The yeast BDF1 gene encodes a transcription factor involved in the expression of a broad class of genes including snRNAs

Zoi Lygerou, Christine Conesa, Pascale Lesage, Robert N. Swanson, Anny Ruet, Marian Carlson, André Sentenac and Bertrand Séraphin

EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany, Service de Biochimie et de Génétique Moléculaire, Centre d’Etudes Nucléaire de Saclay, F-91191 Gif sur Yvette Cedex, France and Department of Genetics and Development and Institute of Cancer Research, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

Received August 22, 1994; Revised and Accepted November 9, 1994

ABSTRACT

While screening for genes that affect the synthesis of yeast snRNPs, we identified a thermosensitive mutant that abolishes the production of a reporter snRNA at the non-permissive temperature. This mutant defines a new gene, named BDF1. In a bdf1-1 strain, the reporter snRNA synthesized before the temperature shift remains stable at the non-permissive temperature. This demonstrates that the BDF1 gene affects the synthesis rather than the stability of the reporter snRNA and suggests that the BDF1 gene encodes a transcription factor. BDF1 is present in single copy on yeast chromosome XII, and is important for normal vegetative growth but not essential for cell viability. bdf1 null mutants share common phenotypes with several mutants affecting general transcription and are defective in snRNA production. BDF1 encodes a protein of 687 amino-acids containing two copies of the bromodomain, a motif also present in other transcription factors as well as a new conserved domain, the ET domain, also present in Drosophila and human proteins.

INTRODUCTION

The nucleus of eukaryotic cells contains small nuclear ribonucleoproteins (snRNPs), particles consisting of small nuclear RNA (snRNA) associated with proteins (1,2). The best characterized snRNPs are probably the U1, U2, U4, U5 and U6 snRNPs which are involved in pre-mRNA splicing (3). The synthesis of these snRNPs has been extensively analyzed in vertebrates (4–6). The primary polymerase II (polII) U1, U2, U4, U5 and U6 transcripts are exported to the cytoplasm where they assemble with a group of eight Sm proteins. This binding triggers the hypermethylation of the m7G cap initially present at the 5' end of all polII transcripts to yield a trimethyl cap structure (3mG), characteristic of most of the U snRNAs. The binding of Sm proteins and, in some cases, the presence of a 3mG cap structure are required for transport of the assembled particle back to the nucleus. Along this complex assembly pathway, snRNAs are covalently modified and assemble with several snRNP specific proteins. Although the majority of the U snRNAs are transcribed by polymerase II, the U6 snRNA is transcribed by polymerase III (polIII). Unlike its counterparts, the U6 snRNA remains in the nucleus where it gets covalently modified internally and at both its 5' and 3' ends. Finally, the U4 and U6 snRNP assemble to form a double U4/U6 snRNP where the two snRNAs are associated through an extensive base pairing interaction (3).

In vertebrates, the promoter regions of the polII and polIII transcribed snRNAs are very similar with the presence of interchangeable distal and proximal sequence elements (DSE and PSE, respectively) upstream of the coding region (4). The main difference between the polII and polIII promoters is the presence of a TATA box in the latter one. However, transcription of both classes of snRNAs requires the TATA box binding (TBP) protein (7–11). Additional proteins interacting with the DSE (12,13) and PSE (10,13,14) elements are required for transcription. As in mammals, plant U1, U2, U4 and U5 snRNAs are polII transcribed while plant U6 is a polIII product. Both polII and polIII promoters contain a TATA box element and an upstream sequence element (USE). However, the spacing between these elements differs between the two classes of promoters [reviewed in (4)]. The S.cerevisiae U1, U2, U4 and U5 snRNAs have a 3mG cap suggesting that they are polII transcribed. Expression of the polIII transcribed yeast U6 snRNA is dependent on TBP but also TFIIIC and its cognate DNA binding sites (15–19). Although these data suggest that the mode of transcription of

*To whom correspondence should be addressed

+Present address Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA
spliceosomal snRNAs is highly conserved throughout evolution, it has recently been found that the U2 snRNA from trypanosomes is a poIIl1 product (20).

We have recently developed a screening procedure for yeast mutants that would affect snRNP synthesis (Z.Lygerou and B.Şeraphin, unpublished data). Analysis of one of the mutants recovered from this screen is presented here. At the non-permissive temperature, this mutant does not accumulate a reporter U4 snRNA. This results from a transcriptional defect rather than a reduction of snRNA stability. The wild type allele has been cloned and sequenced. It encodes a protein with two bromodomains, hence the name of BDF1 (Bromodomain Factor 1). Beside the bromodomain, BDF1 contains a new domain (Extra Terminal domain or ET domain) that is also present in proteins from higher eukaryotes. Our experiments suggest that BDF1 encodes a transcription factor with a broad class of target genes that include snRNAs.

