Recruitment of mRNA cleavage/polyadenylation machinery by the yeast chromatin protein Sin1p/Spt2p

Gitit Hershkovits, Haim Bangio*, Ronit Cohen, and Don J. Katcoff†

The Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, Ramat Gan 52900, Israel

Edited by Roger D. Kornberg, Stanford University School of Medicine, Stanford, CA, and approved May 18, 2006 (received for review March 13, 2006)

The yeast chromatin protein Sin1p/Spt2p has long been studied, but the understanding of its function has remained elusive. The protein has sequence similarity to HMG1, specifically binds crossing DNA structures, and serves as a negative transcriptional regulator of a small family of genes that are activated by the SWI/SNF chromatin-remodeling complex. Recently, it has been implicated in maintaining the integrity of chromatin during transcription elongation. Here we present experiments whose results indicate that Sin1p/Spt2 is required for, and is directly involved in, the efficient recruitment of the mRNA cleavage/polyadenylation complex. This conclusion is based on the following findings: Sin1p/Spt2 frequently binds specifically downstream of many ORFs but almost always upstream of the first polyadenylation site. It directly interacts with Fir1p, a component of the cleavage/polyadenylation complex. Disruption of Sin1p/Spt2 results in foreshortened poly(A) tracts on mRNA. It is synthetically lethal with Cdc73p, which is involved in the recruitment of the complex. This report shows that a chromatin component is involved in 3′ end processing of RNA.

Results
Sin1p Binds Downstream of Many Genes, Just Upstream of a Major Polyadenylation Site. Sin1p is a chromatin component whose presence affects the transcription of a number of genes as is described above. Harbison et al. (16) mapped DNA-binding transcriptional regulators, including Sin1p (Spt2p), to the genome by using a chromatin immunoprecipitation assay. Careful examination of their data indicates that Sin1p often binds downstream of ORFs. In many cases, the binding site is downstream of two ORFs that are on opposing strands, precluding the possibility that Sin1p is binding promoter sequences at those positions. We compared the locations of the Sin1p-binding sites on 12 different genes having high affinity to Sin1p to predicted (20) and actual polyadenylation sites. The actual polyadenylation sites were determined experimentally by using 3′ RACE (see Materials and Methods). It is apparent from the results shown in Fig. 1 that in almost all cases, Sin1p binds ~50 to 100 bp upstream of the first polyadenylation site after the ORF. In some cases, the binding is a bit further downstream, but usually, in the context of chromatin, not more than one nucleosome away from the 3′ end of the coding region. These data suggested that Sin1p might play a role in 3′ processing events of the nascent mRNA transcript.

Sin1p Interacts Directly with Fir1p. We previously reported results from a two-hybrid screen (21, 22) by using Sin1p as bait. In addition to recovering the previously reported colonies containing rtf1 double mutants, Paf1p, Ctr9p and Rtf1p are all members of the PAF complex. In the same study, using ChiP analysis, Spt2p was found preferentially associated with DNA that is actively transcribed, binding to 3′ untranslated regions more frequently than expected. Indeed, careful examination of Sin1p-binding sites on the DNA identified previously (16) shows that Sin1p often binds DNA downstream of ORFs (see Results).

All these data indicated that there might be a role for Sin1p in 3′ end formation of mRNA in yeast. In this paper, we show that Sin1p is associated with Fir1p, a component and positive regulator of the RNA cleavage/polyadenylation complex (17). Furthermore, deletion of Sin1p, or mutations in its C terminus, result in foreshortened poly(A) stretches at the 3′ termini of mRNA. We present a model suggesting that the polyadenylation machinery is recruited to the nascent RNA transcript by interactions with the heavily phosphorylated CTD of RNA polymerase II at serine-2 (18, 19), the PAF complex that accompanies the polymerase (11–13), and Sin1p that is bound to the chromatin just downstream of the ORF of many yeast genes.

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: CTD, C-terminal domain.

