

The inducible *N*-acetylglucosamine catabolic pathway gene cluster in *Candida albicans*: Discrete *N*-acetylglucosamine-inducible factors interact at the promoter of *NAG1*

M. Jyothi Kumar, Md. S. Jamaluddin, K. Natarajan*, Deepinder Kaur, and Asis Datta†

Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi, 110067, India

Communicated by Vulimiri Ramalingaswami, All India Institute of Medical Sciences, New Delhi, India, September 21, 2000 (received for review July 12, 2000)

The catabolic pathway of *N*-acetylglucosamine (GlcNAc) in *Candida albicans* is an important facet of its pathogenicity. One of the pathway genes, encoding glucosamine-6-phosphate deaminase (*NAG1*) is transcriptionally regulated by GlcNAc. Sequence analysis of a 4-kb genomic clone containing *NAG1* indicates that this gene is part of a cluster containing two other genes of the GlcNAc catabolic pathway, i.e., *DAC1*, GlcNAc-6-phosphate deacetylase, and *HXK1*, hexokinase. All three genes are temporally and coordinately induced by GlcNAc suggesting a common regulatory mechanism for these genes. The *NAG1* promoter is up-regulated when induced by GlcNAc in *C. albicans* but not in *Saccharomyces cerevisiae*. *In vivo* analysis of the deletion constructs delineated the minimal promoter to –130 bp and mapped two regions at –200 and –400 bp upstream of +1 (ATG) responsible for GlcNAc induction. Gel mobility-shift assays and “footprinting” (DNase protection method) analyses revealed two regions, 5′-GGAGCAAAAAATGT 3′ (–164 to –150, box A) and 5′-ACGGT-GAGTTG 3′ (–291 to –281, box B), that are recognized and bound by at least two inducible activator proteins directing the regulation of gene expression.

Candida albicans, an opportunistic yeast pathogen of humans, normally exists as a commensal and turns pathogenic when the host is immunocompromised. It can cause a variety of infections, frequently in the gastrointestinal, respiratory, and genital tracts (1). The mucous membranes at the site of infection are rich in aminosugars, e.g., *N*-acetylglucosamine (GlcNAc; ref. 2). The capability to use GlcNAc as a sole carbon source is an attribute of pathogenic *Candida* species (3). A mutant deficient in β -*N*-acetylglucosaminidase, a GlcNAc-responsive enzyme that increases the extracellular availability of GlcNAc, is less virulent (4). The GlcNAc-use pathway is also present in bacteria such as *Escherichia coli*, *Klebsiella pneumoniae*, and *Vibrio* sp. Besides inducing the enzymes of the catabolic pathway, GlcNAc induces changes in cellular morphology from yeast to hyphae, i.e., it forms germ tubes from the yeast phase cells of *C. albicans* (5). Morphogenetic changes like hyphal or pseudohyphal growth enables the cell to propagate into the host tissue as a preliminary manifestation of invasion and spread of pathogenesis. Hence, the inducible GlcNAc catabolic pathway in *C. albicans* may be important in pathogenesis.

To elucidate the role of GlcNAc in pathogenicity, we cloned the gene determining glucosamine-6-phosphate deaminase (*NAG1*), the terminal enzyme of this aminosugar catabolic pathway (6). *NAG1* is transcriptionally induced by GlcNAc. This inducible pathway consists of four enzymes—namely GlcNAc permease, GlcNAc kinase, GlcNAc-6-phosphate deacetylase, and GlcN-6-phosphate deaminase, all of which act sequentially on GlcNAc to generate fructose-6-phosphate that is fed into the glycolytic pathway (2, 3, 6–8). Our laboratory recently cloned two more important genes for virulence factors in *C. albicans*: *ACPR* (*CPH1*), a transcription factor homologous to *STE12* of *Saccharomyces cerevisiae* that regulates the mating pathway (9–11), and *CaSTE7* (*HSTE7*), a mitogen-activated protein

kinase involved in the pseudohyphal formation pathway (12, 13). Mutant strains of *C. albicans* that are defective for hyphal formation are avirulent in animal models suggesting that hyphal formation is necessary for virulence and dissemination (14).

In the present study, we isolated and characterized the genomic clone of *NAG1*. The genes of GlcNAc catabolism-encoding GlcNAc-6-phosphate deacetylase (*DAC1*), GlcN-6-phosphate deaminase (*NAG1*), and a hexokinase (*HXK1*), exist in a cluster in the genomic clone. All three genes are coordinately regulated at the transcriptional level. The organization and regulation of the *NAG1* promoter, a bidirectional promoter, controlling *NAG1* and *DAC1* expression in the opposite orientations are reported. Deletions of the promoter fused to the β -galactosidase gene (*LAC4*) of *Kluyveromyces lactis* (15) delineated the regions necessary for the induction by GlcNAc to two regions: –400 to –195 and –200 to –1 with respect to the *NAG1* start codon. Gel mobility-shift assays (GMSA) of the regions with induced and uninduced protein extracts of *C. albicans* narrowed down the distal region to 214 bp (–408 to –195) and the proximal region to 70 bp (–200 to –131). “Footprinting” (DNase protection method) of both the probes was performed to identify specific protein-binding sequences of uninduced and GlcNAc-induced extract.

Materials and Methods

Cloning and Sequencing of the Deaminase Gene. The 0.8-kb *EcoRI* fragment from plasmid pD6 containing the *NAG1* cDNA (6) was used as a probe to screen a genomic library of *C. albicans* SC5314 in *E. coli* strain with cosmids λ -EMBL3. Clones ($n = 16$) were obtained and characterized with restriction enzyme digestion and Southern hybridization with *NAG1* cDNA. One of the clones with cosmid λ ED14, which contained an insert of ≈ 16.3 kb was purified from plaque for further analysis. A 4-kb *SaII* fragment containing the full-length *NAG1* gene was subcloned into the *SaII* site of plasmid pBluescript II KS(+). The insert in the resultant clone, pED4, was sequenced (GenBank accession no. 6137104, locus AF079804) and analyzed. A homology search of the genomic sequence and amino acid sequence was performed by using the BLAST, the gapped BLAST software (16, 17), and CLUSTALW (18).

