at least 75% of the nucleic acid sample. The volume of the main DNA solution is adjusted to 200 µl with 0.3 M NaOAc, extracted once with phenol/chloroform and chloroform extracted once. Two volumes of 95% EtOH are added to the main DNA solution and precipitated at -20°C overnight.

The DNA is pelleted and rinse once with 70% EtOH. The samples are dissolved with 35 µl water and stored at -20°C until used for ligation and transformation. The ends of the DNA molecules can be filled in with the Klenow fragment of Pol I or used directly for ligation. Usually, 5-15 µl aliquots of the samples were used directly for the ligation reaction. The ligation reaction and the transformation of E. coli with the DNA samples was similar to that previously described (Section 4C).

Transformed colonies were picked and used to inoculate 2.5 ml cultures. Plasmids were prepared by the rapid alkaline lysis method. The DNA samples were screened for inserts by double digests with Hind III and Eco RI. Clones whose plasmids contained inserts were streaked onto 1X YT & Amp plates so as to get single colonies. Single colonies were then picked and used to inoculate 40 ml cultures. These cultures were used to establish glycerol stocks and to prepare more plasmid DNA for further analysis. The plasmid DNA samples were usually digested with Kpn I, which cuts three times within the insert, to confirm it's authenticity.

DNA Sequencing and Sequence Analysis

DNA sequencing was performed by the dideoxynucleotide chain termination method (Sanger et. al., 1977) using double-stranded templates and modified T7 DNA polymerase, Sequenase. The major innovations are DNA sequencing is the use of ^{35}S - isotope, primer extension and Sequenase.

The radioactivity emitted by the ^{35}S - isotope is less harmful than that emitted by ^{32}P. This advantage is greatly magnified by the significant amount of handling of radioactive material inherent in DNA sequencing. During the labelling reaction, radioactive dNTP and "cold" dNTPS are incorporated into the newly synthesized strands. The effect is the extension of the sequencing primer thus increasing the number of nucleotides which can be sequenced. Chemically modified T7 DNA polymerase, Sequenase, is one of several polymerases used in DNA sequencing. Sequenase, unlike the Klenow fragment (Pollk) of E. coli polymerase I, does not require the use of cesium chloride banded DNA templates.

DNA Sequencing

Preparation of the template

Most of the double-stranded DNA templates were prepared by the rapid boiling method with the modification of LiCl precipitating (Pelham, 1985) [Section 3C(ii)].

Template Denaturation

Materials

M13 Reverse Sequencing Primer: Pharmacia Cat. No. 27-1532-01

5X Denaturation Buffer: 1.0 M NaOH; 1.0 mM EDTA, pH 8.0. This solution is prepared fresh before use.

2.0 M NH_4OAc, pH 4.5: 15.4 g Ammonium acetate dissolved in 20 ml water. The solution is adjusted to pH 4.5 with glacial acetic acid. The volume is brought up to 100 ml with water.

Methods

3.0 µg of template is mixed with 4 µl sequencing primer. The volume is adjusted to 16 µl with water. 4 µl of 5X denaturation buffer is added to the side of the microtube and briefly spun to mix the contents. The sample is heated at 85°C for 5 min. The solution is quick-cooled in an ice-bath and 3.0 µl of 2.0 M NH_4OAc is added. The contents were briefly spun and then mixed with 50 µl 95%EtOH. The template is precipitated at -20°C for at least 1 hr. The template is pelleted by centrifuging at 12,000 rpm (10 min) and rinsed twice with 70% EtOH. The tube is air-dried in an inverted position (30-60 min).

Sequencing

There are three steps in the sequencing procedure; annealing of the primer to the template, a labelling reaction and a termination reaction. During the labelling reaction, the ^{35}S-dATP and "cold" dNTPs are incorporated into the newly synthesized strands which have been primed by the annealed sequencing primer. In the termination reaction, the random incorporation of the dideoxynucleotide into the growing strands stops further elongation of the strands.

Materials

Sequenase kit: United States Biochemical Corp. Cat. No. 70700

^{35}S-dATP: Amersham Cat No. S.J. 1304 (>1000 Ci/mmol) or S.J. 304 (>600 Ci/mmol)

Methods

Step 1 - Annealing of Primer

The air-dried pellet is dissolved in 10 µl 1X Sequenase buffer. The primer is annealed to the template at 37°C for 15 min. To ensure dissolving of the template, the solution is mixed by drawing the solution a couple of times into a micropipet tip.

Step 2 - Labelling Reaction

Sequencing reactions are performed with four individual templates simultaneously. The enzyme/label mix for four reactions is prepared by mixing 2 µl water, 4 µl dGTP labelling mix, 4 µl ^{35}S-dATP, 7 µl TE (pH 7.4), 4 µl 0.1 M DTT, and 1 µl Sequenase. The enzyme is added last to the pre-chilled tube. After the primer has annealed to the template, 5.5 µl of the enzyme/label mix is added to each template and incubated at room temperature (5 min).

Step 3 - Termination Reaction

During the annealing reaction, 2.5 µl of each termination mix is aliquotted into marked tubes and prewarmed at 37°C for 5 min. After the labelling reaction is completed, 3.5 µl of a single enzyme/template mix is added to each tube (G, A, T, and C) and incubated at 37°C (15 min). The reaction is stopped by adding 4 µl Stop buffer to each tube. The reaction tubes are stored at -20°C. Prior to electrophoresis, the sequencing reaction tubes are heated to 85°C for 2 min.

Polyacrylamide Gel Electrophoresis

Materials

40% Polyacrylamide: Acrylamide: Bisacrylamide (19:1). This solution is deionized by mixing 0.5 g mixed bed resin (Amberlite MB-1 or Bio-Rad AG 501-X8) per 100 ml solution. Resin and residue are removed by filtration. This solution is stored at 4°C in a dark bottle.

Urea solution: 28.8 g urea dissolved in water for a final volume of 49.9 ml. This solution is deionized by mixing with 0.2-0.5 g mixed bed resin (Amberlite MB-1 or Bio-Rad AG 501-X8) for 30 min. Resin and residue are removed by filtration.

10X TBE: 54 g Tris base; 27.5 g Boric acid; 20 ml 0.5 M EDTA, pH 8.0 dissolved and adjusted to 500 ml.

Sequencing apparatus: Bio-Rad Sequi-Gen Cell (Cat. No. 165-3601)
PAM: Your local super-market

3 MM Whatmann paper

Gel dryer: Bio-Rad Model No. 483

10% APS: Ammonium persulfate dissolved in water and de-gassed under vacuum (10 min). This solution is prepared fresh.

TEMED: Boehringer Mannheim Cat. No. 100 139

X-ray film: Fuji RX X-ray film (35 x 43 cm)

Methods

Preparation of Sequencing Apparatus

The glass plate and the Integral Plate/Chamber (IPC) are washed with a mild detergent (Sparkleen) and rinsed with distilled water. Prior to pouring the polyacrylamide solution, both the glass plate and the IPC are rinsed with 70% EtOH and wiped dry with Kimwipes. PAM is sprayed onto the glass plate of the IPC. Excess PAM is wiped off the plate until a thin film is present. The plate is then air-dried for 30 min. The 0.4 mm spacers are placed on the IPC plate. The outer glass plate is mounted onto the IPC unit and the entire assembly is clamped together.

Preparation of the Polyacrylamide Gel

10.5 ml 40% polyacrylamide, 6.0 ml 10X TBE and 43.5 ml urea solution is mixed in an Erlenmeyer flask. The solution is de-gassed under vacuum for 15 min. Aliquots of 2.0 and 10.0 ml are used to form sealing gels which prevent the leakage of the main polyacrylamide solution. 250 μl 10% APS and 10 μl TEMED are mixed with the 10 ml aliquot of polyacrylamide solution and poured quickly onto a filter paper strip which is placed on the cushion in the casting tray. The assembled glass plate/IPC unit is mounted into the casting tray and the fastened by the screws. 7.2 μl 10% APS and 1.4 μl TEMED are mixed with the 2 ml aliquot and poured between the glass plates with a pasteur pipet. The sealing gels are allowed to polymerize for 15 min. 172.8 μl 10% APS and 33.6 μl TEMED is mixed with the main acrylamide solution. The solution is poured between the glass plates with 10.0 or 25.0 ml pipets. To avoid not being able to remove trapped bubbles that fail to move to the top of the solution, it is advised to pour the solution with the apparatus tilted at an angle and removing bubbles that are at the bottom of the gel before pouring the rest of the gel solution. Bubbles that form while pouring are removed by tapping the glass plates with a blunt instrument e.g. a scissors' plastic handle. The apparatus is placed at a 5° angle from horizontal after the solution has reached the top edge of the glass plate. A blank comb or spacer is inserted between the plates to a depth of 1.0 cm. Bulldog clamps are used to insure a tight seal at the top of the gel. Poured gels are usually stored at room temperature and used between 1-24 hr after polymerization.

Electrophoresis

The assembled unit is detached from the casting tray and the blank comb is removed from the top edge of the gel prior to electrophoresis. The upper buffer chamber of the IPC is rinsed with water so as to remove any traces of polyacrylamide solution that might have been poured into it.

The lower buffer chamber is filled with 500-600 ml 1X TBE. The assembled IPC unit is mounted into the lower buffer chamber and secured by the stabilizing bar. The upper buffer chamber is filled with 1X TBE. The top edge of the gel is rinsed with 1x TBE and the electrodes are connected. The gel is pre-electrophoresed for 30-60 min at ~50 watts (~28 mamps at 2200 V).

Before the samples are applied into the gel, the samples are heated at 85°C (2 - 3 min). The top edge of the gel is rinsed with buffer to remove traces of urea that have leached out during the pre-electrophoresis. The sharktooth comb is inserted into the gel and the wells are marked so as not to confuse the loading pattern. The sequence reaction products of each template is loaded twice in the pattern of G A T G C A T C. This loading pattern allows for the direct comparison of bands between lanes thus minimizing the amount of error in reading the autoradiogram. 1.5-2.0 µl of sample is loaded per lane.

The electrophoretic conditions are the same as in the pre-electrophoresis. Usually , electrophoresis is stopped about 30 min after the Bromophenol Blue marker has migrated off the gel. Using these electrophoresis conditions, 200-250 nts are usually determined. To obtain more sequence information, electrophoresis is continued for 60-90 min after the Xylene cyanol marker has migrated off the gel.

Fixing, Gel drying and Autoradiography

After electrophoresis the glass plate/IPC unit is removed from the lower buffer chamber. The glass plate is carefully separated from the IPC unit so that the gel remains on the glass plate. The gel, on the glass plate, is immersed in a tray containing fixative. The fixative is methanol: glacial acetic acid (10%:10%). The gel is fixed for 10-15 min. The gel is gently rinsed with water. A sheet of 3MM Whatman paper is placed onto the gel. The gel is carefully transferred to the sheet. Trapped air-bubbles are removed by rolling a wet test-tube across the gel so that the bubbles are released at the edges of the gel.

The gel is placed into the vacuum dryer and a film of Saran wrap is layered onto the gel. The gel is dried for 1 hr at 80°C under vacuum. The dried gel/sheet is placed into an X-ray cassette holder in direct contact with an X-ray film. The length of exposure time varies from 18 to 48 hrs. The autoradiogram is developed by the standard method described in Section 5C.

Sequence Analysis

Materials

Hardware: IBM PC and Macintosh SE

Software: Sequence Analysis Package program by Bob Stevens and IBI/Pustell Sequence Analysis System by (IBM PC). DNA Strider Program by Christian Marck (Macintosh).

Methods

Autoradiograms of sequencing gels were placed on a fluorescent light box or on a white background so as to read the sequence. The sequence was read from a 5' to 3' direction (bottom to top) and was written on a quadrille sheet of paper. The reading of the sequence continued for as long as the bands were distinguishable from each other. In the instances where a random termination had occurred, the most intense band was taken to represent that base. This assignment was usually confirmed by another autoradiogram either of that template or of an overlapping template. Each sequence was read twice and attempts were taken to resolve any discrepancies between the readings.

The sequence of each template were loaded into an IBM file using the Sequence Analysis Package (SAP). The printout of that sequence was compared to the hand-written sequence and any input error was immediately corrected. This procedure of reading the sequence, inputting the data, and verifying the file was repeated for each successful sequencing of a template until at least two data files for that template were obtained. Discrepancies between data files were investigated by comparing the autoradiograms of the template and by matching with other sequence data from overlapping templates.

The sequences of the different templates were linked together by the "merger" operation of SAP to give a final sequence output of the 2.9 kbp fragment. Initial homology searches for N. crassa consensus sequences in the 2.9 kbp of sequence were performed with this program for IBM PC.

The MS-DOS format of the sequence data was transformed to the Mac format using the Apple File Exchange program stored on the hard disk of the Macintosh SE. This reconfiguration of the data was necessary to utilize the DNA Strider Program. The utilities of this program are described in the accompanying manual with the program.

Northern Blot and S1 Nuclease Mapping

RNA samples were electrophoresed under denaturing conditions through agarose gels containing formaldehyde (Maniatis et. al., 1982). The RNA molecules were transferred onto nitrocellulose paper in a manner similar to that of the Southern blotting protocol.

Electrophoresis and Northern Blotting

Materials

10X Gel Buffer: 0.4 M morpholinopropanesulfonic acid (MOPS), pH 7.0; 100 mM NaOAc, pH 5.5; 10 mM EDTA, pH 8.0.

Formaldehyde

Formamide: deionized by stirring in mixed bed resin (Amberlite MB-1 or Bio-Rad AG 501-X8). The formamide is collected as the filtrate with a Buchner funnel under vacuum and stored at -20°C.

Loading Buffer: 50% glycerol; 1 mM EDTA; 0.4% Bromophenol blue ; 0.4% xylene cyanol.

RNA ladder: Bethesda Research Laboratories (BRL) Cat. No. 5620SA

Methods

A 1.5% agarose gel containing formaldehyde is prepared as follows. A suspension of 2% agarose in water is heated until the agarose is completely dissolved; 22.6 ml of the agarose solution is then mixed with 3.0 ml 10X gel buffer and 5.4 ml formaldehyde. The solution is quickly poured onto a gel casting tray and allowed to polymerize. After the gel has hardened, it is placed into an electrophoresis chamber filled with 1X gel buffer and 6% formaldehyde.

The RNA samples, dissolved in sterile water, were removed from the -70°C freezer and thawed in an ice-bath. About 5.0 μg of total cellular RNA was dispensed as 4.5 μl aliquots into microfuge tubes. A cocktail solution was prepared so that each RNA sample contained 2.0 μl 10X gel buffer, 3.5 μl formaldehyde, 10.0 μl formamide, 2.0 μl loading buffer, and 6 μg/ml EtBr. The direct addition of the ethidium bromide to the sample allows for visualization of the RNA sample without the need for destaining of the gel. The cocktail was dispensed and mixed with the RNA samples. The samples are incubated at 55°C (15 min), loaded into the gel, and electrophoresed at 60 - 80 volts.

After the electrophoresis is completed, the RNA sample is visualized with an UV transilluminator. The rRNA (28S, 18S, and 5S) and the tRNA molecules are easily distinguished. The mRNA molecules are seen as a smear that extends from slightly above the 28S RNA band to the 5S RNA band. The gel is then placed into 200 ml 10X SSC at room temperature (30 min) and gently rocked. While the gel is being agitated, sheets of Whatmann paper and nitrocellulose membrane is wetted with 10X SSC. The RNAs are transferred onto the nitrocellulose membrane, using 10X SSC as the transfer buffer, in a similar manner as Southern blotting (Section 5A). After the overnight transfer is complete, the membrane is wetted with 5X SSC (5-10 min), air-dried, and then baked at 80°C (2 hr) under vacuum.

The membrane is pre-wetted with 5X SSC and placed into a Seal-a-Meal bag. The membrane is then prehybridized at 42°C (3 hr) in hybridization solution (Section 5C). Radioactive probes, prepared by nick translation, were added to the hybridization solution and allowed to hybridize to the membrane at overnight at 42°C. The next day, the membrane is removed from the bag and washed briefly twice with 100 ml 2X SSPE at room temperature. The membrane is then washed under the following stringency conditions with 200 ml buffer: twice with 2X SSPE, 0.1% SDS at 65°C (30 min) and twice with 0.1X SSPE, 0.1% SDS at room temperature (30 min). The membrane is air-dried and placed within a X-ray film holder. The exposure and development of the X-ray film are similar to that described in Section 5C.

S1 Nuclease Mapping

The S1 nuclease mapping technique has been used successfully to determine the structure of mRNA transcripts (Maniatis et. al., 1982). This technique relies of the single stranded endonuclease activity of the S1 nuclease enzyme. The strategy involves the annealing of DNA molecules to the mRNA and the digestion of the single stranded regions of the DNA which do not hybridize to the mRNA. The single stranded DNA regions represent the sequences of introns which are not transcribed to form the mature mRNA.

Materials

Hybridization Buffer: 85% formamide; 0.4 M NaCl; 1.0 mM EDTA; 40 mM PIPES (Na^+ salt), pH 6.7

S1 nuclease digestion buffer: 0.25 M NaCl; 30 mM KOAc, pH 4.6; 1.0 mM $ZnSO_4$; 5% glycerol; 20 µg/ml ssDNA; 580 U/ml S1 nuclease (Pharmacia). The S1 nuclease is added to the buffer just prior to use.

Termination Mix: 4.0 M NH_4OAC; 0.1 M EDTA, pH 8.0

10 X Alkaline loading buffer: 0.5 M NaOH; 10 mM EDTA; 20% Ficoll; 0.15% bromocresol green.

Yeast tRNA

Methods

DNA from pBN3, prepared by the rapid boiling method, is double digested with Bgl II and Eco RI to release the 2.9 kbp fragment. The fragment was electrophoresed and purified from an agarose gel plug using Geneclean (silica matrix) from Bio 101. The supplier's instructions of using the matrix were followed. A 10 ng aliquot of the fragment was treated with 2 U T4 DNA ligase in a final volume of 10 µl 1X ligase buffer at room temperature for 60 min. The ligation was performed to repair any nicks in the DNA sample; it was presumed that nicks in the DNA sample would result in a smear when the sample was electrophoresed under alkaline conditions.

A 1.0 ng aliquot of the "resealed" fragment was mixed with 50 µg of total cellular RNA. Yeast tRNA is added to the solution to bring the total amount of RNA to 200 µg. The solution is adjusted to 0.3 M NaOAC, pH 5.5. The nucleic sample is then precipitated with 2.5 volumes of 95% EtOH at -70°C (10-15 min). The pellet is rinsed once with 70% EtOH. The air-dried pellet is completely dissolved with 30 µl hybridization buffer and incubated at 75-80°C (15 min) to denature the double stranded DNA. The solution is pulsed in a centrifuge to bring down any condensation and placed into a water bath which is adjusted to 52-54°C. The temperature of the hybridization solution should not be allowed to drop below the annealing temperature. This hybridization temperature was selected on the basis of the G+C content (52%) of the proposed *mtr* ORF. The DNA is annealed to the RNA for 3 hr.

While the tubes are immersed in the water bath, 300 µl of ice-cold S1 digestion buffer (containing 290 U/ml S1) is quickly added to the hybridization solution. The solution is rapidly mixed, pulsed, and incubated at 37°C (30 min). The reaction is stopped by chilling the tubes in an ice-bath and by the addition of 50 µl termination mix.

The S1 nuclease is removed by extracting once with phenol/chloroform and 50 µg yeast carrier tRNA is added to the recovered aqueous phase. The nucleic acid sample is precipitated with a 2.5 volumes 95% EtOH at -70°C (10 min).

The nucleic acid sample is dissolved with 20 µl TE. One half of the sample is mixed with loading buffer and electrophoresed in a 1.5 % neutral agarose gel of 1X TBE. The other half of the sample is mixed with 10X alkaline loading buffer. The alkaline sample is loaded into an agarose gel formed with a 50 mM NaCl, 1.0 mM EDTA solution and electrophoresed in a 30 mM NaOH, 1.0 mM EDTA electrophoresis buffer. After electrophoresis of the samples in the two types of gels is completed, the samples are transferred onto nitrocellulose membranes using 10X SSC as the transfer buffer. The procedures of baking the membrane, preparing P^{32}-labelled probes by nick-translation, hybridizing of the probes, and stringency washing of the membranes are similar to that described for the Southern blotting protocol (Section 5C).

CHAPTER THREE

RESULTS

Transformation of the *mtr* locus with cosmid DNA

The initial transformation and selection experiments employed DNA from a cosmid library in a sib-selection scheme originally proposed by Akins and Lambowitz (1985). The selection scheme used to detect transformants at the *mtr* locus was adapted from Stadler and Kariya (1969) and was used to establish a complementation map of the *mtr* locus (Larimer and DeBusk, 1977).

The ordered cosmid library of N. crassa genomic DNA was constructed by Vollmer and Yanofsky (1986). This genomic library has been a fruitful source for the cloning of numerous genes including the *mtr* gene. A Sal I fragment, which contains the gene for a benomyl-resistant form of N. crassa ß-tubulin, was subcloned into the cosmid. Benomyl is an anti-fungicidal chemical. This gene was used as a cotransformation dominant marker in the transformation experiments of the *mtr* locus. The co-selection of the Bml marker serves two functions; it serves as a positive control for the transformation technique and provides for a "tighter" screening. Cells which appear to be transformants for one marker may be attributed to a high reversion frequency of that particular marker but this probability is dramatically reduced when two unlinked markers are co-selected.

The double mutant his-2; mtr was used in these initial experiments. Strains having a mutant his-2 allele require L-Histidine to grow. This mutant strain exhibits growth on media supplemented with 0.5 mg/ml L-Histidine but shows no growth on media supplemented with 0.5 mg/ml L-Histidine and 5.0 mg/ml L-Arginine (His, 10X Arg). These phenotypes result from the uptake of L-Histidine through the three major amino acid transport systems. A schematic representation of L-Histidine uptake is shown in Fig. 1.1. Cells of this strain transport L-Histidine via the basic (B) and general (G) transport systems when grown on L-Histidine supplemented medium. The strain fails to transport L-Histidine via the neutral (N) transport system because of the mutant *mtr* allele. The strain does not grow on the His, 10X Arg medium because

Table 3.1

Growth of the his-2; mtr double mutant strain

Medium	Growth
Minimal	-
L-Histidine	+
L-Histidine, 10X L-Arginine	-
L-Histidine, 10X L-Arginine, Benom yl	-

Table 3.1 Growth of the his-2; mtr double mutant strain. The concentration of the amino acid, L-Histidine (His) supplemented media is 0.5 mg/ml and the L-Arginine (Arg) supplemented media is 5.0 mg/ml. The concentration of Benomyl (Bml) is at 0.5 μg/ml.

transport of L-Histidine via the B and G systems is competively blocked by the excess concentration of L-Arginine. A summary of growth of this strain on various media is presented in Table 3.1.

Two types of transformants of this host strain were found using the His, 10X Arg, Bml selection medium (Table 3.1). One class of transformants were his-2 transformants and the other class were *mtr* transformants. Delineation of the two classes of transformants is possible by plating the transformants on minimal medium containing benomyl (Stuart et. al., 1988). Transformants of the his-2 locus will grow on both the His, 10X Arg, Bml medium and the minimal, Bml medium. However, transformants of the *mtr* locus will only grow on the His, 10X Arg, Bml medium and not on the minimal, Bml medium.

Transformation of N. crassa can occur by homologous recombination events in which the allelic DNA of the host strain is replaced by the homologous transforming DNA. However, transformation of N. crassa by homologous recombination occurs at a much lower frequency (about 1 to 10% of the transformants) than that by random integration events (Case, 1982).

Three cosmid clones, 5:3B, 5:4H, and 5:10A, were found to transform for the *mtr* locus (Stuart et. al., 1988). The frequency of transformation of the quadruple mutant X-40 with DNA from 5:3B and 5:4H is shown in Table 3.3. This experiment employed the mtr; pmb; pmg; his-2 quadruple mutant X-40 strain; a summary of growth of this strain on various media types is shown in Table 3.2. Earlier studies on transformation of N. crassa reported low transformation frequencies of 10-100 transformants/μg DNA. Present transformation methodologies have increased these frequencies to 100-1,000 transformants/μg DNA (Case, 1982). The increase in the

Table 3.2

Growth of quadruple mutant, X-40, on various media at different temperatures

Medium	Permissive Temp. (25 °C)	Conditional Temp. (34 °C)
Minimal	+	-
Histidine	+	±
His, 10X Arg	+	±
His, 10X Arg, Bml	-	-

Table 3.2 Growth of quadruple mutant, X-40, on various media. The concentration of the amino acid, L-Histidine (His) supplemented media is 0.5 mg/ml and the L-Arginine supplemented media is 5.0 mg/ml. The (±) represents growth but with less vigor than (+). The concentration of Benomyl (Bml) is at 0.5 µg/ml.

transformation frequencies is a result of improvements in the preparation of conidial spheroplasts. The transformation frequencies exhibited by the DNAs of 5:3B and 5:4H compare well with the reported frequencies of other genes (Case, 1982).