*Present address: Procognia Israel, Ashdod 77610, Israel.
†To whom correspondence should be addressed. E-mail: katcoff@mail.biu.ac.il.

© 2006 by The National Academy of Sciences of the USA
Fig. 1. Schematic representation of Sin1p binding site and polyadenylation sites on selected genes. The last 500 nucleotides of the coding region plus 500 nucleotides of the 3' noncoding region are displayed. The numbers on the abscissa are relative to the 3' end of the ORF. All genes selected scored with a high confidence level for Sin1p binding (16). The thick green line denotes the ORF. The black vertical arrow denotes the Sin1p binding site. The black and blue graphs denote predicted polyadenylation sites using the maximum-likelihood (ML) and the discrete state-space (DSM) models, respectively (20). The red arrows denote actual polyadenylation sites determined by 3' RACE. Major polyadenylation sites are marked with a thick arrow, while more minor sites are marked with a thin arrow. A black horizontal arrow denotes the position of the left primer used in the 3' RACE. The gel at the right of each graph shows a sample 3' RACE lane for each gene. Marker sizes are 1000, 900, 800, 700, 600, 500, 400, 300, 250, 200, 150, 100, and 50 bp.
ing Cdc23 and Spp1, we also recovered colonies encoding amino acids 178–876 of Fir1p in the same screen. A direct interaction between the two proteins was confirmed by showing that 35S-labeled, in vitro translated Fir1p (amino acids 173–876) bound GST-Sin1p fusion proteins derived from the N terminus of the protein, whereas GST alone did not bind (Fig. 2). A foreshortened peptide containing amino acids 173–810 bound the GST-Sin1p equally well (data not shown). This data supports the possible involvement of Sin1p in the 3′ processing of nascent mRNA transcripts, because Fir1p was identified recently as a positive regulator of polyadenylation (17).

Several hypothetical functional domains in Sin1p that have been defined based on genetic, structural, and biochemical considerations (6, 9, 21, 23–25). These domains include two HMG1-like domains, (amino acids 26–88 and 98–159), an acidic domain (amino acids 224–304), and a basic C-terminal domain (amino acids 303–333). To address which of these domains interacts with Fir1p, we synthesized portions of Sin1p fused to GST based on the functional domains that have been suggested, bound them to glutathione-agarose beads, and asked whether the radiolabeled Fir1 peptide 173–810 would bind to Sin1p fragments. As can be seen in Fig. 2, peptides containing the first HMG1-like domain (amino acids 41–99) bound the Fir1 peptide, whereas molecules containing the other domains did not bind Fir1p. A fusion peptide containing only a part of this HMG1-like domain (amino acids 1–64) was able to bind Fir1p, but less well than those peptides containing the entire domain. In previous work, we showed that a Sin1p peptide containing this HMG1-like domain could bind four-way junction DNA (9). Taken together, these results demonstrate that the first HMG1-like domain of the protein can interact with both Fir1p and with DNA. Because it is unlikely that this peptide interacts with both DNA and Fir1p simultaneously, we suggest that interaction between Fir1p and Sin1p probably results in Sin1p displacement from the DNA.

**Lack of a Functional Sin1p Results in Foreshortened poly(A)**. To test whether the interaction between the chromatin component (Sin1p) and the polyadenylation regulator (Fir1p) has functional significance, we compared average poly(A) length in yeast strains carrying deletions in one or both of these proteins. In our assay, we isolated whole-cell RNA from each strain and labeled the 3′ termini with [32P]pCp by using RNA ligase. Ten percent of each reaction described in detail in Materials and Methods was loaded per lane. After autoradiography, the phosphorimage was quantitated by using IMAGEJ (see Figs. 6 and 7, which are published as supporting information on the PNAS web site).