Southern Analysis. From *C. albicans* genomic DNA and the *NAG1* genomic clone, λ ED14 was isolated (19), digested with various

Abbreviations: GMSA, gel mobility-shift assay; GlcNAc, *N*-acetylglucosamine.

*Present address: Building 6A, Room B1A-13, Laboratory of Eukaryotic Gene Regulation, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892.

†To whom reprint requests should be addressed. E-mail: adatta@jnu.ernet.in.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.250452997. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.250452997

Table 1. Sequence and position of the primers used in PCR amplification, 5' end mapping, GMSA, and footprinting

Name	Sequence 5' to 3'	Gene and position
N2	TCCAACCCCTAACGATGCTG	<i>NAG1</i> : +19 to +37
N6	CCAGATCTGCGTTGGAAAATATAGCTTG	<i>NAG1</i> : Complementary to +7 to +24, with a 10-bp adapter
N7	TTAAACTCAGCAACATCCTCTGCGG	<i>NAG1</i> : -497 to -473
N9	GTCGGATCCTACCAACACTATGC	<i>NAG1</i> : -1,725 to -1,704
N10	GGTTGATGAGAAATTGGGGTATT	<i>NAG1</i> : Complementary to -1 to -23
N11	GGGGATGACCCGGTTGGAAGG	<i>NAG1</i> : Complementary to +120 to +102
N12	CCATGGCAATATATATATAGGC	<i>NAG1</i> : Complementary to -200 to -179
N13	CCATGGCTGCCACATCACTTGC	<i>NAG1</i> : Complementary to -195 to -217
N16	CTCAAAAACGTGTTACATTG	<i>NAG1</i> : -139 to -119
N17	ATGTCATTTACTAGATTAC	<i>DAC1</i> : +1 to +20
N19	GCGGATGACAAATTCT	<i>DAC1</i> : Complementary to +106 to +91
N30Spl	CCCAGGTGCTAATATTGC	<i>NAG1</i> : -408 to -390

Bases different from the homologous sequence are in bold. Position +1 refers to translation start codon ATG.

enzymes, electrophoresed, and transferred to GeneScreen Plus membrane (NEN Life Science Products). The blots were hybridized with a ³²P-labeled probe derived from pED4 as a *Sall*-*HindIII* 1.15-kb fragment and *NAG1* cDNA. Other blots with digested *C. albicans* genomic DNA were hybridized with the probes derived from pED4 as 1.15-kb *Sall*-*HindIII* (*DAC1*), 1.33-kb *HindIII*-*XbaI* (*NAG1*), and 1.44-kb *XbaI*-*Sall* (*HXK1*) fragments.

Induction of *C. albicans* by GlcNAc. *C. albicans* cells were precultured and induced by GlcNAc as described (6).

Mapping of the Transcription Start Sites. The 5' end of the *NAG1* and *DAC1* mRNA was determined by primer extension. Total RNA from uninduced and GlcNAc-induced *C. albicans* cells was prepared (20) and treated with RNase-free DNase I. For *NAG1*, oligo N11 (Table 1) was used as a primer. The ³²P end-labeled primer (5 × 10⁴ cpm) was annealed to 5 μg of total RNA and reverse transcribed by using rTth polymerase (Perkin-Elmer) according to the manufacturer's instructions. The first strand of cDNA was precipitated, redissolved in loading dye, and resolved on a 9% polyacrylamide sequencing gel (21). In parallel, a sequencing reaction with pED4 as a template with the same primer was loaded. As an internal control, primer N11 and a sense primer N2 (see Table 1) were used to amplify the first strand of cDNA. For *DAC1*, oligo N19 (see Table 1) was used for primer extension and sequencing. Oligos N17 and N19 (see Table 1) were used for PCR verification of the first strands.

Northern Blot Analysis. Total RNA (40 μg each) from glucose-grown (uninduced) or GlcNAc-induced cells was analyzed by Northern blotting. Plasmid pED4 was digested with *HindIII*, *XhoI*, and *XbaI* resulting in three independent probes—namely, a 1.15-kb *XhoI*-*HindIII* fragment containing *DAC1*, a 1.33-kb *HindIII*-*XbaI* fragment containing *NAG1*, and a 1.42-kb *XbaI*-*HindIII* fragment containing *HXK1*. The *C. albicans ACT1* (22) probe used as control was derived from pCActin as a 1.5-kb *Sall*-*EcoRI* fragment. In case of the induction kinetics, the probes used were the 0.8-kb *EcoRI* fragment from the *NAG1* cDNA clone (6), 0.925-kb *BstNI*-*BstXI* fragment from pED4 corresponding to *DAC1*, and a 444-bp *NcoI*-*Sall* fragment from pED4 corresponding to the *HXK1*.

Promoter Fusion, Deletions, and Subcloning in *Candida* Vector. The *NAG1* upstream region was PCR-amplified by using oligos N9 and N10 from pED4 (see Table 1). The PCR product was cloned into the *PvuII* site of pRSLAC4, which contains the complete *K. lactis LAC4* gene-encoding β-galactosidase, to create pRSNAGULAC4. A *BamHI* fragment containing the *NAG1*

promoter and *LAC4* fusion from pRSNAGULAC4 replaced the *BamHI* fragment (containing *C. albicans ACT1* promoter and *LAC4* fusion) in pCL01 (15) to create the plasmid pCL10. To create unidirectional deletions in the promoter, the *BamHI* insert from pRSNAGULAC4 was cloned in pUC19 at the *BamHI* site. This insert was recloned in the *KpnI*-*Sall* site of pBluescript II KS(+) (pNAGULAC4) to facilitate deletions. The deleted promoter *LAC4*-fusion inserts were cloned into *BamHI*-digested pCL01 as for pCL10. All these constructs were transformed into *C. albicans* CAI4 (19).

