Molecular cloning and analysis of *Schizosaccharomyces pombe* Reb1p: sequence-specific recognition of two sites in the far upstream rDNA intergenic spacer

Annie Zhao, Ailan Guo, Zhao Liu and Louise Pape*

Department of Chemistry, New York University, New York, NY 10003, USA

Received August 23, 1996; Revised and Accepted December 16, 1996

**ABSTRACT**

The coding sequences for a *Schizosaccharomyces pombe* sequence-specific DNA binding protein, Reb1p, have been cloned. The predicted *S. pombe* Reb1p is 24–29% identical to mouse TTF-1 (transcription termination factor-1) and *Saccharomyces cerevisiae* REB1 protein, both of which direct termination of RNA polymerase I catalyzed transcripts. The *S. pombe* Reb1p cDNA encodes a predicted polypeptide of 504 amino acids with a predicted molecular weight of 58.4 kDa. The *S. pombe* Reb1p is unusual in that the bipartite DNA binding motif identified originally in *S. cerevisiae* and *Klyveromyces lactis* REB1 proteins is uninterrupted and thus *S. pombe* Reb1p may contain the smallest natural REB1 homologous DNA binding domain. Its genomic coding sequences were shown to be interrupted by two introns. A recombinant histidine-tagged Reb1 protein bearing the rDNA binding domain has two homologous, sequence-specific binding sites in the *S. pombe* rDNA intergenic spacer, located between 289 and 480 nt downstream of the end of the −25S rRNA coding sequences. Each binding site is 13–14 bp downstream of two of the three proposed in vivo termination sites. The core of this 17 bp site, AGGTAAAGGTTATGCAC, is specifically protected by Reb1p in footprinting analysis.

**INTRODUCTION**

Termination of RNA polymerase I transcription at the 3’-end of the eukaryotic ribosomal RNA (rRNA) coding sequences entails association of a sequence-specific DNA binding protein with its binding site(s) in the intergenic region of eukaryotic rRNA genes (1–5). The rRNA gene repeats of higher eukaryotes generally house termination elements at two locations, one location is downstream of the 3’-end of the rDNA coding region and the other is −200 bp upstream of the initiation site at the rDNA gene promoter (2,3,5–10). In the yeast *Saccharomyces cerevisiae*, the promoter proximal Reb1p binding site is in a reverse orientation incompetent for termination (5,11,12).

The sites of transcription termination and recognition sites of the RNA polymerase I termination factors have been extensively studied in mouse, human, *Xenopus* and the yeast *S. cerevisiae* (2,4,5,7–10,13,14). In the mouse rDNA intergenic spacer, the 18 bp sequence that serves as a signal for termination is a highly conserved 5’-AGGTCGACCAGTA/TT/NTCCG-3’, termed the Sal box (2,15), generally preceded by one or more T-rich pyrimidine clusters. While there are eight Sal boxes downstream of the end of the −45S rRNA coding sequences, the unique Sal box located 171 bp upstream of the initiation site may function to prevent dissociation of bound factors by RNA polymerase I molecules transcribing the intergenic rDNA (8,9). The ‘Sal box’ exerts its function by interacting with a sequence specific DNA binding protein, mTTF-1 (16). While the mTTF-1 coding sequences were isolated following partial sequence determination of the purified termination factor, another factor that has as one of its roles termination of polII transcription in yeast had been previously identified as an rDNA enhancer binding factor (REB1p; 17,18).

Molecular cloning of its encoding gene (18) revealed a protein of 809 amino acids with a bipartite DNA binding domain homologous to that of the Myb oncogene (18,19). The *Klyveromyces lactis* Reb1 p homolog was functional in the yeast *S. cerevisiae* and complemented a reb1− mutant strain of *S. cerevisiae* (19). Subsequent analysis demonstrated the activity of the Reb1 protein as an in vitro termination factor (12) for RNA polymerase I catalyzed synthesis. Two in vivo termination sites were mapped (10), with one near the binding site of the Reb1 protein, CCGTAAA, an essential site for termination (5). Another termination site −950 bp 3’ of the end of the −35S pre-rRNA does not contain a Reb1 binding site in the vicinity (10).