MATERIAL AND METHODS

Yeast strains and genetic methods

Saccharomyces cerevisiae strains BY295(MATa ade2 arg4 leu2-3,112 TRPI::SNR14* ura3-52), BY360(MATα ade2 ade5 his3-Δ1 leu2-3,112 TRPI::SNR14* ura3-52), YNN281(MATa trp1-901 his3-Δ200 ura3-52 lys2-801 ade2-101), YNN282(MATα trp1-901 his3-Δ200 ura3-52 lys2-801 ade2-101), MYC829(MATα his3-Δ200 lys2-801 ura3-52), MYC3015(MATα/MA{Tα} leu2-3,112/leu2-3,112 his3-Δ200/Δ200 HIS+ lys2-801/LYS+ ura3-52/ura3-52 ade2-101/Δ/Δ ade5+ suc2/suc2) and MYC3017(MATα leu2-3,112 his3-Δ200 ura3-52 bdf1Δ2::LEU2 U4* were used in this study. The ts mutants are in the BSY295 background. The SNR14* represents the tagged U4 allele. Standard genetic methods were used.

snRNAs analysis

The screen for mutants defective in snRNP assembly will be described in detail elsewhere (Z.Lygerou and B.Şeraphin, in preparation). Briefly, a GAL-regulated tagged U4 (U4*) gene was constructed by introducing a neutral tag in the U4 coding sequence (positions 76–86 were changed to CUCUUAAGCC-AUAG) and fusing the GAL-U1 promoter (21) upstream of the coding sequence as reported previously for the GAL-U2 and GAL-U5 constructs (21,22). This fusion gene uses the proper transcription initiation site (data not shown). The Ul snRNA TATA box region was used because the conservation of the TATA box region in yeast U snRNAs makes them interchangeable (21,23). The fusion construct was integrated in a single copy at the TRP1 locus to give strain BSY295. Strain BSY295 and thermosensitive derivatives thereof were grown at 23°C in YPLactate–gycerol media to OD600 0.5–1.0. They were shifted to 37°C by quickly adding one volume of YP lactate–glycerol media pre-warmed at 51°C. After incubation at 37°C for 30 min, the U4* gene was induced by addition of galactose to 2% final. Two hours later, cells were collected and RNA extracted (24). RNAs were fractionated on 5% non-denaturing polyacrylamide gels, electro-transferred to nylon membrane (GeneScreen) and hybridized at 37°C with 5’ labeled oligonucleotide oli23 (5’ AGCCATATGGCTTAAGAGAATT 3’) complementary to the U4* snRNA and oligonucleotide DT168 (24) complementary to U1 snRNA. To test for the stability of snRNAs in the bdf1 background, cells were grown at 23°C in YPGal media and shifted to 37°C in YPGal media for the indicated time before RNA extraction. As a control an aliquot of the culture of bdf1-1 strain was transferred to YPD at 23°C and RNA collected at the same time points. For the analysis of snRNA production in the bdf1 null mutant, cells were grown in YPD at 23°C and shifted in YPD at 37°C for the indicated time. Primer extensions were done as previously described. The following oligonucleotides were used: U1 [DT168 (24)]; U2 (DT800, 5’ GCCAAAGATGTGATTGAC 3’); U4 (oligo6, 5’ GTATTTCTAAAATCTCAGAATG 3’); U5 (EM19, 5’ GCCCCACAGTCTTTGATGG 3’) and U6 (oligo5, 5’ TCATCTTATGCAGG 3’). Oligo5 was also used for Northern analysis. The U4 and U4* snRNA can be distinguished by primer extension using oligo6 which hybridizes at their 3’ ends because they differ by three nucleotides in length. For primer extension, RNA concentration was calculated from OD260nm reading while RNA quality was ensured by calculating the ratio OD260nm/OD280nm. Equal loading is shown by control extension on unaffected snRNAs. The faint band seen above the U4 and U4* bands can also be used for loading control. This band does not represent a spliceosomal snRNA.

Cloning the BDF1 gene

A bdf1-1 strain was transformed with a yeast genomic library on a LEU2-marked centromeric vector and plated on YPD at 37°C. From independent transformants, three different complementing plasmids were recovered. Restriction analysis indicated that the genomic inserts of these plasmids originated from a single locus (Fig. 4). The BDF1 gene was further localized by subcloning DNA fragments shared by the three original plasmids and assaying for complementation of the thermosensitive growth phenotype (Fig. 4). To check that the cloned fragment covers the BDF1 gene we used it to target integration of a LEU2 marker in the genomic DNA of a wild-type strain. Southern blot confirmed that the integration occurred correctly. The resulting strain was crossed with a bdf1-1 strain and tetrads were dissected. Each of the 10 tetrads analyzed contained two thermoresistant spores that were Leu+ and two thermoresistant spores that were Leu-.

