A complex unidirectional signal element mediates GCN4 mRNA 3' end formation in Saccharomyces cerevisiae.

C M Egli, C Springer and G H Braus
Mol. Cell. Biol. **1995, 15(5):2466.**

MOLECULAR AND CELLULAR BIOLOGY, May 1995, p. 2466–2473 Vol. 15, No. 5
0270-7306/95/$04.00+0

A Complex Unidirectional Signal Element Mediates *GCN4* mRNA 3′ End Formation in *Saccharomyces cerevisiae*

CHRISTOPH M. EGLI, CHRISTOPH SPRINGER, AND GERHARD H. BRAUS*

Institute of Microbiology, Biochemistry & Genetics, Friedrich Alexander University, D-91058 Erlangen, Germany, and Institute of Microbiology, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

Received 14 November 1994/Returned for modification 6 January 1995/Accepted 21 February 1995

The yeast *GCN4* 3′ element represents a class of polyadenylation sites which function unidirectionally and efficiently in test systems in vivo as well as in vitro. A complex signal element is required for polyadenylation activity with a minimal size of 116 nucleotides for the functional element. We subdivided this element into five regions (EL1 to EL5) of 16 to 26 nucleotides each. Each region was characterized by deletion analysis in an in vivo test system. Two TTTTTAT motifs are located in different regions (EL1 and EL4) upstream of the poly(A) site. The 3′ end processing activity was significantly reduced when both motifs were mutated by site-directed mutagenesis and abolished when EL1 and EL4 were deleted. The major poly(A) site is located in EL5, 3 nucleotides downstream of the second TTTTTAT motif. Additional minor poly(A) sites are used in less than 10% of the mRNA 3′ ends. Deletion of EL3 resulted in a changed pattern of mRNA 3′ ends by increased usage of the minor poly(A) addition sites. The major poly(A) site in EL5 can be removed without loss of function when sequences upstream of EL1 are present. The tripartite TAG...TATGT...TTT sequence located downstream of EL5 is not required for function.

In higher eukaryotes, the 3′ ends of mature mRNAs are generated in two different reaction steps. Endonucleolytic cleavage of a precursor transcript at the poly(A) site is followed by addition of up to 200 adenosine (A) residues to the newly produced mRNA 3′ end (reviewed in references 12 and 25). Two signal sequences on the pre-mRNA, one strictly conserved and the other less conserved, are required to perform these reaction steps. The hexanucleotide AAUAAA is located 10 to 30 nucleotides upstream of the poly(A) site and represents the strictly conserved motif. The other motif is a GU- or U-rich sequence located downstream of the poly(A) site. Several different protein factors consisting of various subunits bind to these signal sequences (reviewed in reference 25). The 3′ ends of mRNAs are formed similarly in the yeast *Saccharomyces cerevisiae* (4, 5). Fractionation of yeast cell extracts revealed at least four components required for cleavage and polyadenylation, including at least two cleavage factors, a poly(A) polymerase and a polyadenylation factor (6). The factors derived from *S. cerevisiae* and higher eukaryotes are not interchangeable (reviewed in reference 25). The canonical AAUAAA hexanucleotide is present in only approximately 50% of the yeast transcripts and does not seem to be functional in *S. cerevisiae*. The yeast sequences which have been proposed as signals for mRNA 3′ end formation show no obvious similarity to the hexanucleotide element of higher eukaryotes (13, 16; for a review, see reference 19). Analysis of six different polyadenylation elements in an in vivo test system revealed substantial differences in efficiency and orientation dependence (14). Two classes of elements have been proposed (15), i.e., efficient orientation-dependent sites (unidirectional) and symmetrical polyadenylation sites with reduced efficiency (bidirectional). Efficient polyadenylation elements contain the octanucleotide sequence TTTTTATA (11). The less efficient bidirectional processing elements contain the tripartite sequence motif TAG...TA(T)GT(A)...TTT (26). The importance of this motif and related sequence elements, such as TAG...TA(T)GT(A), TATATA, TACATA, TATGTA, and TACGTA, has been demonstrated for several genes (1, 14, 20, 21). None of these motifs is sufficient for directing mRNA 3′ end formation on its own, but rather they are sufficient only in combination with other sequence elements at or close to the poly(A) site (10, 15, 16, 21). The sequence elements TATATA, TAG ...TATGTA, and TTTTTATA act as upstream elements of the bidirectional *CYC1* polyadenylation signal. The downstream element TTAAGAAC or AAGAA seems to be involved in exact positioning of the poly(A) site (20). It appears that there are different types of *cis*-acting polyadenylation signals in *S. cerevisiae*. These signals include degenerate or redundant sequence motifs and do not fit into a uniform model.

There is barely any knowledge about the efficient and unidirectional polyadenylation signals in *S. cerevisiae*. This prompted us to investigate the *GCN4* 3′ element as a representative of this class of polyadenylation elements. We characterized this polyadenylation element in detail by introducing various *GCN4* 3′ end modifications into an in vivo test construct (Fig. 1A).

MATERIALS AND METHODS

Yeast strains, media, and methods. *S. cerevisiae* RH1631 (*MAT***a** *ura3-52*) was used for transformation of all of the integrative plasmids constructed. Cell extracts for in vitro processing reactions were prepared from strain RH1419 (*matα trp1 pro1-126 prb1-112 pep4-3 prc1-126*). Both yeast strains were derivatives of standard laboratory strain *S. cerevisiae* X2180-1A (*MAT***a** *gal2 SUC2 mal CUP1*). Yeast strains were cultivated in YEPD complete medium or MV minimal medium (18). Yeast transformation (17), DNA isolation (3), and Southern analysis (24) were done as previously described.

Enzymes and oligonucleotides. Restriction enzymes were purchased from Boehringer GmbH (Mannheim, Germany) and New England BioLabs (Schwalbach, Germany). Vent DNA polymerase was purchased from New England BioLabs, and T7 and SP6 RNA polymerases and T7 DNA polymerase were obtained from Pharmacia (Uppsala, Sweden). Nuclease S1 and exonuclease *Bal* 31 were purchased from Boehringer. Oligonucleotides (listed in Table 1) were synthesized by Microsynth (Windisch, Switzerland).

Plasmid construction and cloning. Plasmid pME798 was constructed on the basis of pSP64 (Promega, Madison, Wis.) to obtain an integrative vector. Vector pSP64 was modified by cloning the 1.1-kb *Hin*dIII fragment of *URA3* into the

* Corresponding author. Mailing address: Institute of Microbiology, Biochemistry & Genetics, Friedrich-Alexander-Universität, Staudtstrasse 5, D-91058 Erlangen, Germany. Phone: (49)-9131 858260. Fax: (49)-9131 858254.

