INTRODUCTION
1.1 Gene Expression in Eukaryotes

1.1.1 Overview

Gene expression is operationally defined as the decoding of a DNA sequence information, by the process of transcription, translation and associated processing steps, into finished proteins which in turn act on the cell. The regulation of expression of several prokaryotic genes have come to light. In eukaryotes the regulation is far more complex than that of prokaryotes. Most of the discussion here would be confined to gene regulation in eukaryotes with particular reference to yeast \textit{Saccharomyces cerevisiae}.

In higher eukaryotes, the number of expressed genes is between 10,000 and 15,000 (Lewin, 1985); the analogous estimate for yeast is about 5000 (Lewin, 1985; Struhl, 1989). Not all these genes are expressed continuously; many are expressed at a constant rate, some are expressed only at a particular stage of cell cycle or in specific tissues, and still others are confined to specialized situations which involve certain developmental or environmental cues. How these are brought about is the subject of intense research in this area.
1.1.2 Variety in the Control of Gene Expression in Eukaryotic Cells

Gene expression is mainly regulated at the level of transcription, RNA processing and translation. Regulation by post-translational modifications such as phosphorylation, glycosylation, ubiquitinylation and myristoylation have also come to light.

1.1.2.1 Transcriptional Control

A major point of gene control occurs at the step of initiation of mRNA synthesis (Struhl, 1989). From yeast to humans, the transcriptional machinery and factors involved in transcriptional regulation appear to be conserved. For example, the subunit structure and catalytic properties of RNA polymerase II, sequences of TATAAA box and enhancers are quite conserved. Furthermore, some of the yeast transcriptional activator proteins can function efficiently even in mammalian cells and the converse is also true (Guarente, 1988).

The transcriptional regulatory patterns of individual genes or sets of genes is primarily brought about by sequence-specific DNA-binding proteins that interact at promoter sites (Johnson and McKnight, 1989; Struhl, 1989). These
proteins often have two distinct domains, a DNA binding domain, and an activator domain which interacts with the transcriptional machinery. DNA binding domains have certain structural motifs such as helix-turn-helix, zinc-finger, leucine zipper and certain unclassified DNA-binding motifs (Struhl, 1989; Johnson and McKnight, 1989). A compendium of eukaryotic transcriptional activator proteins has been presented recently (Johnson and McKnight, 1989). Since many activator proteins have similar DNA-binding motif, there should be distinct features by which they recognize specific DNA sequences. The specificity probably comes from distinct amino acid sequences at the DNA-binding domains of these proteins.

Some auxiliary proteins could interact with the activator proteins and confer additional specificity through protein:protein interaction. Such an auxiliary factor could improve the binding specificity of a regulatory protein in one of two ways: It would contribute, from its own polypeptide structure, additional contact surfaces with DNA. Alternately, the role of the auxiliary protein might be indirect. The contact of activator protein with the second regulatory protein might prompt an allosteric change in the protein, causing it to bind certain DNA sequences more strongly. For example, FOS protein (a nuclear transforming
protein) on its own exhibits little or no intrinsic affinity for specific DNA sequences. When mixed with JUN (a nuclear protein encoded by another proto-oncogene) FOS enters into a protein:DNA complex that is more stable than the complex formed by JUN alone. Another example of the role of auxiliary factors have come from studies on Xenopus transcription, TFIIIA and its formation of an active 5S gene transcription complex by interaction with an additional factor TFIIIC (Johnson and McKnight, 1989).

The binding of activator proteins to upstream promoter elements is necessary but not sufficient for stimulating transcription in vivo. The transcriptional activation domain can be an integral part of DNA-binding protein. For example, GAL4 (positive activator of several GAL genes) activates transcription through its C-terminal portion; analysis of GAL4 structure and the structure of several other activator proteins revealed that the activation domain is an acidic region of the protein. The amino acid residues in the activator region presumably interact with TATA-binding protein, RNA polymerase II, or histones. The mode of action of eukaryotic transcriptional activators has been reviewed (Ptashne, 1988).
a) Specific DNA-binding proteins in yeast:

*Saccharomyces cerevisiae* cells contain many specific transcription factors that recognize different DNA sequences. Studies of yeast DNA-binding proteins are most advanced for the GAL4, GCN4, HAP1 and MATα2 proteins.