The BDF1 gene was also cloned independently by screening a yeast chromosomal λgt11 expression library (25) using polyclonal antibodies raised against SDS/polyacrylamide gel-purified 95 kDa polypeptide from a partially purified preparation of TFIIIC (see Discussion). The 4.4 kb EcoRI insert of the largest clone was used for further analysis.

Nucleic acid methods

The 4.4 kb EcoRI fragment that overlaps the BDF1 gene was completely sequenced on both strands by subcloning fragments in appropriate vectors and using the dideoxy chain termination method (26). Sequence analyses and database searches were done using standard programs (27,28). Southern and Northern analyses were performed using standard protocols (26).

Disruption of the BDF1 gene

The pBAC5 plasmid contains the 4.4 kb EcoRI fragment harboring the BDF1 gene in a Bluescript KS backbone. pBA-C5 was cleaved with Smal and SpeI, filled with Klenow and self ligated to give pBAC24. pBAC24 was cleaved by BamHI and BglII. The 2.1 kb BamHI–BglII fragment was replaced by a 1.8 kb BamHI/BamHI fragment that contains the yeast HIS3 gene creating pBAC25. DNA from pBAC25 cleaved with EcoRI was used to transform a his3/his3 diploid yeast strain made by
crossing strains YNN281 (a) x YNN282 (α). Correct integration was verified by Southern analysis. A second disruption, bdfIA2:LEU2 was made in the same way in strain MCY3015 by replacing the 2.1 kb BamHI-BglII fragment of pBAC24 with the 2.9 kb BglII-BglII fragment harboring the yeast LEU2 gene.

Assaying the phenotypes of the bdf1 disruption strain
For the YNN281 x YNN282 derivatives, growth was scored on solid rich medium containing 2% glucose (YPD) or other carbon sources as indicated in the results. Hypertonic growth conditions were monitored on YPD containing 2 M ethylene glycol. With the MCY3015 derivatives, growth on different carbon sources was scored by spotting cell suspensions on solid media. Growth on rich medium containing glucose (2%), galactose (2%), sucrose (2%) or raffinose (2%) was scored anaerobically by incubating the plates in a GasPak disposable anaerobic system (BBL) or scored aerobically in presence of 1 μg/ml of antymycin A. Antimycin A increases the dependence on these sugars by blocking respiration. Growth on rich medium containing glycerol (3%) was scored aerobically. Glucose-repressed and derepressed cells were prepared from exponentially growing cultures as described previously (29). Secreted invertase activity was assayed in whole cells (30).

Nucleotide sequence accession number
The sequence reported here has been entered into the EMBL and GenBank data bases, under the accession number Z18944.

RESULTS
Characterization of a strain defective in U4 snRNA accumulation
We have recently developed a screening procedure for yeast mutants that would affect snRNP synthesis (Z. Lygerou and B. Séraphin, in preparation). This screen makes use of a reporter U4 snRNA gene (hereafter referred to as U4*) with the following properties: i) The reporter U4 gene contains a tag that allows for its easy and specific detection. The tagged U4 snRNA is fully functional and assembles in a U4/6 complex in the same way as the wild-type U4 snRNA. ii) A single copy of the reporter U4 gene is integrated at the TRP1 locus while a functional copy of the wild-type U4 gene is still present at the original locus. iii) The transcription of the reporter gene is driven by the GAL regulatory sequence (21,31). This promoter can be conveniently turned on to initiate synthesis of the U4* snRNA whose assembly into snRNP can then be followed. Without an inducible promoter, the identification of trans-acting mutants defective in snRNP synthesis would have been complicated by the extremely long half-life of the snRNAs in yeast (32) see also refs. (21,22) and B.S. unpublished observation).

A strain carrying the U4* snRNA was mutagenized and temperature sensitive mutants were recovered, as snRNP synthesis is likely to be an essential process (3). Mutants were grown individually, shifted to the non-permissive temperature (37°C) after which the transcription of the U4* snRNA gene was turned on by addition of galactose. RNAs extracted from these strains were fractionated on native gels and analyzed by northern blotting using a probe specific for the U4* snRNA. In a wild type strain, almost all of the resulting U4* snRNA migrates as a U4*/U6 RNA hybrid (e.g., Fig. 1C, lane 2). This indicates that the U4* snRNA was transcribed and assembled into a functional U4* snRNP that is competent to associate with the U6 snRNP. Two main mutant phenotypes were observed. In some cases no signal corresponding to the U4* snRNA is detected. This suggests either that the U4* snRNA is not transcribed or that it is unstable in this mutant. In other mutants, the U4* snRNA accumulates as a free U4* snRNA species. This suggests that the assembly or intracellular transport of the U4* snRNP is defective. Alternatively the mutated gene might be required for the stability of the U6 snRNA or for the joining of the U4* and U6 snRNPs into the U4*/U6 particle. The exact function affected by individual mutants can then be assessed by additional experiments.