A representative result of agarose gel electrophoresis of restriction enzyme digests from all three clones is shown in Fig. 3.1. Restriction enzyme digests of the three clones indicated that 5:3B and 5:10A contained genomic inserts which were similar and that the restriction digest pattern of 5:4H is very dissimilar to that of 5:3B and 5:10A. Transformation with clone 5:10A was not investigated further because of the similarity in restriction enzyme digest patterns of cosmids 5:3B and 5:10A. All three clones carried inserts of approximately 40 kbp.

Table 3.3

Transformation of the mtr-6(r) allele with DNA from cosmids 5:3B and 5:4H

	Cosmid 5:3B				Cosmid 5:4H	
	Expt.1		Expt. 2		Expt. 2	
Enzyme	No. of colonies N	%	No. of colonies N	%	No. of colonies N	%
None	402	100	968	100	2640	100
Eco RI	26	6	96	10	240	11
Bam HI	37	9	-	-	-	-
Hind III	110	27	209	21	522	20
Pst I	32	8	-	-	-	-
Kpn I	41	10	-	-	-	-
Bgl II	80	20	165	17	1760	67
Xho I	41	10	-	-	-	-
Sau 3A	30	7	-	-	-	-
Hpa I	336	84	706	73	1080	41

@ Carried in X-40 spheroplasts.

* Transformations with restriction enzyme digested cosmid DNA were cotransformed with the benomyl-resistant vector pSV50 (Vollmer and Yanofsky, 1986).

A dash (-) indicates that the particular combination of DNA and enzyme was not tested.

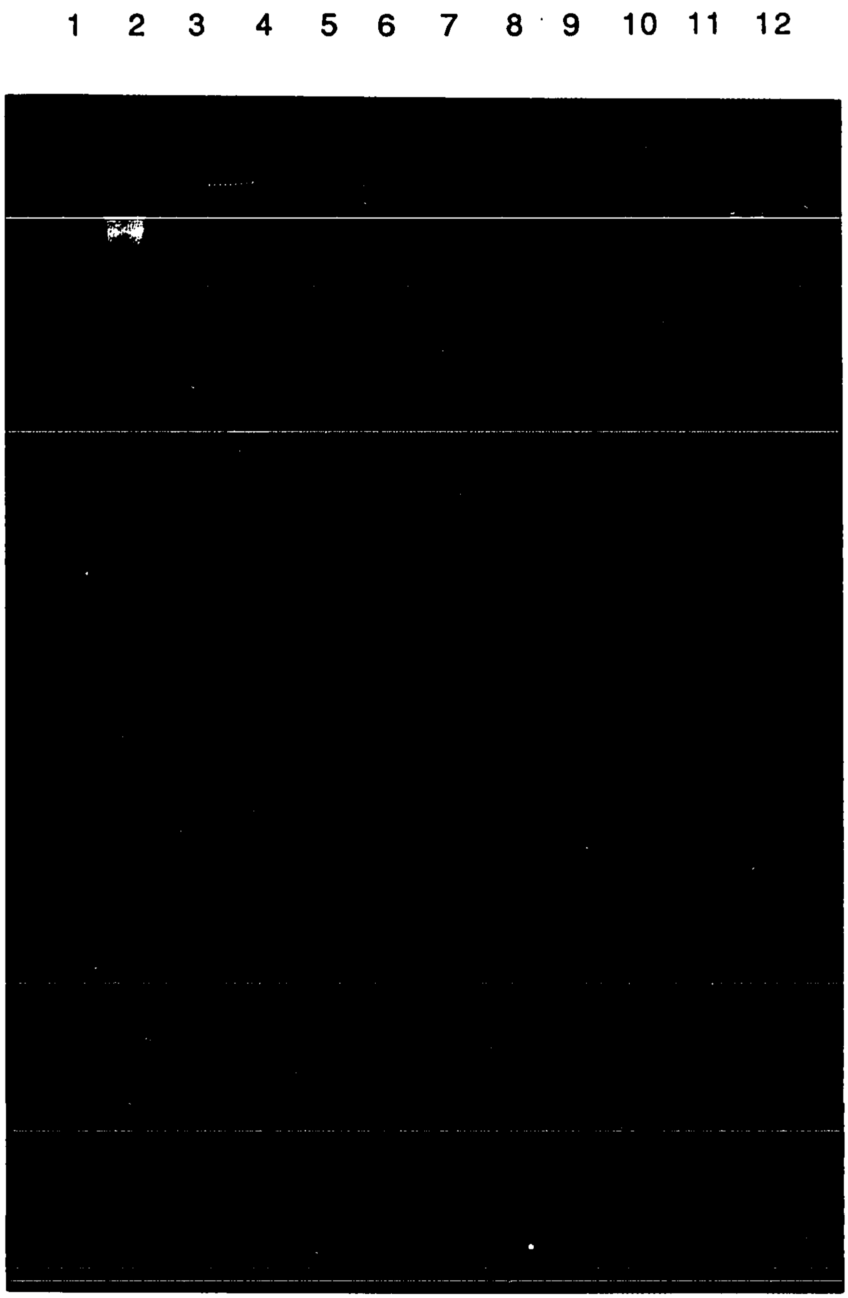


Fig. 3.1 Agarose gel electrophoresis of cosmids 5:3B, 5:4H , and 5:10A. Lanes 1 and 12 - λ Hind III DNA markers; lane 2 - Bgl II 5:4H; lanes 3, 4, and 5 - Eco RI, Eco RI/Xho I, and Xho I digests of 5:3B, respectively; lanes 6, 7, and 8 - Eco RI, Eco RI/Xho I, and Xho I digests of 5:4 H, respectively; lanes 9, 10, and 11 - Eco RI, Eco RI/Xho I, and Xho I digests of 5:10A, respectively.

RFLP Mapping of cosmid clones 5:3B and 5:4H

The restriction fragment length polymorphism (RFLP) kit was developed by Metzenberg et. al., (1984). One version of the kit is composed of two parental strains and 18 selected progeny. One parent carries six morphological and biochemical markers (see Table 2.1) dispersed among the seven linkage groups with a nominal standard laboratory "Oak Ridge" genetic background. The other parent is an exotic wild type "Mauriceville" strain. The progeny were specifically selected for maximum segregation of the conventional genetic markers and of the dispersed, repetitive 5S RNA markers.

RFLPs at a specific site in the genome will tend to segregate with other markers to which the RFLPs are linked. These markers can be conventional genetic markers or RFLP markers which are generated with other restriction enzymes. The segregation of the different genetic markers and of the 5S RNA markers among the eighteen progeny in the RFLP mapping kit is known. The parental strains and progeny, as well as the derived genomic DNAs, are arranged in a consistent manner to provide a distinct segregation pattern of the markers for each linkage group. The random probability of obtaining any one particular segregation pattern is 1 in 2^{18}.

Genomic DNA from each of the twenty strains (2 parental and 18 selected progeny) were prepared and digested with Hind III as described in Materials and Methods. The digested samples were electrophoresed and blotted onto Nytran membrane (Southern, 1975). Hind III digests of DNA from the two parental strains yielded easily distinguishable RFLPs in an initial experiment to compare various restriction enzymes (Fig. 3.2).

Cosmid clones 5:3B and 5:4H, which were able to rescue the *mtr* phenotype by transformation, were labelled with ^{32}P-dCTP by nick-translation (Rigby et. al., 1977).

Fig. 3. 2 Detection of different RFLPs from parental strains 4411 (Oakridge) and 4416 (Mauriceville). P^{32}- labelled 5:3B was used as the probe. Lanes 1, 3, 5, and 7 - genomic DNAs from strain 4411 digested with Bam HI, Eco RI, Hind III, and Xho I, respectively. Lanes 2, 4, 6, 8 - genomic DNAs from strain 4416 digested with the above enzymes and electrophoresed in that sequence. Lanes labelled with (M) represent marker DNA samples.

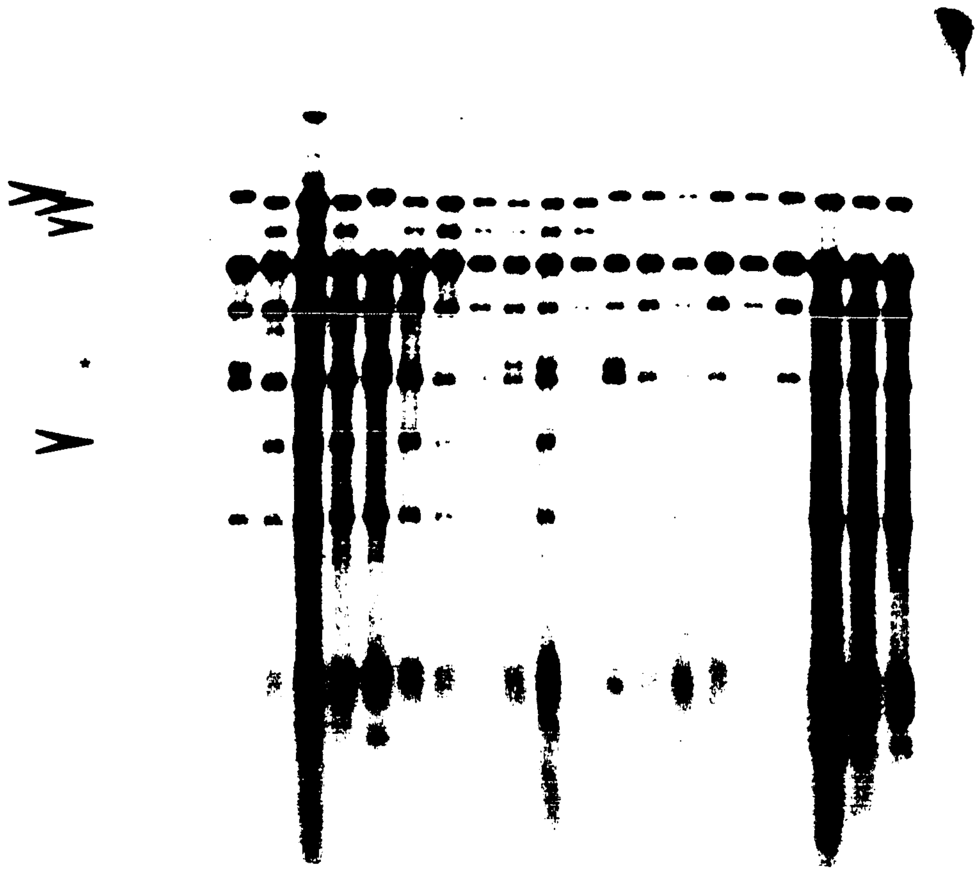


Fig. 3. 3 Restriction fragment length polymorphism (RFLP) mapping of cosmid 5:3B (Stuart et. al., 1988). The long arrows signify the RFLPs which segregate among the twenty strains in the pattern of (O) M M M O (M) M M M M M O O O O O O O O O O which is identical to the previously identified pattern of markers which flank the *mtr* locus on linkage group IVR. The RFLP indicated by the short arrow has a segregation pattern of (O) M M M O (M) M M M M M O O O O O M O O O and is also tightly linked to the *mtr* locus. The RFLP for the benomyl marker is indicated by the asterisk.

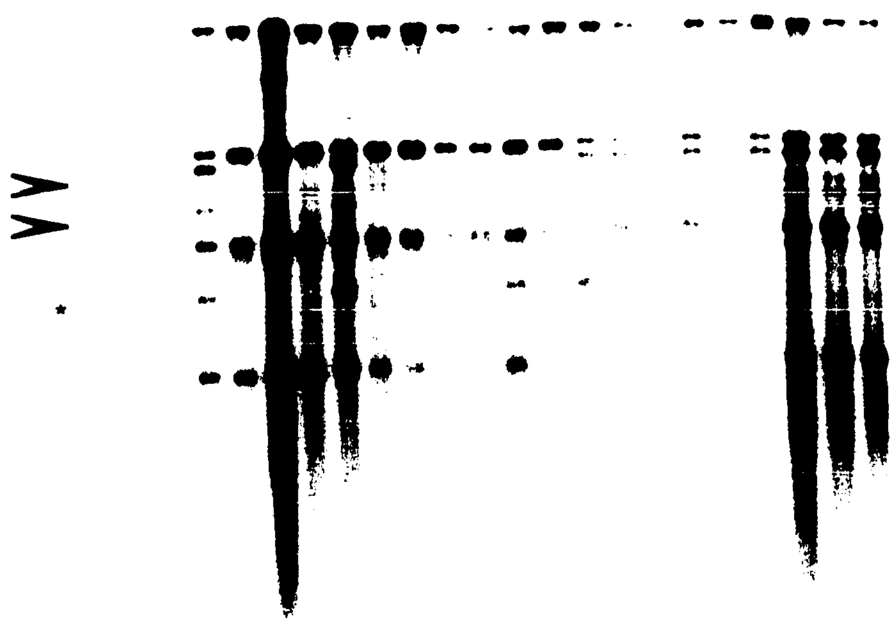


Fig. 3. 4 Restriction fragment length polymorphism (RFLP) mapping of cosmid 5:4H (Stuart et. al., 1988). The arrows signify the RFLPs which segregate among the twenty strains in the pattern of (O) M M M O (M) M M M M M O O O O O O O O O which is identical to the previously identified pattern of markers which flank the *mtr* locus on linkage group IVR. The RFLP for the benomyl marker is indicated by the asterisk.

Both labelled probes were hybridized separately to the filters (4 x 10^5 cpm/ml final concentration). Five restriction fragment length polymorphisms were detected by the 5:3B probe (Fig. 3.3). One RFLP, indicated by the asterisk, segregated in the pattern of (O) M M M O (M) M M O O M O M M M O M M M M. This pattern represents the benomyl resistant gene carried in the cosmid and maps to linkage group VI (Vollmer and Yanofsky, 1986). Three RFLPs, indicated by the long arrows, segregated in the pattern of (O) M M M O (M) M M M M M O O O O O O O O O. This pattern is identical to the previously identified pattern of 5S RNA markers which flank the *mtr* locus on linkage group IVR (Metzenberg et. al., 1984). This finding indicates that the genomic DNA insert in 5:3B exhibits tight linkage to these markers (Stuart et. al., 1988). The segregation pattern of the RFLP shown by the short arrow is (O) M M M O (M) M M M M M O O O O O M O O O and also indicates tight linkage to markers which flank the *mtr* locus.

Three RFLPs were detected by the probe of 5:4H (Fig. 3.4). Like the 5:3B, one of the RFLPs segregated in the pattern, indicated by the asterisk, representative of the benomyl resistant gene. The segregation of the other two RFLPs, shown by the long arrows, is (O) M M M O (M) M M M M M O O O O O O O O O. This segregation pattern is identical to that found with the 5:3B probe and to the identified pattern of markers which flank the *mtr* locus (Stuart et. al., 1988).

As mentioned earlier, restriction enzyme digests patterns of 5:3B and 5:4H are dissimilar. One would expect that clones which rescue an identical genetic marker would carry similar inserts. However, that similarity may be difficult to detect given the size of the insert (~ 40 kbp). Although both clones hybridize to the RFLPs which are linked to the *mtr* locus, the RFLPs detected by the two clones do not appear to be the same when the two autoradiograms were superimposed on each other. The same genomic blot was used for both probes and the autoradiograms were aligned by the RFLP representing the benomyl marker. In addition, it is interesting to note the lack of homology between the genomic inserts of 5:3B and 5:4H at high stringency levels (data not shown).

Subcloning of the *mtr* locus

Together, the results of the transformation experiments and of the RFLP mapping suggested that clones 5:3B and 5:4H contain inserts of DNA which flank the *mtr* locus. The strategy employed to subclone the *mtr* locus makes use of the transformation and selection procedure. Since transformation in N. crassa occurs primarily by heterologous recombination, a high frequency of transformation suggests that the coding region of a gene is intact within a given DNA segment. Conversely, a coding region which is interrupted by the digest with a restriction enzyme would be expected to exhibit a significant decrease in the transformation frequency (Vollmer and Yanofsky, 1986).

DNA of clones 5:3B and 5:4H were digested to completion with a battery of restriction enzymes. The digested DNA samples were combined with pSV50 DNA and used in transformation and selection experiments of N. crassa. The results of such experiments are shown in Table 3.3 (Stuart et. al., 1988). The digests of 5:3B with Hind III, Bgl II, and Hpa I did not decrease the transformation frequency below 10%. Whereas digests with Eco RI, Bam HI, Pst I, Kpn I, Xho I and Sau 3AI resulted in a decrease of transformation frequency to 6 - 10%. Experiment #2 included the digests of 5:4H with Eco RI (to serve as a negative control), Hind III, Bgl II, and Hpa I. Eco RI was the only enzyme which decreased the transformation frequency to about the 10% level. These experiments suggested that digests with Hind III, Bgl II, and Hpa I do not interrupt the coding region of the *mtr* locus.

That the Bgl II digest affected the transformation frequency of 5:4H the least prompted the direction of further analysis. Bgl II digestion of 5:4H resulted in five fragments of about 15.0, 11.0, 9.0, 4.1, and 2.3 kbp. All of these fragments were postulated to be of sufficient length to include the *mtr* gene. Each of the fragments were isolated from LMT agarose gels and the extracted DNAs were used in transformation experiments. Only the 15.0 kbp fragment was able to rescue the *mtr* phenotype (Table 3.4). This fragment was also able to transform the host strain to the benomyl resistant phenotype without the addition of pSV50 DNA. This finding suggested that the 15.0 kbp fragment contained both the benomyl resistant gene and a DNA fragment which rescues the *mtr* phenotype.

The pSV50 vector contains the benomyl resistant gene and a single Bgl II site. The cosmid library was constructed by inserting fragments of a limited Sau 3AI digest into the Bgl II site of pSV50. Theoretically, the Bgl II site of the pSV50 vector should be restored after such insertion events; a 7.9 kbp fragment which represents the vector should be released by a Bgl II digest. This 7.9 kbp Bgl II fragment was not found in 5:4H and this observation evidenced that the Bgl II sites were not restored after the insert was cloned to give 5:4H. This data, together with the finding that the 15.0 kbp fragment contains the benomyl resistance marker, also indicated that the 15.0 kbp fragment could be recircularized to form an intact plasmid. From a ligation and transformation experiment of E. coli strain LE392, six colonies were picked and and the plasmid DNAs were isolated. The DNA samples were initially screened by restriction enzyme digests with Bgl II. All six clones were found to yield a single fragment of 15.0 kbp. The DNA sample of one clone, pBN3, was further characterized by Hind III digests and Bgl II/Hind III double digests. These digests yielded the expected appropriate size fragments; fragments of the pSV50 vector and of the insert. The construction of a

Table 3.4

Transformation of mtr-6(r) allele with DNA fragments of Bgl II digested 5:4H

	Sample	No. of colonies
Expt. 1	> 9.0 kbp fragments + pSV50	344
	4.1 kbp fragment + pSV50	6
	2.3 kbp fragment + pSV50	5
	No DNA	0
	pSV50	83
	5:4H	2016
	Bgl II 5:4H + pSV50	670
Expt. 2	15.0 kbp fragment + pSV50	33
	11.0 kbp fragment + pSV50	2
	9.0 kbp fragment + pSV50	2
	15.0 kbp fragment + pSV50	8
	11.0 kbp fragment + pSV50	0
	9.0 kbp fragment + pSV50	0
	No DNA	0
	5:4H	2128
	Bgl II 5:4H + pSV50	426

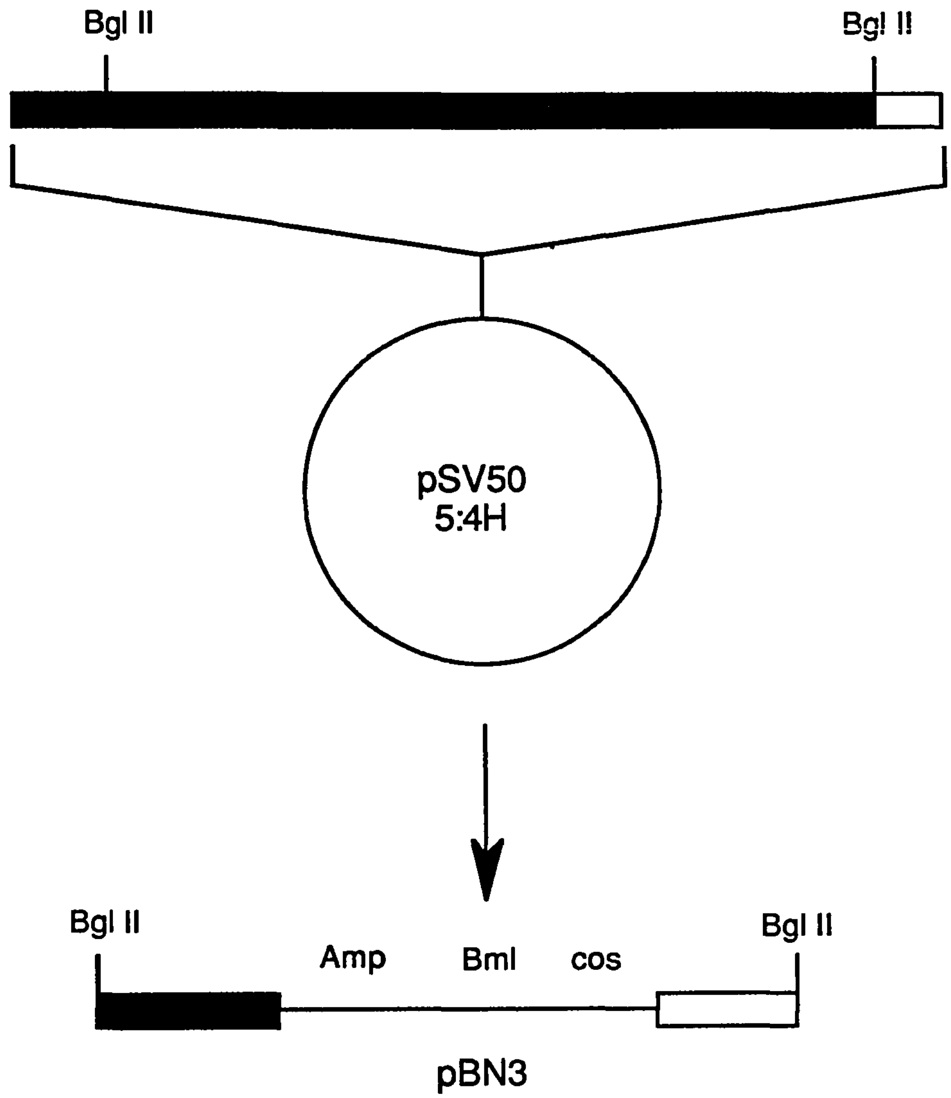


Fig. 3.5 A schematic representation of the construction of pBN3. DNA from cosmid 5:4H was digested with Bgl II and the ~15.0 kbp fragment was recircularized to generate pBN3 - benomyl resistant and neutral (N) transport system. Amp - ampicillin resistance marker; Bml - benomyl resistance marker; cos - cohesive ends from bacteriophage lambda.

representative plasmid , pBN3: benomyl resistant and neutral (N) transport system, is shown in Fig. 3. 5.

The plasmid pBN3 contains an insert of about 7 kbp. This fragment represents the ends of the insert in 5:4H which are flanked by the nearest Bgl II sites. This arrangement suggested that one of the ends contained the *mtr* locus. To determine which end contained the *mtr* locus, both ends of the insert in pBN3 needed to be subcloned. To achieve this goal, restriction enzyme analyses of pBN3 were performed and the results

Table 3.5

Restriction Enzyme Analysis of pBN3

Enzymes	Fragment Size (kbp)
Bam HI	4.9, 4.5, 3.8, 1.8, 0.6
Bgl II/Bam HI	4.5, 3.8, 2.9, 1.8(d), 0.6
Eco RI	9.0, 5.0, 3.6
Bgl II/Eco RI	5.0, 4.4, 3.6, 2.9
Hind III	5.9, 5.4, 4.7, 0.4
Bgl II/Hind III	5.4, 4.7, 2.9, 2.6, 0.4
Pst I	18.0, 2.3, 2.0
Bgl II/Pst I	9.8, 3.1, 2.3, 2.0

Table 3.5 Restriction enzyme analysis of pBN3. The estimates of molecular size of the various DNA fragments were determined by extrapolation on a semi-log plot of R_f vs kbp. The (d) represents a doublet band. Hind III digested λ DNA were used as molecular weight standards.