Measurement of β-Galactosidase Activity. Transformants were grown in GPK (glucose/peptone/KH₂PO₄) and induced in NPK (GlcNAc/peptone/KH₂PO₄). Cell extracts were prepared as described (19). The *lacZ* assay was initiated by adding 0.2 ml of ONPG (*o*-nitrophenyl-β-D-galactoside, 4 mg/ml in Z buffer) and continued until the mixture acquired a pale-yellow color. The reaction was terminated by adding 0.5 ml of 1 M Na₂CO₃. The end product was measured at 420 nm. The specific activity was expressed as nmol per min/mg protein. The plasmid copy number of each transformant was analyzed by Southern blotting and normalized by comparing the signal intensity against the genomic copy of endogenous *NAG1* promoter. The β-galactosidase activity was normalized based on copy number.

GMSA. Cell extracts were prepared with glass beads by using breakage buffer [200 mM Tris-HCl, pH 8.0/10% (vol/vol) glycerol/100 mM (NH₄)₂SO₄/1 mM EDTA/10 mM β-mercaptoethanol/1 mM PMSF]. The homogenate was centrifuged at 125,000 × g for 45 min at 4°C. The supernatant was subjected to a 40–60% (NH₄)₂SO₄ precipitation. The pellet was dissolved in 0.4 ml of protein buffer (23) per gram of cells and dialyzed for 8 h against the same buffer with two changes; the cell extract was partially purified by a DEAE-Sephacel, column and the protein was eluted with a linear gradient of 0.1:1 (vol/vol) M NaCl. The probes for GMSA were derived from pT12BBP containing *NAG1* promoter along with eight N-terminal codons of the *NAG1* ORF. Oligos N7 and N13 (see Table 1) were used to amplify the 303-bp probe of the distal element (-497 to -195) from pT12BBP. A 234-bp probe containing the proximal element (-200 to +24 plus an added 10-base adapter) was prepared by PCR with oligos N12 and N6 (see Table 1) of the same plasmid. To narrow down the binding region, this probe was digested with *AflIII*. The fragments *NcoI*-*AflIII* (70 bp) and *AflIII*-*BglII* (164 bp) were used for competition experiments. The fragments were labeled either by phosphorylating with [γ-³²P]ATP or by end filling with [α-³²P]dATP. Alternatively, when the probes were prepared by PCR, the oligos were labeled by using [γ-³²P]ATP and T4 polynucleotide kinase. For competition, a 54-bp

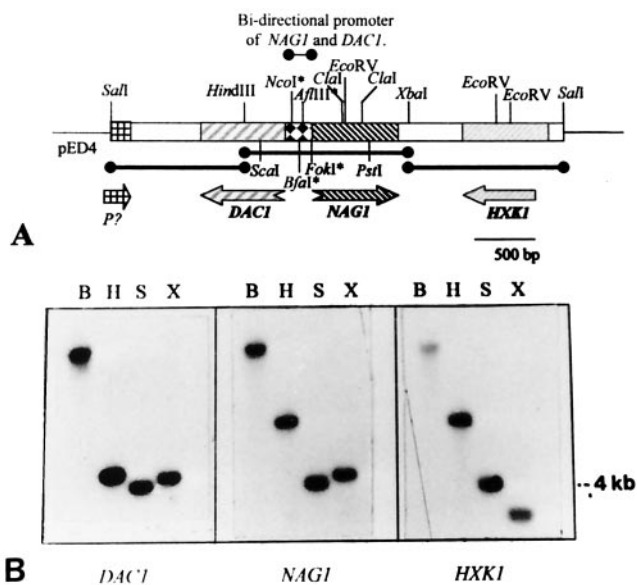


Fig. 1. Genomic organization of GlcNAc catabolic genes and NAG gene cluster in *C. albicans*. (A) Schematic representation of the clustered organization of the 4-kb *SalI* genomic clone fragment in pED4. The hatched boxes represent the three ORFs. Restriction nuclease sites marked with asterisks (*) are nonunique. The shaded arrows at the bottom represent the direction and position of the various reading frames: *DAC1*, *NAG1*, and *HXK1* (arrow depicts a partial sequence). The bar indicates the scale of the map. The dotted box denotes the bidirectional promoter of *NAG1* and *DAC1*. The bold lines represent the fragments used for probes for Northern and Southern analyses. P indicates the partial sequence of putative permease gene. (B) Southern blots of *C. albicans* SC5314 genomic DNA digested with *Bam*HI (B), *Hind*III (H), *Sal*I (S), and *Xba*I (X) were hybridized with the probes as indicated in the three panels. The common 4-kb *Sal*I band detected in all three blots represents the gene cluster.

fragment obtained by digestion of N6-N12 PCR product with *Nco*I and *Bfa*I (−200 to −147) was taken. The binding reaction was carried out in a 20- μ l volume (23) with probe DNA (\approx 1 ng, 10^4 cpm). For competition, unlabeled competitor DNA was added in the binding reaction before the addition of the labeled probe.

DNase I Footprinting. Probes were generated by PCR with pT12BBP as a template and end-labeled with [γ - 32 P]ATP. Two probes spanning the −408 to −195 (214 bp) and −200 to −119 (82 bp) region were PCR amplified by using oligos N30ssPI (see Table 1) and N13, and N12 and N16 respectively (see Table 1). The 82-bp probe contained the 70-bp region of *Nco*I–*Afl*III fragment used for GMSA. DNase I digestion of the bound complex was carried out for 1 min. In case of the 214-bp probe we used 0.08 units of DNase I with varying amounts of poly(dI-dC). The 82-bp probe was digested with 0.11 and 0.04 units of DNase I. The reaction was stopped by the addition of 180–200 μ l of DNase I stop buffer (192 mM sodium acetate/32 mM EDTA/0.14% SDS/11.5 μ g of tRNA), extracted with phenol-chloroform, ethanol precipitated, and dissolved in 10 μ l of sequencing dye. Approximately 5×10^3 cpm was loaded in each lane in a 9% sequencing gel.

