MOLECULAR AND CELLULAR BIOLOGY, Oct. 1987, p. 3386–3393
0270-7306/87/103386-08$02.00/0

Vol. 7, No. 10

Isolation of the Human Gene That Complements a Temperature-Sensitive Cell Cycle Mutation in BHK Cells

MICHAEL ITTMANN, ANGELA GRECO, AND CLAUDIO BASILICO*

Department of Pathology, New York University School of Medicine, New York, New York 10016

Received 8 May 1987/Accepted 30 June 1987

We have cloned the human genomic DNA and the corresponding cDNA for the gene which complements the mutation of tsBN51, a temperature-sensitive (Ts) cell cycle mutant of BHK cells which is blocked in G1 at the nonpermissive temperature. After transfecting human DNA into tsBN51 cells and selecting for growth at 39.5°C, Ts^+ transformants were identified by their content of human AluI repetitive DNA sequences. Following two additional rounds of transfection, a genomic library was constructed from a tertiary Ts^+ transformant and a recombinant phage containing the complementing gene isolated by screening for human AluI sequences. A genomic probe from this clone recognized a 2-kilobase mRNA in human and tertiary transformant cell lines, and this probe was used to isolate a biologically active cDNA from the Okayama-Berg cDNA expression library. Sequencing of this cDNA revealed a single open reading frame encoding a polypeptide of 395 amino acids. The deduced BN51 gene product has a high proportion of acidic and basic amino acids which are clustered in four hydrophilic domains spaced at 60- to 80-amino-acid intervals. These domains have strong sequence homology to each other. Thus, the tsBN51 protein consists of periodic repetitive clusters of acidic and basic amino acids.

The process by which cells progress through the division cycle is still poorly understood. Growth factors appear to play an important role in the initiation of the cell cycle (35, 43), but the events following growth factor stimulation and leading to DNA synthesis are largely unknown. One important role of growth factors is to induce or increase the levels of proteins which are essential for cell cycle progression (41). A number of genes with known products are cell cycle regulated, such as histones (40), thymidine kinase (27), calmodulin (9), and ornithine decarboxylase (19). Other genes whose expression is induced by growth factors or serum and may be cell cycle regulated have been identified (8, 10, 14, 16, 21, 24, 38). These include several cellular proto-oncogenes such as c-*myc*, c-*fos*, c-*ras*, and p53 (8, 14, 21, 38). However, the function of many of these genes is not known and their role in cell cycle progression is still unclear.

One approach to identifying cell cycle genes is through the use of temperature-sensitive (*ts*) cell cycle mutants. Such mutants are blocked in the traverse of specific phases of the cell cycle at the nonpermissive temperature (reviewed in references 1, 2, and 29). This approach has been used successfully with yeasts, in which several critical cell cycle functions have been identified (37) and their genes cloned (3, 5–7, 31). A number of cell cycle mutants have also been isolated in the higher eucaryotes. Two of these *ts* genes have recently been cloned (13, 20). We have recently reported (13) the molecular cloning and preliminary characterization of the gene complementing the defect of the ts11 cell cycle mutant. This gene, which is essential for G1 progression, is induced in mid-G1, encodes a 540-amino-acid polypeptide, and is part of a small gene family.

We report in this paper the identification of the human gene which complements the mutation of tsBN51, a *ts* mutant of the BHK-21 Syrian hamster cell line (32), which is also blocked in G1 progression at the nonpermissive temperature. The tsBN51 gene encodes an mRNA of approximately 2 kilobases (kb) which is expressed in human, hamster, rat, and mouse cells. Biologically active cDNA as well as genomic clones have been isolated. Sequencing of the cDNA reveals that the predicted tsBN51 protein consists of 395 amino acids, with a high proportion of basic and acidic amino acids. The basic and acidic residues are clustered in four major hydrophilic domains, spaced at 60- to 80-amino-acid intervals, which have strong sequence homology to each other.

* Corresponding author.

MATERIALS AND METHODS

Cell culture. Cells were cultured in Dulbecco modified Eagle medium containing 10% calf serum. The tsBN51 cell line was maintained at 33°C; all others were maintained at 39.5°C.

Transfection. BN51 cells were plated 1 day before transfection at 10^6 cells per 100-mm dish and incubated at 33°C. Cells were transfected with calcium phosphate-precipitated DNA by a modification (46) of the method of Graham and Van der Eb (12). For isolation of primary, secondary, and tertiary transformants, 30 μg of high-molecular-weight DNA per plate was transfected. When genomic phages or cDNA plasmids were used, 1 μg of phage or plasmid was transfected with 20 μg of tsBN51 DNA as carrier. The medium was changed twice per week. Surviving (Ts^+) clones were picked after 15 to 18 days of selection at 39.5°C.

Nucleic acid extraction, blotting, and hybridization. DNA and RNA extraction, Southern and Northern (RNA) blotting, hybridization, and autoradiography were carried out as described previously (13).

Molecular cloning. DNA fragments of 15 to 20 kb were isolated from a tertiary transformant by partially digesting the DNA with *Mbo*I restriction endonuclease and size fractionating the DNA by centrifugation on a sucrose density gradient (28). These fragments were ligated with *Eco*RI-*Bam*HI-digested EMBL-3 λ phage vector (11) and packaged with Gigapack packaging as described by the manufacturer (Stratagene, San Diego, Calif.). The library was plated on *Escherichia coli* Q359, and duplicate filters were prepared by the procedure of Benton and Davis (4). Filters were

TABLE 1. Transformation of *ts*BN51 cells by DNA transfection[a]

Donor DNA	No. of Ts^+ colonies per 10^6 cells	No. of Ts^+ colonies/μg of DNA
tsBN51[b]	0.4	0.01
HeLa	4.3	0.14
BHK	3.2	0.11
7B (primary transformant)	1.7	0.06
24A (secondary transformant)	2.0	0.07
λ51/18	330	264
λ51/TM15	303	291
λ51/TM13	<1	<1
pCD51-6A[c]	272	218
pCD51-10A[c]	66	47
pCD51-J2[c]	<1	<1

[a] *ts*BN51 cells growing at 33°C were transfected with 30 μg of high-molecular-weight DNA or 1 μg of plasmid DNA per 100-mm plate (containing approximately 10^6 cells). One to two days later the cells were treated with trypsin and replated at 5×10^5 cells per 100-mm dish and then shifted to 39.5°C 1 day later. Colonies were either fixed and counted or isolated for further study after 14 to 18 days of incubation at 39.5°C.