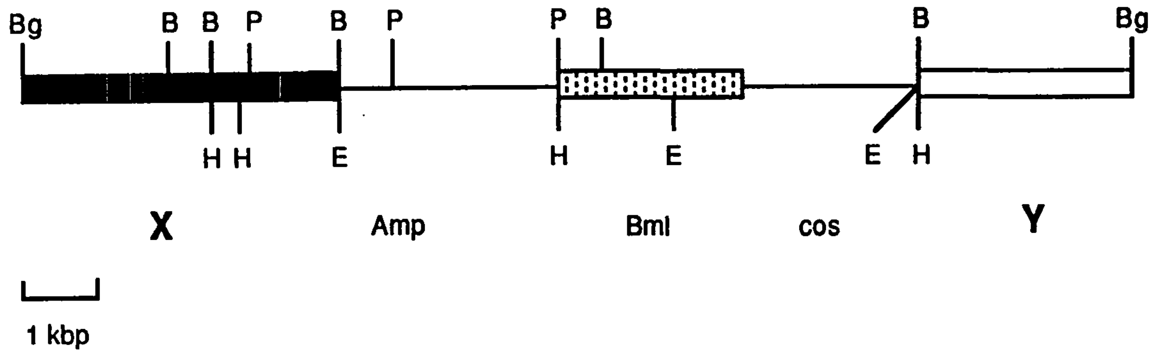


Fig. 3.6 Restriction enzyme map of pBN3. This construct was generated by ligating together Bgl II sites from the insert of 5:4H. The stippled bar which represents the ß-tubulin gene and line represents the pSV50 vector. The ends are denoted by X (the darkened bar) and Y (the open bar). The restriction sites are : Bam HI (B), Bgl II (Bg), Eco RI (E), Hind III (H), and Pst I (P).

of these analyses is shown in Table 3. 5. From these results, a restriction enzyme map of pBN3 was constructed and is shown in Fig. 3. 6; the ends are indicated by X and Y. The map indicated that both ends, X and Y, could be released intact by a Bgl II/Eco RI double digest. The two Eco RI sites in pBN3 are from the polylinker region in pSV50. Both ends were subcloned into the BamHI and Eco RI sites of the plasmid vector pTZ18R.

Eight colonies were picked and the plasmid DNAs were prepared. The DNA samples were initially screened by Bam HI digests. Four of the Bam HI digests were readily interpretable. Two of the clones contained the X end; as judged by yielding a single fragment of about 5.8 kbp with a Bam HI digest. The other two clones contained the Y end; the Bam HI digest yielded three fragments of about 4.9, 1.8, and 0.6 kbp. The lengths of the fragments correspond to the expected lengths shown in the map of pBN3 (Fig. 3.6) and to the calculated size change due to the cloning process.

DNAs prepared from two representative clones of both ends, pX18R2 - X end, pTZ18R vector; and pY18R1 - Y end, pTZ18R vector (later renamed to pCVN 2.9), were used in transformation experiments of N. crassa. DNA from pCVN2.9 - chimeric vector, neutral (N) transport system, 2.9 kbp fragment - was found to be able to rescue the *mtr* phenotype, whereas pX18R2 was not able to transform the *mtr* phenotype (Table 3.6). A comparison of transformation frequency with cosmid 5:4H, pBN3 and pCVN2.9 is shown in Table 3. 8 (Stuart et. al., 1988). The transformation frequency exhibited by the DNA of pCVN2.9 was at about the same frequency of 40% as with pBN3 when compared to 5:4H. This high level of transformation frequency suggested that the insert of pCVN2.9 contained the intact *mtr* coding region.

The results of restriction enzyme digestion analyses of the fragment are shown in Table 3. 7. From these results, a restriction enzyme map of the 2.9 kbp insert was constructed and is shown in Fig. 3.7 (Stuart et. al., 1988).

Table 3.6

Transformation of mtr-6(r) allele with subclones of ends from pBN3

Sample	No. of colonies
pCVN2.9	356
pX18R2	0
No DNA	0
pBN3	608

The effect of restriction digests of the 2.9 kbp fragment with various enzymes were investigated to determine which region of the fragment contain the *mtr* coding region (Table 3.8). The dramatic decrease in transformation frequency with Cla I (16%) and Eco RV (14%) digested DNA suggested that the *mtr* coding region was localized to the right hand end of the restriction map (Stuart et. al., 1988). However, there were some ambiguities in these results as seen in the transformation frequencies with Sph I, Acc I and Xho I digested samples. These samples would have been expected to also reduce the transformation frequency to a level comparable to the Cla I and Eco RV digests. This expectation is borne out due to the proximity of these sites in the restriction map (Fig. 3.7). Thus, it was not clear which region of the 2.9 kbp fragment encompasses the *mtr* locus.

Table 3.7

Restriction Enzyme Analysis of pCVN2.9

Enzymes	Fragment Size (kbp)
Acc I	3.2, 1.7, 0.9
Acc/Eco RI	2.9, 1.7, 0.9, 0.3
Acc I/Xho I	3.2, 1.0, 0.7, 0.6, 0.3
Apa I	5.8
Apa I/Eco RI	5.5, 0.3
Cla I	4.4, 1.5
Cla I/Eco RI	3.4, 1.5, 1.0
Eco RV	5.8
Eco RV/Eco RI	3.2, 2.6
Eco RV/Hind III	5.5, 0.3
Sph I	5.1, 0.7
Sph I/Eco RI	2.9, 2.2, 0.7
Xho I	4.2, 1.6
Xho I/Eco RI	3.2, 1.6, 1.0
Kpn I	4.4, 0.9, 0.6
Kpn I/Eco RI	3.7, 0.9(d), 0.6
Kpn I/Hind III	4.0, 0.9, 0.6(d)
Kpn I/Xho I	4.2, 0.8, 0.6, 0.3, 0.1

Table 3.7 Restriction enzyme analysis of pCVN2.9. The estimates of molecular size of the various DNA fragments were determined by extrapolation on a semi-log plot of R_f vs kbp. The (d) represents a doublet band. Hind III digested λ DNA were used as molecular weight standards. The (d) represents a doublet band.

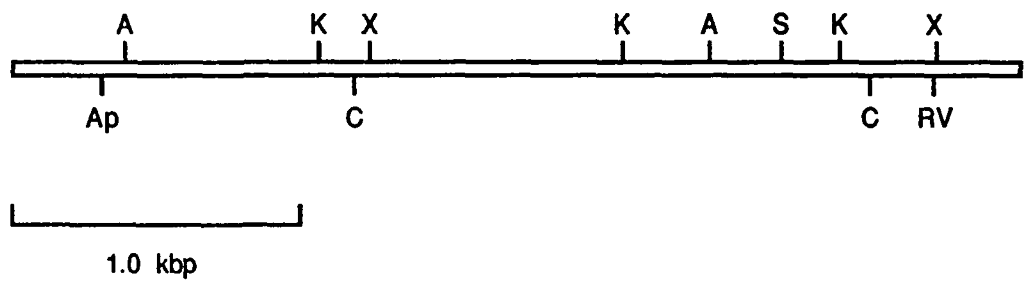


Fig. 3.7 Restriction enzyme map of the 2.9 kbp fragment. The restriction sites are: Acc I (A), Apa I (Ap), Cla I (C), Kpn I (K), Eco RV (RV), Sph I (S), and Xho I (X).

Table 3. 8

Transformation of mtr-6(r) allele with restriction digests of subclone pCVN2.9

Sample	No. of colonies N	% of 5:4H	% of pCVN2.9
pCVN2.9	704	38	100
restricted with:			
Acc I	536	29	76
Pvu II	308	17	43
Xba I	560	30	79
Xho I	488	26	69
Cla I	112	6	16
Apa I	218	12	31
Eco RV	102	5	14
Sph I	728	40	103
Controls			
5:4H	1840	100	-
pBN3	752	41	-
No DNA	2	<1	<1

All samples of pCVN2.9 were mixed with pSV50 DNA and used to co-transform the host strain spheroplasts.

The lack of homology between cosmids 5:3B and 5:4H at high stringency levels was again readdressed after the cloning of pCVN2.9. No homology at high stringency levels was detected in a Southern blot of 5:3B DNA using a P^{32}-labelled pCVN2.9 probe (data not shown).

Resubcloning and Unidirectional Subcloning of the *mtr* locus

Resubcloning

The determination of the nucleotide sequence of the *mtr* gene, and hence its localization within the 2.9 kbp fragment, required preliminary re-subcloning. The use of the Exo III/ S1 nuclease unidirectional subcloning technique necessitated this re-subcloning. The insertion of the 2.9 kbp Bgl II/ Eco RI fragment into the Bam HI and Eco RI sites removed most of the restriction sites within the polylinker which would generate 3' protruding ends. These ends are not substrates for the exonuclease activity of Exo III; thus digests which generate 3' protruding ends allow for exonuclease digestion to occur only in one direction.

The 2.9 kbp fragment is also released from pBN3 with a Bgl II/ Bam HI double digest. Like the Eco RI sites, the Bam HI site is from the polylinker region in pSV50. The Bgl II/ Bam HI fragment was cloned into the Bam HI site of the phagemid vector pTZ18R. There are two restriction enzyme sites on either termini of the insert which when cleaved generate 3' protruding ends. Forty eight colonies of transformed E. coli NM522 cells were picked and screened for containing the appropriate size insert. Four clones were found to have the 2.9 kbp fragment. These clones were named pN807, pN816, pN839 and pN846 - neutral (N) transport system. The inserts of clones pN807 and pN816 are in the same orientation - with respect to the polylinker - as in pCVN2.9, while the insert is in the opposite orientation in clones pN846 and pN839. The clones, with its insert in opposite orientation, were needed to be able to sequence both strands. The orientation were verified by double digestion with Bam HI and Hind III; clones pN807 and pN816 yielded a doublet band of about 2.9 kbp with a Bam HI/Hind III double digestion, whereas clones pN839 and pN846 yielded a single fragment of about 5.8 kbp. The authenticity of the inserts in clones pN807 and p846 were confirmed by restriction digests with Acc I, Apa I, Cla I, Eco RV, Kpn I, and Xho I. The molecular weights of all the resulting fragments corresponded to the expected sizes in the map of pCVN2.9.

Unidirectional Subcloning

Unidirectional subclones or nested deletion subclones of both pN807 and pN846 were generated by the Exo III/S1 method as described in Materials and Methods, Section 9. Exo III/S1 treated DNA samples from each time interval were ligated and used to transform E. coli strain NM522. The resulting colonies were picked and the plasmid DNAs were prepared. The DNAs were initially screened with Hind III digests. DNA samples, which yielded a single fragment of appropriate size with the Hind III digestion, were subsequently screened with a Hind III/Eco RI double digest. The Hind III/Eco RI double digest yielded a 2.9 kbp vector fragment and an insert. Digestion with Eco RI was used to ensure that the all the DNA samples had intact Eco RI sites and possible unchanged sequence primer sites. The Eco RI site within the multiple cloning region is the closest restriction enzyme site to the reverse sequence primer site of pTZ18R.

The insert of all the nested deletion subclones from pN807 and pN846 were checked for authenticity by Kpn I, Kpn I/Hind III, and Kpn I/Eco RI digests. Only those clones which yielded the appropriate size fragments - fragments of the insert and of the vector, were further characterized. Nested deletion subclones of pN846 were named the pNxB series; e.g. pNxB1-3 is the third clone of interval #1 whose insert is flanked by an Exo III/S1 (x) generated end and a Bam HI site. The clones and the insert sizes of the pNxB series are shown in Table 3.9.

Nested deletion subclones of pN807 were named the pNxB/B series with B/B representing the hybrid Bgl II/Bam HI site. These clones and their approximate size inserts are shown in Table 3.10.

Table 3. 9

Nested deletion subclones of pN846

Subclone	Insert Size (kbp)
pNxB1-3	2.2
1-4*	2.2
pNxB2-1	1.9
2-8*	2.0
pNxB8-2 *	1.3
8-3	1.3
pNxB10-8	1.1
10-10*	1.0
10-14	1.1
pNxB3-1	0.8
3-2	0.8
3-6	0.8

Table 3. 10

Nested deletion subclones of pN807

Subclone	Insert Size (kbp)	Subclone	Insert Size (kbp)
pNxB/B1-9	2.7	pNxB/B6-3	1.3
1-10	2.7	6-13	1.8
1-11 *	2.7	6-14	1.8
		6-19	1.8
pNxB/B2-7	2.5	6-21	1.3
2-9	2.6		
2-11 *	2.5	pNxB/B7-5 *	1.6
		7-6	1.6
pNxB/B3-7 ?	2.5	7-8	1.7
3-11	2.5	7-9	1.7
pNxB/B4-2	2.3	pNxB/B8-3	1.5
4-3 *	2.3	8-4 *	1.5
4-8	2.3	8-6	1.3
4-9	2.3		
4-12 *	2.0	pNxB/B9-3	1.2
		9-5	1.2
pNxB/B5-2	1.9	9-6	1.2
5-4 *	1.8	9-12 *	1.3
5-13	1.8		
5-17 ?	1.9	pNxB/B10-2 *	1.0
		10-4	1.0

Restriction Site Subclones

Restriction site subclones pNE/K6 10-2, pNE/E2 10-2, pNE/E4 846, and pNA/E 12 3-1 were generated by the procedures described in Materials and Methods, Section 7. In the screening process for pNE/E4 846, six colonies were picked and the plasmid DNAs were prepared from 2.5 ml cultures. The DNAs were screened with Hind III/Xho I double digestions. Four samples yielded two fragments of about 4.5 kbp and 1.0 kbp. These results suggested that these clones contained authentic inserts derived from the 2.9 kbp fragment. The authenticity of all four clones were confirmed by restriction digests with Kpn I with which three fragments were observed. The molecular weight sizes of the fragments were about 0.6 kbp, 0.86 kbp, and 4.0 kbp; these lengths corresponded to that of Kpn I digests of the 2.9 kbp fragment and to calculated size changes of the vector due to the cloning process.

Twelve colonies were picked in the screening process for pNA/E 12 3-1. The plasmid DNAs were prepared and digested with Hind III/Eco RI. Seven of the DNA samples yielded two fragments of about 2.9 kbp and 0.3 kbp. The cloning procedure, as expected, maintained the Eco RI site after the ligation of the blunt ends. The authenticity of the inserts could not be readily determined by restriction enzyme digests. However, the DNA sequence of pNA/E12 3-1 was homologous to the antiparallel strand at this region and overlapped with the sequence of the preceding template.

Six colonies were picked in the screening process for pNE/E2 10-2. Plasmid DNAs were prepared and digested with a Hind III/Xho I double digestion. All six DNA samples ceded two fragments of about 2.9 and 0.26 kbp. These results suggested that the inserts were authentic in that a Xho I site was present as expected. Further confirmation of the insert's authenticity in pNE/E2 10-2 was provided by the overlap in DNA sequence with the preceding template.

Six colonies were picked in the screening for pNE/K6 10-2. Plasmid DNAs were prepared and digested with a Hind III/ Xho I double digest. Four DNA samples yielded two fragments of about 3.2 and 0.26 kbp. These results indicated that the inserts were authentic in that they contained a Xho I site. The DNA samples of these four clones were further digested with Kpn I, Kpn I/Eco RI double digest, and Hind III/Cla I double digest. The Kpn I digest failed to linearize the plasmid DNA and confirmed the cloning process. The Kpn I/Eco RI linearized the plasmid DNA thus suggesting that the Eco RI site was restored, as expected, in the cloning process. The Hind III/Cla I double digest yielded two fragments of about 3.0 and 0.46 kbp; these results are indicative of a genuine insert.

DNA Sequence of the *mtr* locus

The sequencing strategy used to sequence the 2.9 kbp fragment employed some of the above nested deletion subclones from pN807 and pN846. The deletion subclones used for the sequencing are denoted by asterisks in Tables 3. 9 and 3.10. In a few instances, some of the deletion subclones yielded sequence which did not correspond or overlap with the sequence from the preceding deletion subclone. These sequences start at the 5' termini of the template and have not been observed to start in the interior of the template. The extent of these unexplained sequences may range from 10-100 bp. These deletion subclones with unexplained 5' sequence is denoted by (?) in Table 3. 10 and have not been used in generating the sequence of the 2.9 kbp fragment. This phenomenon has been noted in the development of the Exo III/S1 nested deletion subcloning procedure (Henikoff, 1984).

Some regions required the use of "restriction site subclones" pNE/K6 10-2, pNE/E2 10-2, pNE/E4 846, and pNA/E12 3-1. pNE/K6 10-2 and pNE/E2 10-2 is representative of the strand which was sequenced with the pNxB/B series. pNE/E4 846 and pNA/E12 3-1 represents the complementary strand which was sequenced with the pNxB series. The sequencing strategy is shown in Fig. 3.8.

The sequence generated from each of the deletion subclones was determined and accepted using the following criteria. The autoradiogram of a sequencing reaction had to be "readable". An example of a "readable" autoradiogram is shown in Fig. 3.9. Each template was sequenced until a "readable" autoradiogram of that template was obtained. Each autoradiogram was read twice. In most instances, at least two such autoradiograms were attained before the sequence was accepted to be representative of that template.

Only 2,060 bp of the 2.9 kbp fragment is shown in Fig. 3.10. The sequence is presented in a 5' to 3' orientation, as that of the pNxB/B series; the 5' end of the sequence corresponds to the right hand end of the restriction map (Fig. 3.8). The number of base pairs in the 2.9 kbp fragment is 2,783 bp. The 723 bp of sequence at the 3' end is not shown because of the ambiguous assignment of seven bases. These few uncertainties are probably due to random termination events of the sequencing reaction. However, the following sequence analyses were performed with the entire 2,783 bp of sequence.

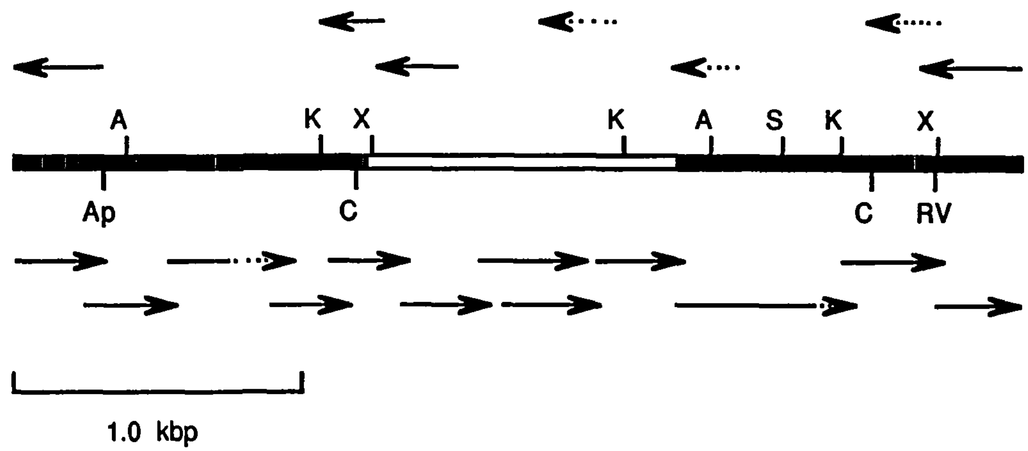


Fig. 3.8 Sequencing strategy of the 2.9 kb genomic fragment. The arrows above and below the restriction enzyme map indicate the extent and direction of the sequencing reactions of the two strands. The solid lines of the arrows signify regions of one strand which have been sequenced at least twice. The dotted lines indicate regions which have been sequenced once. The open bar in the restriction enzyme map indicate the region of the proposed *mtr* ORF. The restriction sites are: A - Acc I, Ap - Apa I, C - Cla I, K - Kpn I, RV - Eco RV, S - Sph I, and X - Xho I.

Fig. 3. 9 Autoradiogram of a "readable" sequencing gel. Three different templates were sequenced and the sequencing products were electrophoresed in 24 lanes; 8 lanes per template. The loading pattern, from left to right, for each template is G A T G C A T C. This loading pattern allows for the direct comparison of bands between lanes and minimizes the error in reading the autoradiogram.

-972 AGATCCGCCTCGCCCCAAGC GCATCCCAACGCGGCGTGCT TATATGTCGCTCTTCCCTCT

-912 CACGTACCTCGCAAGTACCT GTCTCATCTGGCACCCGCCT TCTCCATCCCTCTTCTTCAG

-852 TACTTAATCTGCCCCGGTCC CCCGCAGTTCATCCTGTCTC TCAGACCTTGGATCGTCTGT

-792 TGGTCTTTTCGCTTATTGTC CGTACCCCCATCCATATTTA TTCCTGCCTGGGCCCCCAGA

-732 CAGCATTTTCTCTTTCCCTC CCTCTTTAGCCATCCAAACA GCTTGAGAAGCGAAAATCAA

-672 GCCACTGTCGACAAAGGCCT TCAACAAGGCCTCTTTACCA CCCAACATGGACTCGCAATA

-612 CGAGACAAAAAAGAATGACC CAAACGCCATCATGCCGTAC CCAGAGTCAAACGATGAGCA

-552 TGTTGGCGAGGTCCGCGGCT TGGGCGGCGGCATCATGGAC AAGGAGCCTGAGGCCCAGGA

-492 GGGCATGCCAAGTTCCACCG TCTCGGCTGGAAACGTCTGA CGGTCGTCCTCATCGTCGAG

-432 GCCATTGCCCTCGGCTCTCT CTCGCTTCCCGGCGCCTTCG CTACCCTTGGCATGGTGCCT

-372 GGTGTTATTCTCTCTGTCGG CATGGGACTCATCTGCATCT ACACGGCTCACGTTATCGGA

-312 CAAACCAAGCTCAAGCACCC TGAAATCGCCCACTATGACC GACGTTGGTCGTGTCATGTT

-252 TGGAAGATGGGGATATGAAA TCATCAGCTTCATGTTTGTT CTGCAACTGATCTTCATCGT

-192 CGGCTCCCACGTCCTCACTG GCACCATCATGTGGGGCACC ATCACGGATAACGGCAACGG

-132 TACCTGCTCTCTCGTCTTCG GCATTGTCTCCGCCATCATT CTCTTCCTCCTTGCCATTCC

- 72 TCCCAGTTTCGCCGAGGTTG CCATCCTTGGATACATCGAT TTCGTCTCCATCTGCGCCGC

vvvvvv vvv

- 12 CATCCTCATCACC ATG ATT GCT ACT GGC ATT CGC TCG AGC CAC CAG GAG

Met Ile Ala Thr Gly Ile Arg Ser Ser His Gln Glu

+37 GGT GGT CTC GCT GCT GTT CCC TGG TCT TGC TGG CCC AAG GAG GAC CTT

Gly Gly Leu Ala Ala Val Pro Trp Ser Cys Trp Pro Lys Glu Asp Leu

+85 AGC CTT GCT GAG GGC TTC ATT GCT GTC AGC AAC ATC GTT TTC GCC TAC

Ser Leu Ala Glu Gly Phe Ile Ala Val Ser Asn Ile Val Phe Ala Tyr

+133 AGC TTC GCC ATG TGC CAG TTC AGC TTT ATG GAT GAG ATG CAC ACC CCC

Ser Phe Ala Met Cys Gln Phe Ser Phe Met Asp Glu Met His Thr Pro

+181 TCC GAC TAC AAG AAG TCC ATC GTT GCT CTC GGC TTG ATT GAA ATC TTC

Ser Asp Tyr Lys Lys Ser Ile Val Ala Leu Gly Leu Ile Glu Ile Phe

+229 ATC TAC ACC GTT ACT GGT GGC GTC GTT TAC GCT TTC GTC GGC CCC GAG

Ile Tyr Thr Val Thr Gly Gly Val Val Tyr Ala Phe Val Gly Pro Glu

+277 GTC CAG TCT CCT GCC TTG CTC TCT GCT GGC CCT CTT CTC GCC AAG GTT

Val Gln Ser Pro Ala Leu Leu Ser Ala Gly Pro Leu Leu Ala Lys Val

+325 GCT TTC GGC ATT GCC CTC CCC GTC ATC TTC ATC TCT GGC AGT ATC AAC

Ala Phe Gly Ile Ala Leu Pro Val Ile Phe Ile Ser Gly Ser Ile Asn

+373 ACT GTT GTC GTC AGC AGG TAT CTG ATT GAG CGC ATC TGG CCC AAC AAC

Thr Val Val Val Ser Arg Tyr Leu Ile Glu Arg Ile Trp Pro Asn Asn

+421 GTC ATT CGC TAT GTC AAC ACC CCA GCG GGT TGG ATG GTT TGG CTT GGT

Val Ile Arg Tyr Val Asn Thr Pro Ala Gly Trp Met Val Trp Leu Gly

+469 TTT GAC TTT GGC ATT ACC CTC ATT GCC TGG GTT ATT GCT GAG GCC ATC

Phe Asp Phe Gly Ile Thr Leu Ile Ala Trp Val Ile Ala Glu Ala Ile

```
+517 CCT TTC TTC TCT GAT CTG TTG GCC ATC TGC TCG GCT CTC TTC ATT TCC
     Pro Phe Phe Ser Asp Leu Leu Ala Ile Cys Ser Ala Leu Phe Ile Ser

+565 GGT TTT AGC TTC TAT TTC CCT GCC TTG ATG TAT TTC AAG ATC ACC AGG
     Gly Phe Ser Phe Tyr Phe Pro Ala Leu Met Tyr Phe Lys Ile Thr Arg

+613 AAC GAT GCC AAG AGC CAG GGC AAG AAG TAC TTC TTG GAT GCC CTC AAC
     Asn Asp Ala Lys Ser Gln Gly Lys Lys Tyr Phe Leu Asp Ala Leu Asn

+661 ATG CTC TGC TTC GTC ATC GGC ATG GGC ATT CTT GGT ATT GGT ACC TAC
     Met Leu Cys Phe Val Ile Gly Met Gly Ile Leu Gly Ile Gly Thr Tyr

+709 GCC GCT ATT CAG GAC ATT GTAAGTTTGGCCCGCTTTTCTGTTTACTCTTTGCACAC
     Ala Ala Ile Gln Asp Ile

+765 AAATGCTAACTTGCTTCTCAG  ATG  GAC CGT TAC GAC CAT GGC AAG GTT TCG
                            Met  Asp Arg Tyr Asp His Gly Lys Val Ser

+816 AAG CCT TAT AGC TGT GCG CCC TTG GCT TAA CAGGCCCAACGCACGCTTATGA
     Lys Pro Tyr Ser Cys Ala Pro Leu Ala Stop

+868 TCCTGTTGTTTTTTTTTGGA TGATTTAATTAAAGTTGCGC AGTGATTGACGTCTGTCTTC

+928 ACCCGCGATTGCCCCTTTTG TATACCCCCTCAGACTTGCC GGCCTGGGGAAATGTTTTGA
+988 GTATTTCTATTTTCGGAGTT TCAGGATTTGGCACAAAGCA AACCAGCGCGGAGTTGAAAC
+1048 CGTGGTGTGGTCGCGGTGCG CTGCTGCATTGGTAGTGCTT
```

Fig. 3.10 Sequence of 2,060 bp of the 2.9 kbp fragment. The data base from which the sequence is derived is shown in the Appendix. The proposed *mtr* ORF is translated to the amino acid sequence. The 5' end of the sequence corresponds to the right hand end of the restriction map (Fig. 3.8). The number of base pairs in the 2.9 kbp fragment is 2,783 bp. The (v)s represent the sequence homology to the N. crassa ATG start site. The underlined DNA sequences denote the homologies to sequences typical of introns of N. crassa genes. Four possible capping sites are bracketed upstream from the translation start site. The proposed polyadenylation signal is indicated by the broken box. The amino acid sequences which comprise the six hydrophobic regions are boxed in grey and the RNP octamer motif is boxed.