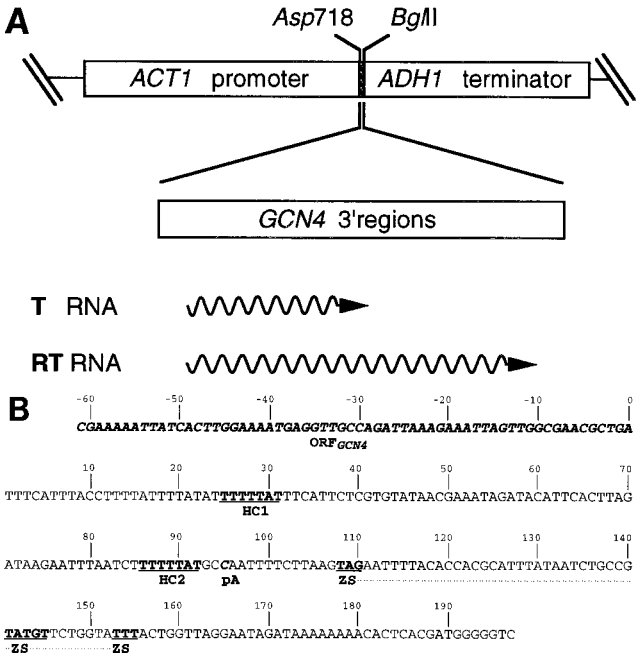

FIG. 1. In vivo test cassette for either wild-type or mutant 3′ processing signals in *S. cerevisiae*. (A) The test cassette consists of the *ACT1* promoter, a multiple cloning site (*Asp* 718-*Cla*I-*Bgl*II; black bar), and the *ADH1* terminator. Functional 3′ processing sites cloned into the multiple cloning site result in short truncated transcripts (T), whereas nonfunctional sites result in long readthrough transcripts (RT). The 259-bp *GCN4* 3′ element and all of the modified *GCN4* fragments (see Fig. 4) were cloned into the multiple cloning site. The wild-type sequence caused mRNA 3′ processing with an efficiency of approximately 90%. (B) The primary sequence of the *GCN4* 3′ UTR and a part of the open reading frame (ORF; in boldface italics) are shown. The two TTTTTAT signals, designated HC1 and HC2, and the tripartite ZS motif TAG...TATGT...TTT, which is located downstream of the poly(A) signal (in boldface and labeled pA), are putative consensus sequences and are underlined and in boldface. The numbers correspond to the assignment of position 0 to the A nucleotide of the TGA stop codon at the end of the *GCN4* open reading frame.

*Xho*I site by inserting the 1.1-kb *Bam*HI fragment of pME729 (16) into the *Bam*HI site of the polylinker and by introducing a multiple cloning site (double-stranded OLCE1-OLCE2) into the *Cla*I site of the 1.1-kb *Bam*HI fragment. The 259-bp *Taq*I fragment of the *GCN4* 3′ terminal part was amplified by using OLCE3 and OLCE6 and cloned into the multiple cloning site after restriction with *Asp* 718 and *Bgl*II. The resulting vector was used for construction of the various *GCN4* 3′ ends.

Plasmids were linearized with *Asp* 718 for the 5′ end deletion set or with *Bgl*II for the 3′ end deletion set and were treated with the *Bal* 31 exonuclease enzyme. The ends of the shortened plasmids were made blunt and ligated with a *Cla*I linker [5′-d(CCAT/CGATGG)-3′]. Deletion fragments were isolated by using restriction enzymes *Asp* 718 and *Cla*I for the 3′ deletion set or *Cla*I and *Bgl*II for the 5′ deletion set. The resulting fragments were cloned into the appropriate linearized polylinker site of plasmid pME798. Additional deletions were constructed by PCR cloning by using oligonucleotides OLCE3, -4, -5, -6, -7, -12, -13, -16, and -17 (listed in Table 1).

Internal deletions were constructed by joining appropriate fragments (see Fig. 4) derived from the 5′ and 3′ deletion sets by using the common *Cla*I site. A second set of internal deletions was generated to analyze possible spacing effects of the deleted fragments. Therefore, a 38-bp spacer DNA was cloned into the *Cla*I site of relevant constructs after filling in the sticky ends of OLCE1/2 and the vectors to become blunt. All constructs were verified by DNA sequencing (23).

For in vitro transcription experiments, the different *GCN4* 3′ end fragments were cloned into the polylinker of the vector pGEM7-Zf(+) (Promega). The 259-bp wild-type 3′ end fragment was cloned after restriction with *Taq*I into the *Cla*I site of the multiple cloning site of pME798, thus obtaining both orientations of the insert. The internal deletion fragment spanning the sequences between positions −5 and +92 was amplified by PCR with OLCE4 and OLCE7 as primers. The resulting fragment was restricted with *Asp* 718 and *Bgl*II and inserted into the corresponding sites of the polylinker.

RNA substrates and processing of synthetic pre-mRNA in vitro. In vitro transcription was carried out with either T7 or SP6 RNA polymerase as previously described (4). The synthetic RNA, which was capped and ^{32}P labeled, was isolated by electrophoresis on a 4% polyacrylamide gel containing 6 M urea. RNA was eluted from the gel and subjected to processing by using yeast cell extracts prepared as previously described (5).

Isolation of total RNA from *S. cerevisiae*. Yeast cells were grown overnight in a 500-ml culture to an optical density at 546 nm of about 2. The cells were spun at 2,500 × *g* for 13 min on one-fifth of a volume of ice and resuspended in 30 ml of PLE buffer [100 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES), 100 mM LiCl, 1 mM EDTA (pH 7.4)]. After centrifugation at 6,000 × *g* for 7 min at 4°C, the cells were resuspended in 3 ml of ice-cold PLE buffer and 1 ml of ice-cold dichloromethane-saturated phenol equilibrated with PLE buffer. Diethylpyrocarbonate (1% [vol/vol]) was added to inactivate RNase. Sterilized glass beads 0.45 mm in diameter were added, and the cells were disrupted by vigorous shaking for two 40-s periods with cooling on ice in between. Nucleic acids were extracted once with 1 volume of dichloromethane-saturated phenol equilibrated with PLE buffer, 0.5 g of bentonite, and 1% (wt/vol) sodium dodecyl sulfate and twice with 1 volume of dichloromethane-saturated phenol equilibrated with PLE buffer. Each time, the mixture was shaken for 10 min at room temperature and centrifuged at 6,000 × *g* for 15 min at 4°C. Total RNA was precipitated by 1.5 volumes of ice-cold isopropanol, and the concentration was determined spectrophotometrically. The precipitated RNA was stored at −20°C. Poly(A) RNA was obtained by enrichment over an oligo(dT) cellulose column (2).

RNA analysis and nuclease S1 mapping. For Northern (RNA) hybridization experiments, approximately 10 μg of total RNA or 5 μg of poly(A) RNA was precipitated, resuspended, and denatured in 30 μl of sample buffer (50% [vol/vol] deionized formamide, 6% [vol/vol] formaldehyde, 1× loading buffer, 10% [vol/vol] 10 mM Tris–1 mM EDTA [TE] buffer) for 5 min at 65°C and put on ice. The RNA was separated on a denaturing formaldehyde agarose gel. The 1.4% (wt/vol) agarose gel (3% [vol/vol] formaldehyde, 20 mM morpholinepropanesulfonic acid [MOPS], 5 mM Na acetate, 1 mM EDTA) was run for 3 h at 120 V in a buffer containing 20 mM MOPS, 5 mM Na acetate, and 1 mM EDTA. The gel