[b] Average of three experiments (range, 0.17 to 0.7 colonies per 10^6 cells).

[c] Plasmids isolated from the Okayama-Berg (33) human cDNA expression library. pCD51-6A and pCD51-10A are described in the text. pCD51-J2 is a slightly shorter cDNA plasmid, with an insert of approximately 1.6 kb.

prehybridized in 50% formamide–5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5× Denhardt solution–0.1% sodium dodecyl sulfate (SDS)–100 μg of sonicated calf thymus DNA per ml. Hybridization was carried out in the same solution except for the addition of sonicated Syrian hamster liver DNA (10 μg/ml). The *Bam*HI fragment of the Blur-8 plasmid (18), ligated into concatamers and nick translated to $>10^8$ cpm/μg, was used as a probe for human AluI sequences. We found that the addition of Syrian hamster DNA helped decrease the background hybridization of hamster repetitive sequences with the human Alu probe. Filters were washed for 30 min in 2× SSC–0.1% SDS at room temperature and then twice with 1× SSC–0.1% SDS at 65°C for 1 h.

Additional genomic phage clones were isolated from the human genomic library constructed by Lawn et al. (25) with the 1.8-kb *Sac*I-*Sal*I fragment from the distal end of the λ51/18 phage used as the probe. The phage were plated in *E. coli* LE392. Hybridization and washing of the filters were similar to the above procedure except that Syrian hamster liver DNA was omitted.

The cDNA clones were isolated from the human cDNA library of Okayama and Berg (33) (kindly provided by P. Berg) as described by Hanahan and Meselson (15) with a 1,200-base-pair (bp) *Cfo*I fragment of the 1.8-kb *Sac*I-*Sal*I fragment from λ51/18.

DNA sequencing. The cDNA was subcloned into the *Bam*HI site of the Bluescript KS vector (Stratagene) in both orientations. Following *Xba*I-*Sac*I double digestion of these plasmids, serial deletions were constructed in both directions by using exonuclease III and mung bean nuclease as described by the manufacturer. These subclones were then sequenced by the dideoxy chain termination method (39) with the T7 promoter primer.

RESULTS

The *ts*BN51 mutant of BHK cells grows at 33°C but when shifted to 39.5°C stops growth after 24 h (32). Flow microfluorometry of cells incubated at 39.5°C indicates that they are arrested with a G1 DNA content (Nishimoto and Basilico, unpublished results); thus, the cell cycle block is within G1. The mutation has a low reversion rate, $<10^{-6}$. This low reversion rate allowed us to clone the human gene complementing this mutation by using a strategy similar to the one we used for cloning the human gene complementing the ts11 mutation (13). In brief, we transfected human DNA into tsBN51 cells, which were then shifted to 39.5°C. Surviving clones were analyzed for the presence of human AluI repetitive DNA sequences (18) by Southern blotting. Two additional rounds of transfection were used to reduce the amount of extraneous human DNA which was present. The gene of interest was then isolated from a recombinant phage library constructed from a tertiary transformant by screening for the AluI sequences which were linked to the complementing DNA sequences.

Isolation of Ts^+ transformants. A summary of the transfection experiments is shown in Table 1. Transfection of tsBN51 cells with HeLa or wild-type BHK DNA caused a slight but significant increase in the number of Ts^+ colonies compared with the transfection with tsBN51 DNA. One of two primary clones analyzed contained a large amount of AluI sequences on Southern blotting (data not shown). A second round of transfection again showed an increased frequency of Ts^+ colonies (Table 1). DNA from one of the AluI-positive secondary transformants (24A) was used for a third transfection cycle. Figure 1A shows the analysis of the DNA of several tertiary transformants by Southern blotting and hybridization to the AluI probe following *Eco*RI digestion. Six of seven clones were AluI positive and all contained a common AluI fragment of approximately 12 kb that was also detected in the secondary transformant. The seventh clone (35C, lane 4) did not contain any AluI sequences and therefore is probably a revertant. The same 12-kb band was also present in an independent secondary transformant and the primary transformant (data not shown). The conservation of this band argues that the complementing gene is linked to the AluI sequences contained in this 12-kb *Eco*RI fragment.

Molecular cloning of genomic DNA. DNA from the tertiary transformant 35L, which appeared to contain the smallest amount of AluI sequences (Fig. 1A), was used to construct a recombinant phage library in the EMBL-3 bacteriophage

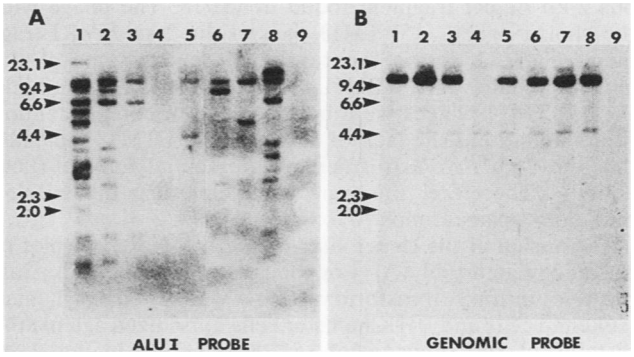

FIG. 1. Southern blots of Ts^+ cell lines with a human AluI probe (A) or a genomic fragment of λ51/18 (B). Ten micrograms of high-molecular-weight DNA was digested with *Eco*RI, separated on a 1% agarose gel, blotted onto nitrocellulose paper, and hybridized to nick-translated Blur-8 plasmid (18) (A) or the 4.4-kb *Hin*dIII-*Sal*I fragment of λ51/18 (B). Lanes: 1, secondary transformant (24A); 2 to 8, tertiary clones (35A, 35B1, 35C, 35D, 35F, 35L, and 35M2, respectively); 9, tsBN51. Fragment sizes (in kilobases) are indicated.