GAL4 protein is required for transcription of GAL genes in *S. cerevisiae*. It binds to upstream activating sequences of several GAL genes. When galactose is added to yeast cells, the levels of its metabolic enzymes and their cognate mRNAs increase rapidly by 5-10 min.

GCN4 protein binds specifically to the promoter regions of many amino acid biosynthetic genes and induces their transcription in response to amino acid starvation. Unlike GAL4, GCN4 level increases about 30-50 fold by a translational regulatory phenomenon.

The HAP1 protein binds to upstream regulatory elements of two cytochrome c genes, *CYC1* and *CYC7* and induces their transcription. Specific DNA binding in vitro and transcription in vivo are stimulated by heme. *CYC1* and *CYC7* binding sites have no obvious similarity and so HAP1 recognizes two different DNA sequences, possibly utilizing a single DNA binding domain.
Mating type in *S. cerevisiae* cells is mainly regulated by three proteins: an activator protein MATα1 and two repressor proteins, MATα2 and MATα1. MATα1 protein is known to be an activator of α-specific genes. Upstream sequences of three α cell-specific genes contain a related sequence. MATα1 binds to DNA in collaboration with another protein by protein-protein interactions. MATα2 regulates yeast cell type by binding specifically to sequences of several α-specific genes and by repressing their transcription. In a/α diploid cells, MATα2 interacts with MATα1 and form a novel regulatory protein which represses haploid cell-specific gene expression. The mating types of cells are intricately regulated by these three MAT-loci proteins as well as by a plethora of other proteins (Herskowitz, 1989).

b) Modulation of Activator proteins: *S. cerevisiae* cells use a variety of molecular mechanisms to regulate the activity of specific transcription factors (Struhl, 1989). (i) The activity of DNA-binding proteins is altered by the binding of a small molecule. For example, HAP protein requires heme for efficient DNA binding in vitro and transcription activation in vivo. (ii) In another mechanism, transcriptional activation is affected by the binding of another protein to the activator protein. The GAL80 protein inhibits transcriptional activation by interacting directly.
with GAL4, such that it masks the primary acidic transcriptional region; it does not affect the GAL4 DNA binding activity. The GAL4-GAL80 interaction is abolished when cells are grown in galactose medium, presumably by an interaction between GAL80 and galactose metabolite (Johnston, 1987). (iii) A third mechanism involves modification of the DNA binding protein. The heat-shock transcription factor is present under all conditions, but during heat-shock it appears to be phosphorylated. As the target genes are expressed only under shock conditions, probably the phosphorylated form represents an active transcription factor; the phosphorylation possibly increases the charge of the acidic domain of the protein.

The above examples involved proteins that are constantly present in cells, but yeast cells also regulate transcription by controlling the amount of a specific DNA-binding factor rather than by modulating its activity of a protein. GCN4 activates transcription only during amino acid starvation (Section 1.1.2.3). Unlike GCN4, whose level changes in response to environmental conditions, the MAT\(\alpha_2\), MAT\(\alpha_1\) and MATa1 proteins regulate transcription of their target genes by being present only in the relevant cell types (Herskowitz, 1989).
1.1.2.2. RNA Processing

When the polypeptide product of a gene is absent from an eukaryotic cell, it is typically due to the fact that the particular gene is transcriptionally inert. However, several recent findings have shown that this is always not true and specific instances of regulation by RNA processing have come to light (Borst et al., 1989).