TABLE 1. Oligonucleotides[a] used in this study

Designation	Sequence (5′ to 3′)	Positions[b]	Strand[c]
OLCE1	d(CTCGAGCCAGATCTGCCATCGATTGGTACCATCTCGAG)		
OLCE2	d(CTCGAGATGGTACCAATCGATGGCAGATCTGGCTCGAG)		
OLCE3	d(CGCGGGTACCCTTGGAAAATGAGGTTGCCAG)	−48/−28	A
OLCE4	d(CGCGGGTACCCGCTGATTTCATTTACC)	−5/+11	A
OLCE5	d(CGCGGGTACCGCCAATTTTCTTAAGTAG)	+93/+110	A
OLCE6	d(GCGCAGATCTCCCCCATCGTGAGTG)	+195/+181	B
OLCE7	d(GCGCAGATCTATAAAAAAGATTAAATTCTTATC)	+92/+70	B
OLCE9	d(GGGAATTCGGTCAATCTTTGTTAAAGAATAGG)		
OLCE12	d(GCGCGGGTACCGATTAAAGAAATTAGTTGGCG)	−28/−8	A
OLCE13	d(GGCGCAGATCTCTACTTAAGAAAATTGGCAT)	+110/+91	B
OLCE14	d(CCGAAGCTCGAGCTGCGGATCC$(T)_{17}$		
OLCE15	d(CCGAAGCTCGAGCTGCGGATCC)		
OLCE16	d(GTTATACACGAGAATGAAGGGCCCAATATCGATAAAAGGTAAATGAAATC)	+49/−1	B
OLCE17	d(GATAAGAATTTAATCTTTTTTATGCCAATTTTCTTAAGTAG)	+70/+110	A

[a] The oligonucleotides used were purchased from Microsynth.
[b] Nucleotide numbers are defined relative to the A residue of the translational stop codon of the *GCN4* open reading frame, which was assigned number 0.
[c] The letters A and B indicate oligonucleotides corresponding to the *GCN4* coding and noncoding strands, respectively.

was soaked twice in 25 mM Na phosphate buffer for 20 min each time, and the RNA was transferred onto a nylon membrane (Amersham) by electroblotting (2 A, 50 V) for 3 to 4 h in 25 mM Na phosphate buffer. After washing in 2× SSC (1× SSC is 0.15 M NaCl plus 0.15 M sodium citrate), drying on 3MM paper, and cross-linking under UV light (254 nm) for 5 min, the membrane with the bound RNA was hybridized at 42°C with a labeled fragment for 24 h at 42°C in 50 ml of a hybridization mixture (50% [vol/vol] formamide, 50 mM Na phosphate [pH 6.5], 800 mM NaCl, 1 mM EDTA, 0.5% [wt/vol] sodium dodecyl sulfate, 10× Denhardt's solution, 150 μg of calf thymus DNA per ml, 500 μg of torula yeast RNA per ml). The fragment, representing the 440-bp *Mlu*I-*Xho*I DNA element of the *ACT1* 5′ region, was randomly radiolabeled as previously described (8). The RNA was visualized by autoradiography. Band intensities from autoradiographs were quantified with a Phosphorimager (Molecular Dynamics, Sunnyvale, Calif.). For 3′ end mapping of transcripts, DNA fragments were labeled with α-^{32}P-labeled nucleotides and avian myeloblastosis virus reverse transcriptase. The fragments were restricted at the *Ava*II site in the *ACT1* promoter and at the *Sph*I site in the *ADH1* polyadenylation element with the *GCN4* polyadenylation element located in between. Fragments were radiolabeled and annealed to 50 mg of total RNA or 10 mg of poly(A) RNA in 30 ml of 80% formamide–0.04 M PIPES (pH 6.4)–0.4 M NaCl–1 mM EDTA by denaturation for 15 min at 85°C followed by hybridization at 42°C (7). A total of 300 ml of ice-cold S1 buffer (0.28 M NaCl, 0.05 M Na acetate [pH 4.6], 4.5 mM $ZnSO_4$, 20 mg of denatured sonicated calf thymus DNA per ml) and 200 U of nuclease S1 were added, and the mixture was incubated for 30 min at 37°C. The reaction was stopped by addition of 75 ml of 2.5 M ammonium sulfate–50 mM EDTA. The nucleic acids were precipitated by ethanol and fractionated on denaturing polyacrylamide gels containing 7 M urea.

Amplification of cDNA ends by PCR. Mapping of the 3′ ends was carried out as described earlier (9, 21). Oligonucleotide OLCE14 (Table 1) was used for the reverse transcription reaction, and OLCE9 and OLCE15 were used for specific amplification of the cDNA. OLCE9 includes a 5′-overhanging *Eco*RI restriction site and binds in the *ACT1* promoter region, and OLCE15 corresponds to the sequence adjacent to the OLCE14 binding site. The amplified fragments were isolated from a 0.8% low-melting-point agarose gel, restricted with the enzymes *Eco*RI and *Bam*HI, and cloned into pGEM7Zf(+) restricted with the same enzymes. DNA sequencing by the dideoxy method served as a control for correct plasmid constructs (23).

RESULTS

The *GCN4* polyadenylation element (Fig. 1B) represents the class of efficient polyadenylation elements which function only unidirectionally in an in vivo test system (15). The 3′ untranslated region (UTR) of the *GCN4* gene contains two TTTT TAT motifs (11), designated HC1 and HC2, and a tripartite sequence motif, TAG...TATGT...TTT, designated ZS (26) located 15 bp downstream of the HC2 motif (Fig. 1B).

In a first step, we tested whether this polyadenylation element is also processed in an orientation-dependent manner in an in vitro system. In a second step, we wanted to define the *cis*-acting elements which are important for *GCN4* 3′ end formation by analyzing mutant forms of the polyadenylation element. Modifications of the *GCN4* 3′ UTR included 5′ and 3′ end and internal deletions, combinations of various fragments, and specific point mutations inserted into the complete element (see Fig. 4). The modified *GCN4* 3′ end elements were cloned into the multiple cloning site of a test gene consisting of the *ACT1* promoter and the *ADH1* terminator (16; Fig. 1A). The test gene was integrated into the chromosome at the *URA3* locus, thereby avoiding multicopy effects. The effects of all modifications were analyzed at the transcript level by performing Northern blot analysis. Functional 3′ processing elements resulted in short truncated transcripts, whereas nonfunctional elements resulted in long readthrough transcripts. The complete wild-type *GCN4* element, spanning the region from position −61 in the *GCN4* open reading frame (relative to the translational stop codon) to position +197, resulted in a 3′ processing efficiency of 85 to 94% truncated transcripts and, accordingly, 6 to 15% readthrough transcripts in various independent experiments in the forward orientation (see Fig. 5, lanes wt+). The efficiency of the inverse orientation (wt−) was below the lowest measurable level (see Fig. 5).