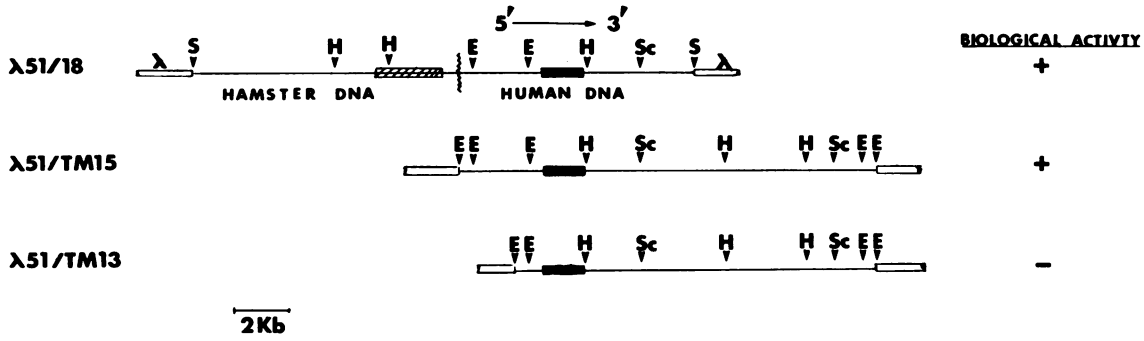

FIG. 2. Restriction map of the DNA inserts in the genomic phage recombinants. The solid box indicates regions containing human repetitive sequences, while the dashed box represents hamster repetitive sequences. The broken line in λ51/18 indicates the approximate site of insertion of the human sequences into the hamster genome. The following restriction enzyme sites are shown: *Eco*RI (E), *Hin*dIII (H), *Sac*I (Sc), and *Sal*I (S). The *Sal*I sites in λ51/18 and the terminal *Eco*RI sites in λ51/TM15 and λ51/TM13 are the insertion sites of the genomic DNA into the phage.

vector (11). A total of 3.2×10^5 plaques were screened as described in Materials and Methods, and a single AluI-positive plaque was identified. A map of the insert in this recombinant phage (λ51/18) was derived by using restriction endonuclease digestion and Southern blotting with probes for human AluI sequences and total hamster DNA (Fig. 2). The 18-kb insert consisted of approximately 10 kb of hamster DNA joined to 8 kb of human DNA near the first *Eco*RI site. A Southern blot of *Eco*RI-digested DNA from tertiary clones hybridized with the 4.4-kb *Hin*dIII-*Sal*I probe derived from the rightmost portion of the insert is shown in Fig. 1B. The probe recognized a single *Eco*RI fragment of ~12 kb corresponding to the common band seen on AluI hybridization. As expected, clone 35C (lane 4), which did not show AluI sequences, did not contain the human 12-kb *Eco*RI fragment. A weakly hybridizing band of approximately 4.4 kb was also seen that was present in all lanes, including tsBN51, and thus probably represents hybridization to the homologous hamster gene.

To obtain the entire 12-kb *Eco*RI fragment, we screened 4.5×10^5 phages of the recombinant library of Lawn et al. (25). The restriction maps from the two phages isolated are shown in Fig. 2. The phage with a longer insert, λ51/TM15, contained the 12-kb *Eco*RI fragment as well as the contiguous 2-kb *Eco*RI fragment found in λ51/18. The phage with the shorter insert, λ51/TM13, lacked the 2-kb *Eco*RI fragment. Table 1 shows the results of transfections of the genomic phages into tsBN51 cells. The original λ51/18 and the λ51/TM15 phages were both capable of complementing the *ts* mutation. The lack of activity of λ51/TM13 indicates that the 2-kb *Eco*RI fragment was essential for biological activity. Therefore, the gene complementing the tsBN51 mutations spans at most 7 to 8 kb of DNA.

Expression of the cloned human gene. A Northern blot of polyadenylated [poly(A)$^+$] or total RNA from several AluI-positive tertiary transformants, as well as from human leukemic cells and BHK hamster cells hybridized against the distal *Sac*I-*Sal*I fragment of λ51/18, is shown in Fig. 3. A strongly hybridizing band of approximately 2 kb was seen in all the tertiary transformants as well as in human RNA. A more weakly hybridizing band of the same size was seen in the BHK RNA and should correspond to the transcript of the homologous hamster gene. A weakly hybridizing band of approximately 5 kb was also present in human and tertiary Ts$^+$ transformant RNA. It is not clear at the moment whether this represents an unspliced transcript or an additional form of mature mRNA. It seems unlikely that it represents a cross-hybridizing RNA, as this band was present in human cells and Ts$^+$ transformants but not in BHK cells. Northern blots of HeLa, primary human fibroblasts (FS4), mouse (BALB/c 3T3), and rat (F2408) RNAs showed transcripts of approximately the same size (data not shown).