Sequence Analysis

The sequence of the 2.9 kbp fragment was initially analyzed with the Sequence Analysis Package by Bob Stevens. The sequence was used to search for homologies to the N. crassa consensus sequences at the ATG start site, and to N. crassa consensus sequences at the exon/intron splice junctions.

Consensus sequence at the ATG site

The N. crassa consensus sequence at the ATG start site is A T C A C/A C/A A T G (Legerton and Yanofsky, 1985; McNally and Free, 1989). Genes of N. crassa which have this sequence are shown in Table 3.11. The DNA sequence of both strands of the 2.9 kbp fragment was analyzed for homologous sequences to this ATG consensus sequence.

The sequence of C A X C A T G, in which X represents any base, was searched for at window setting of 7 base and match setting of 6; i.e., 85.7% homology, the maximum level of homology which can be assigned with a "X" in the search sequence. Six homologous sequences were found at positions -630, -585, -522, -168, -3, and +657 in the strand shown in Fig. 3.10. This strand is referred to as strand #1; whereas the antiparallel strand is termed strand #2. Three of the six sequences exhibited either an A or C at the third position. Only one of the three sequences, position -3, had an A T sequence preceding the search sequence for a 100% homology to the N. crassa ATG consensus sequence. The results of this search are summarized in Table 3.12.

Two homologous sequences were found at the positions -548, and -376 (with respect to the position numbers in Fig. 3.10) in the antiparallel strand, strand #2. Both sequences displayed an A or C at the third position, but did not have the preceding A T sequence, Table 3.12.

Table 3. 11

N. crassa consensus sequence at the ATG start site

Gene	Sequence at ATG site
his-3	A A C A C C ATG
trp-1	A T C A C A ATG
qa-2	A A C A C C ATG
qa-3	A T C A C C ATG
qa-4	T T C G C C ATG
qa-X	G C A G C C ATG
qa-Y	G T C A A G ATG
am-1	T T C A A A ATG
histone-3	A T C A C A ATG
histone-4	A T C A A A ATG
ADP/ATP carrier	A T C A C A ATG
B - tubulin	G T C A A G ATG
grg-1	A C C A A A ATG
Consensus sequence	$A^{8}T^{9}C^{12}A^{11}C^{8}A^{6}$ATG $A^{5}C^{5}$
mtr	A T C A C C ATG

Table 3. 12

Homology search for N. crassa consensus sequence at the ATG start site in strands #1 and #2

Strand #1

Search Sequence	CAXCATG	CAXAATG
Homolgy to the search sequence [position]	CAACATG [-630] CATCATG [-585] CATCATG [-522] CATCATG [-168] CACCATG [-3] CAACATG [+657]	None
Homology at the third base [position]	(CC) CAACATG [- 630] (AT) CACCATG [-3] (CT) CAACATG [+657]	None
100% Homology to the ATG consensus sequence [position]	AT CACCATG [- 3]	None

Strand #2

Search Sequence	CAXCATG	CAXAATG
Homology to the search sequence [position]	CAACATG [- 548] CACCATG [- 376]	None
Homology at the third base [position]	(GC) CAACATG [- 548] (GG) CACCATG [- 376]	None
100% Homology to the ATG consensus sequence [position]	None	None

The sequence of C A X A A T G was also searched for at a window and match setting of 7 and 6, respectively. No homologous sequences were found in either strand of the 2.9 kbp fragment. Like the C A X C A T G search sequence, the probability of finding a homologous sequence to the C A X A A T G search sequence is 1 in 4^6 or once per 4096 bp.

In the entire 2,783 bp of sequence for the two strands, only one sequence, at position -3 in strand #1, was found to be 100% homologous to the N. crassa ATG consensus sequence (see Table 3.12).

Consensus sequences of introns

The N. crassa consensus sequence at the 5' splice boundary of an intron is G T A A/C G T (Legerton and Yanofsky, 1985; Orbach et. al., 1986). In N. crassa, there is also a consensus sequence within the intron, A/G C T A/G A C T/A; at the 3' splice junction of an intron, the consensus sequence is T/C A G. Genes of N. crassa which display these consensus sequences are illustrated in Table 3.13.

The 5' splice junction consensus sequence was searched for with the search sequence G T A X G T. The search parameters were a window setting of 6 and match setting of 5; i.e. a 83.3% homology. Only one sequence, position +727 in strand #1, was found to be homologous to this search sequence. This sequence is a 100% identical to the 5' splice junction consensus sequence, Table 3.14. A similar search was also carried out using strand #2. Four sequences, positions -906, +708, +1239, and +1437, were found to be homologous to the search sequence. Two of the four sequences, positions -906 and +1239 in strand #2, showed a 100% identity to the 5' splice junction consensus sequence (see Table 3.14).

Table 3. 13

<u>N. crassa</u> consensus sequence at exon/intron splice junctions

Gene	Intron 5' Sequence	Intron Internal Sequence	Intron 3' Sequence
B - tubulin			
IVS 1	GTAAGT ------	GCTGACA --- 18 ---	TAG
IVS 2	GTAAGT ------	ACTAACA --- 13 ---	CAG
IVS 3	GTACGT ------	GCTGACA ---- 7 ---	CAG
IVS 4	GTACGT ------	GCTCACC ---- 8 ---	CAG
IVS 5	GTGAGC -----	GCTAGCT --- 14 ---	CAG
IVS 6	GTAAGT ------	ACTGACG ---- 7 ---	CAG
am 1	GTACGT ------	GCTGACT --- 16 ---	CAG
am 2	GTAAGT ------	GCTGACT --- 12 ---	CAG
histone - 3	GTAAGT ------	GCTAACT --- 13 ---	CAG
histone - 4 1	GTAAGT ------	ACTGACT --- 16 ---	CAG
histone - 4 2	GTACGT ------	ACTAACA --- 16 ---	CAG
his - 3	GTAAGT ------	GCTAACT ---- 9 ---	TAG
<u>N. crassa</u> consensus	GTAAGT ------ C	ACTAACA --7-18 -- G G T	CAG
Yeast consensus	GTATGT -----	TACTAACA--18-53--	CAG
Higher Eukaryote consensus	GTAAGT ---------------------------		CAG
mtr	GTAAGT ------	GCTAACT -----7 ---	CAG

Table 3.14

Homology search for 5' splice junction consensus sequences in strands #1 and #2

	Strand #1	Strand #2
Search Sequence	GTAXGT	GTAXGT
Homology to the search sequence [position]	GTAAGT [+727]	GTACGT [- 906] GTAGGT [+ 708] GTACGT [+1239] GTATGT [+1437]
100% homology to the consensus sequence [position]	GTAAGT [+727]	GTACGT [- 906] GTACGT [+1239]

The consensus sequence, A/G C T A/G A C T/A, within the intron was searched for using the search sequence of X C T X A C X. The search parameters were a window setting of 7 and a match setting of 4. Seven sequences, positions -915, -457, -327, -180, +769, +1505, and +1642 in strand #1, were found to be homologous to the search sequence. Only one of the seven sequences, G C T A A C T position +769, showed a 100% identity to the consensus sequence. In addition, the 3' end of this sequence was 7 bp from a 3' splice junction consensus sequence of C A G. This number of bp is within the range of the number of bps (7-18) which separate the internal sequence from the 3' splice junction consensus sequence, Table 3.13. None of the other six sequences exhibited this characteristic (Table 3.15).

A similar search was also carried out on the sequence of the antiparallel strand. Six sequences, positions +114, +387, +732, +1660, +1687, and +1696 in strand

#2, showed homology to the search sequence. One of the six sequences, G C T G A C A position +114, displayed a 100% identity to the consensus sequence. This sequence was followed by a C A G 3' consensus sequence 13 bp downstream. One of other five sequences, A C T T A C A position +732, which did not display a 100% homology was followed by a T A G 3' consensus sequence 10 bp downstream.

The search for internal intron consensus sequences in the entire 2,783 bp of sequence for the two strands yielded two sequences. Both sequences displayed a 100% homology to internal intron consensus sequence and was followed by a downstream 3' splice junction consensus sequence with an acceptable number of bp separating the consensus sequences. One sequence is found in strand #1 and the other sequence is present in strand #2.

Alignment of 5' splice sequence with the internal intron sequence

As stated before, only one sequence was found in strand #1 which exhibited 100% identity to the consensus sequence at the 5' splice exon/intron boundary. This sequence is at position +727. The one sequence in strand #1, which was 100% homologous to the internal intron consensus sequence, occurs at position +769. This sequence was 7 bp apart from a 3' C A G consensus sequence. The above identities, found in strand #1, strongly suggested the presence of a 59 bp intron. The presence of a N. crassa ATG consensus sequence in strand #1 and further sequence analysis (discussed below) supports this finding.

Of the two sequences in strand #2 which was 100% homologous to the 5' splice consensus sequence, only the sequence at position +1239 was followed downstream by

Table 3. 15

Homology search for internal intron consensus sequences in strands #1 and #2

	Strand #1	**Strand #2**
Search Sequence	XCTXACX	XCTXACX
Homology to the search sequence [position]	TCTCACG [- 915] TCTGACG [- 457] GCTCACG [- 327] CCTCACT [- 180] GCTAACT [+ 769] CCTGACT [+1505] CCTCACC [+1642]	GCTGACA [+ 114] GCTGACG [+ 387] ACTTACA [+ 732] TCTGACT [+1660] GCTGACG [+1687] ACTCACT [+1696]
100% homology to the consensus sequence [position]	GCTAACT [+ 769]	GCTGACA [+114]
100% homology and within the range of a 3' consensus sequence [position]	GCTAACT [+ 769]	GCTGACA [+114]

sequence which was 100% homologous to an internal intron consensus sequence at position +114. The lack of a N. crassa ATG consensus sequence in strand #2 and open reading frame (ORF) analysis (discussed below) do not support the possibility of these homologies being that of an intron in this strand.

Open Reading Frame Analysis

The sequence of the 2.9 kbp fragment was analyzed for open reading frames with the DNA Strider Program by C. Marck. This sequence analysis program is on the hard disk of the Macintosh SE computer in the Dept. of Genetics.

An open reading frame (ORF) is defined as a DNA sequence between and including a start and stop codon in frame. For any one strand of DNA sequence, there are three reading frames. It would be expected that start and stop codons to occur on average once per 4^3 or 64 bp. Since most genes are longer than 64 bp, the finding of a long open reading frame suggests the presence of a gene. In eukaryotic organisms, the ATG codon is the start codon and codes for the amino acid methionine; whereas, in prokaryotes, the ATG and GTG (codes for methionine) are used as start codons. The stop codons, TAA (ochre), TGA (opal), and TAG (amber), are used in both eukaryotes and prokaryotes and do not code for an amino acid.

The 2,783 bp of sequence was analyzed for its open reading frame content in both strands. The results of these analyses are illustrated in Fig. 3. 11. As expected, numerous open reading frames were found in all 6 frames. However, a long open reading frame (length 1146 bp) was found in reading frame #2 of strand #1. This ORF contains numerous ATG sites and one of these ATG sites coincides with the one N. crassa ATG consensus sequence found in strand #1 (see Table 3. 12). The 5' splice junction of the proposed 59 bp intron in strand #1 occurs in frame with this ORF. The inclusion of the sequence of intron, with regards to the ORF analysis, results in an in-frame termination at position +864. Computer editing to remove this intron changes the reading frame and results in a stop codon at position +845.

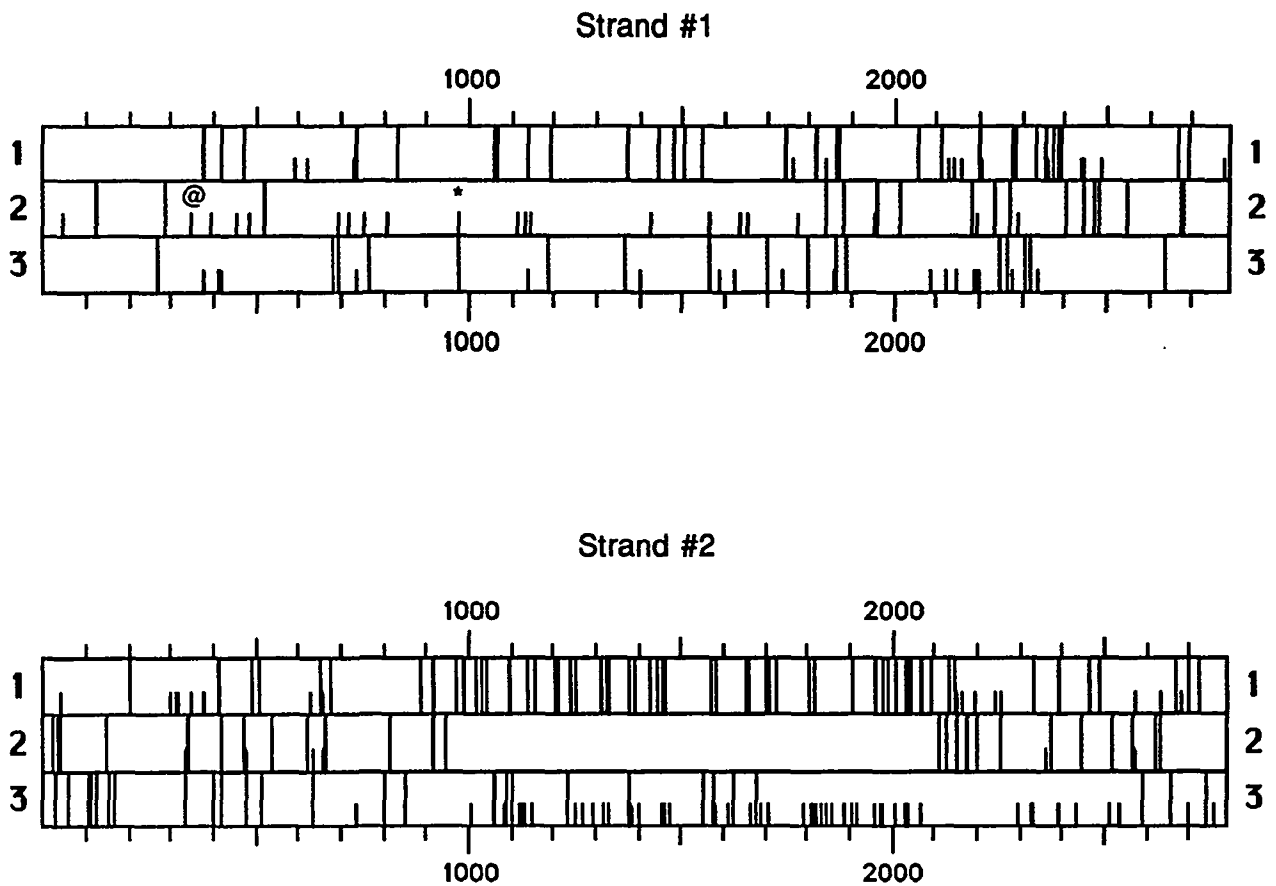


Fig. 3.11 Open reading frame analysis of the two strands of the 2.9 kbp fragment. The left side of the reading frames corresponds to the 5' end of the sequences. The start codons are represented by half length vertical lines. The stop codons are represented by the full length vertical lines. The ATG site of the proposed *mtr* ORF is marked with an asterisk and the ATG site of the proposed upstream open reading frame (uORF) is indicated by @.

Together, the homology to the consensus sequence at the ATG start site, the homology to consensus sequences of the proposed intron, and the ORF analysis of strand #1 strongly supports the idea that proposed *mtr* ORF is 845 bp with a 59 bp intron. The results of ORF analysis of strand #2 and the lack of homology to a consensus sequence at the ATG site do not support the possibility of strand #2 being the coding strand. This analysis of strand #2 also suggest that the homology to the consensus intron sequences found in strand #2 is coincidental.

Analysis for transcription signals of the *mtr* locus

The DNA sequences which flank the proposed *mtr* ORF were examined for consensus sequences commonly found in other eukaryotic genes. These eukaryotic consensus sequences include the "TATA" box (T A T A T/A A T/A), the "CCAAT" box (G G C/T C A A T), the capping site (C/T C A T T C A/G) and the poly A addition signal (A A T A A A). The "TATA" and "CCAAT" boxes have been found in many but not all eukaryotic promoters (Lewin, 1987). The "TATA" box lies between 19 to 27 bp upstream from the initiation site of transcription, the cap site, and the "CCAAT" box is approximately 80 bp upstream from the cap site. The capping site is near the 5' end of the mRNA transcript. The polyadenylation signal is found 10 to 30 bp upstream from the site of poly A addition (Humphrey and Proudfoot, 1988).

No sequence homologies to "TATA" nor "CCAAT" boxes were found within -200 bp upstream from the translation start site. Four sequences, T C A T T C T (-97), C C A T T C C (-79), C C A T C C T (-52) and C C A T C C T (-13), were found to resemble capping sites (Fig. 3. 10). The sequences at positions -97 and -79 resemble the capping site except for the "T" at the last base. The latter two sequences, which differ from the capping site sequence at two bases, are identical to that found in the his-2 gene of N. crassa (Legerton and Yanofsky, 1985). The sequence A A T T A A A is found at 49 bp downstream from the translation stop site at position 894 (Fig. 3. 10). This sequence is identical to the proposed poly A signal of the mammalian multidrug resistance (mdr) gene (Gros et. al., 1986).

Translation and Codon Usage of the *mtr* ORF

The deduced amino acid sequence of the proposed *mtr* ORF is shown in Fig. 3.10. The *mtr* ORF is translated into a polypeptide of 261 amino acids and the predicted molecular weight of the polypeptide is 28,613 daltons. The polypeptide is composed of 164 non-polar amino acid residues and represent 63% of the amino acids.

Codon usage in the *mtr* ORF is biased in that 18 of the 61 sense codons are absent. There is an 72% overall preference for codons ending in G + C residues in codon families for which there is a choice between purines and pyrimidines in the third position. Only two codons, CAA (pro) and GAA (glu), were found to have to an A in the third position. A summary of the codon usage in the *mtr* ORF is presented in Table 3. 16.

Hydropathy Analysis

A hydropathy analysis of the predicted amino acid sequence was performed with the Kyte and Doolittle algorithm (1982). The result of this analysis is shown in Fig. 3.12. The window setting was set at the default value of 11; that is, the hydropathy index of eleven consecutive amino acids was computed and averaged. The maximum value of the ordinate axis is +4.5 to which a polypeptide fragment composed of only isoleucine residues would be assigned. The minimum limit of the ordinate is -4.5 to which

Table 3.16 Codon Usage in the mtr ORF

Amino acid	Codon	Frequency *	Amino acid	Codon	Frequency *
Phe	UUU	4 (20)	Tyr	UAU	5 (42)
	UUC	16 (80)		UAC	7 (58)
Leu	UUA	0 (0)	Stop	UAA	1 (100)
	UUG	6 (27)			
	CUU	5 (23)	His	CAU	1 (33)
	CUC	9 (41)		CAC	2 (67)
	CUA	0 (0)			
	CUG	2 (9)	Gln	CAA	0 (0)
				CAG	5 (100)
Ile	AUU	15 (56)			
	AUC	12 (44)	Asn	AAU	0 (0)
	AUA	0 (0)		AAC	7 (100)
Met	AUG	9 (100)	Lys	AAA	0 (0)
				AAG	10 (100)
Val	GUU	10 (50)			
	GUC	10 (50)	Asp	GAU	4 (40)
	GUA	0 (0)		GAC	6 (60)
	GUG	0 (0)			
			Glu	GAA	1 (13)
Ser	UCU	5 (24)		GAG	7 (87)
	UCC	3 (14)			
	UCA	0 (0)	Cys	UGU	1 (20)
	UCG	3 (14)		UGC	4 (80)
	AGU	1 (5)			
	AGC	9 (43)			
			Trp	UGG	6 (100)
Pro	CCU	5 (38)			
	CCC	7 (54)			
	CCA	1 (8)	Arg	CGU	1 (17)
	CCG	0 (0)		CGC	3 (50)
				CGA	0 (0)
Thr	ACU	3 (33)		CGG	0 (0)
	ACC	6 (67)		AGA	0 (0)
	ACA	0 (0)		AGG	2 (33)
	ACG	0 (0)			
			Gly	GGU	8 (38)
Ala	GCU	13 (48)		GGC	13 (62)
	GCC	12 (44)		GGA	0 (0)
	GCA	0 (0)		GGG	0 (0)
	GCG	2 (7)			

* The percent frequency of a particular codon for an amino acid is bracketed.