We report here the isolation and characterization of a thermosensitive strain, referred to as bdfIA14, in which the transcription of the U4* gene is blocked at the non-permissive temperature. Figure 1A shows that in the bdfIA14 strain the U4*
snRNA does not accumulate at the non-permissive temperature. This is not a general phenotype for thermosensitive strains grown at 37°C (see control ts22 and ts23 strains in Fig. 1A). The absence of U4* snRNA in the bdf1° strain can not be explained by underloading or non-specific RNA degradation as the level of the U1 snRNA is close to normal in that strain (Fig. 1A). The absence of the U4* snRNA in the bdf1° strain is only seen at the non-permissive temperature (data not shown) indicating that the reporter gene is still present and functional. To check if the phenotype observed in the bdf1° strain reflects a general instability of the U4 and/or U6 snRNAs, the original probe was stripped and the filter was hybridized with an oligonucleotide complementary to U6 snRNA. The results (Fig. 1B) show that in the bdf1° strain a substantial amount of endogenous U4 and U6 RNAs are present as U4/U6 hybrid or as free U6 RNA. (Although the amount of free U6 snRNA is slightly reduced in the mutant strain, U6 snRNA is still in excess over free U6 RNA. (Although the amount of free U6 snRNA is slightly reduced in the mutant strain, U6 snRNA is still in excess over free U6 RNA. Nevertheless, to rule out this possibility we followed the induction of the GAL10 mRNA in the bdf1° strain to a wild type strain and assays for the segregation of these phenotypes in the meiotic progeny. The results for two representative tetrads are shown in Figure 1C. In every case the absence of U4* snRNA at the non-permissive temperature co-segregates with the temperature sensitive phenotype. Furthermore the U4* under-accumulation phenotype segregated 2:2 in three tetrad while the temperature sensitive phenotype segregated 2:2 in 6 tetrads. This indicates that a single gene, BDF1, is responsible for both phenotypes.

The bdf1-1 allele might have affected the production of the U4* snRNA by interfering with the induction of the GAL regulatory sequence present upstream of the TATA box of our test gene. This was unlikely because the bdf11 strain grows well on galactose containing media. Nevertheless, to rule out this possibility we followed the induction of the GAL10 mRNA in the bdf1° strain. Additional strains were used as positive (ts33 and wild-type) or negative controls (ts5 that is defective in transcription). These strains were grown at the permissive temperature in lactate-glycerol media, switched to 37°C for 20 min and the GAL10 transcription was induced by addition of galactose. After two further hours of incubation at 37°C the cells were collected, RNAs extracted and analyzed by northern blot using a probe specific for the GAL10 gene. The results shown in Figure 2 indicate that the GAL10 mRNA is induced in the bdf1° strain. We conclude that the mutant gene does not prevent induction via the galactose regulatory sequence.

The reporter U4 snRNA is stable at the non-permissive temperature

The results reported above suggest that the bdf1° strain is defective in the production of the U4* snRNA. One explanation for these results is that in the bdf1-1 background this snRNA is unstable. To test this possibility, isogenic strains, carrying either the bdf1-1 or BDF1 allele, were grown at the permissive temperature in galactose media thereby inducing the transcription of the U4* snRNA. These strains were then transferred to the non-permissive temperature and RNAs extracted at various time points. These RNAs were then analyzed by primer extension using a primer that allows the detection of both U4 snRNAs

To determine if the absence of the U4* snRNA and the temperature sensitive phenotype present in the original strain are linked, we crossed the bdf1° strain to a wild type strain and assayed for the segregation of these phenotypes in the meiotic progeny. The results for two representative tetrads are shown in Figure 1C. In every case the absence of U4* snRNA at the non-permissive temperature co-segregates with the temperature sensitive phenotype. Furthermore the U4* under-accumulation phenotype segregated 2:2 in three tetrad while the temperature sensitive phenotype segregated 2:2 in 6 tetrads. This indicates that a single gene, BDF1, is responsible for both phenotypes.

The bdf1-1 allele might have affected the production of the U4* snRNA by interfering with the induction of the GAL regulatory sequence present upstream of the TATA box of our test gene. This was unlikely because the bdf11 strain grows well on galactose containing media. Nevertheless, to rule out this possibility we followed the induction of the GAL10 mRNA in the bdf1° strain. Additional strains were used as positive (ts33 and wild-type) or negative controls (ts5 that is defective in transcription). These strains were grown at the permissive temperature in lactate-glycerol media, switched to 37°C for 20 min and the GAL10 transcription was induced by addition of galactose. After two further hours of incubation at 37°C the cells were collected, RNAs extracted and analyzed by northern blot using a probe specific for the GAL10 gene. The results shown in Figure 2 indicate that the GAL10 mRNA is induced in the bdf1° strain. We conclude that the mutant gene does not prevent induction via the galactose regulatory sequence.