Results

Organization of the NAG1 Genomic Clone. Sequence analysis of pED4 revealed that it has a 1.725-kb upstream region, 0.747-kb *NAG1* ORF, and a 1.342-kb downstream region (Fig. 1A; GenBank accession no. 6137104, locus AF079804). Although the upstream region of the *NAG1* gene did not have a consensus TATA element, a TATA-like sequence (5′-CCATAAAAGGCC-3′ at position −59 with respect to the *NAG1* translational start codon ATG) was identified by computer analysis. Computational data further re-

vealed a putative Cap signal 5′-CCAATTTC-3′ at −17 and the polyadenylation sequence 5′-AATAAA-3′ at 475 nucleotides downstream of the *NAG1* stop codon at position +1,222 with respect to the *NAG1* translational start codon. A poly(A)-rich sequence 5′-GGAGCAAAAAATGT-3′ (−164 to −150 with respect to the *NAG1* translational start codon) exists in the *NAG1* promoter and is quite similar to the sequence found in the *E. coli* NagC-binding region, box G1 5′-TCCATTTACGAT-GAAAAAATG-3′ (24). NagC is a repressor in the GlcNAc catabolic pathway in *E. coli*. The sequence immediately upstream of the *NAG1* start site contains at least two more poly(A)-rich elements similar to the −164 element and could be an additional upstream regulatory element. In *Candida glabrata*, there is a poly(dA-dT) element, adjacent to a metal responsive element in *AMT1* upstream sequence, a transcription factor that plays a role in transcriptional activation (25).

Identification of Genomic Cluster Containing GlcNAc Catabolic Pathway Genes. Sequence homology searches revealed a *NAG* gene cluster at the *NAG1* locus. The region 766 to 1485 of *NAG* locus showed a strong homology to the GlcNAc-6-phosphate deacetylase gene from *Caenorhabditis elegans*, *Hemophilus influenzae*, and other species. In addition, the 3′ untranslated region of *NAG1*, the region 3,006 to 3,855 of *NAG* locus, showed homology to hexokinase genes of *Kluyveromyces*, mouse, and other species. The translational reading frames of *DAC1* and *HXK1* are in the antisense strand of *NAG1*. Homology to a membrane protein found in the extreme 5′ end of the pED4 sequence could be part of a membrane-associated protein, possibly a permease homologue in the pathway. Comparative analysis of Southern blots of λ ED14 and *C. albicans* genomic DNA probed with *NAG1* upstream sequence and *NAG1* cDNA (data not shown) probed with that of *C. albicans* genomic DNA by using *DAC1* and *NAG1*, as well as *HXK1* (Fig. 1B) revealed that the *C. albicans* genome, as well as λ ED14, contains a single *Sal*I fragment of 4-kb size representing the full-length *DAC1*, *NAG1*, and a partial sequence of *HXK1*. The blot with the 5′ probe suggests that *DAC1* is a single-copy gene, completely contained in λ ED14 DNA and existing within the same locus as the *NAG1* cDNA.

CLUSTALW analysis of the translated sequences of *DAC1* with other GlcNAc-6-phosphate deacetylase sequences, the partial *HXK1* with other hexokinase sequences, and a similar analysis of the *NAG1* sequences substantiated that the three genes are the representative of the respective enzymes in *C. albicans* (18). Moreover, the conserved residues for sugar binding in hexokinase and the potential active-site residues of the deaminase sequences are present in the *Candida* sequences. The remaining sequence of *HXK1* containing the ATP-binding domain is expected to be present in the 3-kb *Xba*I fragment 3′ of the *NAG1* gene.

Primer Extension Analysis. Primer extension analysis of GlcNAc-induced total RNA revealed two transcription start sites in the *NAG1* transcript at −8 and at −29 positions with respect to ATG (data not shown). Although there is a strong −8 stop in the 5′ mapping, it is unclear how the two mapped 5′ ends, −8 and −29, are used *in vivo*. It is likely that *C. albicans* genes could have multiple 5′ ends, similar to other yeasts (26, 27) and unlike most mammalian systems (28). The *DAC1* transcript mapped to an adenine residue (data not shown) at position +1 with respect to the translation start codon. We presume that the atypical multiple transcription start sites could be caused by a bidirectional promoter lacking the TATA consensus.

Coordinated Regulation of the NAG Cluster. To examine the effect of GlcNAc on the transcription of the genes responsible for its catabolism, a Northern analysis of the GlcNAc-induced RNA was performed. DNA fragments corresponding to *NAG1*, *DAC1*, and *HXK1* were radiolabeled and used as probes, and the *C. albicans* actin gene, *ACT1*, was used as control. The results