Isolation of cDNA clones. We constructed a 1,200-bp *Cfo*I probe from the distal *Sac*I-*Sal*I fragment from the λ51/18 phage. This probe, which was free of repetitive DNA sequences and hybridized to the *ts*BN51 RNA, was used to screen 1.5×10^6 clones from the Okayama-Berg human cDNA expression library (33) as described in Materials and Methods. A total of 14 positive clones were isolated and analyzed. The longest of these clones, designated pCD51-6A, was approximately 1.9 kb in length. A slightly shorter clone of approximately 1.8 kb, designated pCD51-10A, with an apparently identical restriction map was also isolated. Since the Okayama and Berg cDNA library is constructed in a mammalian expression vector, we could directly test the biological activity of our cDNA clones by transfection into *ts*BN51 cells (Table 1). The longest cDNA clone, pCD51-6A, was capable of efficiently complementing the *ts*BN51 defect. The slightly shorter cDNA clone, pCD51-10A, also showed

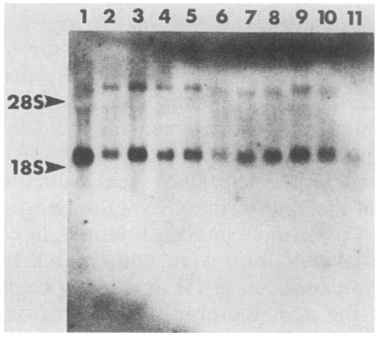

FIG. 3. Expression of the tsBN51 gene in various cell lines. Approximately 1 μg of poly(A)$^+$ RNA, isolated by using Amersham Hybond m-AP paper, or 10 μg of total RNA was separated on a formaldehyde–1.2% agarose gel, blotted onto nitrocellulose paper, and hybridized to the distal 1.8-kb *Sac*I-*Sal*I fragment of λ51/18. Lanes: 1, human poly(A)$^+$ RNA (Fro 2.2 human leukemia cells); 2 to 5, tertiary transformant poly(A)$^+$ RNA (34B, 35B1, 35F, and 35M2, respectively); 6 to 10, tertiary transformant total RNA (34B, 35B1, 35F, 35L, and 35M2, respectively); 11, BHK poly(A)$^+$ RNA.

biological activity, but produced only 20% as many colonies per microgram of DNA transfected. The exact reason for this partial activity is unclear. Since the apparent translation initiation codon was only 51 bp from the 5′ end of pCD51-6A (see below), it is possible that this codon was missing from the slightly shorter pCD51-10A, leading to poor translation of the mRNA transcript or a truncated protein with reduced biological activity.

Orientation of the *ts*BN51 gene. The location of the 5′ and 3′ exons on the genomic DNA was determined by hybridizing Southern blots of the λ15/TM15 phage, digested with various restriction endonucleases, with probes containing only 5′ or 3′ DNA sequences, derived by serial deletions from the pCD51-6A cDNA insert. This indicated that the 5′ end of the cDNA was located within the 2-kb *Eco*RI fragment, in line with the demonstrated requirement of this fragment for biological activity. The 3′ cDNA sequences mapped within the 3-kb *Sac*I-*Hin*dIII fragment (data not shown). Thus, the BN51 gene is transcribed from left to right on the map (Fig. 2), contains 7 to 8 kb of DNA, and must contain at least one intron.

Nucleotide and predicted amino acid sequence of cDNA. The pCD51-6A cDNA was sequenced on both strands, and its nucleotide sequence and predicted protein sequence are shown in Fig. 4. There were 1,881 nucleotides (nt) with a single long open reading frame of 1,185 nt encoding a protein of 395 amino acids. There was a 5′ untranslated region of 51 bases which did not contain a stop codon in any frame. We therefore cannot exclude the possibility that the ATG identified in our sequence is not the true initiation codon. It is unlikely, however, that we missed a significant amount of upstream sequences, since the sum of the cDNA and an average poly(A) tail (150 nt) was very close to the predicted mRNA length obtained from Northern blots. The 5 nt immediately upstream from the putative initiation codon (GCAAC) agree with the proposed initiation codon consensus sequence (CCACC) of Kozak (22) in three of five positions, including the strongly conserved A at position −3. The majority of initiation codons have a similar level of agreement to this consensus sequence (22). A 3′ untranslated region of 645 nt was present. In the 100 nt 3′ of the putative stop codon, there was at least one stop codon in all three reading frames. The consensus polyadenylation signal AATAAA was found 20 bases from the poly(A) tract.

Analysis of the predicted protein primary structure revealed several interesting features. The protein had a size of 44 kilodaltons (kDa) and a high proportion of basic (19%) and acidic (18%) amino acids. Figure 5 shows a hydrophilicity plot of the protein, which indicates the presence of four major hydrophilic domains, i.e., amino acids 62 to 98 (I), 138 to 149 (II), 215 to 230 (III), and 299 to 306 (IV), which were distributed every 60 to 80 amino acids. These domains consisted of clusters of basic and acidic amino acid residues. Domain I (residues 62 to 98) consisted of 17 basic and 10 acidic amino acids within 37 amino acids. It was the only domain with a marked net positive charge. In addition, it was flanked by two triplets of lysine residues at residues 51 to 53 and 115 to 117. The remaining three domains were smaller and had nearly equal numbers of acidic and basic amino acids.

A Diagon plot of the tsBN51 protein against itself is shown in Fig. 6. This indicates that there are four major repeats of sequence homology within the protein, which correspond to the hydrophilic domains described above. The sequence comparisons are shown in Fig. 7. Therefore, the tsBN51 protein contains a periodic repetitive structure characterized by clustered basic and acidic amino acid residues. It can also be noted that the sequence lysine-glutamine-glutamine occurs six times within the tsBN51 protein. Three of these are within the major repetitive domains, while three others are present in areas with less marked hydrophilic character (residues 187 to 189, 203 to 205, and 259 to 261).

DISCUSSION

Identification of the genes required for progression through the cell cycle is an important step in understanding the mechanism and control of cellular proliferation in both normal and neoplastic cells. Although progress has been made in identifying genes whose mRNA level is increased in proliferating cells (10, 16, 24), the role of these genes in the cell cycle is not clear. A potential problem with this method is the possibility that the increased mRNA is a secondary phenomenon and is not directly related to a role for the gene product in progression through G1. In addition, all genes of interest may not have an increased level of mRNA during proliferation. Despite these problems, however, the method holds promise for identifying potentially important genes. Our approach, which involves the molecular cloning of genes which complement *ts* cell cycle mutations, avoids these two difficulties in that the genes are, by definition, necessary for cell cycle progression and do not have to be induced during the cell cycle. In general, we believe that both approaches are needed and should be complementary.

