Initiation of transcription by RNA polymerase II (RNAP II) on Saccharomyces cerevisiae messenger RNA (mRNA) genes typically occurs at multiple sites 40–120 bp downstream of the TATA box. The mechanism that accommodates this extended and variable promoter architecture is unknown, but one model suggests that RNAP II forms an open promoter complex near the TATA box and then scans the template DNA strand for start sites. Unlike most protein-coding genes, small nuclear RNA gene transcription starts predominantly at a single position. We identify a highly efficient initiator element as the primary start site determinant for the yeast U4 small nuclear RNA gene, SNR14. Consistent with the scanning model, transcription of an SNR14 allele with tandemly duplicated start sites initiates primarily from the upstream site, yet the downstream site is recognized with equivalent efficiency by the diminished population of RNAP II molecules that encounter it. A quantitative in vivo assay revealed that SNR14 initiator efficiency is nearly perfect (~90%), which explains the precision of U4 RNA 5' end formation. Initiator efficiency was reduced by cis-acting mutations at −8, −7, −1, and +1 and trans-acting substitutions in the TFIIIB B-finger. These results expand our understanding of RNAP II initiation preferences and provide new support for the scanning model.

Eukaryotes rely on RNA polymerase II (RNAP II) to synthesize all messenger RNAs (mRNAs) and most of the small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) encoded within their nuclear genomes. Efficient and accurate transcription initiation is vital to ensure the proper expression and function of these RNAs. The recruitment of RNAP II to gene promoters is mediated through the assembly of a pre-initiation complex (PIC). RNAP II accessory proteins provide promoter specificity and the structural core for assembly of the PIC. These accessory proteins include the general transcription factors TFIIH, TFIIID, TFIIIB, TFIIIE, and TFIIJ (1–3). Some transcription factors engage in sequence-specific contacts with core promoter elements (4); one of the most fundamental interactions for PIC assembly is between the TATA-binding protein subunit of TFIIID and the TATA box (5–6). In a stepwise model for PIC assembly, TATA-binding protein binding is followed by the addition of TFIIIB, RNAP II-TFIIID, TFIIE, and TFIIH (7).

In metazoans, the assembly of a PIC at the TATA box results in start site selection 25–30 bp downstream (4). The architecture of the PIC is such that the transcription start site is placed precisely within the active center of RNAP II (8–9). In the yeast Saccharomyces cerevisiae, RNAP II initiation typically occurs at multiple sites at variable distances from the TATA box, with most start sites ranging from 40 to 120 bp downstream of the TATA box (10). The initiation mechanism that accommodates this extended and variable promoter architecture is unknown, but it does not appear to be dependent on assembling the yeast PIC in a manner different from that of metazoans. Yeast promoter melting has been shown to begin at the same position as in metazoans, ~20 bp downstream of the TATA box (11). In addition, the ~30-bp distance between the TATA box and RNAP II active center has been confirmed through structural analysis of yeast PICs (8–9). A scanning model for start site selection has been proposed for yeast (11). In this scanning model a PIC assembles at a TATA box, the DNA is melted, and RNAP II translocates downstream searching the template strand for acceptable start sites.

The initial sequence comparisons and mutational analysis of a relatively small set of yeast mRNA genes helped define three related yeast start site consensus sequences, RRYR, TCRA, and Y(A/T)R in the non-template strand, where the initiation site is underlined. Y is pyrimidine, and R is purine (12–14). Recently, an alignment of sequences flanking 4637 yeast transcription start sites has provided a more refined consensus sequence: A(A/R)(A/T)N(N/A)T(N/A)NN(A/R)(A/T)N (15). The DNA sequences encompassing yeast transcription start sites are sometimes termed initiator elements. In metazoans the initiator is defined as a core promoter element distinct from the TATA box that nucleates PIC assembly and is sufficient for accurate transcription (16). Although there is evidence to suggest that some yeast initiators may function in this way (17–18), most appear to play a more limited role in transcriptional control and influence accuracy but not overall efficiency (2).