In vitro termination of RNA polymerase I catalyzed synthesis occurs 16 bp upstream of the *S. cerevisiae* Reb1 protein binding site (20) and 11 bp upstream of the mTTF-1 binding site (21). It was thus of interest that three putative in vivo termination sites were mapped in the *S. pombe* rDNA intergenic spacer several hundred nucleotides downstream of the 3’-end of the mature 25S rRNA (6). We were interested whether *S. pombe* directed termination via a protein homologous to the Reb1p/mTTF-1 polypeptides. In this paper we describe the identification and characterization of the *S. pombe* Reb1p coding sequences and demonstrate that there are two sequence-specific recognition sites downstream of two of these putative termination sites. The *S. pombe* Reb1p is the shortest of any identified thus far but retains a fully functional bipartite rDNA binding domain.

*To whom correspondence should be addressed. Tel: +1 212 998 8444; Fax: +1 212 260 7905; Email: pape101@mcrcr.med.nyu.edu
MATERIALS AND METHODS

Escherichia coli and yeast strains

The bacterial strains used were Escherichia coli M15[pREP4] (Qiagen) (NAL, str, rif, lac, ara, gal, mil, F-, recA, urv4), E.coli XL1-Blue (endA1, hsdR17 (rK, mK)), supE44, thi-1, λ-, recA1, lacY, relA1, lac [F', proAB, lacYIΔM15, Tn10 (tet)], and SUREm (mcrA, Δ(mcrBC-hsdRMS-mrr)], endA1, supE44, thi-a, λ, lacY, galK, lacZΔM15, Tn10, [tet]). The polymerase chain reaction (PCR) amplification reactions contained a 50-fold molar excess of gel purified rDNA. The linearized fragment was treated with calf alkaline phosphatase. The resultant plasmid encoded an N-terminal polyhistidine tag fused to the Reb1 coding sequences. Amplification from residues 147 to 504. Transformants were grown at 37°C to an OD$_{600}$ of 0.6–0.9 and expression of recombinant (His)$_6$Reb1p was induced by addition of the synthetic inducer IPTG to 1 μM for 4 h. The recombinant N-terminal deletion Δ147Reb1p was purified from inclusion bodies under denaturing conditions by affinity chromatographic separation (using a Ni-NTA column from Qiagen; 26). Purified (His)$_6$Reb1p molecules were renatured by stepwise dilution of 6 M urea to 1 M in dialysis buffer containing 20 mM Tris–HCl, pH 7.9, 5 mM MgCl$_2$, 0.1 mM EDTA, 0.02 M DTT (added fresh) and 20% glycerol (26). Protein concentration was estimated using the Bradford assay and comparison with standards using SDS–PAGE and staining with Craven (BioRad; 27).

DNA fragments used for protein–DNA binding assays

A clone containing rDNA intervening sequences from −3240 to −2555 (pDH1H2) was cleaved at a unique XhoI site in the pBS multiple cloning site upstream of the −3240 position of the S.pombe rDNA. The linearized fragment was treated with calf alkaline phosphatase and radiolabeled in the presence of T4 polynucleotide kinase and [γ-32P]ATP. Following a second digestion with HindIII, an 360 bp XhoI–HindIII fragment (containing rDNA from −3240 to −2828) containing three putative termination sites of +276, +338 and +448 (6) was purified following fractionation on a 6% acrylamide, 0.2% bisacrylamide, 0.5× TBE gel.

Gel retardation assays

The electrophoretic mobility shift assays (28) were performed as follows. Binding reactions contained 5–100 ng Reb1 protein, 200 ng poly(dI-dC-dI-dC) or 200 ng poly(dA-dT-dA-dT) or poly(dG-dC-dG-dC) (Pharmacia) (Fig. 3A), 5 fmol $^{32}$P- and end-labeled probe in 20 mM Tris–HCl, pH 7.9, 50 mM KCl, 5 mM MgCl$_2$, 0.1 mM EDTA, 0.2 mM DTT and 10% glycerol. Binding reactions were incubated for 20 min at room temperature and the reaction mixtures were electrophoresed in a 5% acrylamide, 0.13% bisacrylamide, 25 mM Tris–borate, 8.3 mM EDTA gel and electrophoresed at 200 V for 2 h at 4°C. Competitive binding reactions contained a 50-fold molar excess of gel purified rDNA fragments isolated as a XhoI–HindIII fragment (Fig. 4A) or a XhoI–BsrFI fragment with the upstream Reb1 binding site (Fig. 4B) and a BsrFI–HindIII fragment with the 3′ Reb1 binding site, Site 2 (Fig. 4B). The gels were dried and exposed to Kodak XAR-5 autoradiographic film. Quantitation was performed by scanning the gels using a GS-250 Molecular Image™ (BioRad).