(a) Alternate splicing: By the choice of appropriate splice sites, multiple mRNAs can be made from a primary transcript. This is known as alternate splicing and is known to occur in a tissue-specific manner (Breitbart et al., 1987). In Drosophila melanogaster, alternative splicing of RNA is known to regulate sexual differentiation (Boggs et al., 1987). (b) Trans-splicing: Novel combinations of RNA are generated by splicing together exons from different RNAs. Evidences for this trans-splicing have come from trypanosomes and nematodes. In trypanosomes, an exon from one transcript is spliced to the exon of another transcript; the 5'-non coding exon is common to all mRNAs and is derived from the first 35 nucleotides of a separately encoded transcript (Van der Ploeg, 1986). (c) RNA editing: Any process that results in the production of an RNA molecule which differs in nucleotide sequence from that of its DNA template is termed RNA editing (Simpson and Shaw,
RNA editing of pre-mRNAs either re-tailors a non-functional transcript to produce a translatable mRNA, or modifies an already functional mRNA so that it generates a protein of altered amino acid sequence. Such a mechanism has been reported for protozoan mitochondria (Simpson and Shaw, 1989) and plant mitochondria (Gualberto et al., 1989; Covello and Gray, 1989).

1.1.2.3 Translational Control and Post-translational Control

Translational control is less prevalent than transcriptional regulation. General control mechanisms regulating translation of most cellular mRNAs have been well studied; they include the phosphorylation of the subunits of eIF-2 complex (Ochoa, 1983) and phosphorylation of S6 ribosomal protein (Maller et al., 1985). In *S. cerevisiae* a specific instance of translational regulation has been reported. (a) Translational control of GCN4 expression: When the yeast cells are starved for amino acids, the genes involved in amino acid biosynthesis are transcribed. Transcription of many of these genes is activated by the GCN4 protein. It is interesting to note that the transcription of GCN4 gene itself is not regulated by amino acid starvation; its mRNA is made at a constant rate. However, when the cells are cultured under conditions of
amino acid starvation the level of GCN4 protein is increased by an enhanced efficiency of translation. The GCN4 mRNA has an unusually long 5' untranslated region that contains four AUG start codons; apparently these untranslated AUG sites bind ribosomes to prevent other ribosomes from initiating at the correct GCN4 start site (Mueller and Hinnebusch, 1986).

(b) Translational frame-shifting: The process of synthesis of a single protein from two different reading frames of a single mRNA template is known as translational frame shifting, and this phenomenon has been reported from some prokaryotes and eukaryotes (Craigen and Caskey, 1987). In eukaryotes, this phenomenon has been observed in the translation of Rous sarcoma virus gag-pol polyprotein and in the translation of transcripts of Ty transposon in S. cerevisiae (reviewed in Craigen and Caskey, 1987).

(c) The case of a fused protein: In a recent paper, Kanno et al. (1989) have reported that the human erythrocyte glucose-6-phosphate dehydrogenase is synthesized as a fused protein. Here, the C-terminal 479 amino acid residues of the chimeric polypeptide chain are encoded by the human X-chromosomal gene; the N-terminal 55 residues are encoded by a different gene, located on chromosome 6. These N-terminal residues are believed to come from the human mRNA that encodes guanyl monophosphate reductase (Henikoff and Smith, 1989). A splicing intermediate, in this case by
trans-splicing, could not be detected; the non-detectability of fused mRNA have led the authors to speculate the generation of the chimeric protein, by either cross-translation or trans-peptidation (Borst et al., 1989; Kanno et al., 1989). Such fusions would allow unlinked genes to share protein domains without requiring reorganization or duplication of their genes (Henikoff and Smith, 1989).