In this paper we report the isolation and preliminary characterization of the human gene complementing the mutation in the tsBN51 cell cycle mutant. Our approach was essentially the same as used previously in this laboratory for isolating the ts11 cell cycle gene (13). The strategy consisted of transfecting human high-molecular-weight DNA into tsBN51 cells, which were then selected at 39.5°C. The surviving Ts^+ clones were screened for AluI sequences to determine which clones had actually taken up the complementing human DNA. Two additional rounds of transfection reduced the amount of extraneous AluI sequences. The gene was then isolated from a recombinant genomic library constructed from a tertiary Ts^+ transformant by screening for AluI sequences. The recombinant phage was biologically active in complementing the *ts* defect in BN51 cells, as was one of two phages isolated from the Maniatis et al. (25) genomic DNA library. Correlation of the restriction maps of these phages with their biological activity defined a gene within a region of 7 to 8 kb of DNA.

A portion of the genomic clone recognized an mRNA of approximately 2 kb in all cell lines tested, including human, mouse, rat, and hamster, indicating that the gene is evolutionarily conserved. We used this fragment to isolate a biologically active cDNA from the Okayama-Berg human cDNA expression library. The nucleotide sequence of the cDNA predicted a protein of 44 kDa, which is characterized by four periodic, repetitive clusters of basic and acidic amino acids. A search of the National Biomedical Resources Foundation data base of protein sequences with the Fast P program (26) did not find any strong homologies to the predicted tsBN51 protein.

The predicted protein structure is similar to that of the U1 RNA-associated 70-kDa protein. This protein is characterized by a high content of basic (28%) and acidic (18%) amino acids, which tend to be clustered in mixed-charge groups (44), similar to the tsBN51 protein. However, the 70-kDa protein has a more extreme degree of clustering at the carboxyl-terminal half and a larger percentage of arginines

```
                                                1   ACTTCCGCCCGGCGCGAGACCGAAGCTGGCGGCTGGTCGGTTGCAGGCAAC
                                   10                                  20                                  30
 52  ATGTCGGAAGGAAACGCCGCCGGACGCCCAGCACGGCAGGGCCCGGACCTTCTCCTGACTGGGGCCCGGGGGTCATCGGGCGGCGGCGGC
     MetSerGluGlyAsnAlaAlaGlyArgProAlaArgGlnGlyProAspLeuLeuLeuThrGlyAlaArgGlySerSerGlyGlyGlyGly

                                   40                                  50                                  60
142  CTCCCCTCACCCCCGGCCGTTCCCTCCATCCGTTCCAGGGACCTAACCCTCGGGGGAGTCAAGAAGAAAACCTTCACCCCAAATATCATC
     LeuProSerProProAlaValProSerIleArgSerArgAspLeuThrLeuGlyGlyValLysLysLysThrPheThrProAsnIleIle

                                   70                                  80                                  90
232  AGTCGGAAGATCAAGGAAGAGCCCAAGGAAGAAGTAACTGTCAAGAAGGAGAAGCGTGAAAGGGACAGAGACCGACAACGAGAGGGGCAT
     SerArgLysIleLysGluGluProLysGluGluValThrValLysLysGluLysArgGluArgAspArgAspArgGlnArgGluGlyHis

                                  100                                 110                                 120
322  GGACGAGGGCGACGCCGTCCAGAAGTGATCCAGTCTCACTCCATCTTTGAGCAGGGCCCAGCTGAAATGATGAAGAAAAAAGGGAACTGG
     GlyArgGlyArgArgArgProGluValIleGlnSerHisSerIlePheGluGlnGlyProAlaGluMetMetLysLysLysGlyAsnTrp

                                  130                                 140                                 150
412  GATAAGACAGTGGATGTGTCAGACATGGGACCTTCTCATATCATCAACATCAAAAAAGAGAAGAGAGAGACAGACGAAGAAACTAAACAG
     AspLysThrValAspValSerAspMetGlyProSerHisIleIleAsnIleLysLysGluLysArgGluThrAspGluGluThrLysGln

                                  160                                 170                                 180
502  ATCTTGCGTATGCTGGAGAAGGACGATTTCCTCGATGACCCCGGCCTGAGGAACG.CACTCGAAATATGCCTGTGCAGCTGCCGCTGGCT
     IleLeuArgMetLeuGluLysAspAspPheLeuAspAspProGlyLeuArgAsnAspThrArgAsnMetProValGlnLeuProLeuAla

                                  190                                 200                                 210
592  CACTCAGGATGGCTTTTTAAGGAAGAAAATGACGAACCAGATGTTAAACCTTGGCTGGCTGGCCCCAAGGAAGAGGACATGGAGGTGGAC
     HisSerGlyTrpLeuPheLysGluGluAsnAspGluProAspValLysProTrpLeuAlaGlyProLysGluGluAspMetGluValAsp

                                  220                                 230                                 240
682  ATACCTGCTGTGAAAGTGAAAGAGGAGCCACGAGATGAGGAGGAAGAGGCCAAGATGAAGGCTCCTCCCAAAGCAGCCAGGAAGACTCCA
     IleProAlaValLysValLysGluGluProArgAspGluGluGluGluAlaLysMetLysAlaProProLysAlaAlaArgLysThrPro

                                  250                                 260                                 270
772  GGCCTCCCGAAGGATGTATCTGTGGCAGAGCTGCTGAGGGAGCTGAGCCTCACCAAGGAAGAGGAACTGCTGTTTCTGCAGCTGCCAGAC
     GlyLeuProLysAspValSerValAlaGluLeuLeuArgGluLeuSerLeuThrLysGluGluGluLeuLeuPheLeuGlnLeuProAsp

                                  280                                 290                                 300
862  ACCCTCCCTGGCCAGCCACCCACCCAGGACATCAAGCCTATCAAGACAGAGGTGCAGGGCGAGGACGGACAGGTGGTGCTCATCAAGCAG
     ThrLeuProGlyGlnProProThrGlnAspIleLysProIleLysThrGluValGlnGlyGluAspGlyGlnValValLeuIleLysGln

                                  310                                 320                                 330
952  GAGAAAGACCGAGAAGCCAAATTGGCAGAGAATGCTTGTACCCTGGCTGACCTGACAGAGGGTCAGGTTGGCAAGCTACTCATCCGCAAG
     GluLysAspArgGluAlaLysLeuAlaGluAsnAlaCysThrLeuAlaAspLeuThrGluGlyGlnValGlyLysLeuLeuIleArgLys

                                  340                                 350                                 360
1042 TCTGGAAGGGTGCAACTCCTCTTGGGCAAGGTGACTCTGGACGTGACCATGGGAACTGCCTGCTCCTTCCTGCAGGAGCTGGTGTCCGTG
     SerGlyArgValGlnLeuLeuLeuGlyLysValThrLeuAspValThrMetGlyThrAlaCysSerPheLeuGlnGluLeuValSerVal

                                  370                                 380                                 390
1132 GGCCTTGGAGACAGTAGGACAGGGGAGATGACAGTCCTGGGACACGTGAAGCACAAACTTGTATGTTCCCCTGATTTTGAATCCCTCTTG
     GlyLeuGlyAspSerArgThrGlyGluMetThrValLeuGlyHisValLysHisLysLeuValCysSerProAspPheGluSerLeuLeu

1222 GATCACAAACACCGGTAAAATGAGCAGGTGGAGGAGGACGGCGCCTGTGCCCACGCTGCTGCCTGCTCCAGACATTTTGTTCTTGAATCT
     AspHisLysHisArg---

1312 GTGAGACCCAGAAGGGCCCACTGAGCCCACTCACTCCACCTTTGGCAACCATTGTTCCAGGTCCCCCAGGGCTTCCTCCCACAGCAGCTG

1402 TGAATGGCACAGTGACCTTCCTGCAGCGTGGAGATGGCACATCCTTGCTGCTGGGGACTTGGCCCTGCTATTTATTTTTGTATTTATGTC

1492 TTAATCTCTTCCACTGATGCATCCTCCAAGGGTAGATGGGGAGGGTCTGTGTGAAGGGGCCGGCTTCTCTTGGTGCCTGCTGGGTTGCAG

1582 GGGCAGGAAGCGTGTGGACTGCAGCTTCTGCTGGTGCTCCCCCCGTCCTCCTGGAGGCAGTATAGGAGAGAGAGCAAGGATTGAGTCTGA

1672 GACTTAAGCACTCGGTCCCAGCTTGCCAGTTCCTGGTTCTGTGTCCTTGGAAAACTACCTAACCTTTCTGAGCCTCCTATACTATCCGAC

1762 ACAAATGGGGATGATACCTACCTCCAGGGTTGGCGTGAGGATTCATGGGCTATTATAGATGAAAACTGCACAAGGCCAGAACTAGCAGGC

1852 ACTCAATAAACGTTCATGTCCTTTTTCTCTAAAAAAAAAA
```