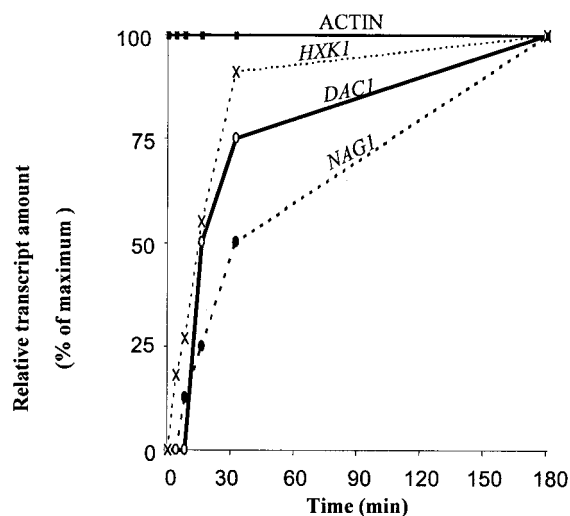


Fig. 2. Induction kinetics of the GlcNAc-induced RNA. RNA was isolated from 4, 8, 16, 32, and 180 min of GlcNAc-induced cells. Four copies of blots were prepared, and the various transcript levels were analyzed by hybridization to individual gene-specific probes. A graph representing the optical quantitation of density from the autoradiogram is presented. The maximum level of each transcript has been taken as 100%, and the various time points are represented as fractions thereof. The 0 time refers to no GlcNAc addition. The signal for actin was taken as control. The transcripts are depicted as *NAG1* (●), *DAC1* (○), *HXK1* (×), and actin (■).

showed that the genes *NAG1*, *DAC1*, and *HXK1* are transcribed in response to GlcNAc induction and remain uninduced when grown in glucose (data not shown). Induction kinetics showed that the *NAG1* transcript appears at 8 min, *DAC1* at 16 min, and *HXK1* at 4 min with respect to induction by GlcNAc (Fig. 2). The difference in appearance of the transcripts may occur, because GlcNAc kinase is the first catabolic pathway enzyme to act on the aminosugar after it has entered the cell. All three transcripts appeared within minutes of each other and were close to a steady state at 30 min. This expression pattern for the *NAG1*, *DAC1*, and *HXK1* establishes a coordinated expression that is induced by GlcNAc.

Deletion Analysis of the *NAG1* Promoter. The *NAG1* promoter was not GlcNAc responsive in *S. cerevisiae* when analyzed using a *lacZ* fusion construct. This result prompted us to use a homologous system to study the promoter. *C. albicans* CAI4 transformants of pCL10 and its deletion clones were assayed for β -galactosidase activity. The transformants showed a high level of induction by GlcNAc, whereas the activity in the transformants of pCL01 (actin promoter fused to *LAC4*) remained unchanged (Fig. 3). Leuker *et al.* (15) reported that pCL01 carries the *C. albicans* autonomously replicating sequence CARS1 and could integrate at chromosomal sites as well as sustain autonomous replication.

The β -galactosidase activities for the constructs containing 1,726, 1,550, 400, 200, and 130 bp of *NAG1* 5'-flanking region were 98.9, 60.4, 12.9, 8.5, and 10.2 Miller units when induced with GlcNAc as compared with 19.1, 59.6, 3.3, 5.6, and 13.1 units, respectively, in an uninduced state. The promoter deleted to -1,550 eliminated GlcNAc induction, indicating the presence of a strong positive regulator sequence in the region between -1,726 and -1,550. Analysis of other deletions through -512 showed no change in GlcNAc inducibility. The region from -1,550 to -512 either has a negative regulatory sequence or is inactive as a promoter element because of the presence of an actively transcribing *DAC1* ORF. The deletion of the promoter through -0.4 kb resulted in a 4-fold induction. This induction

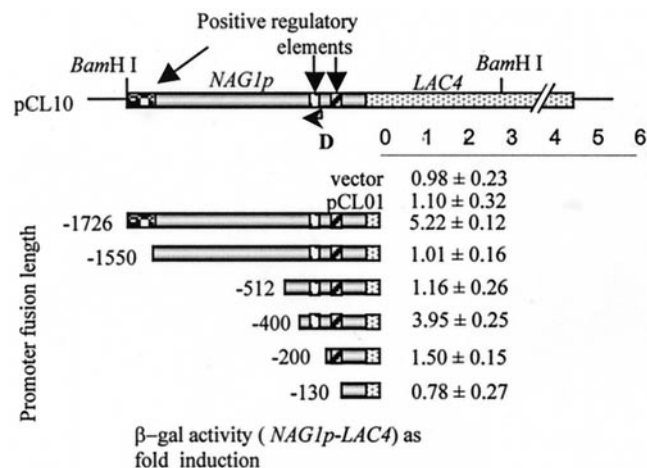


Fig. 3. Analysis of *NAG1* promoter deletion constructs. A schematic representation of the *NAG1* promoter-reporter construct indicating relative positions of three putative positive regulatory elements depicted by one stippled and two hatched boxes. Constructs shown at lower left contain 1,726, 1,550, 400, 200, and 130 bp of *NAG1* 5' flanking region from translation start ATG. The β -galactosidase activities are represented as fold induction by GlcNAc (the ratio of induced to uninduced β -galactosidase activity). The data represent an average of three independent transformants each tested thrice. The putative *DAC1* start site is indicated by arrow D.

indicated the presence of a GlcNAc-responsive repressor element(s) between -512 and -400. On deletion of an additional 200 bp, resulting in a 0.2-kb promoter, inducibility was almost negligible. Further experiments revealed an essential element in the region -200 to -130 that decreased the inducibility and reduced the expression to less than 1-fold, indicating the loss of the minimal promoter sequence.

DNA Protein Interaction. Based on the *in vivo* deletion analysis, the three important upstream regions of the *NAG1* gene responsible for GlcNAc induction (-1,726 to -1,550; -400 to -200; -200 to -1), were chosen for GMSA. As mentioned, the -512 to -400 region has a negative regulatory element, and -400 to -200 contains a positive regulatory element that causes a significant increase in the inducibility with respect to the proximal (-200 to -1) promoter. To examine whether this region binds to any GlcNAc-inducible factors, GMSAs were carried out with an end-labeled 336-bp *NcoI-ScaI* fragment (-535 to -200) or the 303-bp PCR product of N7 and N13. Two distinct protein-DNA complexes were formed with induced cells extract. Competition with excess unlabeled probe DNA diminished the complex formation (Fig. 4A), although nonspecific competitor DNA did not affect binding. Further characterization revealed that a 214-bp region (-408 to -195) is able to form the two complexes, whereas the remaining 89-bp region failed to show any binding (data not shown).