It seems likely that the recognition and efficient utilization of yeast start sites involves a sequence-specific interaction between the yeast initiator element and either RNAP II, an accessory protein, or both. RNAP II and TFIIIB have been shown to dictate the distance from TATA boxes to start sites in yeast (19). TFIIIB substitutions that confer downstream shifts in yeast start site selection map to the “B-finger” domain, which encompasses residues 55–88 of the N-terminal region (9, 20–22). The promoter sequence immediately upstream of yeast start sites can influence the severity with which TFIIIB B-finger substitutions alter start site selection (23). A yeast RNAP II-TFIIIB crystal structure model shows the TFIIB B-finger inserted through the RNA exit pore into the polymerase active site, suggesting that start site selection may be mediated by a direct interaction between the B-finger and promoter DNA (9).

Here we report the characterization of cis- and trans-acting determinants of start site selection at the yeast U4 snRNA gene, SNR14. In contrast to the heterogeneous transcription start site selection exhibited...
at most mRNA genes, yeast snRNAs typically have one major start site, thus providing a model system for the study of accurate initiation. We identified a highly efficient initiator element within the SNR14 promoter, defined the positions most critical for its function in start site selection, and quantified its efficiency relative to other initiator sequences. Substitutions within the TFIIB B-finger genetically interact with SNR14 initiator mutations in a sequence-dependent manner. Quantitative analysis of the utilization of tandemly duplicated initiator elements strongly supports the proposed scanning model for yeast transcription start site selection and demonstrates that scanning by RNAP II is processive.

EXPERIMENTAL PROCEDURES

Plasmid Construction—SNR14 (positions −224 to +701, relative to +1 transcription start site) was cloned by PCR amplification of a genomic DNA template isolated from yeast strain PJ43–2b and ligated into the BamHI site of pRS313 (CEN4, ARS1, HIS3). 5′′ end truncation constructs were generated by the same method but using pRS313-SNR4 as the template. pRS313-SNR4-StDup was created by using QuikChange PCR mutagenesis (Stratagene) to insert 14 bp of DNA (−13 to +1 relative to SNR14 start site) between positions +1 and +2 of pRS313-SNR4, creating an overlapping 20-bp duplication. pRS317 (LYS2)-SUA7 contains the entire S1A7 promoter and coding region (TFIIb gene) and was constructed by ligation of the CiaI/SacI fragment of pRS314 (TRP1)-yIBN (kindly provided by A. Ponticelli, State University of New York at Buffalo) into pRS317 (LYS2). TFIIb expressed from these constructs contains an N-terminal histidine tag. All mutations within pRS313-SNR4-StDup and pRS317-SUA7 were created using the QuikChange method (Stratagene). pRS316 (LIRA3)-SNR4, SUA7 was constructed by ligating a PCR-amplified region of SNR4 (−224 to +701) into the Sall/Xhol sites of pRS316 (LIRA3)-yIBN (kindly provided by A. Ponticelli, State University of New York at Buffalo). Oligonucleotide sequences are available upon request.

Yeast Strains—Genomic DNA isolated from PJ43–2b (MATa trp1 ura3 can1 leu2 his3 ade2 met2 lys2; kindly provided by P. James and E. Craig, University of Wisconsin at Madison) was used for the cloning of full-length SNR14. YKS2 (MATa trp1 his3 ura3 ade2 lys2 snr14::TRP1 YCp50 (LIRA3)-SNR4; kindly provided by K. Shannon and C. Guthrie, University of California at San Francisco) has been described previously (24). JNK1 (MATa trp1 his3 ura3 ade2 lys2 snr14::TRP1 sua7::KANMX4) pRS316 (LIRA3)-SNR4, SUA7 was derived from YKS2 by the following manipulations. YKS2 was transformed with pRS313-SNR4 and pRS317-SUA7 by the lithium acetate procedure (25). Its “Lys” clones were transformed with 5 μg of SUA7::KANMX4 and plated on yeast extract-peptone-dextrose plus 200 mg/μl G418 (Invitrogen). The SUA7::KANMX4 product was obtained by PCR using genomic DNA template from a diploid strain heterozygous for a KANMX4-disrupted allele of SUA7 (Invitrogen). Integrants were confirmed by PCR using genomic DNA isolated from KanR colonies. Integrand strains were transformed with pRS316 (LIRA3)-SNR4, SUA7 and grown on medium lacking uracil. Loss of pRS313-SNR4 and pRS317-SUA7 was confirmed by replica plating onto −his and −lys medium, thus creating JNK1. Plasmid-borne alleles of SNR14 and SUA7 in pRS313 and pRS317, respectively, were transformed into JNK1 by standard plasmid shuffle techniques (26). Loss of pRS316 (LIRA3)-SNR4, SUA7 was selected for on medium containing 0.75 mg/ml of 5-fluoro-orotic acid.