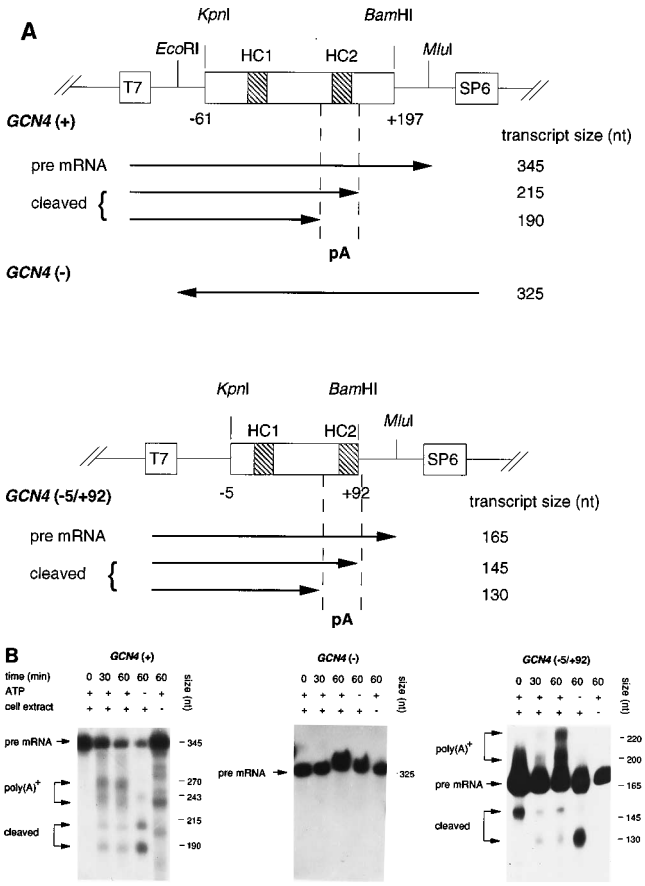

FIG. 2. In vitro processing of synthetic transcripts containing the wild-type and modified *GCN4* polyadenylation elements. (A) All fragments tested were cloned into the *Kpn*I and *Bam*HI restriction sites of the vector pGEM7-Zf(+). In vitro transcripts were expressed from either the T7 or SP6 promoter after linearization downstream of the insert with *Mlu*I and *Eco*RI, respectively. A sense or antisense transcript was generated from the wild-type 259-bp *GCN4* fragment (−61 to +197) containing the distal element with the wild-type poly(A) site. Only a sense transcript was generated from the construct with the internal *GCN4* element from −5 to +92, which contained no distal sequences downstream of the HC2 element. The poly(A) sites are labeled pA. The transcripts were incubated in yeast extracts at 30°C (4). (B) The products of the processing reaction were separated on a 4% polyacrylamide denaturing gel. The pre-mRNA, 5′ cleavage product (cleaved), and polyadenylated cleavage products [poly(A)$^+$] are indicated. The incubation conditions are shown above the lanes. Sizes were calculated by linear regression and are shown in nucleotides (nt).

Pre-mRNAs containing the *GCN4* 3′ end formation element are processed in an orientation-dependent manner in vitro. The *GCN4* element is functional in only one orientation in an in vivo test system (15), whereas polyadenylation elements derived from the *CYC1*, *GAL1*, *GAL7*, and *GAL10* genes caused processing in vitro and in vivo in both orientations (4, 22). The 259-bp fragment containing the *GCN4* polyadenylation element (Fig. 1B) was transcribed in both orientations by using either the T7 or SP6 RNA polymerase, leading to primary transcripts of 345 and 325 bp, respectively (Fig. 2). In contrast to that of other polyadenylation elements, processing of the *GCN4* antisense transcript was not detectable in a partially fractionated yeast cell extract. The sense transcript was cleaved at position +95 and at an additional site upstream of the HC2 motif (Fig. 2). The *GCN4* 3′ end which is generated in vitro was subsequently polyadenylated with a poly(A) tail length of approximately 70 A residues. Processing of a different *GCN4* fragment spanning the region from positions −5 to +92

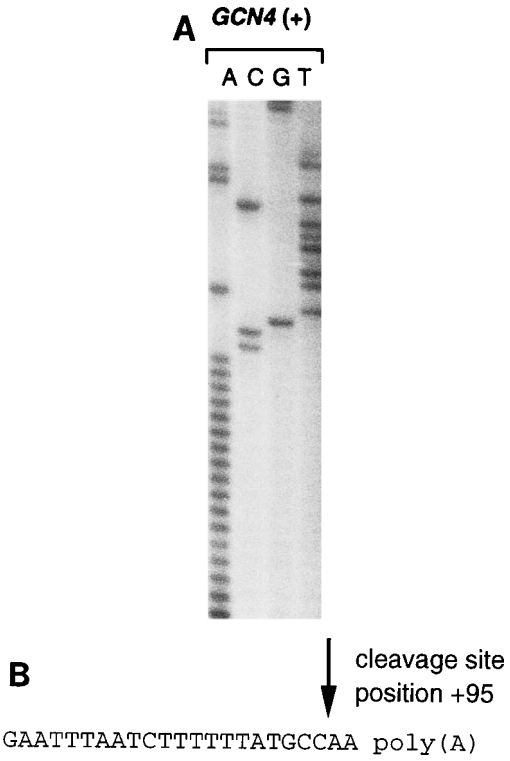

FIG. 3. Determination of the major poly(A) site of the *GCN4* wild-type 3′ element. (A) Sequence analysis of the cloned cDNA derived from RNA of the test gene carrying the wild-type *GCN4* 3′ element. One cDNA sequence of the nine analyzed sequences is shown as an example. The cDNAs were amplified as previously described (21), cloned into an *Eco*RI-*Bam*HI-cleaved pGEM77Zf(+) vector, and sequenced. The poly(A) site is located at position +95 in all nine cDNAs. The poly(A) tail lengths varied between 17 and 41 A residues, with an average value of 24 A residues. (B) Primary sequence of the cDNAs analyzed. The poly(A) addition site at position +95 is relative to the A nucleotide of the translational stop codon.