DNase I footprinting assay

Aliquots of 5 fmol $^{32}$P-end-labeled DNA fragment were bound to 5–100 ng Reb1 protein and DNase I footprinting analysis was performed as previously described (29). Following extraction with phenol and precipitation, 4 μl formamide loading buffer was
The sequencing gels were 6% acrylamide, 0.3% bisacrylamide, proteins (1) and mouse and human DNA fragment for fine mapping the protected region of DNA. The sequencing gels were 6% acrylamide, 0.3% bisacrylamide and were boiled for 2 min. Maxam–Gilbert procedures or by eledroporation using a BioRad GenePulser (31). Bacterial cells were transformed by either standard CaCl2/PEG procedures or by eledroporation using a BioRad GenePulser (31).

DNA sequencing and homology analyses

Sequence analysis of S.pombe Reb1p coding sequences was performed using Sequenase Version 2.0 as well as other reagents from USB/Amersham. The two primers designed for sequencing the borders of the two introns in the S.pombe Reb1 genomic DNA were RebS1 (5'-GAAGATGAAAGAATTAAG) and RebS2 (5'-CAGCAATGACGAAAGGC); the template was plasmid Sp Reb1ΔN147. The template for sequence analysis of the 5'-untranslated region and N-terminal half of Sp Reb1 was Sp Reb1ΔC291. Programs utilized to align the predicted S.pombe Reb1p sequence with that of S.cerevisiae and K.lactis Reb1 proteins (1) and mouse and human TTF-1 proteins included the Genetics Computer Group programs Fasta and Bestfit, on the NYU Medical Center Research Computing Resources, and the Intelligenesics GeneWorks™ protein alignment program. The first nucleotide of the S.pombe Reb1 cDNA was determined to be 45 nt upstream of the initiating AUG; the first intron was determined to be 41 nt in length and the second intron 51 nt in length.

Transcription reactions

The templates p3′Δ3+1, pΔXH1:3′Δ+31 and pΔXH2:3′Δ+31 were assayed for the ability to support RNA polymerase I catalyzed transcription as described (30). The 5′-end-labeled, single-stranded probe was labeled at a XbaI site at +340, present in the multiple cloning site of the vector sequences in p3′Δ+1 and treated with HindIII, separating the strands and isolating the template strand labeled on its 3′-end at –3262 and extending to –2555.

Escherichia coli transformations

Bacterial cells were transformed by either standard CaCl2/PEG procedures or by electroporation using a BioRad GenePulsers (31).

RESULTS

One of the roles of the sequence-specific DNA binding factor REB1/mTTF-1 lies in termination of RNA polymerase I transcription (5,16). The termination sites lie at the end of the 37–45S RNA coding sequences and near the beginning of the rDNA promoter in the yeast S.cerevisiae and in the higher eukaryotes Xenopus laevis, mouse and human (2,4,5,8,9,15). To enable molecular analysis of this factor and its recognition sites in the borders of the two introns in the S.pombe rDNA intergenic spacer, the Reb1 cDNA and genomic sequences were isolated.

The initial strategy for isolation of encoding sequences was amplification of S.pombe cDNA sequences using degenerate primers. The primers were designed to amplify sequences homologous between S.cerevisiae and K.lactis REB1 coding sequences and among the S.cerevisiae and K.lactis REB1 and mouse TTF-1 factors. Only one pair of primers, Reb1F5 (encoding QCRYKW) and pACTR, produced a specific band containing coding sequences homologous to the Reb1 gene (see Materials and Methods). Analysis revealed that the putative Reb1 genomic DNA sequences had been sequenced as part of the S.pombe sequencing project and listed in the GenBank/EMBL database (accession no. U33010; 24). Figure 1A displays the predicted S.pombe Reb1p, which is composed of 504 amino acids.