Figure 2. The GAL10 mRNA is induced in a bdf1° strain at 37°C. Induction of the GAL10 mRNA in strains ts5, bdf1°, ts33 and wild type (WT) was assayed by Northern blotting. Normal induction is seen for all the strains except ts5 that is defective in transcription. The small variation in the level of the GAL10 mRNA in lanes 2--4 are due to loading as shown by probing the same gel with a probe for the RP29 ribosomal protein mRNA.

Figure 3. The reporter U4 is stable in a bdf1° strain at 37°C. The levels of the U4*, endogenous U4 and U2 snRNAs were assayed by primer extension. The bdf1° strain, pre-grown in galactose media, was switched either to 37°C in galactose media or kept at 23°C in glucose media for the indicated time. As a control the wild type (WT) strain was also switched to 37°C. The two U4 snRNA species differ by three nucleotides in length.

Figure 4. Organization of the BDF1 locus on chromosome XII. Some restriction sites are indicated. The location and direction of transcription of the SFP1, BDF1 and SIK2 genes are indicated. The presence of the 3' end of a putative open reading frame upstream of BDF1 is depicted (ORF1, 105 amino-acids). Interestingly, the ORF1 encoded product shows significant sequence similarity with the C-terminal region of a protein of unknown function from Bacillus subtilis (Accession number: D26185). The extents of the genomic DNA present in the three plasmids recovered by complementation of the temperature sensitivity of the bdf1-1 strain are indicated below the map (first three lines). The structure of the subclones used to localize the BDF1 gene is also indicated below (four last lines). + and - signs indicate if the corresponding plasmid complemented or not a bdf1-1 mutant. The central EcoRI fragment was sequenced.
owing to their size difference. As control we simultaneously used a primer specific for U2 snRNA (Fig. 3) and for U1, U5 and U6 snRNA (data not shown). The U4* snRNA, that represents about half of the starting U4 snRNA, is still present at high levels in the mutant cells even 24 h after the shift to the non-permissive temperature (lanes 1—5). In contrast, when the same strain is kept at permissive temperature but shifted to glucose media (to repress the GAL promoter) the U4* snRNA decreases to a level undetectable by 24 h (lanes 6—9). These results indicate that the U4* snRNA is stable for long periods at the non-permissive temperature in a bdfl1 strain. Because instability of the U4* snRNA cannot explain why the newly synthesized reporter snRNA is not detected in the bdfl1 strain at 37°C, we conclude that the BDFl gene is required for the transcription of the U4* snRNA. However, we cannot formally rule out that BDFlp is required for the stability of nascent transcripts. This hypothesis seems unlikely because of the presence in BDFlp of motifs characteristic of transcription factors and because our results suggest that BDFlp interacts (directly or indirectly) with DNA (see below).

Cloning and sequencing of the BDFl gene

The BDFl gene was cloned by complementation of the bdfl-1 thermosensitive phenotype using a library of genomic fragments from a wild type yeast strain inserted in a centromeric vector. The three different plasmids that were recovered originated from a single locus (Fig. 4). The cloned DNA directed integration of a LEU2 marker to the BDFl locus demonstrating that the cloned fragment is allelic with the bdfl-1 mutation. The complementing activity was further localized to a XhoI—EcoRI fragment by subcloning and testing the resulting constructs for complementation of the thermosensitive phenotype.

A 4.4 kb EcoRI fragment covering the BDFl gene was entirely sequenced on both strands (Fig. 4) uncovering a long open reading frame of 2061 base pairs. The DNA upstream of the first ATG in the open reading frame is extremely AT rich with two imperfect TATA boxes located at positions −80 and −115. Northern analysis with a probe internal to the BDFl reading frame revealed the presence of a major mRNA of 2.4 kb and a minor mRNA of 3 kb (data not shown). Southern analyses indicates that there is a single copy of the BDFl gene in the yeast genome (data not shown). Comparison of the sequence obtained from one extremity of one of the original plasmids indicated that the BDFl gene is tightly linked to the SFP1 gene (Fig. 4). Because the SFP1 gene has been mapped to chromosome XII between ura4 and ilv5 (33), we concluded that BDFl lies in the same region. Consistent with this, a BDFl probe hybridized to one of the largest yeast chromosomes (IV or XII, data not shown) in genomic Southern blot. While this work was in progress, Widner and Wickner reported the sequence of the SKI2 gene (34). Database search indicated that the SKI2 gene lies immediately downstream of the BDFl gene (Fig. 4). Consistent with our data, the SKI2 gene has been genetically and physically mapped to the ura4−ilv5 interval on chromosome XII (34).