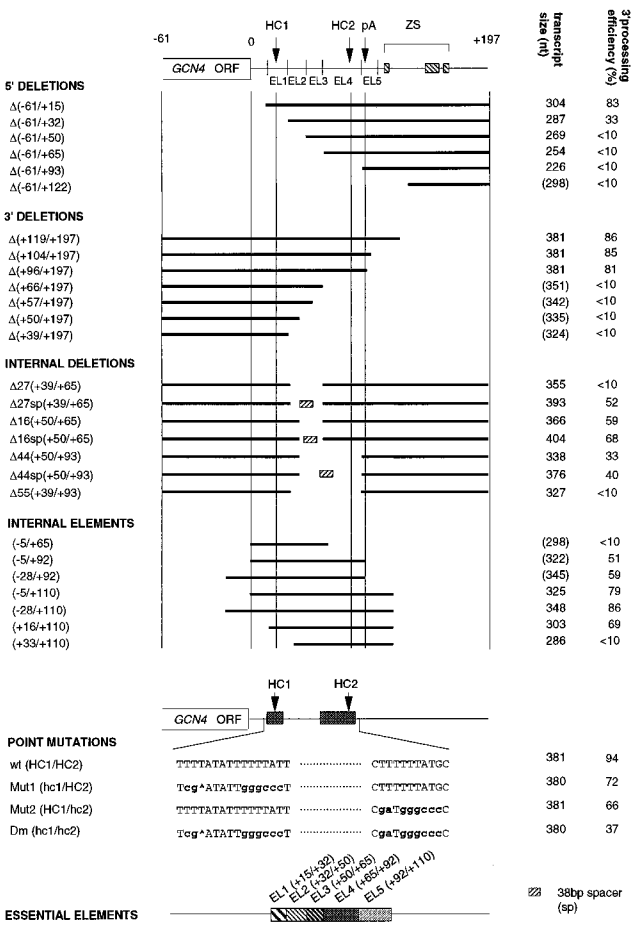

FIG. 4. Schematic representation of the mutations constructed in the *GCN4* 3′ UTR, the resulting transcript sizes, and the effects on 3′ processing efficiency (portion of truncated transcripts relative to total transcripts of the test genes). The transcript sizes were calculated from the 5′ cap site to the major poly(A) site at position +95. Parentheses indicate that the major poly(A) site was deleted, and the numbers indicate the calculated size from the 5′ cap site to the 3′ end of the cloned *GCN4* sequence. The 5′ and 3′ deletions were obtained by using the *Bal* 31 exonuclease assay and a *Cla*I linker. Internal deletions were constructed by connecting appropriate 5′ and 3′ deletions by using the added *Cla*I restriction site. Internal elements and point mutations were generated by PCR. The numbers after the delta indicate the numbers of deleted base pairs, whereas the numbers in parentheses indicate the positions of the boundaries in relation to the A nucleotide of the translational stop codon TGA, which was numbered 0. The 38-bp multiple cloning site which was used for construction of the test gene was used as a spacer element (sp) and does not show any 3′ processing activity. The nucleotide exchanges in the point mutations are indicated with small boldface letters, whereas a caret indicates deletion of a single nucleotide. nt, nucleotides.

also resulted in two cleavage sites. S1 mapping (see Fig. 6) and rapid amplification of cDNA ends-PCR mapping (Fig. 3) revealed that the poly(A) site at position +95 relative to the A residue of the translational stop codon TGA was the only major 3′ end detected in vivo. Additional minor poly(A) sites were used in less than 10% of the mRNA 3′ ends. Therefore, this in vivo poly(A) site at position +95 is referred to as the major poly(A) site in the following text.

Deletion analysis delimits the unidirectional *GCN4* polyadenylation element. The *GCN4* polyadenylation element has been subdivided into five regions (EL1 to EL5; Fig. 4) to define the sequences required for 3′ end formation. 5′ deletion up to position +15 relative to the A of the translational stop codon TGA (deletion Δ−61/+15 in Fig. 5A) resulted in a 3′ processing efficiency (83% truncated transcripts) similar to that of the complete wild-type element spanning the region from position −61 in the *GCN4* open reading frame to position +197 (85 to 94% truncated transcripts). Further deletion to position +32 reduced the 3′ processing activity to 33% (deletion Δ−61/+32). Therefore, the upstream boundary for a completely functional element was located in the +15 to +32 region. This region also included the TTTTTAT motif HC1 and was named EL1.

Deletion of 18 additional nucleotides up to position +50 (deletion Δ−61/+50) completely abolished 3′ end formation. We designated the +32 to 50 region, containing two overlapping TGTATA/TATAA motifs, EL2 as depicted at the bottom of Fig. 4.

An internal deletion of EL2 and 11 additional nucleotides within the complete *GCN4* polyadenylation element [Δ27(+39/+65)] showed complete loss of activity. Insertion of a 38-bp polylinker spacer DNA fragment into this internal deletion [Δ27sp(+39 to +65)] resulted in partial recovery of the activity (52% of the truncated transcript) of the *GCN4* polyadenylation element. The spacer DNA alone was not able to provide any polyadenylation activity in the in vivo mRNA 3′ end formation test system (data not shown).

The internal deletion Δ16(+50/+65) which was constructed by linking the appropriate 5′ and 3′ deletion fragments together reduced *GCN4* 3′ processing activity (59% truncated transcripts) compared with the complete (−61 to +197) wild-