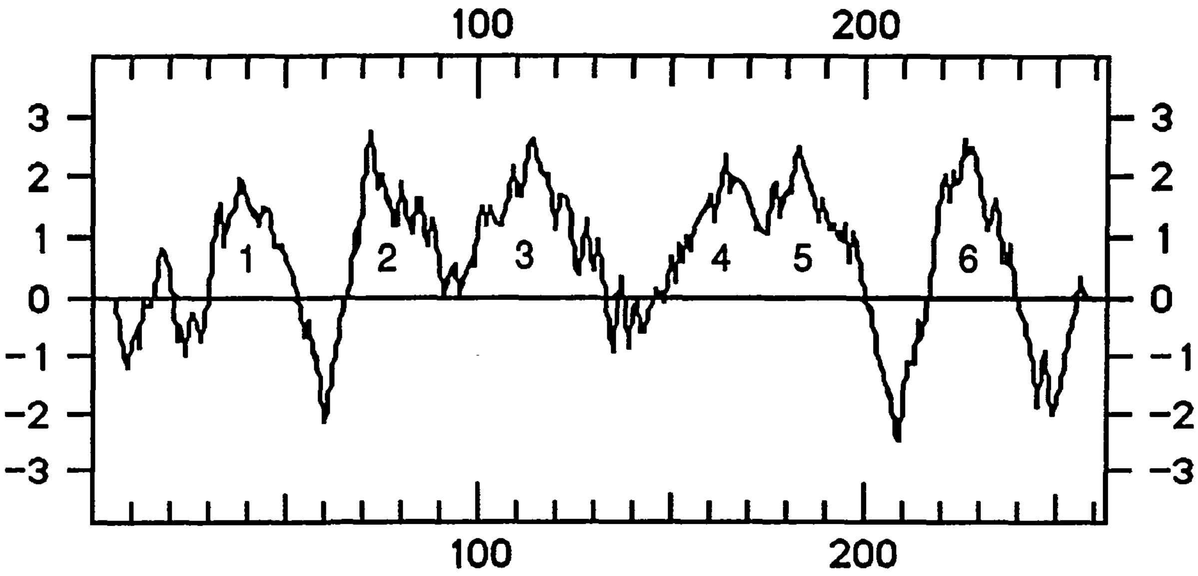


Fig. 3.12 Hydrophobicity analysis of the proposed *mtr* polypeptide by the algorithm of Kyte and Doolittle (1982). Hydrophobic regions are shown above the X-axis. All six hydrophobic regions are 20 - 30 amino acids long, sufficient to the cross the lipid bilayer, and may represent transmembrane domains.

arginine is assigned. Hence, portions of the graph above the median line of zero signify regions of greater hydrophobic content.

The hydropathy profile indicates six highly hydrophobic regions. The extent of the amino acid sequences which results in the maxima of the graph was determined. Region 1 extends from Ser_{29} to Phe_{53}, region 2 - Lys_{65} to Val_{84}, region 3 - Leu_{104} to Val_{128}, region 4 - Trp_{151} to Glu_{170}, region 5 - Leu_{172} to Tyr_{193}, and region 6 -Tyr_{214} to Ile_{239}. The average length of these six regions is 23 amino acid residues and all six regions are of sufficient length to span the membrane (Kaback, 1986).

Secondary Structure Analysis of the *mtr* polypeptide

The deduced amino acid sequence of the *mtr* polypeptide was analyzed for possible secondary structures. The analysis was performed with the IBI/Pustell Sequence Analysis system. The results of the secondary structure analyses by the Chou-Fasman (CF) algorithm (1973), by the Robson (RG) method (Garnier et. al., 1978), and by an integrated version of both methods (CFRG) are shown in Fig. 3. 13. An evaluation of the Chou-Fasman and Robson-Garnier algorithms suggests that the methods are 55% and 60%, respectively, accurate in predicting secondary structure (Taylor, 1987). The integrated version of the methods (CFRG) is an attempt to increase the accuracy of these predictions. The stippled regions indicate the propensity of the included amino acids to form particular secondary structures. Since both algorithms were developed from data bases of soluble, globular proteins, it was considered that analyses of only the hydrophilic regions of the polypeptide (Fig. 3.12) would be reasonable.

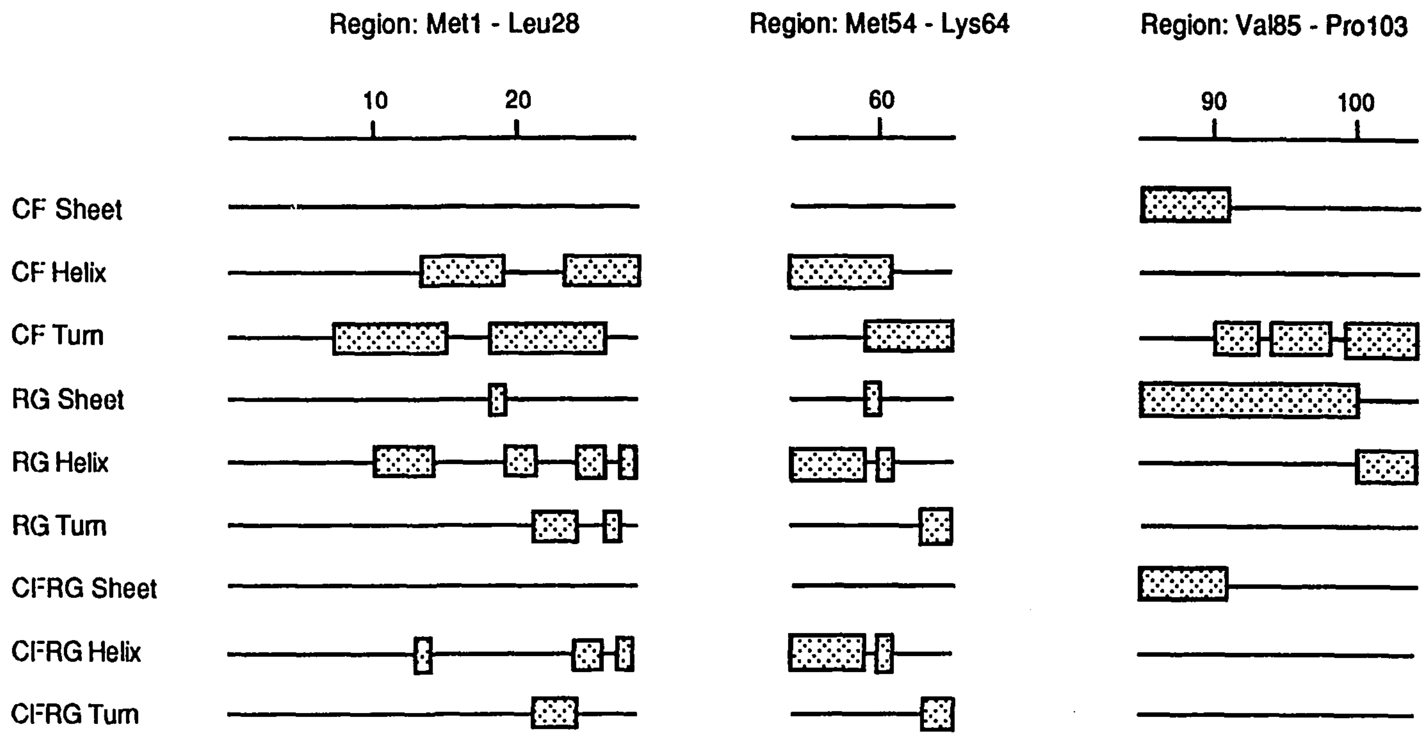


Fig. 3. 13 Secondary structure analyses of hydrophilic regions Met $_{1}$- Leu $_{28}$, Met $_{54}$ - Lys $_{64}$, and Val $_{85}$ - Pro $_{103}$. CF - Chou-Fasman method; RG - Robson-Garnier method; CFRG - the integrated version of the Chou-Fasman and Robson-Garnier algorithms. The regions enclosed by the stippled boxes indicate the propensity of the incuded amino acid residues to form the particular secondary structures.

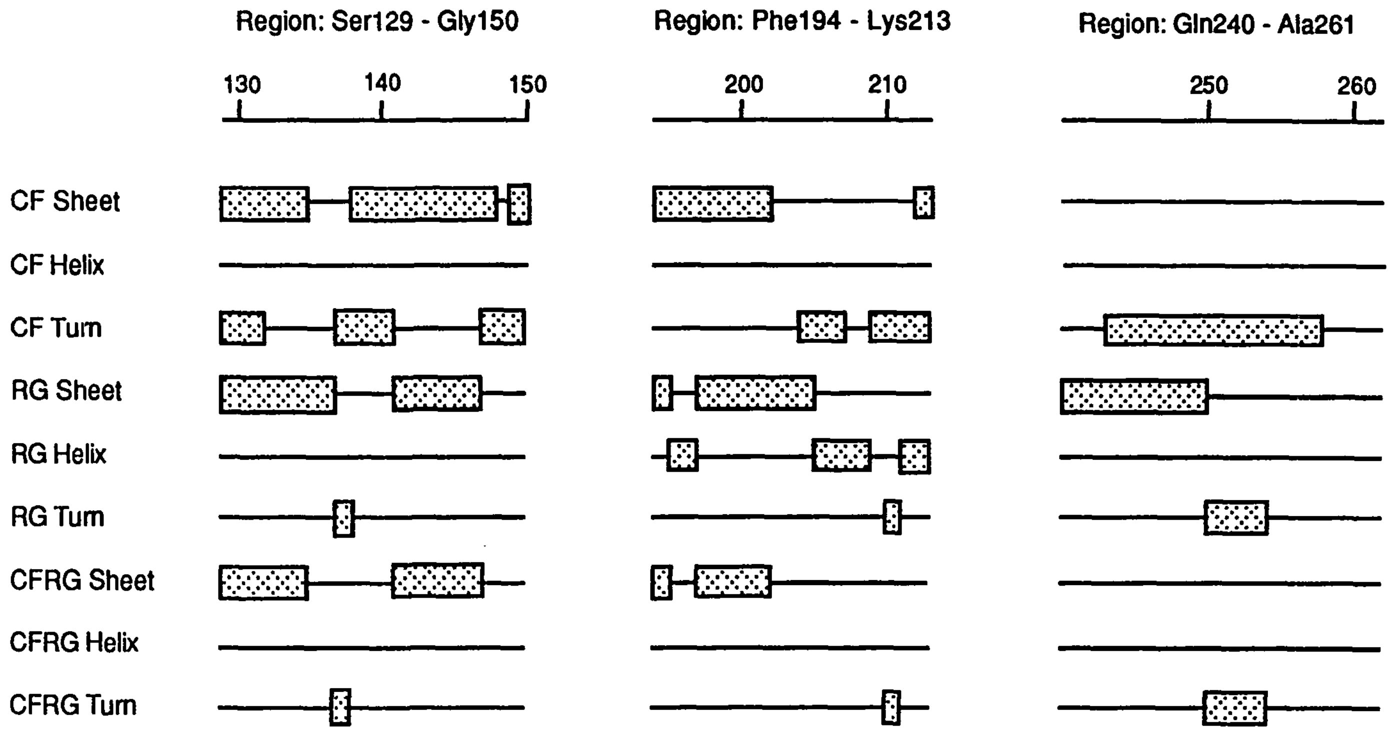


Fig. 3. 14 Secondary structure analyses of hydrophilic regions Ser $_{129}$ - Gly $_{150}$, Phe $_{194}$ - Lys $_{213}$, and Gln $_{240}$ - Ala $_{261}$. CF - Chou-Fasman method; RG - Robson-Garnier method; CFRG - the integrated version of the Chou-Fasman and Robson-Garnier algorithms. The regions enclosed by the stippled boxes indicate the propensity of the incuded amino acid residues to form the particular secondary structures.

Amino Acid consensus and Signal sequence analysis

Amino Acid consensus sequence analysis

Homology searches for consensus amino acid sequences were performed on the predicted *mtr* amino acid sequence. A summary of the amino acid consensus sequences which were investigated is shown in Table 3. 17. The primary source of the amino acid consensus sequences was Doolittle (1986).

The underlined amino acid residues were used as initial search sequences. Only six of the seventeen search sequences were found in the predicted *mtr* amino acid sequence. It is interesting to note the lack of homology with the ATP or nucleotide binding site consensus sequence in this initial and non-stringent search. The six search sequences included that of sulfhydryl proteases, hydroxy kinases, lactamases, substrate of cAMP protein kinases, Asn-linked carbohydrate, and RNP motif. Four of the six initial sequence homologies lacked the amino acids needed to conform with the particular consensus sequence. However, one of these four sequences (K S Q G K K) resembled the sequence substrate of cAMP protein kinases. The sequence R/K X G R/K R/K is predicted to form a ß-bend (Maiden et. al., 1987). Two of the initial sequence homologies, lactamases and RNP octamer motifs, contained additional amino acid residues which conformed to some degree to the consensus sequences. Two sequence homologies were found for the lactamase consensus sequence and are found at S_{61} to K_{64} and at S_{209} to K_{212}. The sequence homology to the RNP octamer motif consensus sequence is found at V_{84} to P_{91} and occurs in the hydrophilic region between hydrophobic regions 2 and 3 (Fig. 3. 10).

Table 3.17 Amino Acid Consensus Sequences

Enzyme Active Sites	Sequence
Common serine proteases	GDSGG
Subtilisin-type serine proteases	TMSA
Sulfhydryl proteases	G X C W/Y
Acid proteases	U D T G
Hydroxy-kinases	D F G $(X)_n$ A P E
Guanido-kinases	T C P $(X)_n$ N L G T
Phosphoglyceraldehyde dehydrogenase	CTTNC
Triose phosphate isomerase	U A Y E P
Phosphoglucomutase	TASHD
Phospholipase A	CCXXHD
Lactamases	SXXK
Substrate sequence of cAMP protein kinase *	R/K X G R/K R/K
Binding sites	
Asn-linked carbohydrate	N X S/T
ATP or nucleotide binding	G/A $(X)_{3-4}$ G K S/T
Cell adhesion	RGD
Cytochrome c thioester	CXXCH
RNP octamer motif @	ba G ar A/G F al po F

* from Maiden et. al., 1987

@ from Query et. al., 1989

Table 3.18

Comparison of the *mtr* octamer sequence

homology to the consensus RNP octamer motif

Yeast PABP#4	K	G	F	G	F	V	C	F
hnRNP-A1/HDP/UP1#2	R	G	F	A	F	V	T	F
Human La	K	G	S	I	F	V	V	E
T4 gp32	K	V	F	K	Y	R	F	G
mtr	V	V	Y	A	F	V	G	P
Conserved sequence	ba 24	G 22	ar 19	A/G 23	F 24	al 26	po 1	F 17
	24 2 al 1 ac	22 2 al 2 ac 1 ba	19 5 al 3 po	26 1 ba	27 -	26 1 ba	23 3 al 1 ar	23 2 po 1 ac 1 al

No. of amino acid differences	0	1	2	3	4	6
No. of proteins	2	12	7	4	1	1

ba - basic, al - aliphatic, ar - aromatic, ac - acidic, po - polar

The homology to the RNP octamer motif is not 100% percent. A comparison of the *mtr* amino acid sequence to a condensed version of the amino acid sequences which form the data base for the RNP octamer motif is presented in Table 3. 18 (Query et. al., 1989). The frequency and type of amino acid residue at each position of the octamer is tabulated. The number of amino acid differences in the motif per protein is recorded against the number of proteins which exhibit those differences. Two proteins which display the least number of differences (0) to the consensus sequence are the yeast PABP#4 and the hnRNP-A1/HDP/UP1#2. The human La and T4 gp32 proteins exhibit 4 and 6 amino acid differences to the consensus sequence, respectively. The *mtr* amino acid sequence is in agreement with the consensus sequence at five consecutive amino acid residues; three of the five residues are highly conserved. The finding of homology to a five amino acid sequence is significant. This significance is reinforced by the lack of homology with "shorter" amino acid sequences such as those shown in Table 3. 17.

Signal sequence analysis

Signal sequences are found at the N-terminal of most secreted proteins. These sequences are recognized by the signal recognition particle (SRP) and are cleaved off by a leader peptidase while the precursor protein crosses the membrane (Walter and Blobel, 1982). Unlike the analyses for amino acid homologies, the search for a possible signal sequence in the predicted *mtr* amino acid sequence was performed manually. The lack of a consensus amongst the numerous signal sequences predetermined this manual analysis. However, it has been shown that signal sequences share structural similarities. A "typical" eukaryotic signal sequence consists of 18-20 amino acids at the N-terminal of a secreted protein. A signal sequence consists of a positively charged

N-terminal (n) of 6 amino acids, a central hydrophobic region (h) of 8 residues, and a more polar C-terminal (c) of about 5 amino acids (von Hiejne, 1985).

The analysis of the 18-20 amino acids at the N-terminal of the proposed *mtr* polypeptide reveals the absence of positively charged amino acids (Arg, His, and Lys) in the first 6 residues. As confirmed by the hydropathy analysis (Fig. 3. 12), there is a lack of highly hydrophobic residues in the following 8 residues. There is no clear evidence of polar residues which would comprise a "typical" C-terminal of a signal sequence. These results suggest that the proposed *mtr* polypeptide does not possess a signal sequence.

Northern Blot and S1 Nuclease Analysis

Northern Blot

In a preliminary Northern blot analysis, 5.0 µg of total RNA from a wild type strain (FGSC #987)and from four *mtr* mutant strains [alleles flr354(s); flr271(s); 21(r); and 6(r)] were electrophoresed and blotted onto nitrocellulose paper. The blot was probed with P^{32}-labelled plasmid DNA from clone pN807. The probe hybridized to a ~ 2.3 kbp fragment in all five samples (data not shown). This finding was unexpected because of the above DNA sequence analyses which indicated that the *mtr* ORF is 845 bp with a short 59 bp intron.

The possibility of comigration of the mRNA with the rRNA molecules in the total RNA preparation was addressed by the enrichment of poly A^+ RNA samples. The above Northern blot experiment was repeated using poly A^+ RNA samples from the wild type strain and from the mtr-6(r) mutant strain. Again, the pN807 probe hybridized to a ~2.3 kbp mRNA transcript (Fig. 3. 15).

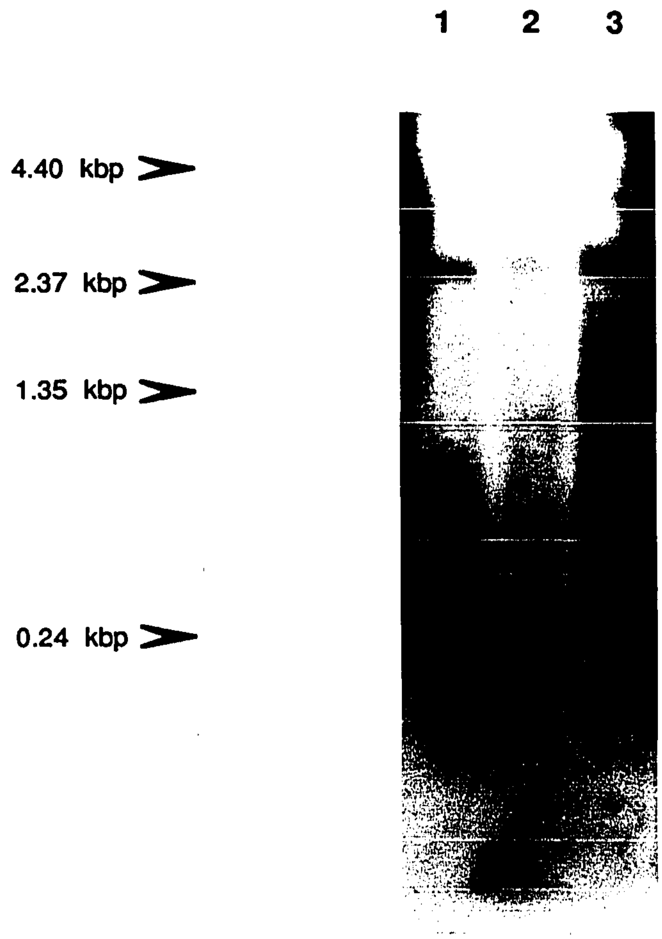


Fig. 3. 15 Northern blot analysis of poly A+ mRNA. Lane 1 - 1.0 µg poly A+ mRNA from wild type strain 74-OR23-1A, lane 2 - 1.0 µg poly A+ mRNA from mtr-6(r) strain, and lane 3 - 5.0 µg poly A+ mRNA from mtr-6(r) strain. The blot was probed with P^{32}-labelled DNA from plasmid pN807. The molecular weights (kbp) of the RNA markers are shown on the left. The probe hybridizes to a transcript of ~2.3 kbp.

S1 Nuclease Mapping

The *mtr* mRNA transcript was mapped by S1 nuclease analysis. The 2.9 kbp DNA fragment was purified from clone pBN3, treated with T4 DNA ligase, and 1.0 ng of the fragment was hybridized to 50 µg total RNA from the wild type and mtr-6(r) strains. The samples were digested with S1 nuclease, precipitated and electrophoresed under alkaline denaturing conditions. S1 nuclease preferentially digests single-stranded DNA and RNA regions; the non-hybridized region of a DNA/RNA hybrid which represents an intron is a suitable substrate for the S1 nuclease.

The samples were blotted onto nitrocellulose paper and probed with P^{32} -labelled pN807 DNA (Fig. 3.16). The primary DNA fragments of the DNA/RNA hybrids which were protected are ~ 1.8 kbp and ~0.5 kbp. These results confirm the Northern blot analyses which indicate that the *mtr* mRNA transcript is ~2.3 kbp. The S1 nuclease data also indicate the presence of a small intron and confirm the DNA sequence analysis that predicted a 59 bp intron in the *mtr* ORF.

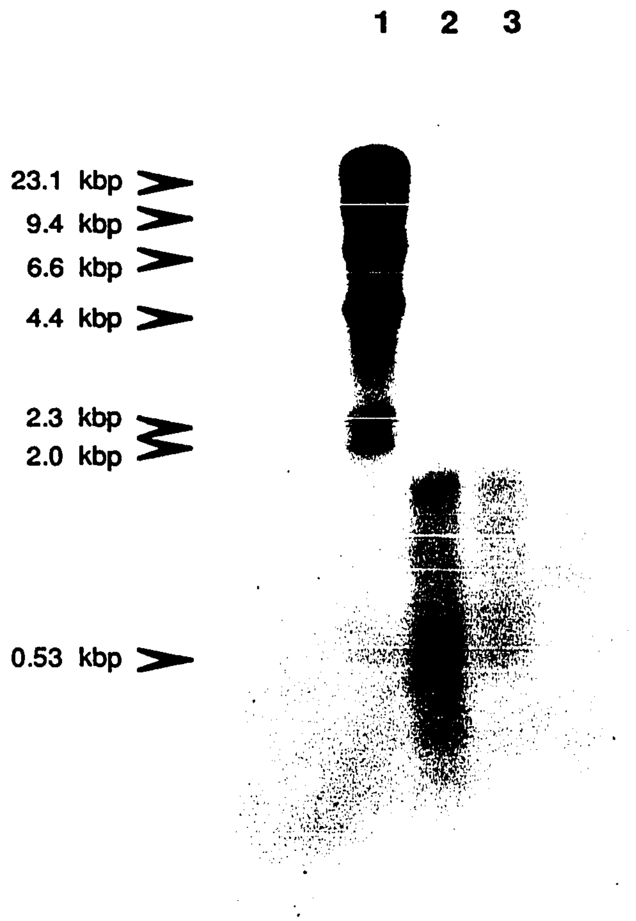


Fig. 3. 16 S1 nuclease mapping analysis of mRNA transcript. Lane 1 - 100 pg Hind III digested lambda DNA, lane 2 - 1.0 ng of 2.9 kbp fragment hybridized to 50 μg total cellular RNA from wild type strain 74-OR23-1A and digested with S1 nuclease, lane 3 - same as in lane 2 except for the substitution for total cellular RNA from the mutant mtr-6(r) strain. The blot was probed with P^{32}-labelled DNA from plasmid pN807. The molecular weights of the lambda Hind III DNA markers are shown on the left. The probe hybridizes to S1 resistant DNA fragments of ~1.8 kbp and ~0.5 kbp.

CHAPTER FOUR

DISCUSSION

The specific goal of the present research effort is the cloning and determination of the structure and sequence of the *mtr* gene of Neurospora crassa. The gene product of the *mtr* locus participates in the N (neutral aliphatic and aromatic amino acids) transport system in N. crassa (Perkins et. al., 1982). Mutants of the *mtr* locus have been selected for resistance to the poisonous amino acid analog 4-methyl-tryptophan, 4-MT, (Stadler, 1966; Lester, 1966). Through the determination of the DNA sequence, the amino acid sequence of the *mtr* polypeptide can be predicted. A possible secondary structure of the *mtr* polypeptide can be proposed based upon hydrophobicity and secondary structure analyses of the polypeptide. The determination of the secondary and tertiary structures of the *mtr* protein is critical to an understanding of the molecular mechanisms by which amino acids are transported across the cell membrane.

Transformation of Neurospora crassa

Neurospora crassa is employed as a model organism to address many fundamental questions in biology. General procedures to investigate the genetics, biochemistry, and physiology of this "pink bread mold" are well documented (Davis and DeSerres, 1970). Several hundred loci have been mapped to all seven linkage groups with their relative map distances and gene order determined. The genome complexity of N. crassa has been estimated to be 2.7×10^7 bp (Krumlauf and Marzluf, 1979).

Despite the wealth of information, advances in molecular investigations of the biology of N. crassa have previously been limited to a few genes. Various approaches have been used to successfully clone genes of N. crassa. These strategies include complementation of E. coli mutants (Vapnek et. al., 1977; Dunn-Coleman, 1984; Legerton and Yanofsky, 1985), hybrid selection (Arends and Sebald, 1984; Berlin and Yanofsky, 1985), and screening of libraries with synthetic DNA probes (Munger et. al., 1985). However, these approaches have inherent limitations.