DNase I Chromatin Footprinting—Chromatin footprinting was performed as previously described (27) using the yeast strain PJ43–2b. After digestion of lysed yeast cells or purified genomic DNA with DNase I (Invitrogen), cleavage sites on the non-template strand of the SNR14 promoter were mapped by primer extension using 32P-labeled oligo U4–14C, which is complementary to non-template strand residues +32 to +51. Sequencing ladders were generated by primer extension of genomic DNA using 32P-labeled oligo U4–14C and a dNTP mix containing deoxy-ATP or -GTP.

RNA Analysis—Total cellular RNA was isolated using the guanidinium thiocyanate method including a 65 °C phenol extraction (28). Reverse transcription for the determination of the Sec3 mRNA-processing sites was performed in a 50-μl reaction volume containing 5 μg of total RNA from strain PJ43–2b, 50 μM Tris-HCl (pH 8.3), 8 mM MgCl2, 50 mM NaCl, 11 mM dithiothreitol, 1 mM dNTPs, 40 units of RNasin (Promega), 250 pmol of T16-EcoR1 oligo, and 37.5 units of avian myeloblastosis virus reverse transcriptase (United States Biochemical). cDNA synthesis proceeded at 42 °C for 1 h. Ten μl of the 50-μl RT reaction was used as a template for PCR in a 100-μl volume containing 20 mM (NH4)2SO4, 50 mM Tris-HCl (pH 9.0), 0.75 mM MgCl2, 50 pmol of T16-EcoR1 oligo, 100 pmol of SEC3-RT-PCR oligo, and 1 unit of MasterAmp™ T1 DNA polymerase (Epipcentre). Each PCR cycle consisted of a denaturation at 94 °C for 30 s, annealing at 42 °C for 30 s, and elongation at 72 °C for 1 min. A total of 30 cycles was performed with an additional extension at 72 °C for 5 min. RT-PCR products were gel-purified and ligated into the BamHI/EcoRI site of pRS316. Recovered plasmids were sequenced using an M13F oligo.

RNA immunoprecipitations were done as described previously (29). Each immunoprecipitation mixture contained 20 μl of swollen protein A-Sepharose CL-4B beads (Amersham Biosciences) that had previously been incubated with 5 μg of 2,2,7-trimethylguanosine Ab-1 antibody (Calbiochem), 50 units of RNasin (Promega), and 5 μg of total cellular RNA in a 200-μl volume. The precedent for Ab-1 recognition of both 7-methylguanosine and 2,2,7-trimethylguanosine caps came from Bochnig et al. (30).

 Primer extension analysis of 5 μg of total cellular RNA was carried out using 32P-labeled oligonucleotide U4–14B (complementary to nucleotides 140–159 of yeast U4 RNA) or SCR1 (complementary to nucleotides 75–92 of yeast sc1 RNA) (31). Sequencing ladders were generated using the Sequitherm EXCEL II DNA sequencing kit (Epipcentre). The cDNA products were electrohoresed on 6% polyacrylamide, 8.3 M urea gels. Gels were visualized with a Storm PhosphorImager (Amersham Biosciences), and data were quantitated with Amersham Biosciences ImageQuant software (Version 5.2).

RESULTS

Conserved Sequence Elements Upstream of the Yeast U4 snRNA Gene, SNR14—To begin characterizing SNR14 promoter architecture, we used comparative sequence analysis to identify conserved elements upstream of the transcription start site and downstream of the 5′-adja-
cent gene, SEC3 (Fig. 1A). An alignment of sequences upstream of SNR14 in four different species of Saccharomyces (32) helped identify several conserved elements (Fig. 1B). The most strikingly conserved regions include the sequence immediately upstream of the transcription start site, a TATA box located 100 base pairs upstream of the start site, a T-stretch just upstream of the TATA box, and a region located 31–44 base pairs upstream of the TATA box. The most upstream conserved region may be an upstream activating sequence and in S. cerevisiae exactly matches the consensus binding site of the transcriptional activator Abf1 (33–34).