Analysis of the BDFl encoded protein

The BDFl encoded protein (BDFlp) is 687 amino-acids long (calculated molecular weight of 77 084 Da) and acidic (predicted pl 5.6). However, in vitro expression of the BDFl gene yielded a single abnormally migrating polypeptide that comigrated with a 94 kDa marker protein (data not shown). BDFlp contains several domains of biased composition (proline-rich: residues 124—151, basic domain: residues 486—525 and acidic regions: residues 420—454 and 674—687). Putative nuclear localization signals can also be found in BDFlp, consistent with the result of a recent study (35) showing that fragments of BDFlp can target a fused β-galactosidase reporter protein to the nucleus.

BDFlp shows sequence similarity to Ring3, fsh and Orfx proteins, however, this similarity is limited to three domains. First, BDFlp contains two copies of a repeated sequence of 87 amino-acids (Fig. 5A and B) rich in proline and aromatic residues called bromodomain (36,37). To date, bromodomains have only been observed in proteins involved in transcription or in proteins with no known function. This motif occurs twice in BDFlp, fsh, CCG1, RING3 and YK107 proteins and once in the other proteins (Fig. 5B). The two BDFl bromodomains share some distinctive characteristics with the bromodomains of Drosophila fsh and human Ring 3 as well as the bromodomains present in the Orfx product that is encoded by a randomly sequenced human cDNA. These include a 2 amino-acid insertion at positions 15—16 as well as some specifically conserved residues (e.g., D at position 10, Y at position 59, M at position 71 and P at position 87).

Besides the bromodomains, BDFlp contains a new domain named the ET domain (for Extra Terminal domain) which shows sequence similarity with the Ring3, fsh and Orfx protein sequence. The ET domain is located at the extreme C-termini of the proteins except for the long form of the fsh protein (Fig. 5A). The amino terminal part (residue 1—64, Fig. 5C) of this 167 amino-acid long domain shows significant sequence conservation (also noted by Widner and Wickner (34)). In contrast the carboxy-terminal part of this domain is less conserved in primary sequence but is highly enriched in serine, threonine and charged residues.

Disruption of the BDFl gene

The BDFl gene was disrupted in two different diploid strains by replacing part of its coding sequence with selectable markers. Upon sporulation at 30°C, all asci analyzed gave four viable spores. Thus the BDFl gene is not essential for viability. However, the two spores carrying the disrupted bdfl allele were distinctively smaller than the wild-type spores. bdflΔ1 carrying haploid strains presented morphological and growth defects: they grew 2.5 times slower than the wild-type haploid at 30°C and were thermosensitive and cryosensitive at 37 and 16°C, respectively. The disruption conferred a flocculent phenotype and the mutant cells looked larger and rounder than the wild-type cells. The bdflΔ1 haploids showed defects in utilizing various carbon sources as well as in resistance to hypertonic media. The cells exhibited no growth on YPD + 2 M ethyleneglycol, onYP + 2% galactose, YP + 2% ethanol or YP + 2% melibiose, and very slow growth on YP + 2% glycerol and YP + 2% maltose. The bdfl null mutants share several common phenotypes with previously isolated mutants in glucose repression (e.g., (29)), however direct measurement of invertase levels indicated that BDFlp does not have a major role in SUC2 expression. Thus BDFlp is probably not directly involved in glucose repression.