**Fig. 2.** Association between Sin1p and Fir1p peptides in vitro. GST or GST/Sin1p bacterially produced fusion proteins were immobilized on 200-μl glutathione agarose beads. Fir1p peptides labeled with [35S]methionine were produced in the Promega TNT-coupled transcription-translation system. Ten microliters of the reaction was incubated with the glutathione agarose beads. After washing, bound proteins were eluted with glutathione, and three-quarters of the eluted volume was subjected to SDS-PAGE and autoradiography. The multiple bands represent partial Fir1p molecules that result from premature transcriptional and translational termination, internal transcrip-
tional initiation, and RNA and protein degradation. The arrow shows the full-length protein.

**Fig. 3.** Poly(A) length determination of total cell RNA. End-labeled RNA was digested with RNase and loaded on a 15% polyacrylamide sequencing gel. Ten percent of each reaction described in detail in Materials and Methods was loaded per lane. After autoradiography, the phosphorimage was quantitated by using IMAGEJ (see Figs. 6 and 7, which are published as supporting information on the PNAS web site).
strains (17). Our in vivo results agree with those in vitro results, showing that fir1 deletions result in poly(A) tracts that are somewhat foreshortened (Fig. 3, compare lanes 1 and 4). Strikingly, a strain carrying the sin1 deletion together with a fir1 deletion had significantly shorter poly(A) tracts than the fir1 deletion alone, suggesting that the Sin1p might contact additional polyadenylation components in addition to Fir1p.

Neither Sin1p nor Fir1p Affect poly(A) Site Selection. Because our results intimated that Sin1p is involved in polyadenylation, we asked whether it or Fir1p are involved in selection of the site at which the mRNA is cleaved and polyadenylated. This question was addressed by performing 3′ RACE on RNA samples that had been isolated from wild-type, sin1 and fir1 strains, and a sin1, fir1 double mutant. Of the 21 reactions performed, four sample reactions are displayed in Fig. 4. No significant differences were seen between the samples, indicating that neither Sin1p nor Fir1p participate in determination of the cleavage and polyadenylation site.

Discussion

Almost all mRNA transcripts in eukaryotes are polyadenylated at their 3′ ends by a complex of proteins that are largely conserved throughout evolution. Polyadenylation has been shown to be important for mRNA stability, its protection from ribonucleases, transcription termination, export from the nucleus, and translation (26). Polyadenylation occurs in conjunction with transcriptional termination in several steps. Before or concurrent with termination, the transcript is cleaved at specific sites, a tract of adenylate residues is added, and it is then trimmed in a transcript-specific manner. These reactions are highly coordinated and are carried out by a host of factors. In yeast, these factors include poly(A)-binding protein (Pab1p), cleavage and polyadenylation factors CFIA, CFIB, CPF, and poly(A) nuclease PAN (17). Each factor with the exception of CFIB is encoded by multiple polypeptides. Additional regulatory factors are associated with these complexes. These factors include Fir1p (876 aa) that positively regulates mRNA polyadenylation and Ref2p that regulates it negatively.

Recent progress has been made that addresses how the cleavage/polyadenylation complex is recruited to the nascent RNA. The CTD of RNA polymerase II, which becomes progressively more phosphorylated at serine-2 as it progresses during elongation, interacts with Pcf11 (18, 19) and serves as a termination factor (27).

Sin1p is synthetically lethal with cdc73, paf1, and ctr9 (10, 15), all components of the PAF elongation complex. Recent data indicate that the PAF complex facilitates the linkage of transcriptional and posttranscriptional events, particularly polyadenylation (12). Although the PAF complex associates with RNA polymerase II during initiation and elongation, it is required for proper polyadenylation of at least a subset of transcripts (13). In the absence of Cdc73p, the remaining PAF components are not associated with the transcribing polymerase, and polyadenylation is foreshortened. It has been suggested that the PAF complex functions in 3′ end formation by recruitment of the 3′-end-processing factors (11). The synthetic lethality of Sin1p with components of this complex, and the fact that cdc73 and sin1 mutants individually result in foreshortened poly(A) tracts, imply that Sin1p and components of the PAF complex have required overlapping functions.