In addition to promoting SNR14 transcription, another likely function for conserved sequences in this intergenic region is to direct cleavage and polyadenylation of Sec3 mRNA. RT-PCR was used to identify

14120 JOURNAL OF BIOLOGICAL CHEMISTRY

VOLUME 281 • NUMBER 20 • MAY 19, 2006
the predominant Sec3 mRNA 3′-processing sites. The most efficiently recovered site, which appeared in 12 of 21 clones, mapped to the middle of the putative Ab1 binding site (Fig. 1B). Six other nearby sites were represented by 1–2 clones each. This result implies that the sites of SEC3 transcription termination and SNR14 PIC assembly overlap.

DNase I chromatin footprinting was used to complement comparative sequence analysis in the search for potential SNR14 promoter elements. This procedure probed \textit{in vivo} assembled chromatin by lysing yeast cells directly into a solution of DNase I. For comparison, purified genomic DNA was digested with DNase I. Cleavage sites were detected by primer extension (27). No obvious DNase I footprint was detected by primer extension (27). No obvious DNase I footprint was observed between the SNR14 TATA box and start site (Fig. 1C) despite the fact that the gene is single-copy and highly transcribed. This finding suggests that there is no high occupancy protein binding site in the region of the promoter that separates the location of PIC recruitment from that of transcription initiation. Rather, the subtle changes in DNase I protection and enhancement suggest partial occupancy. In addition, some subtle changes in DNase I sensitivity were observed at the putative upstream activating sequence, TATA box, and initiator region, consistent with partial occupancy. We could not make any conclusions regarding the protein occupancy of the T-stretch given that it was not efficiently cleaved by DNase I. Because a scanning RNAP II complex is unlikely to provide sufficient promoter occupancy for detectable DNase I protection, the footprinting results obtained for the SNR14 promoter are consistent with this model for transcription initiation.

\textbf{The Conserved SNR14 TATA Box Is Not a Determinant of Start Site Position \textit{In Vivo}}—Functional upstream SNR14 promoter elements were roughly mapped using 5′-truncation analysis of a plasmid-borne allele (Fig. 2A). Inserts were tested in both orientations within the vector to control for effects of plasmid sequences. Primer extension of U4 RNA synthesized from its chromosomal locus showed the single major tran-
The effects of promoter truncations varied somewhat depending on the orientation of \( \text{SNR14} \) alleles within the plasmid, with the reverse orientation yielding more severe effects on growth and U4 RNA synthesis. Because the differences appeared to be limited to transcription efficiency yielding more severe effects on growth and U4 RNA synthesis.

RNAP III and was not immunoprecipitated with an anti-cap antibody, showing that immunoprecipitation is specific for capped transcripts. The scR1 RNA served as a negative control since it was synthesized by RNAP III as an uncapped transcript. The positions of alternate start sites within upstream plasmid sequence and the U4 coding region were mapped to the nucleotide (data not shown) and are indicated on the right. No IP, no immunoprecipitation. A sequence context of alternative \( \text{SNR14} \) start sites, with non-template strand DNA aligned from \(-10\) to \(+10\) relative to the site of initiation.

Thus, transcription initiation of \( \text{SNR14} \) from the normal \(+1\) start site is remarkably resistant to deletion of conserved upstream promoter elements. Much like what has been observed previously at mRNA genes, the \( \text{SNR14} \) TATA box primarily affects the frequency of RNAP II transcription and not the position of initiation (10). Alternative \( \text{SNR14} \) initiation sites minimally include a purine preceded by a pyrimidine and most often by two cytosines. These results suggest that start site selection at \( \text{SNR14} \) is directed by an initiator-like element, potentially in combination with a downstream (intragenic) element.

**A Polar Effect on \( \text{SNR14} \) Start Site Selection Supports a Unidirectional Scanning Model**—A scanning model for yeast transcription start site selection posits that after melting DNA near the TATA box, RNAP II scans the upstream sequence of the TSS. Initiation sites minimally include a purine preceded by a pyrimidine and most often by two cytosines. These results suggest that start site selection at \( \text{SNR14} \) is directed by an initiator-like element, potentially in combination with a downstream (intragenic) element.
is driven by a random collision of RNAP II or another initiator-binding protein with the DNA, we should detect roughly equal utilization of the two sites. If the start site is defined by its distance from factors bound upstream or by a polymerase scanning from the TATA box, we should detect primarily upstream starts. In contrast, if the start site is measured by dividing its signal intensity by the total signal intensity of products from all detectable start sites (~8u, +1u, −8d, +1d, +7d). For simplicity, the efficiency of the +7d start site was assumed to equal 100%. RNAP II flux is defined as the relative number of polymerases encountering a given start site and was arbitrarily assigned an initial value of 100 units (U); C, calculating start site efficiency for SNR14-StDup A+1uT (panel A, lane 2).