A bdfl null allele is defective in snRNA production

The conditional bdfl-1 allele recovered from our screen is involved in the transcription of the U4* snRNA. We tested whether the BDFl gene is also required for transcription of the endogenous snRNAs. A strain carrying the bdflΔ1 null allele as well as a wild type control strain were grown at 23°C in glucose containing media. Cells were shifted to 37°C at time zero
Figure 5. Bromodomains and ET domains. (A) Structure of the BDFl, Ring3, OrfX and fsh proteins. The short and long form of the fsh protein result from alternative splicing. The long form contains all the amino acids present in the short form. Other events of alternative splicing make additional small changes in these proteins. The bromodomain and ET domain are indicated as dark and gray areas respectively. (B) Alignment of the bromodomains. The protein and species names are given. For proteins with two bromodomains, these are indicated by the protein name followed by a dash and 1 or 2. The number of amino acids preceding and following the bromodomain in a given protein are indicated before and after the protein sequence, respectively. For the fsh protein, the numbers of amino acids following the second bromodomain in the short and long form are both indicated. For proteins with two bromodomains, the number of amino acids present between the two domains is indicated after the first domain and before the second domain either followed or preceded by a dash. The consensus sequence represents the positions that are more than 60% conserved. The Swissprot, EMBL or Genbank accession number for the protein sequences, or for the nucleic acid sequences from which peptide sequences were derived, are as follows: CE0201 (T01020), SPT7 (L2537, X76294, M87651), GCN5 (Q03330), BDFl (this study, Z18944), Ring3 (D26156), OrfX (D26362), Fsh (P13709), R10E11.1 (Z29095), p300 (U01877), CREB-BP (S66385), CREB-BP (S66385), CCG1 Human (P21675), CCG1 Hamster (D26114), TFHD-230 (S61883), YK107 (002206), HBrm/hSNF2a (X72889, D26155), BRG-l/hSNF20 (S66910/D26156), Brm (P25439), NPS1/STH1 (P32597), Prgn (M91585). (C) Alignment of the ET domains. Conventions are as in part B except that the consensus represents amino acids that are conserved between the four sequences. Amino acids indicated in small characters in the fsh sequence (positions 77-81) are absent in some alternative splicing variants. In that case they are replaced by a single histidine residue.
and collected at various time points. RNAs were extracted and analyzed by primer extension using primers specific for the U1, U2, U4, U5 and U6 snRNAs. The results, depicted in Figure 6, indicate that the U2 and U6 snRNAs were the most affected by the bdflΔl null mutation but levels of the U1 and U4 snRNA decreased also significantly. In contrast, the level of the U5 snRNA was barely affected. No effect of the temperature shift on snRNA levels could be detected in the wild type control strain. Comparison of the level of U2 snRNA (quantified using a PhosphorImager) and of the growth of cells (followed by optical density) after the temperature shift indicates that the decrease in the level of U2 snRNA is strictly accounted for by dilution of the pool of snRNAs present at the time of the temperature shift by cell division (data not shown). Given the very long half life of snRNAs in yeast ([32], see also Fig. 3) this strongly suggests that the BDF1 gene is involved in the transcription of the U2 snRNA and possibly also of the U1, U4 and U6 genes. The bdfl-1 allele recovered from our screen affected the transcription of the U4* but not the transcription of the endogenous U2 and U4 snRNAs (see Fig. 3 and data not shown). This probably results from an incomplete block of BDF1 function with this allele. This is supported by the fact that strains carrying the bdfl-1 allele grow well on galactose and do not display a cryosensitive phenotype.

Figure 6. snRNA levels in a bdflΔ1 strain. The levels of the U1, U2, U4, U5 and U6 snRNA in a strain carrying a disruption of the BDF1 locus (bdflΔ1) and a wild type (BDF1) control were assayed by primer extension. The strain were grown in YPD and switched to 37°C for the indicated time.

DISCUSSION

We have shown that the wild-type BDF1 function is required for the production of a reporter U4 snRNA. Furthermore our data demonstrate that the absence of this RNA in the bdfl-1 strain after shift to galactose is not due to its instability. This indicates that BDF1 is required for the transcription of our reporter gene. In wild type cells, BDF1p is probably involved in the transcription of the U2 snRNA and possibly also U1, U4 and U6 snRNAs. However, the multiple phenotypes (thermosensitive and cryosensitive growth, flocculation and absence of, or reduced growth on several carbon sources) observed with the bdfl deletion mutants suggest that BDF1 affects the transcription of many genes. These phenotypes are consistent with BDF1p being a transcription factor as similar pleiotropic phenotypes have been reported for mutations affecting general transcription factors ([38]). For example mutations in SPT7 also result in clumpiness ([39]) while deletion of SNF2/SWI2 results in defective growth on several carbon sources ([39],[40]). How does BDF1 affect transcription in yeast? We have shown that the galactose induction pathway is functional in a bdfl-1 strain by analyzing the induction of the GAL10 gene. However the activation signal is not transmitted from the upstream GAL4 binding site to the downstream gene in the case of the U4* snRNA reporter, even though it contains an identical GAL UAS. The main difference between these two genes resides in the TATA box and transcription initiation region. As for other snRNAs, the TATA box present upstream of the U4 reporter gene is highly conserved in sequence and position ([23]). This is rather unusual for yeast promoters. One might thus speculate that BDF1p is specifically required for bridging upstream activators and the TFIIID complex present on the snRNA TATA box. In that context, it is noteworthy that two other bromodomain containing proteins (CCG1 and SPT7, Fig. 5 and see below) have been shown to interact physically or genetically with the TBP subunit of the TFIIID transcription factor ([39],[41]—[43]). An alternative possibility is that BDF1p affects chromatin structure. The specific effect of the bdfl-1 mutation on our reporter gene would then result from its chromosomal location, that differs from the GAL locus. An effect of BDF1 on chromatin could also explain why the different snRNAs (e.g., U2 and U5) are affected to a different extent by the disruption of the BDF1 locus. This hypothesis is not unreasonable since the bromodomain containing protein encoded by the SNF2/SWI2 gene (SNF2p) has been shown to affect chromatin structure ([44]). Both of these models are consistent with the nuclear location of BDF1p ([35]).