FIG. 4. Nucleotide and predicted amino acid sequence of the pCD51-6A cDNA clone. The putative start codon (position 52), stop codon (position 1237), and polyadenylation signal (positions 1856 to 1861) are underlined.

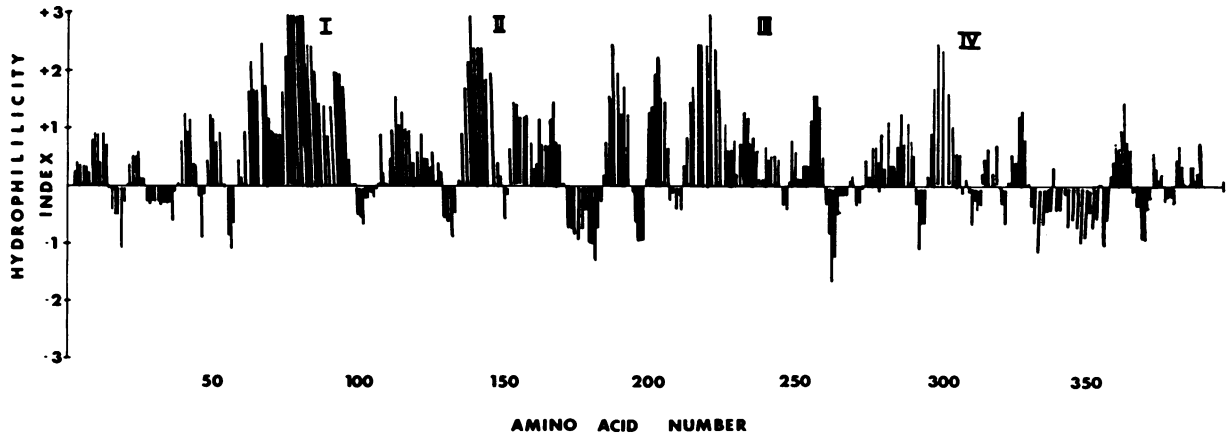

FIG. 5. Hydrophilicity plot of the predicted protein of the *ts*BN51 gene. Hydrophilicity was calculated by the method of Hopp and Woods (17) with a span of 6 amino acids. Similar results were obtained by the method of Kyte and Doolittle (23) (data not shown). The repetitive hydrophilic domains are indicated as I through IV.

than the tsBN51 protein. A direct sequence comparison with the algorithm of Lipman and Pearson (26) revealed some homology (15% identity, 48% conservative replacements) between amino acids 40 to 270 of the tsBN51 protein and amino acids 364 to 594 of the 70-kDa protein. These regions contain hydrophilic domains I to III of the tsBN51 protein and the markedly hydrophilic carboxyl-terminal domains of the 70-kDa protein, respectively. A comparison of the tsBN51 protein with the randomly shuffled 70-kDa protein sequences (n = 100) revealed that the Z score of the initial and aligned overlaps was 2.9 standard deviations above the mean, which is of borderline significance (26). Thus, we cannot be certain at this time whether there is some distant relationship between the two proteins or whether the similarity is fortuitous.