A general method of cloning genes in N. crassa is the transformation and complementation of N. crassa mutants. Such methods, originally devised to clone genes in E. coli, have been extended to other organisms such as yeast (Struhl, 1983), Drosophila and humans. The intent of these technically varied methods is the introduction of DNA into the organism and the selection of transformants. From a genetic viewpoint, the devised selection scheme is paramount; it provides for a ready evaluation of the DNA introduction methodology and more importantly it furnishes the framework that the DNA molecule contains the gene of interest. Unlike yeast, transformation of N. crassa generally results in non-homologous integration of the transforming sequences (Case, 1982; Akins and Lambowitz, 1985).

The results of the transformation experiments with 5:3B and 5:4H indicate that these clones contain sequences which suppress the mtr-6(r) mutation (Table 3.3). The mtr-6(r) mutation has been hypothesized to be either a small intragenic deletion or inversion. These cloned sequences can be allelic to the mtr-6(r) mutation and suppress the mutation by complementation or non-allelic and suppress the mutation by some unknown manner. Thus, a viable interpretation of the transformation results is that the clones contain sequences which are allelic to the mtr-6(r) mutation. Since transformation of N. crassa generally results in non-homologous integration of the DNA, the high level of transformation exhibited by cosmid 5:4H suggests that the sequences are not only allelic but may also represent the entire *mtr* gene.

RFLP Mapping

The DNAs of clones 5:3B and 5:4H suppress the mtr-6(r) mutation. The results of the RFLP mapping demonstrate that both clones carry inserts which are linked to DNA markers flanking the *mtr* locus (Figs. 3.3 and 3.4). Together, the results of the transformation and RFLP mapping experiments suggested that clones 5:3B and 5:4H contain sequences allelic to the mtr-6(r) and probably encompass the intact *mtr* gene.

The RFLPs detected by the 5:3B do not appear to be identical to those detected by 5:4H when the two autoradiograms are superimposed on each other. The lack of homology between the inserts of 5:3B and 5:4H was detected at high stringency levels when examined by Southern blot analysis is disturbing. One would expect to observe homology at the DNA level between the clones since both clones rescue the same genetic marker. Presently, a simple genetic interpretation of these results is not obvious.

The subcloning of the *mtr* locus, described in the following section, was performed on the DNA from cosmid 5:4H based primarily on the higher transformation frequency exhibited by 5:4H DNA over that displayed by the DNA of 5:3B (Table 3. 3).

Subcloning the *mtr* locus

Digests of 5:3B and 5:4H DNA with Hind III, Hpa I and Bgl II did not reduce the transformation frequency to 1-10% of that of the undigested samples (Table 3.3). Transformation frequencies of ≤10% suggest that the restriction enzyme digests interrupt or cleave at the coding region of a gene and that the transformants are due to homologous integration events (Vollmer and Yanofsky, 1986). It has been demonstrated that transformation with linear DNA molecules of intact coding sequences exhibit a decrease in transformation frequency when compared to circular DNA molecules (Case, 1982). Hence, the decrease in transformation frequency exhibited by the Hind III, Hpa I, and Bgl II digests appears to agree with this observation. The results of these experiments suggested that digests with Hind III, Hpa I and Bgl II did not interrupt the *mtr* gene.

The 15.0 kbp fragment from the Bgl II digest of 5:4H was found to rescue the *mtr* marker and was also able to transform the host strain to benomyl resistance without the co-transformation of pSV50 DNA (Table 3.4). The fragment was recircularized at the Bgl II sites to generate pBN3, benomyl resistant and neutral (N) transport system (Fig. 3.5). Thus, the Bgl II sites at the ends of the 15.0 kbp fragment originate from the interior of the 5:4H insert and are the most proximal Bgl II sites adjacent to the polylinker site of pSV50.

The plasmid pBN3 contains a ~ 7.0 kbp insert which encompasses the ends of 5:4H, X and Y, which are flanked by the Bgl II sites; a restriction map of pBN3 is shown in Fig. 3.6. This configuration suggested that either the X or Y end contain the *mtr* gene. Both ends were released intact as Bgl II/ Eco RI fragments; the Eco RI sites originate from the polylinker site of the pSV50 vector. The fragments were subcloned and the DNAs were used to transform the quadruple mutant host strain. The results of this experiment (see Table 3.6) indicated that the *mtr* gene was contained within a 2.9 kbp fragment, the Y end, which was cloned as pCVN2.9.

The transformation frequency exhibited by pCVN2.9 is comparable to that displayed by pBN3 (Table 3.8). However, the transformation frequencies of pCVN2.9 and pBN3 is about 40% that of 5:4H. These results are in agreement with the observation that plasmids which carry large inserts of an intact gene transform more efficiently than plasmids which harbor smaller inserts of that gene (Case, 1982).

A restriction enzyme map of the 2.9 kbp fragment was determined and is shown in Fig. 3.7. The decrease in transformation frequency displayed by the Eco RV (5%) and Cla I (6%) digests of pCVN2.9 and the proximity of these sites in the restriction map suggested that the *mtr* coding region was localized at the left hand side of the map. Digests with Sph I, Acc I, and Xho I would have been expected to exhibit a similar decrease in transformation frequencies due to the nearness of these sites to the Cla I and Eco RV sites. Hence, the totality of the transformation and restriction enzyme mapping results of pCVN2.9 could not be readily interpreted to localize the *mtr* coding region within the 2.9 kbp fragment.

Resubcloning, Nested deletion subcloning, and Sequencing

Results of the transformation and RFLP mapping experiments demonstrated that the 5:4H cosmid clone carried a genomic insert which encompassed the intact *mtr* gene. The *mtr* gene was subcloned from 5:4H into an intermediate clone, pBN3, and was further subcloned as a 2.9 kbp fragment in pCVN2.9.

At the time of generating pCVN2.9, the query of which end, X or Y of pBN3 (Fig. 3.6), contained the *mtr* gene was best addressed by releasing the ends intact with a Bgl II/ Eco RI double digestion. However, the construct of pCVN2.9 has two drawbacks which do not readily allow for the eventual sequencing of the 2.9 kbp fragment. First, pCVN2.9 is a directional clone and as such only one strand can be sequenced. One approach to circumvent this obstacle would have been to clone the Bgl II/Eco RI fragment into the "sister" vector of pTZ18R, pTZ19R, which has the multiple cloning site (MCS) in reverse orientation to the lac Z gene as in pTZ18R. Second, pCVN2.9 as a replacement construct resulted in the removal of restriction sites within the MCS. These sites were needed for an effective use of the Exo III/ S1 unidirectional subcloning technique in generating nested deletion subclones (Henikoff,1984).

The 2.9 kbp fragment was recloned as a Bgl II/ Bam HI fragment from pBN3 to generate pN807, pN816, pN839 and pN846. This strategy results in the cloning of the fragment in two orientations, unlike pCVN2.9, and thus enables the sequencing of both strands. The resulting constructs are insertional clones with the fragment being bordered with restriction sites which can utilized in the Exo III/S1 nested deletion protocol. Clones pN807 and pN846 were used to generate nested deletion subclones (Tables 3.9 and 3.10) and various restriction site subclones (see Results section). Some of these clones were employed in a sequencing strategy to sequence the 2.9 kbp fragment (Fig. 3.8). Only the sequence of 2,060 bp is shown in Fig. 3.10. The 5' end of the sequence corresponds to the left hand side of the restriction map.

DNA Sequence Analysis

The *mtr* gene was cloned as a 2.9 kbp fragment DNA and numerous subclones were generated to sequence the entire fragment. Sequence analyses were performed on the entire sequence of 2,783 bp which represents the 2.9 kbp fragment. Theoretically, the finding of a long ORF suggests the possible coding region of a particular gene. However, in consideration of the ease by which sequence can be misread, this approach is difficult to implement in practice. Another approach of finding a possible coding region of a gene is the examination for homologies to consensus sequences. This approach is less likely to be affected by misreading of the sequence. The main drawback with this method is that the gene of interest may lack significant homology to the consensus sequences and thus be overlooked. However, the homology search can be established at varying stringency levels and the detected homologies can then be evaluated. To examine the sequence of the 2.9 kbp fragment for a proposed *mtr* open reading frame (ORF), the sequence of both strands was initially analyzed for homologies to DNA consensus sequences.

The consensus sequence at the N. crassa ATG translation start codon is A T C A C/A C/A A T G; some N. crassa genes which have this sequence are shown in Table 3. 11. The consensus sequences that typify introns of N. crassa are: G T A A/C G T (at the 5' end), A/G C T A/G A C T/A (within the intron), T/C A G (at the 3' end). Some introns of N. crassa genes which display these sequences are shown in Table 3.13. Other sequences commonly found in eukaryotic genes include the capping site (C/T C A T T C A/G) (cited in Legerton and Yanofsky, 1985), the "TATA" box (T A T A T/A A T/A), the "CCAAT" box (G G C/T C A A T), and the polyadenylation signal (A A T A A A) (Lewin, 1987). The search for the N. crassa consensus sequence at the ATG start site detected only one homologous sequence with 100% identity (Table 3.12). This sequence, A T C A C C A T G, starts at position -5 in strand #1 (Fig. 3.10). The detection of this homology is significant when one considers that the random probability of finding a particular 9 bp sequence is 1 in 4^9 or 1 in 262,144.

The analysis for the consensus sequence at the 5' end of N. crassa introns revealed three homologous sequences. One of the three homologies (G T A A G T) occurs in strand #1 at position +727 (Table 3.14). Two sequences were found to be homologous to the internal intron consensus sequence; one of the two homologies occurs in strand #1 at position +769. Both sequences were followed downstream, with an acceptable number of base pairs, by sequences homologous to the 3' end consensus. The homology to the internal consensus sequence found in strand #1 was readily aligned to the 5' splice junction sequence at position +727. These results suggest the presence of 59 bp intron in strand #1 (Fig. 3. 10).

The homologies to the consensus sequences at the N. crassa ATG start site and to the sequences that exemplify introns of N. crassa genes strongly suggest that the *mtr* ORF is 845 bp with a 59 bp intron near the 3' end of the gene (Fig. 3.10). The finding of a long open reading frame which includes the above homologous sequences supports this proposal (Fig. 3.11). The observed homologies of the *mtr* ORF intron suggest that these sequences are probably recognized by a spliceosome and participate in the formation of a lariat intermediate in cis- splicing (Lewin, 1987).

Similar homology searches of other consensus sequences were performed for the proposed *mtr* ORF. No sequence homologies to the "TATA" and "CCAAT" boxes were detected within 200 bp upstream from the translation start site. These sequences are commonly found in the promoter regions of many eukaryotic genes. However, the lack of homology to these "boxes" have been reported for other genes of N. crassa which have been cloned (Legerton and Yanofsky, 1985; Schechtman and Yanofsky, 1983). Thus, it would appear that other possible consensus sequences are used in the promoter regions of N. crassa genes.

Four sequences, at positions -97, -79, -52, and -13, were found to resemble the capping site consensus sequence. The former two sequences do not coincide with the consensus sequence at the last 3' nucleotide position. The latter two sequences, which differ from the capping site sequence at two bases, are identical to that found in the his-2 gene of N. crassa (Legerton and Yanofsky, 1985). From the present data, it is difficult to state which of these sequences represent the capping site of the *mtr* ORF.

A possible approach to determine the capping site of the *mtr* ORF is to perform primer extension analyses of the mRNA transcript (Ausubel et. al., 1987). In this approach, a DNA primer or fragment which complements the 5' end of the mRNA is prepared. The primer is end-labelled, annealed to the mRNA, and is then extended with cold dNTPs using a reverse transcriptase. The reaction product is resolved in a sequencing gel alongside a sample of a sequencing reaction using the identical primer. The migration of the reverse transcribed product should coincide with one of the bands of the sequencing reaction (Ausubel et. al., 1987). The results of these experiments should reveal the last base(s) of the transcript at the 5' end. The capping site would be expected to be 1-2 bp upstream from the 5' end of the mRNA (Lewin, 1987).

The sequence A A T T A A A is found at 49 bp downstream from the translation stop site at position 894. This sequence, although not identical to the consensus poly-adenylation signal, is identical to the proposed poly A signal of the mammalian multidrug resistance (mdr) gene (Gros et. al., 1986). One approach to determine the site to which the poly A tail is added is to clone and sequence a cDNA of the *mtr* mRNA transcript and compare it to the genomic sequence shown in Fig. 3.10.

The possible sequence homologies to the commonly found transcriptional signals and the approaches to investigate their biological significance should be addressed with caution. The reason for this caveat is the finding that the 2.9 kbp fragment hybridized to a poly A+ transcript of ~2,300 nts in a Northern blot (Fig. 3. 15). From the above DNA sequence analysis, it was expected that the *mtr* mRNA would be ~900-1,200 nts. This estimate is based on the spectrum of the "typical" lengths of 5' nontranslated leader sequences (36-64 nts), 3' nontranslated region (53-631 nts), and poly A tails (40-200 nts) (Lewin, 1980) added to the 786 bp (845 bp ORF minus the 59 bp intron) of the *mtr* translated sequence.

Two possible genetic interpretations can be proposed from the above results. The simplest interpretation is that the 5' and 3' nontranslated regions and poly A tail of the *mtr* mRNA transcript are much longer than those normally observed in other genes. The other possibility is that a trans-splicing event, versus cis-splicing which joins exons, occurred to give rise to a mature transcript. Trans-splicing has been observed in the mRNA processing events in trypanosomes (Laird, 1989). Trans-splicing involves the joining of two different mRNA molecules, coded by distinct and separate exon sequences, to form a mature mRNA.

To address these two interpretations and to confirm the presence of the proposed 59 bp intron, a S1 nuclease mapping analysis of the mRNA was performed (Fig. 3.16). Two primary DNA fragments, ~1,800 nts and 500 nts, were protected from the S1 nuclease. The simplest explanation of the results is that the 1,800 nt fragment represents the translated sequence of exon 1 (726 bp) and 5' nontranslated sequences. Likewise, the 500 nt fragment represents the exon 2 sequence (60 bp) and 3' non-translated sequences. These results also confirm the Northern blot analysis in that the 2.9 kbp fragment hybridizes to a ~2,300 nt mRNA. The data also support the proposal of the short intron in the transcript. However, the possibility of a trans-splicing event is not supported by these results; one would have expected to observe the hybridization to fragments of lower molecular weight. Hence, the most plausible interpretation is that the 5' and 3' nontranslated regions of the mRNA are longer than expected.

It has been observed that some *mtr* mutations have been mapped to regions that extend from position -743 to -404 (Dr. David R. Stadler, personal communication). This observation is unexpected with regards to the proposal of a 845 bp *mtr* ORF. However, the mutant mapping analyses appear to correlate well with the mRNA analyses. To address this discrepancy, an interesting speculation is the possibility of an eukaryotic polycistronic mRNA transcript. This idea is contrary to the popular view that eukaryotic mRNAs are monocistronic and polycistronic mRNAs are only observed in prokaryotes. The observations supporting the possibility of a *mtr* polycistronic mRNA are: (1) the 5' nontranslated mRNA sequence (~1,000 nt) is substantially longer than what is expected (36-64 nts) (Lewin, 1980), (2) mutations which affect the function of *mtr* have been mapped to regions substantially upstream from the proposed *mtr* ORF, and (3) an upstream open reading frame (uORF) has been identified.

It may be argued that the first two observations can be explained by the novelty of the *mtr* gene having longer than expected 5' "nontranslated" sequences and that perturbations of upstream regulatory sequences result in the disruption of *mtr* gene expression. However, an uORF was discovered in the same reading frame (reading frame 2 of strand #1, Fig. 3. 11) as that as the proposed *mtr* ORF. The sequence at the ATG start site of this uORF, at position -630, possesses some homology to the N. crassa ATG consensus sequence (Table 3. 12) and the uORF stops at position -453. This uORF (174 bp long) could possibly code for a hydrophilic polypeptide of 57 amino acids. Presently, it is not known whether this uORF is translated or what function it may possess in relation to the proposed *mtr* ORF. Translational read-through of the sequences between the uORF and the *mtr* ORF results in an additional deduced polypeptide sequence of 151 amino acid residues. Similar findings of uORFs were reported for the mRNA of the N. crassa arg-2 gene (Sachs, 1989) which resembles an uORF in the homologous S. cerevisiae gene, and for the mRNA of the cpc-1 gene (Plamann and Yanofsky, 1989) of N. crassa. The mRNA transcript of the cpc-1 gene contains two short ORFs in the 5' leader region which are required for cpc-1 function. The ubiA locus of C. elegans also codes for a polycistronic mRNA which contain 11 tandem repeats of ubiquitin sequence (Graham et. al., 1989).

Amino Acid Analysis

The finding of a long open reading frame, and significant sequence homologies to the known consensus sequences of N. crassa genes and to other sequences commonly found in eukaryotic genes support the proposal that the *mtr* ORF is 845 bp with a small 59 bp intron near the 3' end.

The amino acid sequence of the *mtr* ORF was deduced from the DNA sequence. The codon usage of the *mtr* ORF is somewhat biased in that 18 of the 61 sense codons are not used. The bias in codon usage exhibited by the proposed *mtr* ORF is very similar to that previously determined for other genes of N. crassa (Schechtman and Yanofsky, 1983; Legerton and Yanofsky, 1985; and Orbach et. al., 1986) and supports the authenticity of the DNA sequence and deduced amino acid sequence of the proposed *mtr* ORF.

It has been proposed that codon usage bias may reflect a high level of gene expression (Orbach et. al., 1986). This proposal is supported by the absence of 32 codons in the H3 and H4 histone genes (Woudt et. al., 1983), the absence of 22, and 19 codons in the *am* (Kinnaird and Fincham, 1983) and ß-tubulin (Orbach et. al., 1986) genes, respectively. Whereas, the *his-3* (Legerton and Yanofsky, 1985) and *trp-1* (Schechtman and Yanofsky, 1983) genes use 58 and 57 codons, respectively. In consideration of the above proposal, it is tempting to speculate that the *mtr* ORF is expressed at levels comparable to the ß-tubulin gene. As the N transport system is constitutively expressed (Wolfinbarger, 1980), this speculation appears plausible.

The inferred *mtr* polypeptide (261 amino acids) has a molecular weight of 28,613 daltons. The polypeptide contains 164 non-polar amino acid residues which represent 63% of the amino acids. The hydropathy analysis (Fig. 3.12) indicates the presence of six highly hydrophobic regions. The high content of non-polar amino acids suggests that the *mtr* polypeptide is an integral membrane protein. This suggestion is reinforced by the fact that the average length of these six hydrophobic regions is 23 amino acids. This average number of hydrophobic amino acid residues is highly indicative of membrane spanning domains (Doolittle, 1986). Hence, it appears that the *mtr* polypeptide crosses or is imbedded in the lipid bilayer at six regions along its length.

Hydropathy analysis of the additional amino acid residues, due to possible translational read-through of the sequences between the uORF and the mtr ORF, indicates that this additional polypeptide sequence is composed of a hydrophilic N-terminal region and two distinct hydrophobic regions. Each of the additional hydrophobic regions is composed of ~20 amino acids. Presently, it is not known whether the opal codon (TGA) between the uORF and the *mtr* ORF is read-through or suppressed. The distinct possibility that an error in sequencing of that region (sequenced in one strand, see Appendix) can best be addressed by sequencing the second strand.

The proposed *mtr* polypeptide lacks a distinguishable N-terminal signal sequence. Several other integral membrane permease proteins also exhibit a similar absence of a signal sequence. These proteins include the S. cerevisiae arginine permease (Ahmad and Bussey, 1986), the S. cerevisiae histidine permease (Tanaka and Fink, 1985), the E. coli lactose permease (Buchel et. al., 1980), the human and murine glucose transporters (Muekler et. al., 1985; Birnbaum et. al., 1986), and the his M, his P, and his Q proteins of the high affinity histidine permease system in S. typhimurium (Higgins et. al., 1982).

Presently, the mechanism(s) by which these integral membrane proteins are targeted to the membrane is unknown. The sequence of membrane spanning domains and their secondary structures are probably prominent features to be considered in designing experiments and models to address this exciting question. An undefined apparatus which may participate in the localization process is a possibility that cannot be excluded.

The *mtr* polypeptide has been proposed to be a glycoprotein similar to that found for the gene product of the pmb locus (Stuart and DeBusk, 1971; Stuart and DeBusk, 1975). No amino sugars were detected in the pmb gene product; this suggests that commonly found N-glycosidic linkages do not occur in this glycoprotein. However, it has been proposed that there are a maximum of 17 sites (serine and threonine) residues to which neutral sugars can be attached (Stuart and DeBusk, 1975). The proposed *mtr* polypeptide does not contain asparagine-linked carbohydrate binding sequences (Table 3. 17). This finding compares well with the absence of N-glycosidic linkages in the pmb gene product. There are 10 serine, 6 lysine, and 4 threonine residues (Fig. 3. 10) in the hydrophilic regions of the proposed *mtr* polypeptide at which potential O-linked glycosylation can occur. Again, this observation is comparable to the finding of the pmb gene product.

One commonly found source by which energy is coupled to the active transport process is the hydrolysis of the high energy $\beta-\gamma$ phosphate bond of ATP or from some other source e.g. acetylphosphate. Another energy source is the simultaneous transport of an ion (Na^+ or H^+) with the metabolite. It is worthy to note the absence of homology to the nucleotide or ATP binding sequence in the proposed *mtr* polypeptide. This finding suggests that the N transport system is not energized by the hydrolysis of a phosphate bond.

In the absence of an ATP-binding site in the *mtr* polypeptide, the possibility that the N transport system is energized by the coupling of a H^+ ion is plausible. This speculation is supported by an earlier study that demonstrated that the proton uncoupler, sodium azide NaN_3, inhibits the transport activity of the N system (DeBusk and DeBusk, 1965). Since the above metabolic poisoning experiments were performed with intact cells, it is difficult to exclude the possibility that the observed inhibition may be a result of a general disruption of the energy transfer system. Similar experiments using right side out (RSO) vesicles may assist in substantiating the hypothesis that the N transport system is energized by proton coupling. This hypothesis would also suggest that the N transport system is a secondary transport system, as defined by the chemiosmotic hypothesis (Maloney, 1987), in which the solute together with the H^+ is transported in response to the proton electrochemical gradient (Wolfinbarger, 1980). A proton can be co-transported with the metabolite via two possible mechanisms. One method may involve the protonation of the metabolite prior to the uptake of the substrate and the other mechanism may include a concurrent binding and transport of the proton by the permease molecule.

It has been demonstrated that the transport of phenylalanine by the *mtr* (N) system in N. crassa is inhibited by preincubation of conidia with RNase A (Stuart and DeBusk, 1975) and that the arginine-affinity purified material contained significant 260 nm absorbing material (Stuart and DeBusk, 1971). The inhibition of transport was also demonstrated when the RNase was covalently attached to polyacrylamide beads (Stuart and Woodward, 1975); this observation indicated that the RNase enzymatic activity was at the cell surface. These results suggested that RNA molecules (ext RNA) were located on the membrane surface. These RNA molecules are heterogeneous in size (150-380 nts) and possess some secondary structure as indicated by biphasic melting profiles (Stuart and Woodward, 1977). Similar research on phenylalanine-affinity purified material from cell surface extracts also suggested the association of RNA with the purified material (Koo and Stuart, 1984).

The homology in the proposed *mtr* polypeptide to the RNP octamer motif (Query et. al., 1989) supports the contention that the *mtr* polypeptide is associated with a RNA molecule. The finding that the homologous sequence occurs in the hydrophilic region between hydrophobic regions #2 and #3 (Fig. 3. 12) further reinforces this idea. The above finding would be difficult to interpret if the observed homology was located within a hydrophobic domain when one considers the highly charged nature of a RNA molecule . It is also worthy to note that the N-terminal valine residue of the *mtr* octamer sequence (Fig 3. 10) does not coincide with the basic (ba) residue of the consensus RNP octamer motif. Perhaps, the selective force of having a non-polar amino acid residue participate in the formation of a membrane spanning domain contributes to the lack of homology at this amino acid site.

The question as to what role does the RNA molecule(s) have in the amino acid transport process is best considered in light of other reported RNA/protein (RNP) complexes involved in transport. A cytoplasmic 7S RNA molecule, complexed with six polypeptides, was shown to be required for the proper functioning of the signal recognition particle (SRP) (Walter and Blobel, 1982). The interaction of the SRP with microsomal membranes allows for the translation and translocation of nascent secretory proteins across the endoplasmic reticulum membrane. The SRP-RNA, by itself, does not exhibit any translational or translocating activity. Reconstitution experiments with the SRP-RNA and the six different polypeptides suggest that the RNA functions as a structural element (Walter and Blobel, 1983).