1.2 C. albicans: Morphogenesis Pathogenesis and Molecular Genetics

1.2.1 Overview

C. albicans, an yeast-like fungus, is a normal commensal of human beings, other mammals and birds. It is considered an opportunistic pathogen because under certain predisposing conditions it becomes virulent and causes candidiasis. C. albicans has aroused considerable interest not only because of its pathogenic potential, but also because it can undergo a morphological transition; the latter feature has made it an useful organism to study cellular differentiation. All these aspects have been discussed in detail in several reviews covering C. albicans morphogenesis (Odds, 1985; Shepherd et al., 1985; Soll, 1986; Odds, 1988; Datta et al., 1989; Paranjape and Datta, 1990), antigenic variation (Poulain et al., 1985), pathogenicity (Shepherd et al., 1985; Dei-Cas and Vernes, 1986; Bodey and Fainstein,
1985; Odds, 1988; Datta et al., 1989), genetics (Shepherd et al., 1985; Datta et al., 1989) and also host defense mechanisms (Waldorf, 1986; Datta et al., 1989). So only a brief account of the various aspects would be presented here.

1.2.2 Morphogenesis

*C. albicans* is a polymorphic fungus. It displays four different cellular and a variety of colony morphologies. It grows as blastospores, pseudohyphae or hyphae and it can also form chlamydospores. All these forms can be induced in vitro under defined growth conditions. Yeast to hypha conversion occurs through an intermediate germ tube stage; the morphology change is controlled by environmental factors, in which the relative extent and timing of apical and general cell wall synthesis determine overall cell shape. By varying different parameters such as carbon source, pH, ions and temperature, hyphae can be induced. Germ tubes are induced when blastospores are incubated at a temperature between 33° and 42°, in a medium containing amino acids such as proline, glutamine and arginine (Odds, 1988), or amino sugars such as GlcNAc and ManNAc (Simonetti, et al., 1974; Shepherd et al., 1980; Sullivan and Shepherd, 1982) or ethanol (Pollack and Hashimoto, 1985). Presumably, inducers bind to cell surface
receptor(s) and produce an intracellular signal, which primes the cells for germ tube formation (Shepherd and Sullivan, 1983).

C. albicans can switch its colony morphology at high frequency (Pomes et al., 1985; Slutsky et al., 1985) and this provided an interesting system to study morphogenesis (reviewed in Datta et al., 1989). The cellular basis for the differences in colony morphologies has been worked out for white to opaque transition (Slutsky et al., 1987; Anderson and Soll, 1987). White and opaque cells differ in their shape, size and budding pattern. The opaque cell surface has a pimpled or punctate pattern, and this pattern is correlated with the appearance of a cell surface protein (Anderson and Soll, 1987). A hypha-specific protein is also expressed on the budding opaque cell surface. Thus opaque phenotype seems to result from a temporally modulated expression of bud-, hypha- and opaque-specific genes (Anderson and Soll, 1987). Two aspects of colony morphology switching are most striking: firstly it occurs at a high frequency and secondly it results in a large number of phenotypic changes. Thus switching is probably due to a change in the expression of a master control gene. This gene would regulate the expression of many other genes, which would result in diverse phenotypes.
Attempts to unravel the mechanism that controls yeast to germ tube conversion have not been greatly successful. But specific protein phosphorylation has been observed and this is believed to be a component of an intracellular signalling pathway (Paranjape and Datta, 1990). Both germ tube-specific and bud-specific phosphorylation of different proteins have been observed (Paranjape and Datta, manuscript in preparation); it is possible that these phosphorylated proteins would either directly control alteration in the phenotypes or would in turn modulate other events.

On the other hand, if morphogenesis is directly regulated by specific genes, then it would involve differential expression of certain morphology-specific proteins. Finney et al. (1985) using two dimensional gel electrophoresis, could detect two cytoplasmic polypeptides, one being bud-specific and the other hypha-specific. Some proteins are specifically expressed on the hyphal cell surface (reviewed in Datta et al., 1989). If the genes coding for the hypha-specific and opaque cell-specific proteins are cloned, then specific mutants can be obtained by directed mutagenesis technique (Kelly et al., 1987); the cloned genes and mutant strains should be useful to further understand the role of these proteins in regulation of morphogenesis.
1.2.3 Pathogenesis

The pathogenic potential of *C. albicans* has made it a medically important organism. Other species responsible for candidiasis in humans are *C. tropicalis*, *C.pseudotropicalis*, *C. guilliermondii*, *C. krusei*, *C. parapsilosis* and *C. stellatoidea*; however, *C. albicans* is the predominant species identified. Pathological studies have demonstrated three types of infections: superficial, locally invasive and systemic (Luna and Tortoledo, 1985); of these the systemic candidiasis is the most serious.