In fact, the upstream start site (+1u) is preferentially utilized (Fig. 4A, lane 2), consistent with upstream recruitment and a directional scanning model. When the initiating non-template nucleotide of the upstream start is changed from A to T, the downstream start (+1d) is more heavily utilized, confirming that the start sites are in competition and that precise spacing from an upstream recruitment site is not required (Fig. 4A, lane 3). Changing the downstream start from A to T essentially abolishes its usage entirely (Fig. 4A, lane 4). When both start sites are changed from A to T, weak alternative sites are used at −8d and +7d (Fig. 4A, lane 5). Weak initiation at −8 and +7 is also observed in the wild-type allele. Interestingly, initiation at −8u, the most upstream observed start site in the SNR14-StDup allele, does not increase when the +1u and +1d start sites are mutated. This finding is also consistent with the directional scanning model, in which read-through of a site should not affect initiation at sites upstream.

Yeast Start Site Efficiency Is an Intrinsic, Quantifiable Property of an Initiator Sequence—To obtain a quantitative estimate of initiation sequence preference, we assumed that a homogeneous population of initiation-competent polymerases scan unidirectionally through the −8u to +7d interval of the SNR14-StDup allele in search of a good match to the ideal initiator consensus. Start site efficiencies were calculated by dividing the relative yield of a start site product by RNAP II flux at that site. The relative yield of a transcript from a given start site was determined by dividing its signal intensity by the total signal intensity of products from all detectable start sites (~8u, +1u, −8d, +1d, +7d). RNAP II flux was defined as the relative number of polymerases encountering a given start site and was arbitrarily assigned a value of 100 units at the −8u site. Because 2% of the U4 cDNA ends at the −8u position, 2 units of RNAP II must have initiated at this site, and 98 units at the +7d site. When RNAP II flux was determined without considering flux, the +1u and +1d start site efficiencies differ by about 10-fold (87% versus 8%). When RNAP II
flux at the two positions is included in the calculation, the +1u and +1d start site efficiencies are found to equal at 89% (Fig. 4B). Even when the efficiency of the +1u site is reduced more than 10-fold by the A+1uT mutation, the efficiency of +1d and other downstream sites remains about the same when flux is considered (Fig. 4). These results indicate that initiator efficiency is an intrinsic property that is largely independent of start site position. We can, therefore, use this value to classify initiator strength. For example, we can deduce that the +1 initiator sequence of the U4 gene is a nearly perfect initiator, with an efficiency of 89–92%.

Mutations in SNR14-StDup Define Preferred Sequences at −1 and −8 of the Initiator—Having developed a quantitative assay for in vivo start site selection, we next tested the sequence requirements for RNAP II initiation through site-directed mutagenesis of phylogenetically conserved base pairs in the SNR14 major initiator. In addition to the +1 position, the nucleotide identities of positions −8, −7, −5, −4, −2, and −1 of SNR14 are conserved across the Saccharomyces genus (Fig. 1). A double transversion mutation at positions −8u, −7u (A→T) or −2u, −1u (C→G) of the SNR14-StDup allele significantly shifted initiation toward downstream start sites, reducing efficiency of the +1u site by about 5-fold (Fig. 5A, lanes 2 and 5). Upon separation of the −8u/−7u double mutation into single point mutations, it became clear that the −8u mutation contributes more to the initiation defect than the −7u mutation (Fig. 5A, lanes 3 and 4). A similar dissection of the −2u/−1u double mutation showed that the change at position −1u accounted for all of the downstream shift (Fig. 5A, lanes 6 and 7). Transversions at −5u or −4u on their own had little if any effect on start site selection (Fig. 5B, lanes 8 and 9).