The 5′ start of the cloned Reb1 cDNA was 45 nt upstream of the initial ATG and the genomic Reb1 sequences were found to be interrupted by two introns, 41 and 51 nt in length (see Fig. 1A).
Expression and purification of the DNA binding domain of *S. pombe* Reb1p

A subclone coding for the C-terminal 358 amino acids of *S. pombe* Reb1p (from residues 147 to 504) was constructed, designed to contain the putative rDNA binding domain and an N-terminal polyhistidine tag. This (His)_6Reb1p was expressed at high levels in bacteria and was affinity purified from the insoluble fraction obtained following bacterial lysis. Figure 2 shows the polypeptide profile of bacterial proteins extracted from a transformant expressing *S. pombe* (His)_6Reb1p (lane 1) and of fractions containing affinity purified (His)_6Reb1p (lanes 2 and 3). The apparent molecular weight is ~38 kDa, which is smaller than the predicted size of 43.9 kDa for unknown reasons.

Two electrophoretically retarded complexes formed upon *S. pombe* Reb1p binding

The ability of the recombinant (His)_6Reb1p to form a specific complex with *S. pombe* rDNA intergenic sequences downstream of the 3’-end of the ~37S rRNA coding sequences was assessed in an electrophoresis mobility shift assay (28). This region of the intergenic rDNA spacer has been shown to contain putative in vivo termination sites (6). We hypothesized that the binding site of *S. pombe* Reb1p would lie ~12–20 nt downstream of true intergenic DNA. The faster migrating Complex I, is proposed to contain one Reb1p bound to the rDNA fragment, while the slower migrating complex, Complex II, is proposed to contain two molecules of Reb1p.
bound per probe molecule. Figure 3B illustrates formation of a single rDNA-Reb1p complex at lower concentrations of Reb1p (lane 2), while increasing concentrations of Reb1p result in formation of the slower migrating Complex II (lanes 3 and 4). This retarded complex results from specific association of \textit{S.pombe} Reb1p with binding sites within the \textit{S.pombe} rDNA intergenic region from −3240 to −2828, since the complex is no longer apparent when the binding reactions are performed in the presence of a fragment containing both (see Fig. 4A) or either of the two sites (see Fig. 4B). In addition, we have estimated the equilibrium dissociation constant for the binding reaction of Reb1p to this rDNA fragment containing both of the recognition sites in their natural positions. This $K_d$ is estimated to be $8.7 \times 10^{-11}$ M (data not shown).

**Reb1p protects sites starting at −3165 and −2992 from nuclease digestion**

To localize the internal position(s) of \textit{S.pombe} Reb1p binding, a DNase I footprinting assay was performed using the same rDNA fragment containing the Reb1 binding sites used in Figure 3. The two recognition sites that result in protection of the rDNA from DNase I cleavage in the presence of bound Reb1p are located at Site 1, between positions −3165 and −3153, and at Site 2, in the rDNA intergenic region containing the homologous sequence and positioned between −2992 and −2976 (see Fig. 5). It should be noted that nucleotides upstream of −3156 and downstream of −3153 are not cleaved by DNase I and thus the exact boundary of protection may extend beyond these borders. Binding of Reb1p also results in a striking hypersensitive site at position −3145 and at a site downstream of the −2992/−2976 region as well.

The rDNA intergenic region containing the Reb1p binding sites blocks non-specific RNA polymerase I transcription

The rDNA intergenic region containing the Reb1p binding sites was tested for its ability to block read-through transcription in an \textit{in vitro} transcription assay. Transcription reactions contained three different templates: template pAΔ31, bearing a full rDNA promoter from −243 to +31; pΔXH1:+31, containing intergenic rDNA sequences from −3959 to −3241 upstream of the promoter; pΔXH2:+31, containing the rDNA intergenic region with the Reb1p binding sites and extending from −3240 to −2555 upstream of the rDNA promoter (30). Transcripts were detected utilizing a specific probe made by end-labeling at position +430 on the template strand of pAΔ31 (see Fig. 6B for a diagram of the templates). An S1-protected fragment of −585 nt representing read-through transcripts (R) is seen in lane 2 (Fig. 6A). Since both templates pΔXH1:3ΔA+31 and pΔXH2:3ΔA+31 differ from the probe at position −243, any read-through transcripts they support are detected. However, sequences upstream of the promoter in pΔXH2:3ΔA+31 serve to abolish read-through transcripts, as seen in lane 3. This same region has been shown to confer 3'-end formation in \textit{in vivo} (6), with the mapped 3'-ends formed at 13–14 nt upstream of the Reb1p binding sites located in this rDNA intergenic region.