*C. albicans* is a normal commensal of the respiratory and gastrointestinal tracts, vagina and less frequently the skin of healthy persons. An important feature of *C. albicans* pathogenesis is that when the host defense mechanism breaks down it becomes infectious which results in candidiasis. Hence *C. albicans* is called an opportunistic pathogen. Host defense mechanisms to *C. albicans* infections are skin-mucosal barriers, immune system and hormone levels. Breakdown of any of these defense mechanisms predispose the host to *C. albicans* infections (Waldorf, 1986; Odds, 1988; Datta et al., 1989). Its ability to change cellular and colony morphologies, its capacity to adhere to host tissues, and to secrete hydrolytic enzymes are considered to help the organism to become virulent. The ability of pathogenic
yeasts such as *C. albicans* to utilize N-acetylglucosamine (Singh et al., 1979), and also to secrete N-acetylglucosaminidase (Sullivan et al., 1982) might confer an adaptive value. The mucous membranes, which are the common sites of infection are rich in amino sugars; so the ability to utilize GlcNAc would help *C. albicans* to survive at infected sites.

1.2.4 Molecular genetics

*C. albicans* is naturally diploid and no sexual cycle has been found. The base composition of the genome is about 35% G+C (Stenderup and Lethbak, 1969). The DNA content per cell (diploid) is about 37 femtograms (Riggsby et al., 1982). Though a varied estimate of chromosome number, between five and ten, have been reported from various laboratories (reviewed in Datta et al., 1989), Magee et al. (1988) have provided evidence for the presence of seven chromosomes in *C. albicans*; a total of 14 genes have been assigned to the chromosomes using a combination of contour-clamped homogenous field electrophoresis (Chu et al., 1986) and Southern blotting techniques.

Though no natural sexual cycle has been reported in *C. albicans*, hybridization at a low frequency between *C. albicans* and *C. guilliermondii* was reported (Suzuki
et al., 1986). In these cases, the hybrids had apparently resulted from partial transfer of genetic material, rather than by true karyogamy. However, protoplast fusion technique has been helpful in the parasexual genetic analysis of *C. albicans*. Five linkage groups have been demonstrated in *C. albicans* by treatment of hybrids obtained by protoplast fusion with either ultraviolet irradiation or heat shock, the former to induce mitotic recombination and the latter to induce chromosome loss (reviewed in Datta et al., 1989).

Many strains of *C. albicans* are known to be heterozygous at certain loci. On UV irradiation these strains become homozygous by mitotic recombination, and yield a limited set of mutants (Datta et al., 1989). Other *C. albicans* mutants were isolated by either chemical or UV mutagenesis. Genetic analysis of the mutants obtained after strong mutagenesis is often complicated because they may harbor multiple, independent and unknown lesions, and thus such mutants are of limited value.

A DNA-mediated transformation system for *C. albicans* was reported (Kurtz et al., 1986). *C. albicans* ADE2 gene, obtained by complementation of *E. coli* and *S. cerevisiae* mutants was used to transform *C. albicans*. The transformation frequency was low and the transforming DNA was found
integrated in the genome. Subsequently, autonomously replicatory sequences were isolated from *C. albicans* (Kurtz et al., 1987); an ade2 strain was transformed to prototrophy at a high frequency, with a plasmid containing ADE2 gene and *C. albicans* genomic sequences conferring autonomous replicating function. Gene disruption technique devised for *S. cerevisiae* (Rothstein, 1983) was extended to *C. albicans* (Kelly et al., 1987). *C. albicans* URA3 gene (Gillum et al., 1984) was disrupted in vitro using cloned ADE2 gene. This construct was used to replace the chromosomal URA3 by integrative transformation.