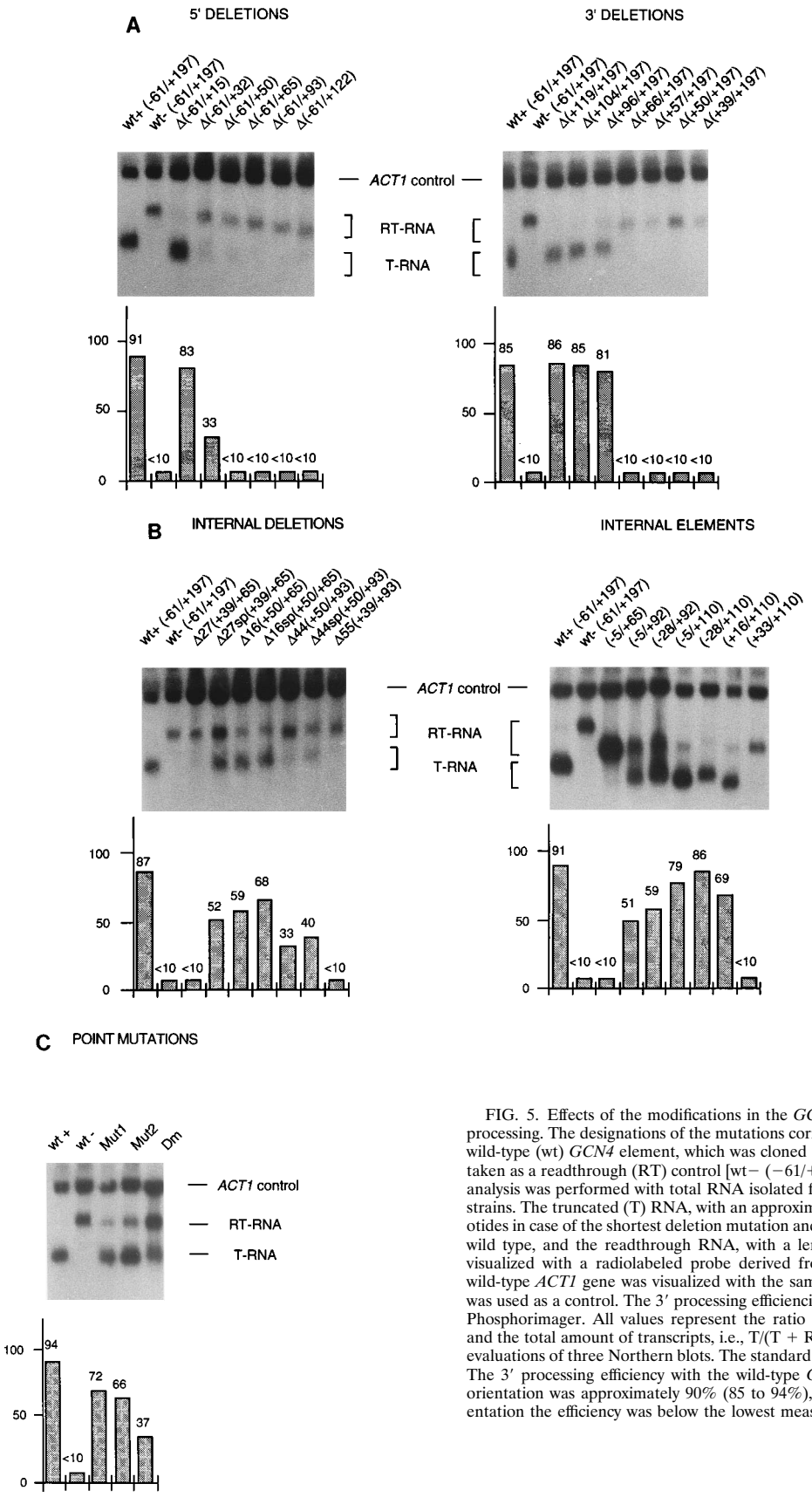

FIG. 5. Effects of the modifications in the *GCN4* 3′ elements on mRNA 3′ processing. The designations of the mutations correspond to those in Fig. 4. The wild-type (wt) *GCN4* element, which was cloned in the inverse orientation, was taken as a readthrough (RT) control [wt− (−61/+197)]. Northern hybridization analysis was performed with total RNA isolated from the wild-type and mutant strains. The truncated (T) RNA, with an approximate length between 300 nucleotides in case of the shortest deletion mutation and 380 nucleotides in case of the wild type, and the readthrough RNA, with a length of 720 nucleotides, were visualized with a radiolabeled probe derived from the *ACT1* promoter. The wild-type *ACT1* gene was visualized with the same radio labeled fragment and was used as a control. The 3′ processing efficiencies were determined by using a Phosphorimager. All values represent the ratio between truncated transcripts and the total amount of transcripts, i.e., T/(T + RT), and each is the average of evaluations of three Northern blots. The standard deviation did not exceed 20%. The 3′ processing efficiency with the wild-type *GCN4* element in the forward orientation was approximately 90% (85 to 94%), whereas with the inverse orientation the efficiency was below the lowest measurable level.

type activity (85 to 94% truncated transcripts). Insertion of the same 38-bp spacer DNA fragment as described above [Δ16sp(+50 to +65)] did not greatly increase polyadenylation efficiency (68% truncated transcripts). The region between positions +50 and +65 was designated EL3. It contained the motif TAGATA, which represents a modification of a previously described hexanucleotide consensus signal in *S. cerevisiae* (14).

The role of the DNA sequences downstream of EL3 in mRNA 3′ end formation have been analyzed by 3′ deletions starting at position +197 of the −61 to +197 wild-type element. 3′ deletions downstream of position +96 did not significantly reduce polyadenylation efficiency (Fig. 5A). Deletion of the major poly(A) addition site [Δ(+92/+197)] also did not reduce polyadenylation efficiency (data not shown). 3′ end formation was completely abolished in the deletion Δ(+66/+197). Therefore, we named the region downstream of EL3, spanning the DNA between positions +50 and +92, EL4. Region EL4 contains, in addition to the HC2 motif, the overlapping sequences TAGATA and AAGAA, which had been proposed to be involved in 3′ processing and poly(A) site positioning (14, 20).

The internal deletion of the EL3 and EL4 regions within the complete wild-type polyadenylation element [Δ44(+50/+93)] reduced 3′ processing activity to 33% truncated transcripts. The processing activity of the deletion mutant was not strongly improved by insertion of the 38-bp spacer DNA (Δ44sp [+50 to +93]; 40% truncated transcripts; Fig. 5B). 3′ processing activity fell below the measurable level when EL2, EL3, and EL4 were simultaneously deleted [Δ55(+39/+93); Fig. 5B].

In summary, the deletion analysis narrows the *GCN4* polyadenylation element to the region between position +15 and position +92, subdivided into regions EL1 to EL4. The region from positions +92 to +110 contained the major poly(A) addition site and was named EL5. The deletion analysis did not suggest that this region is required for efficiency of polyadenylation.

The tripartite sequence motif TAG...TATGT...TTT in the *GCN4* 3′ UTR is not essential for 3′ end formation. The tripartite sequence motif TAG...TATGT...TTT (ZS) is located downstream of the poly(A) site, and such sequences had been shown to be functional in several other polyadenylation elements (15, 26). We tested whether the ZS sequence motif of *GCN4* could be functional for 3′ processing in vivo or in vitro to further delimit the 3′ boundary of the *GCN4* polyadenylation element.

The 3′ deletion mutants Δ(+119/+197) and Δ(+104/+197) contain only a partial or no ZS motif and resulted in only truncated but no readthrough transcripts (Fig. 5B). Deletion of all sequences upstream of the tripartite sequence motif resulted in no truncated transcript [Δ(−61/+93)]. Addition of regions EL3 and EL4 or EL1 as elements upstream of the ZS motif did not restore any polyadenylation activity [Δ(−61/+50) and Δ55(+39/+93)]. Partial 3′ processing activity (33% truncated transcripts) was restored when the downstream region with the ZS sequence motif was combined with EL1 and EL2 as upstream sequences [Δ44(+50/+93)]. Variation of the spacing between the two fragments in mutation Δ44sp(+50/+93) hardly increased activity.

The 5′ deletion mutant Δ(−61/+93), including the complete ZS sequence, was tested in an in vitro assay. The fragment was expressed by a phage T7 promoter, and the resulting RNA was incubated with a functional cell extract (5). No cleavage and polyadenylation of the in vitro-generated transcript were observed (data not shown). This corroborated the finding that the tripartite motif is neither functional on its own nor involved in *GCN4* 3′ processing.

A minimal *GCN4* polyadenylation element consisting of 116 nucleotides is fully active. The deletion analysis suggested that DNA regions EL1, EL2, EL3, and EL4 are required for 3′ processing activity. Various internal fragments (Fig. 4) were tested for in vivo polyadenylation in the test system to define a minimal fully active polyadenylation element. The internal fragment (−5 to +65) including EL1 to EL3 was not sufficient to provide any detectable polyadenylation activity. Addition of EL4 (−5 to +92) provided only partial polyadenylation activity (51% truncated transcripts) in comparison with the fully active (−61 to +197) wild-type fragment (85 to 94% truncated transcripts). A larger fragment which is extended by sequences located further upstream (−28 to +92) did not significantly improve polyadenylation activity. A fragment which contained the EL5 region, including the poly(A) addition site (−5 to +110), restored the wild-type polyadenylation activity. Therefore, the smallest fragment which was fully functional for 3′ processing (79% truncated transcripts) spanned 116 nucleotides between positions −5 and +110 and included EL1, EL2, EL3, EL4, and EL5 (Fig. 4 and 5B). The larger fragment, from −28 to +110, did not significantly increase 3′ end formation activity, indicating that no part of the open reading frame is strictly required for 3′ end formation.

Mutations in EL1 and EL4 decrease *GCN4* 3′ end formation efficiency. The analysis of the polyadenylation activity of the internal fragments pointed to a specific role of regions EL1 and EL4 (containing TTTTTAT motifs HC1 and HC2) in polyadenylation activity. EL1, including the HC1 motif, seems to be required for 3′ processing because the internal element (+33 to +110) lacking the EL1 region is not functional. Addition of 17 nucleotides including EL1 restored the 3′ processing activity of the element (+16 to +110; Fig. 5B). A comparison of the activities of the two internal fragments from −5 to +92 and from −5 to +65 suggested an essential role of EL4 in polyadenylation. Whereas the fragment spanning the region from positions −5 to +92 was partially functional, the processing activity was abolished in the internal fragment from −5 to +65, which lacked most of EL4, including the HC2 motif (Fig. 5B).

Therefore, both TTTTTAT motifs (HC1 and HC2) located in EL1 and EL4 were mutated in the complete wild-type element (−61 to +197) to test their influence on *GCN4* mRNA 3′ processing. The HC1 motif, including the 5′ flanking AT-rich TTTATATTTTTTAT sequence, was mutated to CGATATTG GGCCC in mutant Mut1 (Fig. 4). The HC2 motif was changed by several point mutations from the primary sequence TTTTT TATG to GATGGGCCC (Mut2 in Fig. 4). Both Mut1 and Mut2 resulted in similar decreases in polyadenylation efficiency (Fig. 5C). The combination of Mut1 and Mut2, resulting in the double mutation Dm, had a significant effect on 3′ end formation and resulted in only 37% truncated transcripts.