It has also been shown that a cytoplasmic RNA moiety, complexed with other protein factors and with mitochondrial matrix precursor proteins, is required for the translocation of the precursor proteins across the outer mitochondria membrane (Firgaira et. al., 1984). Like the signal recognition particle, large amounts of RNase was required to disrupt the function of the RNA/protein complex. This finding suggests that the RNA is highly protected by the protein factors. However, the characterization and biochemical function of the RNA were not determined.

In consideration of the above reports of RNA/protein complexes being necessary for the translocation of proteins across membranes, it is reasonable to propose that the RNA molecule which is associated with the *mtr* polypeptide (Koo and Stuart, 1984) functions as a structural element. One structural function of the RNA will be discussed in the following presentation of a possible tertiary structure of the *mtr* polypeptide.

How does a highly charged RNA molecule cross the hydrophobic lipid bilayer? It is interesting to speculate that the transcribed RNA is physically attached to the *mtr* polypeptide as a ribonucleoprotein (RNP) in the cytoplasm. With the insertion of the protein into the membrane, the RNA crosses the membrane because of the pre-formed attachment of the RNA to the external hydrophilic region of the protein via the RNP octamer motif.

Perspectives and Models

Secondary Structure

A possible secondary structure of the *mtr* polypeptide is presented in Fig. 4.1. In the proposal of a possible secondary structure of the *mtr* polypeptide, several factors were considered. Two commonly used algorithms to determine possible secondary structures of soluble globular proteins are the Chou-Fasman (Chou and Fasman, 1973)

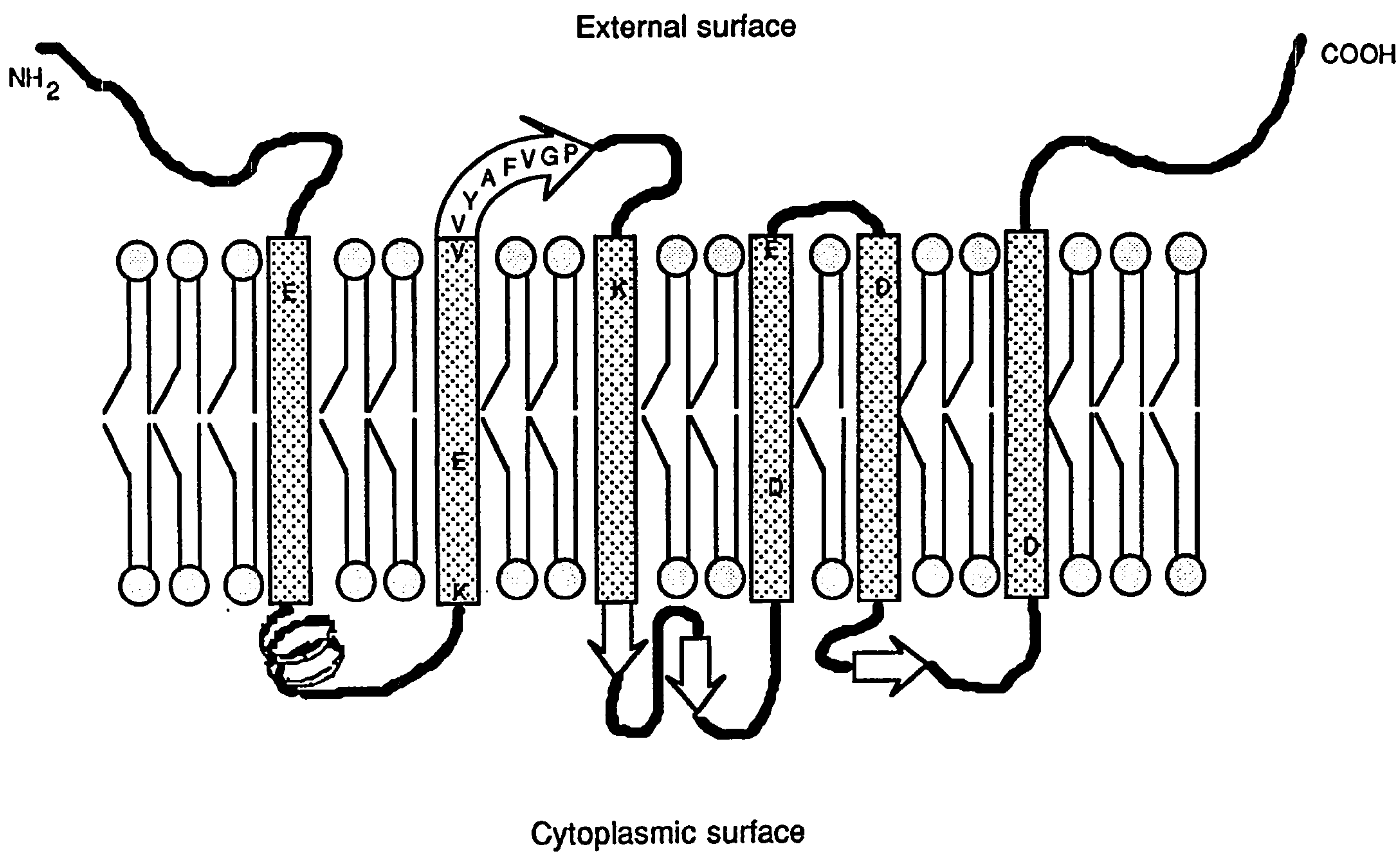


Fig. 4.1 Schematic representation of a proposed secondary structure of the *mtr* polypeptide. The membrane spanning domains are depicted as cylinders and represent alpha helices. The secondary structures of the hydrophilic regions are based on the secondary structure analyses of Figs. 3. 13 and 3. 14. An alpha helix is shown between membrane spanning domains #1 and #2. The arrows represent beta-sheet structures and the solid lines indicate the absence of a distinct secondary structure. The relative positions of charged amino acids (D - aspartic acid, E -glutamic acid, K - lysine) within the membrane spanning domains are indicated. The *mtr* amino acid sequence (V V Y A F V G P) homologous to the RNP octamer motif is positioned on the external surface of the membrane.

and the Robson methods (Garnier et. al., 1978). The Chou-Fasman (CF) and Robson (RG) algorithms were evaluated to be 55% and 60% accurate in the predicting secondary structures (Taylor, 1987). The integrated version (CFRG) of the two algorithms is an attempt to increase the accuracy of the predicted secondary structures (Figs 3.13 and 3.14). The proposed secondary structures of the hydrophilic regions of the *mtr* polypeptide are based solely on the results of the integrated version. However, it must be noted that the secondary structure predictions from the integrated version have a 18% chance (0.4 x 0.45) of being inaccurate. Since these algorithms were based on information of soluble globular proteins, only the hydrophilic regions of the *mtr* polypeptide were analyzed.

Many membrane spanning domains of other proteins have been proposed to have an alpha helical structure (Eisenberg, 1984; Kaback, 1986; and Lodish, 1988). Most of the proposed membrane spanning domains have been based on the distinctive number of amino acids which comprise these hydrophobic regions. Theoretically, the number of amino acids which comprise the membrane spanning domains has been estimated to be 21 amino acids in an alpha helical conformation (Eisenberg, 1984). However, the observed number of amino acids of other membrane spanning domains compares well with the estimate of 24 ± 4 (Kaback, 1986). Based on stereochemical and thermodynamic considerations, it has been proposed that the alpha helix configuration is the preferred form by which proteins are inserted into the membrane (Engelman and Steitz, 1984).

The proposed orientation of the hydrophilic regions, that is which regions are exposed to the external or cytoplasmic environments, is based on the idea that the RNA and the hydrophilic region to which it is attached is exposed to the external surface of the membrane. The proposal that the RNA and this hydrophilic region is exposed to the external surface is based on the earlier research on extrinsic RNA molecules by Stuart (Stuart and DeBusk, 1975; Stuart and Woodward, 1975). The alternate crossing of the hydrophobic regions through the membrane predisposes the orientation of each successive hydrophilic region.

The proposed secondary structure of the *mtr* polypeptide (Fig. 4. 1) is based only on the sequence data, the analyses of that data and inferences from earlier research efforts. Physical data, e.g. circular dichroism, high resolution X-ray diffraction and nuclear magnetic resonance (NMR) spectroscopy, of the *mtr* polypeptide is needed to support or modify this proposal. As a prerequisite to these studies, isolation and purification methods of the *mtr* polypeptide are needed. These experimental activities are not trivial.

Current views of determining protein structure point towards the use of improved 2D-NMR spectroscopy techniques. These techniques do not require the "growing " of crystals of the *mtr* polypeptide which would be needed for X-ray diffraction studies. Such 2D-NMR studies include NOESY (Nuclear Overhauser Effect SpectroscopY) and COSY (COrrelation SpectroscopY) experiments (Bax and Lerner, 1986). NOESY experiments determine "through space" resonances between hydrogens less than or equal to 0.5 nm apart and COSY experiments determine "through bond" resonances between hydrogen less than or equal to three bonds apart (Campbell, 1988). These types of experiments have been applied to short (20-30 amino acids) synthetic hydrophobic oligopeptides (e.g. melittin) which span lipid bilayers. The results of these experiments, performed in polar organic solvents e.g. methanol and trifluoroethanol, indicate that the oligopeptides adopt alpha helical configurations (Campbell, 1988).

Tertiary Structure

A possible tertiary structure of the *mtr* protein is shown in Fig. 4. 2. This proposal is based on the above proposed secondary structure and genetic data which indicate that the *mtr* polypeptide functions as a homopolymeric protein (Larimer and DeBusk, 1977). The genetic data is the complementation between mutant *mtr* alleles which indicate that the *mtr* gene is monocistronic. The structure of the *mtr* protein is proposed to be a dimer. In consideration that many other membrane transport proteins contain 12-24 transmembrane domains (Lodish, 1988), this proposal also seems reasonable. From the present data, the possibility of other polymeric forms of the *mtr* polypeptide cannot be excluded.

A possible structural function of the extrinsic RNA molecule is the binding together of the two *mtr* polypeptides via the RNA molecule. One approach to test this proposal is the use of mobility shift assays (Query et. al., 1989), in which the *mtr* polypeptide is incubated with and without the ext RNA. The samples are electrophoresed in non-denaturing native gels; the expectation is that the RNA (+) sample will be electrophoretically retarded because of the dimerization of the polypeptide and due to the binding of the ext RNA. The monomeric form in the RNA (-) sample would be expected to migrate faster than the dimer.

Conclusions

Mutants of the *mtr* locus of <u>Neurospora crassa</u> have been selected for resistance to the poisonous amino acid analog 4-methyl-tryptophan (4-MT) (Stadler, 1966; Lester, 1966). The gene product of the *mtr* locus participates in the transport of neutral aliphatic and aromatic amino acids across the plasma membrane. Three cosmids

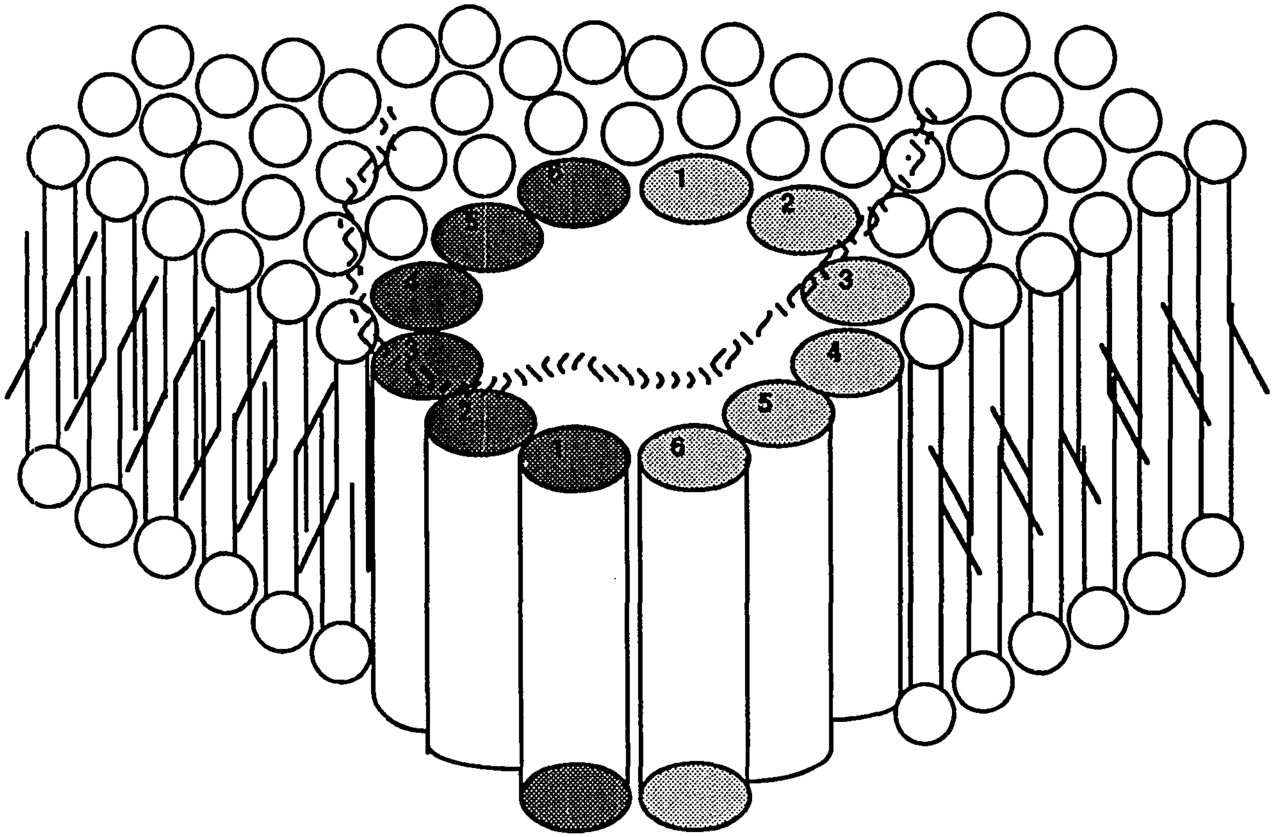


Fig. 4.2 Schematic representation of a proposed three dimensional view of the *mtr* protein. The alpha helices of the membrane spanning domains of the monomeric *mtr* polypeptide are represented by cylinders (1 through 6). The *mtr* protein is proposed to consist of two monomeric subunits. The criss-crossed line represents an extrinsic RNA molecule, attached to the hydrophilic region between membrane spanning domains # 2 and #3, which link the monomers.

were cloned from a genomic N. crassa library which complements the mtr-6(r) mutant allele. Two of the cosmids contains inserts which are tightly linked to restriction fragment length polymorphisms (RFLPs) (Figs. 3. 3 and 3.4) which flank the *mtr* locus on linkage group IV (Stuart et. al., 1988).

The *mtr+* gene was subcloned as a 2.9 kbp fragment and a preliminary restriction enzyme map of the fragment was produced (Fig. 3. 7) (Stuart et. al., 1988). Nested deletion and restriction site subclones were produced and used to sequence the 2.9 kbp fragment (Fig. 3. 10). Sequence analyses indicate the presence of a 845 bp open reading frame (ORF) with a small 59 bp intron near the 3' end of the ORF. The proposed *mtr* ORF contains the sequence A T C A C C A T G which is a 100% homologous to the N. crassa consensus sequence at the ATG start site (Table 3. 11). The intron has sequences which have 100% identity to N. crassa consensus sequences at exon/intron splice junctions and to internal intron sequences (Table 3. 13).

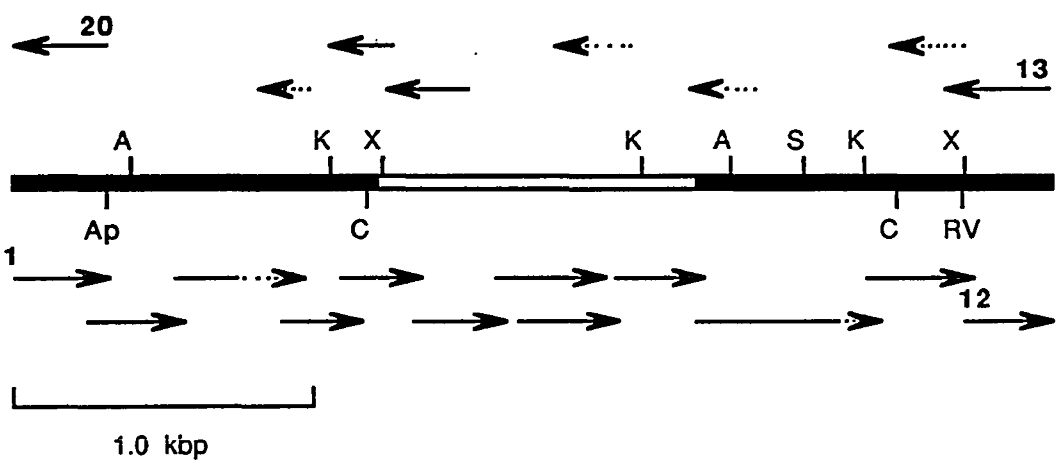
The 2.9 kbp fragment hybridizes to poly A+ mRNA transcript of ~2,300 nt (Fig. 3.15). S1 nuclease analysis of the transcript indicated that two DNA fragments, ~1,800 nt and ~500 nt, were protected (Fig. 3. 16). These results confirm the size of the mRNA transcript and the presence of the intron which was proposed by the sequence analysis. The findings that some *mtr* mutant alleles map far upstream (Dr. David R. Stadler, personal communication) from the proposed *mtr* ORF, that the 5' "non-translated" sequences are longer than expected, and that an ORF is found in the same reading frame as the *mtr* ORF support the proposal that the mRNA transcript contains an upstream open reading frame (uORF) and the *mtr* ORF.

The *mtr* ORF is predicted to translate a 261 amino acid polypeptide (mol. wt. 28,613 daltons). Codon usage of the *mtr* ORF is biased with 18 of 61 sense codons being absent. This bias suggests that the *mtr* ORF may be expressed at levels comparable to the ß-tubulin gene. The polypeptide is composed of 164 non-polar amino acids which represent 63% of the amino acids. Hydropathy analysis of the predicted polypeptide indicates the presence of six hydrophobic transmembrane domains (Fig. 3. 12 and 4. 1) with an average length of 23 amino acid residues. The transmembrane domains are proposed to be in alpha helical conformation (Eisenberg, 1984; Lodish, 1988). The finding of the six transmembrane domains supports the idea that the *mtr* polypeptide functions as the permease molecule in translocating neutral aliphatic and aromatic amino acids across the plasma membrane.

The *mtr* polypeptide does not have a signal sequence; a finding which has previously been observed for many other transport membrane proteins (e.g. Tanaka and Fink, 1985; Mueckler et. al., 1985; and Ames, 1986). Amino acid sequence homology to a RNP octamer motif (Table 3. 18) supports the idea that the *mtr* polypeptide functions as a ribonucleoprotein particle (Koo and Stuart, 1984). A possible secondary structure of the *mtr* polypeptide (Fig. 4.1) is proposed which is based on computer analyses of the hydrophilic regions and on the proposal that the RNP binding site is exposed to the external surface of the membrane. The *mtr* polypeptide is also proposed to function as a dimeric protein (Fig. 4. 2).

APPENDIX

The sequences, obtained from the template of individual deletion subclones and restriction site subclones, are presented in the following pages. The sequence shown in Fig. 3. 10 is based upon the data base of these individual sequences. The sequences are aligned to display the regions of "overlap" between the various templates. The numbers which have been placed at the 5' end of each sequence denote the specific subclone: **1**- pN807, **2**- pNxB/B1-11, **3**- pNxB/B2-11, **4**- pNxB/B4-3, **5**- pNxB/B4-12, **6**- pNxB/B5-4, **7**- pNxB/B7-5, **8**- pNxB/B8-4, **9**- pNxB/B9-12, **10**- pNxB/B10-2, **11**- pNE/K6 10-2, **12**- pNE/E2 10-2, **13**- pN846, **14** - pNE/E4 846, **15** - pNxB1-4, **16**- pNxB2-8, **17**- pNxB8-2, **18**- pNxB10-10, **19**- pNxB3-1, and **20**- pNA/E12 3-1. The sequences derived from the complementary strand (strand #2) are underlined.



```
1  AGATCCGCCTCGCCCCAAGC   GCATCCCAACGCGGCGTGCT   TATATGTCGCTCTTCCCTCT

   CACGTACCTCGCAAGTACCT   GTCTCATCTGGCACCCGCCT   TCTCCATCCCTCTTCTTCAG
       GTACCT G A GTA CT   GTCTCATCT GCACC GCCT   TCTCCATCCCTCTTCTTCAG

   TACTTAATCTGCCCCGGTCC   CCCGCAGTTCATCCTGTCTC   TCAGACCTTGGATCGTCTGT
                                                                2  TGT
   TACTTAATCTGCCCCGGTCC   CCCGCAGTTCATCCTGTCTC   TCAGACCTTGGATCGTCTGT

   TGGTCTTTTCGCTTATTGTC   CGTACCCC ATCCATATTTA   TTCCTGCCTGGGCCCC AGA
   TGGTCTTTTCGCTTATTGTC   CGTACCCCCATCCATATTTA   TTCCTGCCTGGGCCCCCAGA
   TGGTCTTTTCGCTTATTGTC   CGTACCCCCATCCATATTTA   TTCCTGCCT  20
                                                           Apa I

   CAGCATTT CTCTTT
   CAGCATTTTCTCTTTCCCTC   CCTCTTTAGCCATCCAAACA   GCTTGAGAAGCGAAAATCAA

   GCCACTGTCGACAAAGGCCT   TCAACAAGGCCTCTTTACCA   CCCAACATGGACTCGCAATA
       Acc I

   CGAGACAAAAAAGAATGACC   CAAACGCCATCATGCCGTAC   CCAGAGTCAAACGATGAGCA
                                                              3  GAGCA

   TGTTGGCGAGGTCCGCGGCT   TGGGCGGCGGCATCATGGAC   AAGGAGCCTGAGGCCCAGGA
   TGTTGGCGAGGTCCGCGGCT   TGGGCGGCGGCATCATGGAC   AAGGAGCCTGAGGCCCAGGA

   GGGCATGCCAAGTTC AC G   T
   GGGCATGCCAAGTTCCACCG   TCTCGGCTGGAAACGTCTGA   CGGTCGTCCTCATCGTCGAG

   GCCATTGCCCTCGGCTCTCT   CTCGCTTCCCGGCGCCTTCG   CTACCCTTGGCATGGTGCCT

   GGTGTTATTCTCTCTGTCGG   CATGGGACTCATCTGCATCT   ACACGGCTCACGTTATCGGA

   CAAACCAAGCTCAAGCACCC   TGAAATCGCCCACTATGACC   GACGTTGGTCGTGTCATGTT
                                              4   GACGTTGGTCGTGTCATGTT
         AAGCTCAAGCACCC   TGAAATCGCCCACTATG CC   GACGTTGGTCGTGTCATGTT

   TGGAAGATGGG ATATGAAA   TCATCAGCTTCATGTTTGTT   CTGCAACTGATCTTCAT
   TGGAAGATGGGGATATGAAA   TCATCAGCTTCATGTTTGTT   CTGCAACTGATCTTCATCGT
                                                                   CGT
   TGGAAGATGGGGATATGAAA   TCATCAGCTTCATGTTTGTT   CTGCAACTGATCTTCATCGT

   CGGCTCCCACGTCCTCACTG   GCACCATCATGTGGGGCACC   ATCACGGATAACGGCAACGG
   CGGCTCCCACGTCCTCACTG   GCACCATCATGTGGGGCACC   ATCACGGATAACGGCAACGG
   CGGCTCCCACGTCC  19                                            Kpn I
```

```
TACCTGCTCTCTCGTCTTCG   GCATTGTCTCCGCCATCATT   CTCTTaCTCCTTGCCATTCC
                5 CG   GCATTGTCTCCGCCATCATT   CTCTTCCTCCTTGCCATTCC
TACCTGCTCTCTCGTCTTCG   GCATTGTCTCCGCCATCATT   CTCTTCCTCCTTGCCATTCC
                                                   TCCTCCTTGCCATTCC

TCCCAGTTTCGCCGAGGTT    CC
TCCCAGTTTCGCCGAGGTTG   CCATCCTTGGATACATCGAT   TTCGTCTCCATCTGCGCCGC
TCCCAGTTTCGCCGAGGTTG   CCATCCTTGGATACATCGAT   TTCGTCTCCATCTGCGCCGC
TCCCAGTT CGCCGAGGTTG   CCATCCTTGGATACATCGAT   T CGTCTCCATCT CGCCGC
                                 Cla I

CATCCTCATCACC ATG ATT GCT ACT GGC ATT CGC TCG AGC CAC CAG GAG
CATCCTCATCACC ATG ATT GCT ACT GGC ATT CGC TCG AGC CAC CAG G  18
CATCCTCATCACC ATG ATT GCT ACT GGC ATT CGC TCG AGC CAC CAG GAG
                                          Xho I

GGT GGT CTC GCT GCT GTT CCC TGG TCT TGC TGG CCC AAG GAG GAC CTT
GGT GGT CTC GCT GCT GTT CCC TGG TCT TGC TGG CCC AAG GAG GAC CTT

  AGC CTT GCT GAG GGC TTC ATT GCT GTC AGC AAC A
6 AGC CTT GCT GAG GGC TTC ATT GCT GTC AGC AAC ATC GTT TTC GCC TAC
  AGC CTT GCT GAG GGC TTC ATT GCT GTC AGC AAC ATC GTT TTC GCC TAC

TCC GAC TAC AAG AAG TCC ATC GTT GCT CTC GGC TTG ATT GAA ATC TTC
TCC GAC TAC AAG AAG TCC ATC GTT GCT CTC GGC TTG ATT GAA ATC TTC

ATC TAC ACC GTT ACT GGT GGC GTC GTT TAC GCT TTC GTC GGC CCC GAG
ATC TAC ACC GTT ACT GGT GGC G  17

GTC CAG TCT CCT GCC TTG CTC TCT GCT GGC CCT CTT CTC GCC AAG GTT
                                      7 CCT CTT CTC GCC AAG GTT

GCT TTC GGC ATT GCC CTC CCC GTC ATC TTC ATC TCT GGC AGT ATC AAC
                                          8 T GGC AGT ATC AAC

ACT GTT GTC GTC AGC AGG TAT CTG ATT GAG CGC ATC TGG CCC AAC AAC
ACT GTT GTC GTC AGC AGG TAT CTG ATT GAG CGC ATC TGG CCC AAC AAC

GTC ATT CGC TAT GTC AAC ACC CCA GCG GGT TGG ATG GTT TGG CTT GGT
GTC ATT CGC TAT GTC AAC ACC CCA GCG GGT TGG ATG GTT TGG CTT GGT

TTT GAC TTT GGC ATT ACC CTC ATT GCC TGG GTT ATT GCT GAG GCC ATC
TTT GAC TTT GGC ATT ACC CTC ATT GCC TGG GTT ATT GCT GAG GCC ATC
```