The product of the BDF1 gene was fortuitously found to copurify with a preparation of the TFIIIC factor ([45]). When polyclonal antibodies raised against the gel purified 95 kDa TFIIIC subunit were used for screening a yeast chromosomal λgt11 expression library, the only clones recovered contained the BDF1 gene. Recombinant BDF1 protein was shown to be recognized by antibodies raised against the 95 kDa component of yeast TFIIIC (data not shown). These antibodies alter the migration of TFIIIC–DNA complexes in mobility-shift assays and to inhibit tRNA synthesis in an in vitro transcription system ([45]). However, fragments of peptide sequences that have been obtained for the 95 kDa subunit of TFIIIC were not found in the polypeptide encoded by the BDF1 gene ([46]). These data indicate that the BDF1 gene encodes an immunodominant contaminant that comigrates with the 95 kDa subunit of TFIIIC.
These data suggest that BDF1p is a DNA binding protein since it copurifies with the yeast TFIIC transcription factor during four chromatographic steps including a DNA affinity column. Alternatively, BDF1p might interact indirectly with DNA through a tight association with a DNA binding protein. Both of these possibilities are consistent with the involvement of BDF1p in transcription.

To test whether BDF1 could function directly in transcriptional activation, we constructed LexA-BDF1 and GAL4-DNA binding domain-BDF1 (G<sub>bb</sub>-BDF1) fusion constructs under the control of the constitutively active ADH1 promoter. Complementation analyses indicate that the LexA-BDF1 fusion provides full BDF1 function while the G<sub>bb</sub>-BDF1 is only partially functional. The LexA-BDF1 or G<sub>bb</sub>-BDF1 protein fusions were unable to activate the transcription of reporters carrying the corresponding DNA binding sequences including the U4* snRNA (data not shown). Thus, we were unable to detect transactivation at promoters to which BDF1p was targeted by fusion to two specific DNA binding domains. Although these negative results could be explained by the nature of the fusions made, this strategy has often been successful. This suggests that some bromodomain containing proteins, including BDF1p, might not be able to directly activate transcription (44).

Two bromodomain motifs are found in the BDF1 product. Database search reveals that bromodomains are widespread, being present in proteins from yeast, animals and plants (the partial sequence of a plant bromodomain can be found in a randomly sequenced expressed sequence tag from rice, accession number OS1690A). We recovered 20 bromodomain containing proteins making a total of 28 bromodomains as one or two bromodomains are present in individual proteins. The location of this motif within individual proteins is highly variable (Fig. 5B). The functional significance of the bromodomain is not known. However all proteins of known function that harbor a bromodomain are involved in some aspect of transcription. This includes the SPT7 protein that has been shown to genetically interact with the SPT15 gene encoding the TBP subunit (39). GCNS is necessary for the activity of at least two sequence specific transcriptional regulators (47). The two Drosophila proteins encoded by the maternally expressed and alternatively spliced fsh transcripts are necessary for the activation of the Ubx gene (48,49). The product of the human CCG1 gene is a TFHD subunit, as are its homologues of the two bromodomains and of the presence of the ET domain at its C-terminus. This ET domain is restricted to the human Ring3, Orfx and the short form of the fsh protein. Data base search indicates that the ET domain is restricted to the four proteins. Interestingly these four proteins contain two bromodomains. Furthermore the bromodomains of these proteins show more sequence similarity to each other than to other bromodomains (see Results and Figure 6). Even greater conservation is observed if the four N-terminal and C-terminal bromodomains are aligned separately. The particular conservation of the two bromodomains and of the presence of the ET domain makes it likely that BDF1 represents the yeast homologue of the Drosophila fsh and human Ring3 and Orfx even though sequence similarity between these proteins is mostly restricted to these three regions. Further studies will be required to pin-point the role of BDF1 in transcription as well as the function of bromodomains in transcription factors.

ACKNOWLEDGEMENTS

Z.Lygerou, C.Conesa and P.Lesage contributed equally to this work. We thank C.Marc, T.Gibson and C.Ouzounis for help with bioinformatic, M.Snyder for the gift of the lambda expression library and J.Erriaux for providing yeast strains and plasmids. We are also grateful to G.Dujardin, F.Forquinon, B.Leblanc, I.Mattaj and S.Taraviras for comments and/or discussions. This work was supported by EMBL (Z.L. and B.S.) and NIH grant GM47259 to M.C. P.L. was supported by an 'Association pour la Recherche contre le Cancer' post-doctoral fellowship. B.S. is on leave from CNRS.

REFERENCES