Deletion of upstream sequences can induce increased usage of cryptic poly(A) addition sites. The in vitro results obtained with the fragment from −5 to +92 showed that a precursor transcript was cleaved in the absence of the major poly(A) addition site and that cleavage took place at the same distance from upstream elements (e.g., HC2) as in the wild-type situation. The 3′ ends of various transcripts were analyzed to find determinants for poly(A) site selection (Fig. 6).

Nuclease S1 analysis revealed that one major poly(A) site is used in the fragment from −5 to +110 (Fig. 6), which confirmed that the required determinants were located on the minimal polyadenylation element. The 3′ end was determined

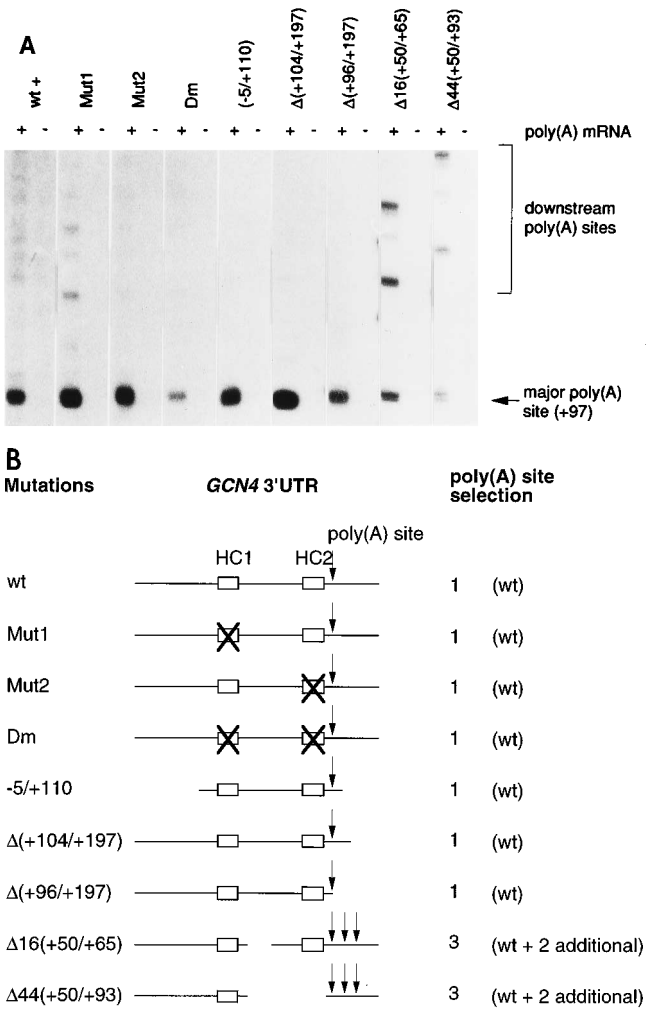

FIG. 6. Determination of the positions of *GCN4* poly(A) addition sites. (A) Nuclease S1 mapping of the 3′ ends of the truncated transcripts. Poly(A)$^+$-enriched RNAs from *S. cerevisiae* strains with the integrated *GCN4* 3′ end mutations were hybridized to a complementary 3′ end-labeled DNA fragment (ranging from the *Ava*II site in the *ACT1* intron to the *Sph*I restriction site in the *ADH1* terminator). Nuclease S1-resistant hybrids were analyzed on denaturing polyacrylamide gels by using sequence ladders from the corresponding constructs and *Hin*fI- or *Hpa*II-digested plasmid pBR322 as size standards. As a negative control (−), torula yeast RNA was used. The numbers correspond to the assignment of position 0 to the A nucleotide of the TGA stop codon. The single lanes were cut and aligned according to the poly(A) site at position +95 to facilitate comparison. All sizes were calculated by linear regression. The similar patterns of the downstream poly(A) sites correspond to identical positions in the *GCN4* 3′ UTR. (B) Schematic representation of the poly(A) site selection in the different mutations. The poly(A) sites where the primary transcripts are cleaved and polyadenylated are indicated with arrows. wt, wild type.

at the same position as in the *GCN4* wild-type mRNA major poly(A) site.

The 3′ deletions Δ(+104/+197) and Δ(+96/+197), lacking parts of the EL5 region, were analyzed to test whether downstream sequences are required for poly(A) site positioning. The 3′ ends of both transcripts were mapped as unique signals at the same position as the wild-type major poly(A) site.

We further analyzed whether the HC motifs are involved in the positioning. In the mutants carrying Mut1, Mut2, and Dm, the same wild-type major poly(A) site was preferentially used. Long exposure of the autoradiograph showed the pattern of the additional minor 3′ ends of larger size than the major wild-type transcript.

The internal deletion mutant Δ16(+50/+65) was used to determine whether EL3 was involved in poly(A) site positioning. Three signals of similar intensities suggested that the major poly(A) site is partially shifted to the poly(A) addition sites located more downstream which correspond to the minor 3′ ends described above (Fig. 6). These data suggested that EL3 contains determinants for poly(A) site selection.

The EL4 region contains a motif (AAGAA) located 17 nucleotides upstream of the major poly(A) site. This motif resembles the sequences TTAAGAAC and AAGAA, which had been shown to be determinants for poly(A) site positioning in the yeast *CYC1* gene (21). The poly(A) site of internal deletion mutant Δ44(+50/+93), lacking EL4 in addition to EL3, was mapped to determine whether this region was also involved in poly(A) site positioning in the *GCN4* polyadenylation element. The transcript ends were at positions similar to those in deletion mutant Δ16(+50/+65). These results suggested that the major determinants for poly(A) site positioning are located in EL3.

DISCUSSION

This report provides the first extensive analysis of a highly efficient orientation-dependent polyadenylation signal. Synthetic transcripts carrying the 259-bp *GCN4* 3′ end sequence are processed only in the sense but not in the antisense orientation in vitro. A minimal *GCN4* element of 116 bp with high A and T residue contents was constructed which still provided full polyadenylation activity of the test mRNA in vivo. The tripartite TAG...TATGT...TTT consensus sequence ZS located downstream of the major poly(A) site was not required for the minimal *GCN4* 3′ processing element. This ZS motif was not able to direct 3′ processing of the test gene either in vivo or in vitro, although it resembles other yeast 3′ processing signals which are functional in the same test systems (16). Even in the presence of most of the additional 3′ UTR sequences which have been shown to be important for 3′ processing, the tripartite sequence was not functional. Partial restoration of 3′ processing activity was achieved only when the ZS sequence motif was combined with EL1 and EL2 in a construct missing EL3 and EL4. One possibility is that the close proximity to regions EL1 and EL2 can activate the function of the ZS sequence.

For refined analysis, we subdivided the minimal polyadenylation element into regions EL1 to EL5. This subdivision was the result of extensive deletion analysis and examination of several DNA fragments in the test system which seemed to be promising candidates for a functional minimal element. EL5, which contains the major poly(A) addition element, was required for construction of a functional minimal element but could be removed by 3′ deletions without loss of polyadenylation function when *GCN4* sequences upstream of the minimal polyadenylation element were present. These sequences remain to be identified.