We next determined whether the template pΔXH2:3ΔA+31 supported 3'-end formation that corresponded to those detected in \textit{in vivo} analyses (6) and whether addition of recombinant \textit{S.pombe} Reb1p was required for this formation. As seen in Figure 7, the 3'-ends of RNAs from the template pΔXH2:3ΔA+31 correspond to 3'-end formation upstream of the Sites 1 and 2 protected by recombinant Reb1p (Fig. 5). The size of the protected fragments, −82 and 256 nt, corresponded to 3'-end formation as seen in \textit{in vivo} (6), −12–15 nt upstream of the Reb1p binding sites, Sites 1 and 2. In addition, formation of these 3'-ends was increased by addition of Reb1p (compare lanes 4 versus 3 and 2 versus 1, Fig. 7). The 3'-ends present in the absence of recombinant Reb1p (lanes 1 and 3) are presumably formed due to endogenous Reb1p present in the S-100 extract, and addition of competitor DNA bearing Reb1p binding sites in \textit{in trans} abolishes their formation (data not shown). These data support the role of \textit{S.pombe} Reb1p in termination of RNA polymerase I catalyzed transcription, analogous to the role of homologous \textit{S.cerevisiae} Reb1p and mouse TTF-1. This 3'-end formation was not apparent when the Reb1p binding sites were in the reverse orientation (data not shown).

**DISCUSSION**

The \textit{S.pombe} sequence-specific rDNA binding factor Reb1p has two high affinity binding sites downstream of the end of the −37S
rRNA coding sequences. These are located 13 and 14 bp upstream of non-specific rRNAs (lane 3, pAXH2:3'4+31). (B) Diagram of the templates used for transcription analysis. Each template contains a full rDNA promoter extending from —243 to +31. Upstream rDNA intergenic sequences are noted for pΔXH1:3'4+31 and pAXH2:3'4+31 (30).

While the mouse intergenic rDNA has eight termination sites downstream of two of the three proposed in vivo termination sites (‘+267’ and ‘+448’, previously called Sites 1 and 3 (6); no Reblp binding was detected near the ‘+338’ site; 6). In addition, transcription of a template bearing this rDNA intergenic region upstream of the gene promoter displays a block in read-through transcription, presumably due to the presence of these Reblp binding sites, and 3'-end formation due to added Reblp occurs at positions seen in vivo (6) that are 12–15 nt upstream of the Reblp binding sites. A recombinant protein bearing the C-terminal two thirds of the S.pombe Reblp protein is active in sequence-specific rDNA binding, similar to the activity of the C-terminal halves of mTTF-I and S.cerevisiae Reblp (19,32). Reblp protected a core region in the duplicated sites, 5'-AGGTAAGGTAATGCAC, and produced a striking nuclease hypersensitive site 3' of this site. However, this minimal site was insufficient for high affinity binding (data not shown).

Figure 6. The rDNA intergenic spacer containing the S.pombe Reblp binding sites blocks read-through transcription. (A) Correctly initiated transcripts supported by S.pombe rDNA mini-genes containing the full rDNA promoter (lane 1, p3'4+31) result in a 340 bp S1 protected fragment (30). When the template contains this same promoter but with upstream sequences extending from —3959 to —3241, a divergent S1 fragment of —590 bp (R) is apparent, since transcripts reading into the promoter from upstream sequences diverge in sequence from the specific end-labeled probe —250 nt upstream of the initiation site (lane 2, pΔXH1:3'4+31). A template containing Reblp binding sites in the rDNA segment from —3240 to —2555 upstream of the promoter no longer displays this read-through S1-resistant band (R), presumably due to termination of non-specific rRNAs (lane 3, pAXH2:3'4+31). (B) Diagram of the templates used for transcription analysis. Each template contains a full rDNA promoter extending from —243 to +31. Upstream rDNA intergenic sequences are noted for pΔXH1:3'4+31 and pAXH2:3'4+31 (30).