Interaction of GlcNAc-Inducible Factors at -200 to -130 of *NAG1*. The proximal domain of the *NAG1* upstream region covering a major portion of the bidirectional promoter between *NAG1* and *DAC1* was tested for binding to any GlcNAc-inducible factor(s). GMSA was carried out with an end-labeled 234-bp PCR-amplified fragment from positions -200 to +34. Protein-DNA complex was formed only with induced cell extract (Fig. 4B, lane I) but not with uninduced extracts (Fig. 4B, lane U). Excess unlabeled probe effectively competed for the complex (Fig. 4B; Lane I+C), whereas addition of a nonspecific competitor like linear pUC19 did not diminish complex formation (data not shown). To characterize the region responsible for binding to the protein(s)

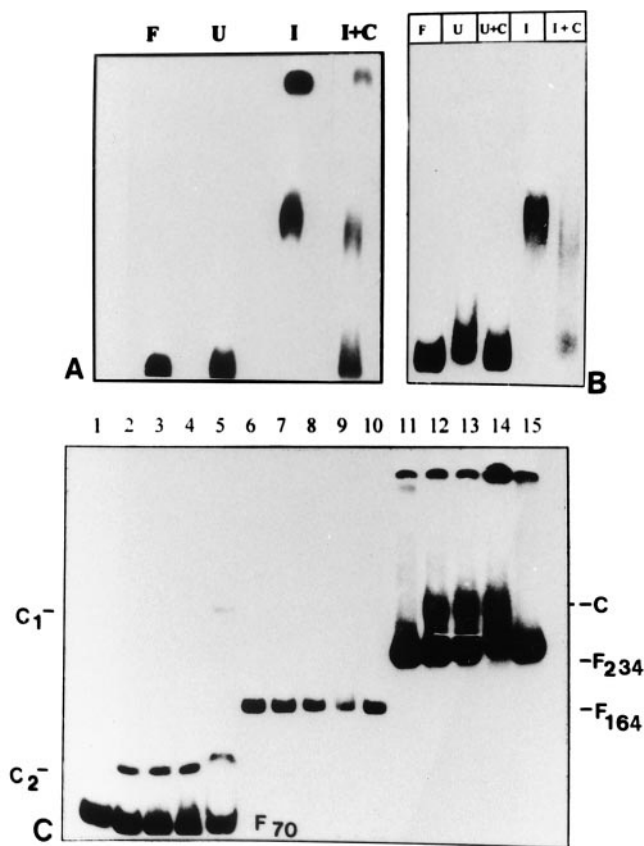


Fig. 4. DNA GMSA of different regions of *NAG1* promoter. (A) The -497 to -195 region of the *NAG1* promoter was analyzed in a GMSA with $20 \mu\text{g}$ of crude induced (I) and crude uninduced (U) extract. C_1 and C_2 indicate the two complexes formed by the probe with induced extract. I + C indicates the competition assay with 25-fold molar excess of the unlabeled probe DNA. (B) The assay of the proximal region of the *NAG1* promoter (-200 to $+34$) was performed with crude cell extracts. Then $20 \mu\text{g}$ of crude induced (lanes I and I + C) or uninduced (lanes U and U + C) cell extracts were incubated with an end-labeled 234-bp probe either in the absence (lanes U and I) or in the presence (lanes U + C and I + C) of unlabeled specific-competitor DNA. The reaction mixture was resolved on a 6% nondenaturing PAGE. Lane F indicates free probe. (C) DNA GMSA with probes 70 bp (-200 to -131 ; lanes 1–5), 164 bp (-130 to $+34$; lanes 6–10), and 234 bp (-200 to $+34$; lanes 11–15). The amounts of proteins used were $20 \mu\text{g}$ (lanes 2, 7, and 12), $40 \mu\text{g}$ (lanes 3, 8, and 13), and $60 \mu\text{g}$ (lanes 4, 9, and 14). Lanes 1, 6, 11, and 15 contain free probe as indicated by F_{70} , F_{164} , and F_{234} . Protein ($20 \mu\text{g}$) partially purified through DEAE-Sephacel column was used in lanes 5 and 10. Protein–DNA complexes (C_1 , C_2) shown on both sides of the gel were resolved on an 8% nondenaturing gel. Although the amount of the 164-bp probe is much less than that of either the 70-bp or 234-bp probes, a longer exposure of the gel did not reveal any new bands.

further, the 234-bp DNA fragment was digested with *Afl*III to obtain 70 bp (-200 to -131) and 164 bp (-130 to $+34$) fragments. When assayed with crude extract, the 70-bp probe formed only one complex (C), although the same probe resulted in two complexes (C_1 and C_2) with partially purified extract of *C. albicans* (Fig. 4C).

The levels of complex C_2 increased with increasing concentration of the crude extract, and C_1 appeared only high-protein concentrations (data not shown). When the binding reaction with $60 \mu\text{g}$ of crude extract was preincubated with a 25-fold and 50-fold molar excess of unlabeled 70-bp competitor DNA, complex formation was abolished. When challenged with 100-fold molar excess of unlabeled nonspecific DNA, there was no significant reduction in the formation of either C_1 or C_2 . The