The minimal *GCN4* polyadenylation element is covered with sequence motifs which are identical or similar to proposed sequences involved in mRNA 3′ end formation (14, 20). The specific role of any of these elements as specific *cis*-acting sequence motifs remains to be elucidated.

In addition, effects on 3′ processing activity by transcript size have to be considered. Most of the *GCN4* fragments analyzed differ in the distance between the 5′ cap and the mapped poly(A) site (Fig. 4). We did not find any obvious correlation between transcript size and 3′ processing efficiency, but we

cannot completely rule out the possibility that the size of the transcript is a parameter which plays a role in 3′ processing.

The importance of RNA secondary structures in 3′ processing activity is unclear. Regions EL1 and EL4 contain the TTTTTAT sequence motif, which was previously described (11), and mutations within these motifs decrease *GCN4* polyadenylation efficiency. The changes introduced in HC1 and HC2 theoretically have the potential to form RNA secondary structures. Thus, e.g., the Mut1 sequence ATTGG(G-bulge)C could base pair directly with the GCCAAT sequence at the poly(A) site. Since the binding of protein factors might strongly affect and change any RNA structure and since it is not clear which proteins bind to which sites of the *GCN4* 3′ UTR, the role of RNA secondary structures is difficult to demonstrate.

Region EL3 seems to carry determinants required for correct positioning of the actual poly(A) site relative to the upstream elements. EL3 determines the correct distance of polyadenylation even when the authentic nucleotides at this position are missing. Deletion of EL3 results in increased usage of minor poly(A) addition sites located downstream of the major poly(A) site. There may be cellular mechanisms which allow the activation of these cryptic poly(A) sites.

Further investigation will identify the sequences on EL3 which are required for poly(A) site positioning. For the *CYC1*, *GAL1*, and *GAL7* elements, it has been shown that short regions of 20 nucleotides or less which are located downstream of the polyadenylation sites are sufficient for accurate processing (22). A TTAAGAAC or AAGAA sequence is required for exact poly(A) site positioning of the *CYC1* polyadenylation element. The AAGAA sequence motif in the *GCN4* polyadenylation element which is located in EL4 does not seem to be involved in poly(A) site positioning.

In conclusion, the *GCN4* polyadenylation signal represents a complex element which does not easily fit into the current 3′ processing models, in which mainly three components are required for proper 3′ end formation in *S. cerevisiae*. These sequences are an upstream enhancer-like element, a downstream element which is important for positioning of the poly(A) site, and the poly(A) site itself. Future experiments have to elucidate whether the high 3′ processing efficiency of the *GCN4* element is the result of the high number of redundant consensus motifs representing an increased number of binding sites for protein factors and whether different classes of polyadenylation sites differ in the specific *trans*-acting factors that recognize the poly(A) site and its environment.

ACKNOWLEDGMENTS

We thank Ralf Hütter for generous support. We are especially grateful to Markus Künzler, Stefan Irniger, Katrin Düvel, and Chris Berens for helpful discussion.

This work was supported by the Deutsche Forschungsgemeinschaft and by grants from the Swiss Federal Institute of Technology.

REFERENCES

1. **Abe, A., Y. Hiraoka, and T. Fukasawa.** 1990. Signal sequence for generation of mRNA 3′ end in the *Saccharomyces cerevisiae GAL7* gene. EMBO J. **9:**3691–3697.
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin mRNA by chromatography on oligothymidylic cellulose. Proc. Natl. Acad. Sci. USA **69:**1408–1412.
3. **Braus, G., R. Furter, F. Prantl, P. Niederberger, and R. Hütter.** 1985. Arrangements of genes *TRP1* and *TRP3* of *Saccharomyces cerevisiae* strains. Arch. Microbiol. **142:**383–388.
4. **Butler, J. S., and T. Platt.** 1988. RNA processing generates the mature 3′ ends of yeast *CYC1* mRNA *in vitro*. Science **242:**1270–1274.
5. **Butler, J. S., P. P. Sadhale, and T. Platt.** 1990. RNA processing in vitro produces mature 3′ ends of a variety of *Saccharomyces cerevisiae* mRNAs. Mol. Cell. Biol. **10:**2599–2605.
6. **Chen, J., and C. Moore.** 1992. Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. Mol. Cell. Biol. **12:**3470–3481.
7. **Favoloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65:**718–749.
8. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137:**266–267.
9. **Frohman, M. A., M. K. Dush, and G. R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:**8998–9002.
10. **Heidmann, S., B. Obermaier, K. Vogel, and H. Domdey.** 1992. Identification of pre-mRNA polyadenylation sites in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:**4215–4229.
11. **Henikoff, S., and E. H. Cohen.** 1984. Sequences responsible for transcription termination on a gene segment in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4:**1515–1520.
12. **Humphrey, T., and N. J. Proudfoot.** 1988. A beginning to the biochemistry of polyadenylation. Trends Genet. **4:**243–245.
13. **Hyman, L. E., S. H. Seiler, J. Whoriskey, and C. L. Moore.** 1991. Point mutations upstream of the yeast *ADH2* poly(A) site significantly reduce the efficiency of 3′ end formation. Mol. Cell. Biol. **11:**2004–2012.
14. **Irniger, S., and G. H. Braus.** 1994. Saturation mutagenesis of the polyadenylation signal reveals a hexanucleotide element essential for mRNA 3′ end formation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **91:**257–261.
15. **Irniger, S., C. M. Egli, and G. H. Braus.** 1991. Different classes of polyadenylation sites in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **11:** 3060–3069.
16. **Irniger, S., H. Sanfaçon, C. M. Egli, and G. H. Braus.** 1992. Different sequence elements are required for function of the cauliflower mosaic virus polyadenylation site in *Saccharomyces cerevisiae* compared with in plants. Mol. Cell. Biol. **12:**2322–2330.
17. **Ito, H., Y. Jukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:**163–168.
18. **Miozzari, G., P. Niederberger, and R. Hütter.** 1978. Tryptophan biosynthesis in *Saccharomyces cerevisiae*: control of the flux through the pathway. J. Bacteriol. **153:**163–168.
19. **Proudfoot, N.** 1991. Poly(A) signals. Cell **64:**671–674.
20. **Russo, P., W. Z. Li, Z. Guo, and F. Sherman.** 1993. Signals that produce 3′ termini in *CYC1* mRNA of the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:**7836–7849.
21. **Russo, P., W. Z. Li, D. M. Hampsey, K. S. Zaret, and F. Sherman.** 1991. Distinct cis-acting signals enhance 3′ endpoint formation of *CYC1* mRNA in the yeast *Saccharomyces cerevisiae*. EMBO J. **10:**563–571.
22. **Sadhale, P. P., and T. Platt.** 1992. Unusual aspects of in vitro RNA processing in the 3′ regions of the *GAL1*, *GAL7*, and *GAL10* genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:**4262–4270.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
24. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
25. **Wahle, E., and W. Keller.** 1992. The biochemistry of 3′ end cleavage and polyadenylation of messenger RNA precursors. Annu. Rev. Biochem. **61:** 419–440.
26. **Zaret, K. S., and F. Sherman.** 1982. DNA sequences required for efficient transcription termination in yeast. Cell **28:**563–573.