The directed mutagenesis technique should facilitate construction of stable mutants, if cloned genes are available. Several genes have been cloned so far; the approaches taken to clone genes have been either genetic complementation of *E. coli* or *S. cerevisiae* mutants (Gillum et al., 1984; Rosenbluh et al., 1985; Jenkinson et al., 1988; Magee et al., 1988; Singer et al., 1989) or hybridization with heterologous gene probes (Smith et al., 1988; Magee et al., 1988) or screening with antibodies of either genomic DNA expression libraries (Matthews and Burnie, 1989; K. Ganesan, A. Banerjee and A. Datta, manuscript in preparation) or a cDNA expression library (this work).
1.3 Aminosugar Metabolism

The most common aminosugars in nature are glucosamine (GlcN), N-acetylglucosamine (GlcNAc), and galactosamine (GalN). Glucosamine is widely distributed as a constituent of mucopolysaccharides and mucoproteins, such as hyaluronic acid and heparin. The chief component of cell wall of fungi and of the shells of crustaceae (lobsters, crabs etc.) is chitin; it is a homopolymer of GlcNAc joined by (1→4) glucosidic linkages. Galactosamine occurs in a group of sulfated mucopolysaccharides present in chondroproteins found in cartilage, adult bone, cornea etc.

1.3.1 Regulation of Aminosugar Metabolism in E. coli

N-acetylglucosamine is a component of the peptidoglycan found in cell walls of E. coli. Peptidoglycan is a repeating polymer of GlcNAc and N-acetylmuramic acid cross linked by peptide chains. The outer membrane of E. coli contains complex lipopolysaccharides of which GlcNAc is a basic component.

E. coli can utilize GlcNAc and GlcN as carbon sources, producing growth rates comparable to glucose. The GlcNAc and GlcN catabolic pathway is well studied in E. coli (Holmes and Russell, 1972). When E. coli cells are grown in
a medium containing GlcNAc, it is transported into the cells through a phosphotransferase system (White, 1979) and by the GlcNAc-specific transporter, Enzyme II\textsuperscript{nag}; the latter is coded by the nagE gene (Rogers et al., 1988; Peri and Waygood, 1988). The Enzyme II\textsuperscript{nag} is a transmembrane protein and helps in the translocation and phosphorylation of the sugar. White (1968) isolated \textit{E. coli} mutants defective in utilization of GlcNAc and the mutations were mapped to nagA and nagB genes, encoding GlcNAC-6-P deacetylase and GlcN-6-P deaminase respectively.

The deacetylase (White and Pasternak, 1967) and the deaminase were purified from \textit{E. coli} (Comb and Roseman, 1962; Calcagno et al., 1984). In \textit{E. coli}, two enzymes are known to catalyze interconversion of fructose-6-P and GlcN-6-P (Wu and Wu, 1971; Sarvas, 1971).

(i) \[
\text{Fructose-6-P + GlcN-6-P} \quad \xrightarrow{\text{GlcN-6-P synthetase}} \quad \text{GlcN-6-P + L-glutamic acid}
\]

(ii) \[
\text{Fructose-6-P +NH}_3 \quad \xrightarrow{\text{GlcN-6-P deaminase}} \quad \text{GlcN-6-P + Water}
\]

The reaction (ii) was detected in the direction of GlcNAc-6-P formation from fructose-6-P and ammonia, when it was coupled with acetylase to form GlcNAc-6-P in the presence of
coenzyme A (Ghosh et al., 1960). It was believed that deaminase does not function anabolically in *E. coli*, since mutants lacking GlcN-6-P synthetase die in the absence of exogenous aminosugar (Wu and Wu, 1971; Sarvas, 1971). Recent work has established that in *E. coli*, deaminase is induced only when GlcN or GlcNAc is supplied (Rogers et al., 1988). So the earlier conclusion that deaminase does not catalyze GlcN-6-P formation may not be true, because under the conditions used, deaminase may not be present at all.