Nourani et al. (15) have shown that Sin1p is frequently associated with the coding region of actively transcribed genes and that its presence is important for the maintenance of histone H3 at transcribed regions. In addition, sin1Δ mutants display increased intrachromosomal recombination in actively transcribed regions (15). These data suggested that the nonsequence specific association of Sin1p with the transcribed DNA helps maintain the integrity of the chromatin during transcription.

We propose that as the transcription complex approaches a potential cleavage/polyadenylation site, the cleavage/polyadenylation complex is recruited by three components, all associated with the RNA polymerase II (Fig. 5): (i) The CTD of RNA polymerase II, which becomes increasingly phosphorylated at serine-2 (18, 28), interacts with Pcf11 in the CFIA component of the polyadenylation complex (19). (ii) Members of the PAF complex that are associated with the RNA polymerase II interact with as-yet-undetermined members of the polyadenylation complex (suggested by ref. 13). (iii) Sin1p, which may accompany the elongation complex (15), specifically binds a DNA sequence in the 3′ noncoding region, near the cleavage/polyadenylation site, and interacts with Fir1p in the cleavage/polyadenylation complex. If any of these three interactions is ablated (by gene disruption), polyadenylation is less efficient, and average cellular poly(A) is shorter than in the wild type. If any two of the “recruiters” are missing, polyadenylation is curtailed, and the result is lethality. This idea is supported by reported synthetic lethalities: (i) between ctk1, whose gene product phosphorylates the CTD at serine-2, and any of the PAF complex components (11); and (ii) between cdc73, paf1, or ctr9, whose gene products are part of the PAF complex, and Sin1 (10).

Although our data indicate that Sin1p participates in the recruitment of the mRNA processing/polyadenylation machinery and ChIP analysis (16) shows that it preferentially binds DNA just upstream of the polyadenylation site, Sin1p is not required to identify the polyadenylation site.

Hpr1p, which appears to be involved in chromatin remodeling, has been shown to be associated with the PAF complex (14). At 37°C, a strain lacking Hpr1p and expressing a semidominant sin1Δ-2 allele (29) is unable to grow, whereas strains with a sin1Δ deletion grow almost normally. It was suggested in that paper that in the wild type, Hpr1p might function to facilitate the
removal of Sin1p from the chromatin and that in its absence, in the sin1-2 strain, sin1-2p may be bound to the chromatin more tightly, causing toxicity. That interpretation is consistent with another observation that overexpression of the C-terminal of Sin1p (where the sin1-2 mutation is found) is toxic (8), but that this toxicity is suppressed in spt4, spt5, and spt6 strains. Each of these proteins, like Hpr1p, has been implicated in modification of chromatin structure. The Spt4/Spt5 complex interacts genetically and physically with Spt6, presumably to mediate transcription through chromatin (30, 31). Interestingly, recent data have shown a functional connection between Spt2p and Spt6p (15), and spt6 mutations can alter mRNA 3’ end formation at least in one documented case (32).

Sin1p has been isolated in a protein complex that includes Pub1p, a poly(A)+ RNA-binding protein, Npl3, a poly(A)+ RNA-transport protein, TIF4631 and TIF4632, proteins that associate with the poly(A)-binding protein Pab1p, and Sgn1p, another RNA-binding protein (33). These data also support Sin1p involvement in RNA processing.

We propose that, concomitant with its recognition by the polyadenylation complex, Sin1p is removed from the chromatin, allowing the transcription apparatus to pass with the recruited 3’ processing machinery in tow. The notion that Sin1p is removed from the chromatin is supported by the fact that the same short domain of Sin1p binds both DNA and Fir1p, and it is unlikely that both can interact with Sin1p simultaneously.