The tsBN51 protein is also similar to that of the high-mobility-group (HMG) proteins (34, 45), which contain high proportions of basic and acidic residues, with two domains derived by partial duplication (36). Unlike the tsBN51 protein, there is a highly acidic C-terminal region. A direct comparison of the sequences revealed no homology, so that it is unlikely that the tsBN51 is part of the HMG gene family, although a similar function is possible based on the structural analogies. Another protein with a similar structure is myosin (30). The structure of myosin is characterized by repetitive 28-amino-acid sequences with a high proportion of basic and acidic amino acids, which are probably involved in the self-association of myosin rods (30). Again, there is no strong homology to the tsBN51 protein by direct comparison of the sequences. Thus, we cannot assign the tsBN51 protein to a known gene family or superfamily, but it would seem likely by analogy to proteins of similar structure that the periodic charged domains are involved in binding of the protein to either nucleic acid or another protein.

The availability of both genomic and cDNA clones for the tsBN51 gene should facilitate the study of the cellular localization of its product, the regulation of its expression, and the effects of its constitutive expression on the cell's phenotype. These studies should help elucidate the biochem-

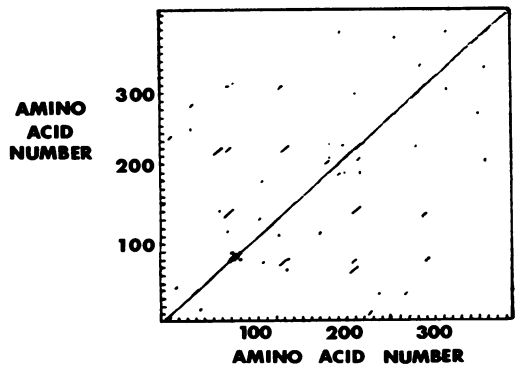

FIG. 6. Diagon plot of predicted protein of the tsBN51 gene against itself by the program of Staden (42). A proportional algorithm with a span of 11 amino acids and a minimum score of 132 was used.

Domain	Amino Acids	Alignment
I	68-87	PKEEVTVKKEKREREDRDRQR
II	131-150	PSHIINIKKEKRETDEETKQ
I	63-86	KIKEEPKEEVTVKKEKREREDRDRQ
III	215-23	KVKEEPRDEEEEAKMKAPPKAARK
I	68-74	KREREDRD
IV	299-305	KQEKDRE
II	135-149	INIKKEKRETDEETK
III	214-228	VKVKEEPRDEEEEAK
II	141-149	KRETDEETK
IV	299-307	KQEKDREAK

FIG. 7. Comparison of the amino acid sequences of the hydrophilic domains of the predicted protein of the tsBN51 gene. The indicated domains are as shown in Fig. 5. Identical amino acids are indicated by two dots, and conservative substitutions are shown by a single dot.

ical function of the tsBN51 protein and its role in cell cycle progression.

ACKNOWLEDGMENTS

We thank V. Levytska, E. Deutsch, and E. Nonon for their assistance.

This investigation was supported by Public Health Service grant CA45268 from the National Cancer Institute. M.I. was the recipient of Public Health Service fellowship GM-11401. A.G. was supported by a fellowship from the Associazione Italiana per La Ricerca Sul Cancro.