*E. coli* genes encoding EnzymeII*nag* (nagE) (Rogers et al., 1988; Peri and Waygood, 1988), and the deacetylase (nagA) and deaminase (nagB) (Plumbridge, 1987; Rogers et al., 1988) have been cloned, sequenced and studied. nagE, nagB and nagA are closely linked in the *E. coli* genome. The expression of the genes are tightly regulated and are induced only on growth in GlcN or GlcNAc containing medium (Rogers et al., 1988; Plumbridge, personal communication). nagC and nagD are also involved in GlcNAc metabolism. nagE belongs to one operon and nagB, nagA, nagC and nagD constitute another operon. These two operons are divergently transcribed. Role of nagC and nagD genes in the GlcNAc metabolism is not clear but since nagC gene has features of a transcriptional repressor, it is presumed to be a regulatory gene (Plumbridge, personal communication).
1.3.2 N-acetylglucosamine Catabolic Pathway in *C. albicans*.

*C. albicans* strains can utilize GlcNAc as a sole carbon source (Bhattacharya et al., 1974a, b). Other pathogenic yeasts that can utilize GlcNAc are *C. krusei*, *Torulopsis candida* (*Candida famata*) and *C. tropicalis*; non-pathogenic yeasts that cannot utilize are *C. utilis*, *S. cerevisiae*, *S. fragilis* and *Schizosaccharomyces pombe* (Singh and Datta, 1979a; K. Natarajan, unpublished). Thus it is believed that aminosugar metabolism is characteristic of pathogenic yeasts. Aminosugars are present in mucous membranes, which are the sites of colonization by *C. albicans*. Although aminosugars are not present in free form, they are constituents of glycoproteins such as mucin. Since *C. albicans* has to survive on the mucous membrane by utilizing a carbon source, the ability to utilize GlcNAc could be an adaptation to the pathogenic life.

GlcNAc is converted to fructose-6-phosphate by the sequential action of a set of enzymes (Fig. 1): GlcNAc is transported by a membrane-associated permease (Singh and Datta, 1979a; Singh et al., 1980a); the transported sugar is metabolized by the sequential action of GlcNAc kinase (Bhattacharya et al., 1974a, b; Singh and Datta, 1979c; Rai et al., 1980; Shepherd et al., 1980), GlcNAc-6-P deacetylase (Gopal et al., 1982; Rai and Datta, 1982) and GlcN-6-P
Fig. 1. **N-acetylglucosamine catabolic pathway in C. albicans.** GlcNAc induces the synthesis of all the enzymes of the pathway. It is transported into the cells by GlcNAc permease (A) and further sequentially catabolized by GlcNAc-kinase (B), GlcNAc-6-P deacetylase (C), GlcN-6-P deaminase (D), and is converted to fructose-6-P. N-acetylmannosamine induces ManNAc-2-epimerase (E) which converts ManNAc to GlcNAc. Glucosamine is transported by a general sugar permease (F) and is apparently converted to GlcN-6-P.
deaminase (Singh and Datta, 1979b,c; Gopal et al., 1982). Synthesis of all these enzymes is tightly regulated; activity of these enzymes are absent in cells grown on glucose and are induced only when GlcNAc is supplied.

N-acetylmannosamine induces ManNAc-2-epimerase and all the GlcNAc catabolic pathway enzymes (Biswas et al., 1979; Sullivan and Shepherd, 1982). GlcN also supports the growth of C. albicans (Torosantucci and Cassone, 1983; Natarajan et al., 1984). It is transported presumably by a general sugar permease (Fig. 1). GlcN induces deaminase but not kinase and deacetylase (Natarajan et al., 1984). In C. albicans glucose does not repress the synthesis of GlcNAc catabolic pathway enzymes (Singh and Datta, 1978), but in E. coli these enzymes are repressed in presence of glucose (Rogers et al., 1988).