Sin1p was initially discovered as a suppressor of mutations in the SWI/SNF complex (2, 7). Later, Perez-Martín and Johnson (8) showed that the C terminus of Sin1p physically associates with components of the SWI/SNF complex and that this interaction is blocked in the full-length Sin1p protein by the N-terminal half of the protein. Recently, it has been reported that SWI/SNF can stimulate transcription even after promoter clearance (34) during elongation. Pollard and Peterson (1) reported that members of the ADA/GCN5 complex that can acetylate both histones and Sin1p are synthetically lethal with swi1, a component of the SWI/SNF complex. Mutations in Sin1 restore viability. Taking all these data together, we propose that the SWI/SNF complex, the ADA/GCN5 complex (via acetylation of Sin1p), and Hpr1p each facilitate the removal of Sin1p from the chromatin at a site downstream of the ORF of many genes. Continued research is necessary to elucidate the precise role of each of these factors in mRNA transcriptional termination and processing. The removal of Sin1p from the chromatin might be a prerequisite for the passage of the transcription apparatus, without which there can be no further processing of the mRNA.

We show that a chromatin component is required to complete mRNA processing, providing a link between chromatin structure and RNA processing. As details of the gene expression machinery are explored, it is becoming apparent that sometimes unexpected functional relationships exist between the different components (35–38). Further research will be necessary to determine the details of how the mRNA processing machinery is recruited and activated.

Materials and Methods

Yeast Strains. SLL5 (wild type with respect to Sin1), SLL7 (sin1Δ), and SLL101 (sin1 1-324) were gifts of L. Lefebvre (University of British Columbia, Vancouver; ref. 6). YIF3 is fir1::HIS3, leu2Δ, trp1Δ, ura3-52, his3-Δ200, ade2-101, lys2-801. YIF9 is YIF3 with the addition of sin1::LEU2, Cy110 is sin1::TRP1, leu2Δ, trp1Δ, ura3-52, his3-Δ200, ade2-101, lys2-801.

Two-Hybrid Screen and Verification of Protein Interactions. The two-hybrid screen by using Sin1p as a bait was carried out as in Shpunig et al. (22). A plasmid termed p82 was isolated from the libraries and was found to encode amino acids 173–876 of Fir1p. Plasmids p178-876 and p178-746 were constructed by PCR amplification of the partial Fir1 region using p82 as template and inserting them between the BamHI and PstI sites.
of pBluescript (Stratagene). Coupled transcription-translation of [35S]methionine-labeled Fir1p peptides was accomplished by using the Promega TNT system. Plasmids encoding the GST/Sin1p fusions were described in ref. 9. The in vitro-binding assay between the Sin1p and Fir1p peptides was carried out as in Novoseler et al. (9).

**poly(A) Length Determination.** The determination of poly(A) length for total cell RNA was done essentially as described in ref. 39. Briefly, [32P]pCp was ligated to the 3′ end of 0.5 μg of RNA in a 30-μl reaction by using 12 units of T4 RNA ligase (Fermentas, Burlington, ON, Canada). After a 24-h incubation at 4°C, 40 μg of yeast tRNA was added, and all but the poly(A) was digested by a mix of 80 units of RNase T1 and 32 units of RNase A at 37°C in a final volume of 80 μl. Finally, the enzymes were neutralized by proteinase K and phenol/chloroform extraction and ethanol precipitated. The samples then were run on a 15% polyacrylamide sequencing gel and autoradiographed. Quantitative evaluation of the autoradiogram was done by analysis of phosphorimage images by using IMAGEJ.

**3′ RACE.** 3′ RACE was performed on total cell RNA by using the FirstChoice RLM-RACE kit (Ambion, Austin, TX) according to the instructions of the manufacturer. The PCR uses an RNA template and amplifies a DNA fragment between a primer specific to a gene (see list in Table 1) and the 5′ end of the poly(A) segment at the 3′ end of the RNA molecule.

We thank L. Lefevbre for strains. This work was supported by internal Life Sciences Faculty funding at Bar Ilan University.