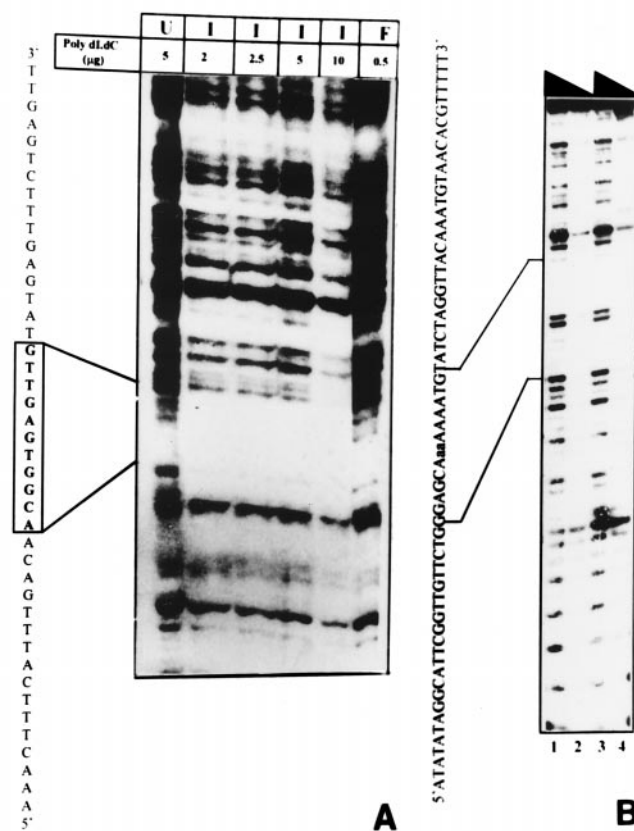


Fig. 5. DNase I footprinting assay. (A) DNase I footprinting assay of the 214-bp probe with crude uninduced (U) and induced (I) cell extracts. The footprinting region (Box B) corresponds to the nucleotide sequence as shown to the left. F indicates DNase I-digested free probe. (B) DNase I footprinting assay of the 82-bp probe performed with crude induced cell extract ($140 \mu\text{g}$). DNase I used was 0.11 units (lanes 1 and 3) and 0.04 units (lanes 2 and 4). The sequence shown on the left indicates the position of the footprint (Box A).

proximal 70-bp probe (-200 to -131) used for competition in an assay with the distal probe (-408 to -195) and vice versa did not result in any competition effect (data not shown). This result suggests that the two elements are bound by different sets of proteins.

Nucleotide Sequence for Complex Formation in -400 and -200 Elements. To localize the nucleotide sequence responsible for complex formation, induced cell extract was footprinted with an end-labeled 214-bp probe (-408 to -195) by using DNase I. The footprinting experiment revealed an 11-bp protected region in the *NAG1* coding strand 5'-ACGGTGAGTTG-3' (-291 to -281 ; Fig. 5A). However, the footprint falls in the region expected to be in the reading frame of *DAC1*.

DNase I footprinting was also performed with induced crude cell extracts and an end-labeled 82-bp probe (-200 to -119) encompassing a 70-bp region that was used in GMSA. The results revealed a 15-bp protected region in the *NAG1* coding strand 5'-GGAGCAAAAAAATGT-3' (-164 to -150 ; Fig. 5B). This sequence is quite similar to the sequence found in the *E. coli* NagC binding region, box G1 5'-TCCATTTACGATGAAAAAATG-3' (24, 29). The window in the footprint is interrupted by two DNase I hypersensitive adenine bases at positions 7 and 8.

Discussion

GlcNAc induces numerous changes in the *C. albicans* morphology as well as physiology. Apart from inducing its own catabolic

pathway, GlcNAc also serves as a constituent of cell wall chitin and a part of the carbohydrate moiety of various glycoproteins. Moreover, in the presence of GlcNAc the cell adapts to use the aminosugar as a sole carbon source (3). GlcNAc induces germ tubes in *C. albicans* within 3 h of induction. Various enzymes in the GlcNAc pathway, including GlcNAc-kinase, GlcNAc-deacetylase, and Nag1, reach steady-state levels within the same period (3, 7). Such a transition usually involves transcriptional regulation. Thus, a rapid transcriptional activation is expected for Nag1 and other genes in the GlcNAc catabolic pathway. Northern blots showed that transcription of all three GlcNAc catabolic pathway genes was activated within minutes of each other (Fig. 2). Interestingly, the *NAG1* transcript is undetectable in uninduced RNA, appears within 8 min after induction by GlcNAc, and approaches a steady-state level in less than 1 h. Therefore, transcription factor(s) might be induced or activated, which in turn activates the expression of all genes in the GlcNAc catabolic pathway in a coordinated manner. A possible regulatory mechanism involving either a single transcription factor or a master switch might activate the transcription of the GlcNAc catabolic genes simultaneously.

Clusters of functionally related genes are a general feature of prokaryotes and are less prevalent in eukaryotes. However, metabolic pathways in several fungi have been found organized in such clusters, e.g. proline and ethanol use in *Aspergillus*, penicillin biosynthesis in *Penicillium*, and mycotoxin biosynthesis in *Fusarium* (30). Nutrient use pathways increase metabolic versatility by enabling organisms to use a variety of complex compounds and increase the efficacy of the biochemical apparatus. These genes are functionally regulated and activated in a coordinate manner. Use of GlcNAc in *C. albicans* is an alternate pathway, and in such systems bidirectional transcription is known to be a potential mechanism to coordinate the expression of adjacent genes. Such a strategy is often seen in the regulation of fungal metabolic clusters where specific pathway regulators have been identified. In *E. coli*, GlcNAc and the *N*-acetylgalactosamine (GalNAc) metabolic pathway, genes are also organized in clusters and possibly have common regulatory mechanisms (24, 31). To the best of our knowledge, such a gene cluster in *Candida* has not been reported thus far. The *NAG* genes (*NAG1*, *DAC1*, and *HXX1*) exist in a single locus and are subject to a transcriptional activation in response to a single inducer GlcNAc.