GlcNAc has another role in C. albicans. It can induce germ tube formation (Simonetti et al., 1974; Shepherd et al., 1980). GlcNAc can serve as a precursor for chitin synthesis. It is first converted to GlcNAc-6-P by the catabolic pathway enzymes (Fig. 1). GlcNAc-6-P is polymerized into chitin, by sequential reactions catalyzed by phosphoacetyl glucosamine mutase, UDP-GlcNAc pyrophosphorylase and chitin synthase (Gopal et al., 1982).
For induction of catabolic pathway enzymes, transport of GlcNAc is essential; when immobilized GlcNAc is added to the cells, it induces germ tube formation, but not the catabolic enzymes (Sullivan and Shepherd, 1982). Thus the mechanism of induction of enzymes presumably involves a direct interaction of GlcNAc with some intracellular target(s).

It is apparent that there are several differences between the *E. coli* and *C. albicans* pathways: the glucose repression phenomenon is absent in *C. albicans*; the transport and phosphorylation of GlcNAc in *C. albicans* are catalyzed by two different enzymes, whereas in *E. coli* a single enzyme has both the functions. In *E. coli* GlcN induces all the pathway enzymes, whereas in *C. albicans* it induces only deaminase (J. Plumbridge, personal communication).

There are some apparent similarities between the GlcNAc catabolic pathway of *C. albicans* and galactose catabolic pathway of *S. cerevisiae*: (1) The kinetics of induction of the two pathway enzymes are similar (see Section 2.7.1 for deaminase induction) indicating that, like in *S. cerevisiae*, pre-existing factors presumably control the rapid transcriptional response. (2) Since intracellular GlcNAc
is essential for induction of enzymes, like in GAL gene regulation, where a galactose metabolite is believed to be directly involved, it appears that GlcNAc would have an analogous function.

1.4 Aim and Scope

The objective of this work has been to study gene regulation in C. albicans, a pathogenic yeast. C. albicans has become an important organism because of two distinct features: Firstly, it is a human pathogen. Normally, it is a commensal in healthy human beings, mammals, birds; but when host defense mechanism breaks down it becomes pathogenic, causing candidiasis. Secondly, it can undergo cellular and colony morphology changes that has made it a chosen system to study cellular differentiation.

C. albicans and other pathogenic yeast can utilize aminosugars as sole carbon source for its growth. aminosugar such as GlcNAc induces the synthesis of the catabolic pathway enzymes permease, kinase, deacetylase and deaminase. Uptake of the inducer has been shown to be necessary for induction of enzymes. An interesting feature is that a single inducer GlcNAc, induces all the enzymes.
Molecular studies of the *C. albicans* GlcNAc catabolic pathway is important for following reasons: (i) it can serve as another model system to study genetic regulatory circuits controlling gene expression in eukaryotes. (ii). *C. albicans* is a medically important organism due to its pathogenicity. Several groups are attempting to understand the mechanism of *C. albicans* virulence. Molecular approaches would be useful to understand virulence. (iii) Since aminosugar utilization is an attribute of pathogenic yeasts it is imperative to study in detail the fundamental aspects of gene regulation.

Cloned genes are necessary to study its structure and expression. Moreover if aminosugar metabolism is necessary for *C. albicans* infection, mutants are required. Since *C. albicans* is diploid and asporogenous mutants are difficult to obtain. However if cloned genes are available mutant strains can be constructed by in vitro gene disruption and directed mutagenesis. Towards this end glucosamine-6-P-deaminase, one of the enzymes of the pathway, has been chosen, since it is the terminal enzyme and would be required for metabolism of both GlcN and GlcNAc. Deaminase has been purified, characterized and antibody against it was made in rabbits. $\lambda gt11$ cDNA expression was constructed using GlcNAc-induced poly(A)$^+$ RNA. Using antibody, the library
was screened for clones expressing fusion proteins out of seven clones obtained, three have been identified to be putative deaminase cDNA clones. Using cloned DNA, expression was studied by Northern blot analysis which indicated that the deaminase gene is regulated at the transcriptional level.