CCT TTC TTC TCT GAT CTG TTG GCC ATC TGC TCG GCT CTC TTC ATT
CCT TTC TTC TCT GAT CTG TTG GCC ATC TGC TCG GCT CTC TTC ATT TCC

GGT TTT AGC TTC TAT TTC CCT GCC TTG ATG TAT TTC AAG ATC ACC AGG
GC TTC TAT TTC CCT GCC TTG ATG TAT TTC AAG T ACC A G

AAC GAT GCC AAG AGC CAG GGC AAG
9 C CAG GGC AAG AAG TAC TTC TTG GAT GCC CTC AAC
AAC GAT GCC AAG AGC CAG GGC AAG AAG TAC TTC TTG GAT GCC CTC AAC

ATG CTC TGC TTC GTC ATC GGC ATG GGC ATT CTT GGT ATT GGT ACC TAC
ATG CTC TGC TTC GTC ATC GGC ATG GGC ATT CTT GGT **16** Kpn I

GCC GCT ATT CAG GAC ATT GTAAGTTTGGCCCGCTTTTCTGTTTACTCTTTGCACAC

AAATGCTAACTTGCTTCTCAG ATG GAC CGT TAC GAC CAT GGC AAG GTT TCG

AAG CCT TAT AGC TGT GCT CCC TTG GCT TAA CAGGCCCAACGCACG TTATGA
10 CAGGCCCAACGCACGCTTATGA
GCg CC TT GCc TAt CAaGCC AACGCACGCT ATGA

TCCTGTTGTTTTTTTTTGGA TGATTTAATTAAAGTTGCGC AGTGATTGACGTCTGTCTTC
TCCTGTTGTTTTTTTTTGGA TGATTTAATTAAAGTTGCGC AGTGATTGACGTCTGTCTTC
C TGT GTTTT G A T ATT A T AA GTTGCGC AGTGATTGACGTCTGTCTTC

ACCCGCGATTGCCCCTTTTG AT CCCCTCAGACTTGCC GG CTGGG AA TGTT GA
ACCCGCGATTGCCCCTTTTG TATACCCCCTCAGACTTGCC GGCCTGGGGAAATGTTTTGA
ACC GCGATTGCCCCTTT G TATACCCC TCAGACTTGCC GGCCTGGGGAAATGTTTTGA
Acc I

GTATTTCTATTT
GTATTTCTATTTTCGGAGTT TCAGGATTTGGCACAAAGCA AACCAGCGCGGAGTTGAAAC
GTATTTCTATTTTCGCAGTT TCAGGATTTGGCACAAAGCA AACCA C CGGAGTTGAAAC

CGTGGTGTGGTCGCGGTGCG CTGCTGCATTGGTAGTGCTT gTTCCAGGTTTTTGTTTGGT
CGTGGTGTGGTCGCGGTGCG CTGCTGCATTGGTAGTGCTT TTTCCAGGTTTTTGTTTGGT

GGTTTGGATGCGTGCACCAC TTTTTTTTTAACGTTTtAT tGCATGCATGTATTATATGG
GGTTTGGATGC **15** Sph I

GAAAGTCATGGGACATGGCA ACTATaCGAAcCGACGCAAA GATAGGATGGGATGGATGAT

GGATGGACGTACGATCCAAC GCGCTGGGGACTGGACTGAA CGGAATTAGGACGGACGGGA

```
CAgGTAACcTAGGTACCTAA  TGACCGGAATATGTTTACAA  ATCATTGTTTAGTGCGGGTG
              11 CTAA  TGACCGGAATATGTTTACAA  ATCATTGTTTAGTGCGGGTG
             Kpn I

ACCGGCAATAGAGACGATGG  GCACAGGAATATCGATAGAT  GCTACCTATAC
ACCGGCAATAGAGACGATGG  GCACAGGAATATCGATAGAT  GCTACCTATACCCTAAAGAA
                   G  GCACAGG  TATCGATAGAT  GgTACCTATACC TAAAGAA
                                  Cla I

CTCTATAGGTATAATATTCG  CTGAACATACCTTGCCCAAA  AAACAAGAGAACACCCATGG
CTCTATAGGTATAATATTCG  CTGAACATACCTTGCaCAAc  AAACAAGAGAACACCCATGG

TTATGAAATCATCCTGTTGT  TGTGCCATAATTTCCATCCT  GACTCCCATGCCTTCCTGCT
TTATGAAATCATCCTGTTGT  TGTGCCATAATTTCCATCCT  GACTCCCA GCCTTCCTGCT

TTTTTTTTTTTTTTTTCTTG  GCCACGCGCCTCCGATATCT  CGAGTTTTTGAAGGATTCTC
                                     12 ATCT  CGAGTTTTTGAAGGATTCTC
TTTTTTTTTTTTTTTTCTTG  GCCACGCGCC 14
                CTtG  gCCAaGCGCCTC GATATCT  CGAGTTTT GAAGGATTCTC
                                 Eco RV  Xho I

GTGTTGGGTGGA CTTTTCT  CAACAATCCTGGGG
GTGTTGGGTGGAGCTTTTCT  CAACAATCCTGGGGCTTCGA  AACCCTCCACCAGACCTCAC
GTGTTGGGTGGAGCTTT CT  CAACAATCCTGGG CTTCGA  AACCCTCCACCAGACCTCAC

CCCAGCAGTCAGAGTTTAGC  CCGCCAGCCAGCCCGTCAGC  CAAGTGAGTTCTAAGATTAA
CCCAGCAGTCAGAGTTTAGC  CCGCCAGCCAGCCCGTCAGC  CAAGTGAGTTCTAAGATTAA

TCTCGACTCCTTGACAAGGC  TTTGCGTGGCCACGTCTCCT  CTCAACACGCAAAACTTTTG
TCTCGACTCCTTGACAAGGC  TTTGCGTGGCCACGTCTCCT  CTCAACACGCAAAACTTTTG

TCATTGTTACTACACTACAG  GTTACCGTGTCGATGTTCCA  GAC
TCATTGTTACTACACTACAG  GTTACCGTGTCGATGTTCCA  GAC 13
```

BIBLIOGRAPHY

Akins, R. A., and A. M. Lambowitz (1985). General method for cloning Neurospora crassa nuclear genes by complementation of mutants. Mol. Cell. Biol. 5: 2272-2278.

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson (1983). *Molecular Biology of the Cell.* Garland Publishing, Inc.

Ames, G. F. L. (1986). Bacterial periplasmic transport systems: Structure, mechanism, and evolution. Ann. Rev. Biochem. 55: 397-425.

Arends, H., and W. Sebald (1984). Nucleotide sequence of the cloned mRNA and gene of the ADP/ATP carrier from Neurospora crassa. EMBO 3: 377-382.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (1987). *Current Protocols in Molecular Biology*, John Wiley & Sons.

Bax, A., and L. Lerner (1986). Two-dimensional nuclear magnetic resonance spectroscopy. Science 232: 960-967.

Bayer, M. E., and C. C. Remsen (1970). Structure of Escherichia coli after freeze-etching. J. Bacteriol. 101: 304-313.

Berlin, V., and C. Yanofsky (1985). Isolation and characterization of genes differentially expressed during conidiation of Neurospora crassa. Mol. Cell. Biol. 5: 849-855.

Birnbaum, M. J., H. C. Haspel, and O. M. Rosen (1986). Cloning and characterization of a cDNA encoding the rat brain glucose-transporter protein. Proc. Natl. Acad. Sci. USA. 83: 5784-5788.

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

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331.

Buchel, D., B. Gronenborn, and B. Mueller-Hill (1980). Sequence of the lactose permease gene. Nature 283: 541-545.

Campbell, I. D. (1988). The structure and dynamics of membrane spanning helices by high resolution NMR and molecular dynamics. in *Transport through Membranes: Carriers, Channels and Pumps.* eds A. Pullman, J. Jortner, and B. Pullman. Kluwer Academic Publishers.

Case, M. E. (1982). Transformation of Neurospora crassa utilizing recombinant plasmid DNA. In *Genetic Engineering of Microorganisms for Chemicals.* (eds. A. Hollaender, R. D. DeMoss, S. Kaplan, J. Konisky, D. Savage, and R. S. Wolfe). Plenum Publ. Corp.

Chothia, C. (1984). Principles that determine the structure of proteins. Ann. Rev. Biochem. 53: 537-572.

Chou, P. Y., and G. D. Fasman (1973). Structural and functional role of leucine residues in proteins. J. Mol. Biol. 74: 263-281.

Cronan, J. E., R. B. Gennis, and S. R. Maloy (1987). Cytoplasmic membrane. In *Escherichia coli and Salmonella typhimurium vol 1.* Chapter 5. (editor-in-chief F. C. Neidhardt). American Society for Microbiology.

Davis, R. H., and F. J. DeSerres (1970). Genetic and microbiological research techniques for Neurospora crassa. Methods Enzymol. 27A:79-143.

DeBusk, B. G., and A. G. DeBusk (1965). Mc!ecular transport in Neurospora crassa I. Biochemical properties of a phenylalanine permease. Biochim. Biophys. Acta. 104: 139-150.

DeBusk, R. M., and A. G. DeBusk (1980). Physiological and regulatory properties of the general amino acid transport system of Neurospora crassa. J. Bacteriol. 143: 188-197.

Doolittle, R. F. (1986). *Of URFS and ORFs.* University Science Books

Dunn-Coleman, N. S. (1984). Cloning and preliminary characterization of a molydenum cofactor gene of Neurospora crassa. Curr. Genet. 8: 589-595.

Eisenberg, D. (1984). Three-dimensional structure of membrane and surface proteins. Ann. Rev. Biochem. 53: 595-623.

Engelman, D. M. and T. A. Steitz (1984). On the folding and insertion of globular membrane proteins. In *The Protein Folding Problem* (ed. D. B. Wetlaufer). AAAS Selected Symposia Series. Westview Press, Inc.

Ferguson and Williams (1988). Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Ann. Rev. Biochem. 57: 285-320.

Firgaira, F. A., J. P. Hendrick, F. Kalousek, J. P. Kraus, and L. E. Rosenberg (1984). RNA required for import of precursor proteins into mitochondria. Science 226: 1319-1322.

Furlong, C. E. (1987). Osmotic-shock sensitive transport systems. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology vol I.* Chapter 51. (editor-in-chief F. C. Neidhardt). American Society for Microbiology.

Garnier, J., J. D. Osguthorpe, and B. Robson (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97-120.

Graham, R. W., D. Jones, and E. P. Candido (1989). UbiA, the major polyubiquitin locus in Caenorhabditis elegans, has unusual structural features and is constitutively expressed. Mol. Cell. Biol. 9: 268-277.

Gros, P., J. Croop, and D. Housman (1986). Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell 47: 371-380.

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

Higgins, C., P. Haag, K. Nikaido, F. Ardeshir, G. Garcia, and G. Ames (1982). Complete nucleotide sequence and identification of membrane components of the histidine transport operon of Salmonella typhimurium. Nature 298: 723-727.

Holmes, D. S., and M. Quigley (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114: 193-197.

Horak, J. (1986). Amino acid transport in eucaryotic microorganisms. Biochim. Biophys. Acta. 864: 223-256.

Humphrey, T., and N. J. Proudfoot (1988). A beginning to the biochemistry of polyadenylation. TIG 4: 243-245.

Kaback, H. R. (1986). Active transport in Escherchia coli: Passage to permease. Ann. Rev. Biophys. Biophys. Chem. 15: 279-319.

King, J. (1989). Deciphering the rules of protein folding. C & EN

Kinnaird, J. H., and J. R. S. Fincham (1983). The complete nucleotide sequence of the Neurospora crassa am (NADP-specific glutamate dehydrogenase) gene. Gene 26: 253-260.

Koo, K., and W. D. Stuart (1984). The *mtr* gene product in Neurospora crassa is associated with a cell-surface, phenylalanine-binding ribonucleoprotein particle. Genetics 107S: 557-558.

Krumlauf, R., and G. A. Marzluf (1979). Characterization of the sequence complexity and organization of the Neurospora crassa genome. Biochemistry 18: 3705-3713.

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

Laird, P. W. (1989). Trans splicing in trypanosomes - archaism or adaptation? TIG 5: 204-208.

Larimer, F. W., and A. G. DeBusk (1977). Genetic and kinetic analysis of Neurospora crassa *mtr* mutants. J. Bacteriol. 129: 1636-1638.

Legerton, T. L., and C. Yanofsky (1985). Cloning and characterization of the multifunctional his-3 gene of Neurospora crassa. Gene 39; 129-140.

Lehninger, A. (1975). *Biochemistry.* Worth Publishers, Inc.

Lester, G. (1966). Genetic control of amino acid permeability in Neurospora crassa. J. Bacteriol. 91: 677-684.

Lewin, B. (1980). *Gene Expression II Eucaryotic Chromosomes.* Chapter 23. John Wiley & Sons.

Lewin, B. (1987). *Genes III.* John Wiley & Sons.

Lin, E. C. C. (1987). Dissimilatory pathways for sugars, polyols, and carboxylates. In *Escherichia coli and Salmonella typhimurium vol 1.* Chapter 18. (editor-in-chief F. C. Neidhardt). American Society for Microbiology.

Lodish, H. F. (1989) Multi-spanning membrane proteins: How accurate are the models? Trends Biochem. Sci. 13: 332-334.

Maiden, M. C. J., E. O. Davis, S. A. Baldwin, D. C. M. Moore, P. J. F. Henderson (1987). Mammalian and bacterial sugar transport proteins are homologous. Nature 325: 641-643.

Maloney, P. C. (1987). Coupling to an energized membrane: Role of the ion-motive gradients in the transduction of metabolic energy. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology vol I.* Chapter 17. (editor-in-chief F. C. Neidhardt). American Society for Microbiology.

Mandel, M. and A. Higa (1970). Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53: 159-162.

Maniatis, T., E. F. Fritsch, and J. Sambrook (1982). *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Publications, N. Y.

McNally, M. T., and S. J. Free (1989). Isolation and characterization of a Neurospora glucose-repressible gene. (in press).

Metzenberg, R. L., J. N. Stevens, E. U. Selker, and E. Morzycka-Wroblewska (1984). A method for finding the genetic map position of cloned DNA fragments. Neurospora Newsletter 31: 35-39.

Mueckler, M., C. Caruso, S. A. Baldwin, M. Panico, I. Blench, H. R. Morris, W. J. Allard, G. E. Lienhard, and H. F. Lodish (1985). Sequence and structure of a human glucose transporter. Science 229: 941-945.

Munger, K., U. A. Germann, and K. Lerch (1985). Isolation and structural organization of the Neurospora crassa copper metallothionein gene. EMBO 4: 2665-2668.

Orbach, M. J., E. B. Porro, and C. Yanofsky (1986). Cloning and characterization of the gene for ß-tubulin from a benomyl-resistant mutant of Neurospora crassa and its use as a dominant selectable marker. Mol. Cell. Biol. 6:2452-2461.

Oxender, D. L., E. J. Collarini, M. A. Shotwell, C. D. Lobaton, A. Moreno, and G. S. Campbell (1985). Regulation and genetics of amino acid transport. Ann. N. Y. Acad. Sci. 456: 404-416.

Pelham, H. (1985). Cleaning up plasmid minipreps with lithium chloride. TIG 1:6

Perkins, D. D., A. Radford, D. Newmeyer, and M. Bjorkman (1982). Chromosomal loci of Neurospora crassa. Microbiol. Rev. 46: 426-570.

Plamann, M., and C. Yanofsky (1989). Analysis of cpc-1 expression and function. Fungal Genetics Newsletter 36:10.

Postma, P. W. (1987). Phosphotransferase system for glucose and other sugars. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology vol 1.* Chapter 11. (editor-in-chief F. C. Neidhardt). American Society for Microbiology.

Query, C. C., R. C. Bentley, and J. D. Keene (1989). A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell 57: 89-101.

Rao, E. Y. T., T. K. Rao, and A. G. DeBusk (1975). Isolation and characterization of a mutant of Neurospora crassa deficient in general amino acid permease activity. Biochim. Biophys. Acta. 413: 45-51.

Reinert, W. R., V. B. Patel, and N. H. Giles (1981). Genetic regulation of the *qa* gene cluster of Neurospora crassa: Induction of *qa* messenger ribonucleic acid and dependency on *qa*-1 function. Mol. Cell. Biol. 1: 829-935.

Richardson, J. S. (1984). What do the folds in proteins look like? In *The Protein Folding Problem.* (ed. D. B. Wetlaufer). AAAS Selected Symposia Series. Westview Press, Inc.

Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg (1977). Labeling deoxyribonucleic acid to high specific activity *in vitro* ny nick translation with DNA polymerase I. J. Mol. Biol. 113: 237-251.

Roess, W. B., and A. G. DeBusk (1968). Properties of a basic amino acid permease in Neurospora crassa. J. Gen. Microbiol. 52: 421-432.

Rosen, B. P. (1987). ATP-coupled solute transport systems. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology vol I.* Chapter 50. (editor-in-chief F. C. Neidhardt). American Society for Microbiology.

Sachs, M. S. (1989). Structural and regulatory features of the N. crassa arg-2 gene. Fungal Genetics Newsletter 36: 10.

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

Schechtman, M. G., and C. Yanofsky (1983). Structure of the trifunctional trp-1 gene from Neurospora crassa and its aberrant expression in E. coli. J. Mol. Appl. Genet. 2: 83-99.

Sefton, B. M., and J. E. Buss (1987). The covalent modification of eukaryotic proteins with lipid. J. Cell. Biol. 104: 1449-1453.

Shuman, H. A. (1987). The genetics of active transport in bacteria. Ann. Rev. Genetics 21: 155-177.

Singer, S. J., and G. L. Nicolson (1972). The fluid mosaic model of the structure of cell membranes. Science 175: 720-731.

Southern, E. M. (1975). Detecetion of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517.

Stadler, D. R. (1966). Genetic control of the uptake of amino acids in Neurospora. Genetics 54: 677-685.

Stadler, D. R., and B. Kariya (1969). Intragenic recombination at the *mtr* locus of Neurospora with segregation at an unselected site. Genetics 63: 291-316.

Stein, W. D. (1986). *Transport and diffusion across cell membranes.* Academic Press.

Struhl, K. (1983). The new yeast genetics. Nature 305: 391-397.

Struhl, K. (1985). A rapid method for creating recombinant DNA molecules. Biotechniques 3: 452-453.

Stuart, W. D., and A. G. DeBusk (1971). Molecular Transport I. In vitro studies of isolated glycoprotein subunits of the amino acid transport system of Neurospora crassa conidia. Arch. Biochem. Biophys. 144: 512-518.

Stuart, W. D., and A. G. DeBusk (1973). Transport of arginine by an *in vitro* system. Biochem. Biophys. Res. Comm. 52: 1046-1050.

Stuart, W. D., and A. G. DeBusk (1975). Genetic alterations of ribonuclease-sensitive glycoprotein subunits of amino acid transport system in Neurospora crassa conidia. Arch. Biochem. Biophys. 166: 213-222.

Stuart, W. D., and D. O. Woodward (1975). Effects of ribonuclease A on amino acid transport in Neurospora crassa. J. Bacteriol. 122: 41-46.

Stuart, W. D., and D. O. Woodward (1977). New class of ribonucleic acid in Neurospora associated with the outer cell envelope. J. Bacteriol. 129: 395-399.

Stuart, W. D., K. Koo, and S. J. Vollmer (1988). Cloning of mtr, an amino acid transport gene of Neurospora crassa. Genome 30: 198-203.

Tanaka, J., and G. R. Fink (1985). The histidine permease gene (HIP1) of Saccharomyces cerevisiae. Gene 38: 205-214.

Taylor, W. R. (1987). Protein structure prediction. in *Nucleic acid and protein sequence analysis* (Chapter 12) eds. M. J. Bishop and C. J. Rawlings. IRL Press.

Travis, G., L. Wolfinbarger, W. D. Stuart, and G. DeBusk (1972). Disruption of an amino acid transport mutant of Neurospora crassa by KCl. J. Bacteriol. 109: 933-935.

Ullu, E., S. Murphy, and M. Melli (1982). Human 7SL RNA consists of a 140 nucleotide middle-repetitive sequence inserted in an Alu sequence. Cell 29: 195-202.

Vapnek, D., J. A. Hautala, J. W. Jacobson, N. H. Giles, and S. R. Kushner (1977). Expression in Escherchia coli K-12 of the structural gene for catabolic dehydroquinase of Neurospora crassa. Proc. Natl. Acad. Sci. USA 71: 3508-3512.

Vollmer, S. J., and C. Yanosky (1986). Efficient cloning of genes of Neurospora crassa. Proc. Natl. Acad. Sci. USA. 83: 4869-4873.

von Heijne, G. (1985). Signal sequences: The limits of variation. J. Mol. Biol. 184: 99-105.

Walter, P., and G. Blobel (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature 299: 691-698.

Walter, P., and G. Blobel (1983). Disassembly and reconstitution of signal recognition particle. Cell 34: 525-533.

Wiley, W. R. (1970). Tryptophan transport in Neurospora crassa: a tryptophan-binding protein released by cold osmotic shock. J. Bacteriol. 103: 656-662.

Wilson, D. B. (1978). Cellular transport mechanisms. Ann Rev. Biochem. 47: 933-965.

Wolfinbarger, L., and A. G. DeBusk (1971). Molecular transport: I. *In vivo* studies of transport mutants of Neurospora crassa with altered amino acid competition patterns. Arch. Biochem. Biophys. 144: 503-511.

Wolfinbarger, L. (1980). Transport and utilization of amino acids by fungi. In *Microorganisms and Nitrogen Sources*. pp. 63-87. (ed. J. Payne). John Wiley & Sons.

Woudt, L. P., A. Pastink, A. E. Kempers-Veenstra, A. E. M. Jansen, W. H. Mager, and R. J. Planta (1983). The genes coding for histone H3 and H4 in Neurospora crassa are unique and contain intervening sequences. Nucl. Acids Res. 11:5347-5360.

Yim, L. K. C., and W. D. Stuart (1983). Isolation of para-fluorophenylalanine resistant mutants from HeLa cell cultures. Biochem. Genet. 21: 443-452.