Another interesting feature is that the reading frames of the genes *NAG1* and *DAC1* are in opposite orientation, which indicates that the intervening promoter acts bidirectionally (Fig. 1A). In *C. albicans*, the organization of the *NAG* genes is quite

different from that in *E. coli* (29, 32). Moreover, *NAG1* and *DAC1* share a divergent promoter, whereas a third gene, *HXX1*, downstream of *NAG1*, has a separate promoter. Deletion analysis suggested that the active *NAG1* promoter extends to at least -400 bp upstream of the ATG with transcription starting at -8 and -29. There seems to be a synergistic effect of both complexes on transcriptional activation when the distal (Box B) and the proximal (Box A) regulatory regions are present upstream of the reporter gene. The removal of Box B and Box A abolished induction, whereas removal of only Box B reduced transcription ≈50%. GMSA and the footprinting experiments strongly supported the conclusion that these regions contain elements interacting with DNA-binding proteins. It is likely that there are two different GlcNAc-inducible factors binding to the promoter at either of the two regulatory sites studied. Each region shows two complexes with the induced extracts (Fig. 4A and C). The proteins that would interact at the distal regulatory region, Box B, form very strong complexes, easily detectable in crude preparations, indicating a possible abundance of the proteins or strong activation of preexisting factors in response to GlcNAc. In the case of Box B, the unique feature is that it falls downstream of the *DAC1* ATG in the opposite strand. The binding of regulatory proteins to this region is significant, because it can be concluded that the proteins, in conjunction with the proteins binding to the downstream regulatory site, confer more inducibility to the promoter. This binding may be either because of direct interaction with the protein or via bridging proteins. The sequence analyses revealed at least two more regions similar to Box A immediately upstream of the *NAG1* start site. These poly(A) elements could represent additional binding sites for the protein(s), although we could not find any binding in the conditions used.

Our results show that GlcNAc regulates all three genes, *DAC1*, *HXX1*, and *NAG1*, at the transcriptional level, and the transcript level reaches a steady state, maximum around 30 min after induction. It can be inferred that these three genes are regulated in a coordinated manner. The GlcNAc regulation of the *NAG1* pathway genes is mediated by the involvement of at least two inducible putative transcription factors. These factors may themselves be regulated by various protein-protein interactions and other mechanisms in response to induction by GlcNAc.

We thank J. F. Ernst for providing the plasmid pCL01 and pRSLAC4, and K. Ganesan, Subhra Chakraborty, and Niranjan Chakraborty for critical reading of the manuscript. This work was supported by grants from Department of Biotechnology, Council of Scientific and Industrial Research, and University Grants Commission India.

- Odds, F. C. (1988) *Candida & Candidosis: A Review & Bibliography*, Balliere (Tindall, London), 2nd Ed.
- Datta, A., Ganesan, K. & Natarajan, K. (1989) *Adv. Microb. Physiol.* **30**, 53–88.
- Singh, B. & Datta, A. (1979) *Biochim. Biophys. Acta* **557**, 248–258.
- Jenkinson, H. F. & Sheperd, M. G. (1987) *J. Gen. Microbiol.* **133**, 2097–2106.
- Mattia, E., Carruba, G., Angiolella, L. & Cassone, A. (1982) *J. Bacteriol.* **152**, 555–562.
- Natarajan, K. & Datta, A. (1993) *J. Biol. Chem.* **268**, 9206–9214.
- Biswas, M., Singh, B., Rai, Y. P. & Datta, A. (1982) *Indian J. Exp. Biol.* **20**, 829–834.
- Singh, B. & Datta, A. (1979) *Biochem. J.* **170**, 427–431.
- Liu, H., Kohler, J. & Fink, G. R. (1994) *Science* **266**, 1723–1726.
- Malathi, K., Ganesan, K. & Datta, A. (1994) *J. Biol. Chem.* **269**, 22945–22951.
- Singh, P., Ganesan, K., Malathi, K., Ghosh, D. & Datta, A. (1994) *Biochem. Biophys. Res. Commun.* **205**, 1079–1085.
- Clark, K. L., Fledmann, P., Dignard, D., Larocque, R., Brown, A. J., Lee, M. G., Thomas, D. Y. & Whiteway, M. (1995) *Mol. Gen. Genet.* **249**, 609–621.
- Singh, P., Ghosh, S. & Datta, A. (1997) *Gene* **190**, 99–104.
- Lo, H.-J., Köhler, J. R., DiDomenico, B., Loebenberg, D., Cacciapuoti, A. & Fink, G. R. (1997) *Cell* **90**, 939–949.
- Leuker, C. E., Hahn, A. M. & Ernst, J. F. (1992) *Mol. Gen. Genet.* **235**, 235–241.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
- Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
- Rose, M. D., Winston, F. & Hieter, P. (1990) *Methods in Yeast Genetics: A Laboratory Course Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
- Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1994) *Current Protocols in Molecular Biology* (Wiley, New York).
- Losberger, C. & Ernst, J. F. (1989) *Nucleic Acids Res.* **17**, 9488.
- Company, M., Adler, C. & Errede, B. (1988) *Mol. Cell Biol.* **8**, 2545–2554.
- Plumbridge, J. (1995) *EMBO J.* **14**, 3958–3965.
- Zhu, Z. & Thiele, D. J. (1997) *Cell* **87**, 459–470.
- Maicas, E., Shago, M. & Freisen, J. D. (1990) *Nucleic Acids Res.* **18**, 5823–5828.
- Slusher, L. B., Gillman, E. C., Martin, N. C. & Hopper, A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9789–9793.
- Kelley, D. E., Coleclough, C. & Perry, R. P. (1992) *Cell* **29**, 681–689.
- Plumbridge, J. & Kolb, A. (1998) *Nucleic Acids Res.* **26**, 1254–1260.
- Keller, P. N. & Hohn, T. M. (1997) *Fungal Genet. Biol.* **21**, 17–29.
- Reizer, J., Ramseier, T. M., Reizer, A., Charbit, A. & Saier, M. H., Jr. (1996) *Microbiology* **142**, 231–250.
- Vogler, A. P. & Lengeler, J. W. (1989) *Mol. Gen. Genet.* **219**, 97–105.