Figure 7. Reblp causes 3'-end formation upstream of Sites 1 and 2. In vitro transcription reactions supported by the template pΔXH2:3'4+31, bearing the Reblp binding sites upstream of the rDNA promoter, were conducted as in Figure 6, either in the absence (—) or presence (+) of added recombinant S.pombe Reblp (25 ng). The 3'-ends of transcripts were assessed by S1 analysis using a single-stranded probe labeled at position —3262 on its 3'-end and extending to the HindIII site at —2555 of the template strand. The S1 protected fragment of —82 nt represents 3'-end formation upstream of Site 1, analogous to the putative in vivo termination Site 1 (6). The S1-protected fragment of —256 nt represents transcripts with 3'-ends upstream of the 3' Reblp binding site. A longer exposure of Site 1, from lanes 3 and 4, is shown to the right.

Although the Reblp binding site was initially localized within the S.cerevisiae rDNA enhancer, this site is not essential or necessary for enhancement of rDNA transcription (13,33–36). However, lesions in both of the Reblp binding sites in the S.cerevisiae intergenic spacer, one in the rDNA enhancer and one in the upstream promoter, result in a drastic reduction in transcriptional initiation in vivo (33), supporting a role for Reblp site occupancy both in the 5'- and 3'-region of the intergenic rDNA spacer in activation of an adjacent rDNA transcription unit.

The highest levels of homology are found in the DNA binding domains of the S.cerevisiae and K.lactis REB1 proteins (19), mTTF-Ip (32), hTTF-Ip (37) and c-myb (Fig. 1A; see Fig. 1B for regions of homology among the S.pombe, S.cerevisiae and K.lactis Reb1 proteins and Fig. 1C for between S.pombe Reb1p and mTTF-I). Schizosaccharomyces pombe Reb1p lacks a conserved tryptophan residue, Trp688 (m1-1144), that is required for mTTF-Ip sequence-specific DNA binding activity and that is conserved among mTTF-I, S.cerevisiae Reb1p and K.lactis REB1 (32). However, S.pombe Reb1p has a conservative substitution of a hydrophobic residue, Phe, in its place (Fig. 1C).

Saccharomyces cerevisiae Reb1p associates with recognition sites in regulatory regions of multiple genes transcribed by RNA polymerase II and this interaction affects expression of these genes (38–42). Thus, S.pombe Reb1p and mTTF-I might also...
recognize additional sequence-specific binding sites in promoters or upstream activating sites of RNA polymerase II transcribed genes as well. A search of the S. pombe database uncovered several potential Reb1p binding sites, although none are in sequenced RNA polymerase II promoter regions. One, however, is in the coding region of the S. pombe L3 ribosomal protein gene (13/14 nt are conserved in the L3 sequence AGGTAAAGGTTAACG, starting 688 nt after +1; 43; accession no. U00798). It will be of interest to determine whether this represents a Reb1p binding site and whether it has an effect on expression of the L3 ribosomal protein gene.

It has been hypothesized that Reb1p may have a role in maintenance of genomic structure, affecting expression of the rDNA locus (33), in formation of nucleosome-free domains (40,44) and in expression of diverse protein coding genes (38–42). The functions of the S. pombe Reb1 protein in addition to involvement in association with the intergenic rDNA spacer sequences and in termination of RNA polymerase I catalyzed transcription remain to be investigated. Characterization of the S. pombe Reb1 coding sequences and demonstration that it has two sequence-specific binding sites in the rDNA intergenic spacer of S. pombe facilitate future investigation into the functional activity of this protein that is conserved from yeasts to man.

ACKNOWLEDGEMENTS

This work was supported by NSF grant MCB-9219220 to L.K.P. and in part by NYU Research Challenge Fund grants. The NSF is also thanked for its support of the NYU Medical Center Research Computing Resources through grant BIR-9318128. We thank Drs Jon Warner and Walter Lang for informative discussions and Dr Warner for suggestions on the manuscript and we thank Mr Seung Lee for his help in this project.

REFERENCES