LITERATURE CITED

1. **Baserga, R.** 1985. The biology of cell reproduction, p. 69–80. Harvard University Press, Cambridge, Mass.
2. **Basilico, C.** 1977. Temperature-sensitive mutations in animal cells. Adv. Cancer Res. **24**:233–266.
3. **Beach, D., B. Durkacz, and P. Nurse.** 1982. Functionally homologous cell cycle control genes in budding and fission yeast. Nature (London) **300**:706–709.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
5. **Birkenmeyer, L. G., J. C. Hill, and L. B. Duman.** 1984. Saccharomyces cerevisiae CDC8 gene and its product. J. Mol. Biol. **4**:583–590.
6. **Breter, H., J. Ferguson, T. Peterson, and S. Reed.** 1983. Isolation of three genes which function at start, the controlling event of the *Saccharomyces cerevisiae* cell division cycle: *CDC36*, *CDC37*, and *CDC39*. Mol. Cell. Biol. **3**:881–891.
7. **Camonis, J., M. Kalekine, B. Gondre, H. Garreau, E. Boy-Marcotle, and M. Jacquet.** 1986. Characterization, cloning and sequence analysis of the CDC25 gene which controls the cyclic AMP level of Saccharomyces cerevisiae. EMBO J. **5**:375–380.
8. **Campisi, J., H. Gray, A. Pardee, M. Dean, and G. Sonenshein.** 1984. Cell-cycle control of c-myc but not c-ras expression is lost following chemical transformation. Cell **36**:241–247.
9. **Chafouleas, J., L. Lagace, W. Bolton, A. Boyd, and A. Means.** 1984. Changes in calmodulin and its mRNA accompany reentry of quiescent (G0) cells into the cell cycle. Cell **36**:73–81.
10. **Cochran, B. H., A. C. Reffel, and C. D. Stiles.** 1983. Molecular cloning of gene sequences regulated by platelet-derived growth factor. Cell **33**:939–947.
11. **Frischauf, A. M., H. Lehrach, A. Poustka, and N. Murray.** 1983. Lamba replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**:827–842.
12. **Graham, F. L., and A. J. Van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
13. **Greco, A., M. Ittmann, and C. Basilico.** 1987. Molecular cloning of a gene that is necessary for G1 progression in mammalian cells. Proc. Natl. Acad. Sci. USA **84**:1565–1569.
14. **Greenberg, M. E., and E. B. Ziff.** 1984. Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. Nature (London) **311**:433–438.
15. **Hanahan, D., and M. Meselson.** 1983. Plasmid screening at high colony density. Methods Enzymol. **100**:333–342.
16. **Hirschhorn, R. R., P. Aller, Z. Yvan, C. W. Gibson, and R. Baserga.** 1984. Cell-cycle-specific cDNAs from mammalian cells temperature sensitive for growth. Proc. Natl. Acad. Sci. USA **81**:6004–6008.
17. **Hopp, T., and K. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
18. **Jelinek, W. R., T. P. Toomey, L. Leinwand, C. H. Duncan, P. A. Bizo, P. V. Choudary, S. M. Weissman, C. M. Rubin, C. M. Houck, P. L. Deininger, and C. W. Schmid.** 1980. Ubiquitous, interspersed repeated sequences in mammalian genomes. Proc. Natl. Acad. Sci. USA **77**:1398–1402.
19. **Kahana, C., and D. Nathans.** 1984. Isolation of cloned c-DNA encoding mammalian ornithine decarboxylase. Proc. Natl. Acad. Sci. USA **81**:3645–3649.
20. **Kai, R., M. Ohtsubo, M. Sekiguchi, and T. Nishimoto.** 1986. Molecular cloning of a human gene that regulates chromosome condensation and is essential for cell proliferation. Mol. Cell. Biol. **6**:2027–2032.
21. **Kelly, K., B. H. Cochran, C. D. Stiles, and P. Leder.** 1983. Cell-specific regulation of the c-myc gene by lymphocytes, mitogens and platelet-derived growth factor. Cell **36**:603–610.
22. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
23. **Kyte, J., and R. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
24. **Lau, L. F., and D. Nathans.** 1985. Identification of a set of genes expressed during the G0/G1 transition of cultured mouse cells. EMBO J. **4**:3145–3151.
25. **Lawn, R., E. Fritsch, R. Parker, G. Blake, and T. Maniatis.** 1978. The isolation and characterization of linked α- and β-globin genes from a cloned library of human DNA. Cell **15**:1157–1174.
26. **Lipman, D., and W. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **117**:1435–1441.
27. **Liu, H. T., C. W. Gibson, R. R. Hirschhorn, S. Rittling, R. Baserga, and W. E. Mercer.** 1985. Expression of thymidine kinase and dihydrofolate reductase genes in mammalian ts mutants of the cell cycle. J. Biol. Chem. **260**:3269–3274.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 282–286. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Marcus, M., A. Fainsod, and G. Diamond.** 1985. The genetic analysis of mammalian cell-cycle mutants. Annu. Rev. Genet. **19**:389–421.
30. **McLachlan, A., and J. Karn.** 1982. Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle. Nature (London) **299**:226–231.
31. **Nasmyth, K. A., and S. Reed.** 1980. Isolation of genes by complementation in yeast: molecular cloning of a cell-cycle gene. Proc. Natl. Acad. Sci. USA **77**:2119–2123.
32. **Nishimoto, T., and C. Basilico.** 1978. Analysis of a method for selecting temperature-sensitive mutants of BHK cells. Somat. Cell Genet. **4**:323–340.
33. **Okayama, H., and P. Berg.** 1982. High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2**:161–170.
34. **Pentecost, B., J. Wright, and G. Dixon.** 1985. Isolation and sequence of c-DNA clones coding for a member of the family of high mobility group proteins (HMG-T) in trout and analysis of HMG-T-mRNAs in trout tissues. Nucleic Acids Res. **13**:4871–4887.
35. **Pledger, W. J., C. D. Stiles, H. N. Antoniades, and C. D. Scher.** 1977. Induction of DNA synthesis in BALB/c 3T3 cells by serum components: reevaluation of the commitment process. Proc. Natl. Acad. Sci. USA **74**:4481–4485.
36. **Reeck, E. R., P. J. Isakson, and D. C. Teller.** 1982. Domain structure in high molecular weight high mobility group nonhistone chromatin proteins. Nature (London) **300**:76–78.
37. **Reed, S.** 1980. The selection of S. cerevisiae mutants defective in the start event of cell division. Genetics **95**:561–577.
38. **Reich, N., and A. Levine.** 1984. Growth regulation of a cellular tumor antigen, p53, in nontransformed cells. Nature (London) **308**:199–201.
39. **Sanger, R., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
40. **Schumperli, D.** 1986. Cell-cycle regulation of histone gene expression. Cell **45**:471–472.
41. **Smith, J. C., and C. D. Stiles.** 1981. Cytoplasmic transfer of mitogenic response to platelet-derived growth factor. Proc. Natl. Acad. Sci. USA **78**:4363–4367.
42. **Staden, R.** 1980. An interactive graphic program for comparing and aligning nucleic acid and amino acid sequences (DIAGON). Nucleic Acids Res. **8**:3673–3694.
43. **Stiles, C. D., G. T. Capone, C. D. Scher, H. N. Antoniades, J. J. Van Wyk, and W. J. Pledger.** 1979. Dual control of cell growth by somatomedins and platelet-derived growth factor. Proc.

Natl. Acad. Sci. USA **76**:1279–1283.

44. **Theissen, H., M. Etzerodt, R. Reuter, C. Schneider, F. Lottspeich, P. Argos, R. Luhrmann, and L. Philipson.** 1986. Cloning of the human c-DNA for the U1 RNA-associated 70K protein. EMBO J. **5**:3209–3217.
45. **Walker, J., K. Gooderham, J. Hastings, E. Mayes, and E. Johns.** 1980. The primary structure of non-histone chromosomal proteins HMG1 and 2. FEBS Lett. **122**:264–270.
46. **Wigler, M., A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, and L. Chasen.** 1979. DNA-mediated transfer of the adenine phosphoribosyltransferase locus into mammalian cells. Proc. Natl. Acad. Sci. USA **76**:1373–1376.