The base preference at positions −8u and −1u was explicitly tested by creating all possible base substitutions. The A−8uT mutation reduced usage of the +1u start site by about 2-fold, an effect that was slightly greater than that observed for A−8uG and A−1uC (Fig. 5B, lanes 2–4). The C−1uG mutation reduced usage of the +1u start site by at least 5-fold, an effect that was followed closely by C−1uA (Fig. 5B, lanes 5 and 6). The C−1uT mutation had a very minor effect on start site selection (Fig. 5B, lane 7). Overall, it appears that the strength of the SNR14 start site is dependent on a purine at +1, a pyrimidine at −1, and an adenine at −8.

There are other sequences within the RNAP II scanning window between the SNR14 TATA box and +1 start site that resemble start sites but at which initiation does not efficiently occur. We predicted that changing these sites to match the bases preferred at positions −8, −7, and −1 would contribute to more efficient start site usage. Weak initiation occurs at −8u in a wild-type SNR14-StDup allele (Fig. 5C, lane 1). The C−16u/15uA or T−9uC mutations on their own increase usage of the −8u start site toward a higher efficiency further demonstrates the
importance of the \(-8, -7,\) and \(-1\) positions in \textit{SNR14} initiator function.

The efficiencies of 26 different wild-type and mutant yeast initiators calculated from various \textit{SNR14-StDup} alleles are shown in Table 1, with values ranging from about 89 to 4\%. The wild-type \textit{SNR14} \(+1\) site (A\(+1d\), A\(+1u\)) is the most efficient, and divergence from this sequence reduces start site efficiency. Changes at the \(-8, -7, -1,\) and \(-1\) positions reduce start site efficiency anywhere from about 2- to 15-fold. Although the efficiency of the wild-type A\(-8u\) start was improved 10-fold by changing the \(-8, -7, -1,\) and \(-1\) positions toward a more preferred initiator sequence, it is interesting to note that the efficiency of this site is still around 2-fold less than the wild-type A\(+1\) WT start site. The \(-8u\) and \(-8d\) start site efficiencies differ by 6-fold (4\% versus 24\%) even when the flanking sequences are identical from -9 to +7, as is the case in the \textit{StDup-A}+1\textit{u} allele (Table 1). Taken together, these results indicate that there are positions other than \(-8, -7, -1,\) and \(+1\) at which nucleotide identity influences initiator efficiency.

**Substitutions in the TFIIB B-finger Exacerbate the Effect of Initiator Mutations at \(-8\) and \(-1\)—One possible cause of the start site selection defects exhibited by \textit{SNR14} initiator mutants is a disruption of direct amino acid/nucleotide contact(s) made between a protein component of the yeast PIC and the initiator. To analyze the role of TFIIB as the potential trans-acting component of the PIC that interacts with the yeast initiator element, we generated a double knock-out strain that has disrupted chromosomal copies of the \textit{SNR14} and \textit{SUA7} (TFIIB) genes and carries wild-type copies of these genes on a \textit{URA3}-marked plasmid. Standard plasmid shuffle protocols were used to introduce mutant alleles of \textit{SNR14} and \textit{SUA7}. The effect of TFIIB B-finger substitutions previously shown to alter initiation on protein-coding genes (22–23) was tested in the context of the \textit{SNR14-StDup} allele. For the most part,

### Table 1

| Start Site (StDup allele)* | -13 | -8 | +1 | +7 | Efficiency\(^b\) | Range | \(n\) |
|---------------------------|-----|----|----|----|-----------------|-------|-----|
| +1d (WT)                  | TCCCTTAATACCTCCATCCTTA | 89 | 74 – 92 | 18 |
| +1u (WT)                  | TCCCTTAATACCTCCATCCTTA | 87 | 84 – 91 | 8  |
| +1u (A-5uT)               | TCCCTTAATACCTCCATCCTTA | 85 | 1   |
| +1u (C-2uG)               | TCCCTTAATACAGCAGATCCTTA | 84 | 1   |
| +1u (C-4uG)               | TCCCTTAATACAGCAGATCCTTA | 83 | 1   |
| +1u (C-1uT)               | TCCCTTAATACATACCATCCTTA | 79 | 1   |
| +1u (A-7uT)               | TCCCTTAATACATACCATCCTTA | 70 | 1   |
| +1u (A-8uG)               | TCCCTTAATACCTCCATCCTTA | 62 | 1   |
| +1u (A-8uG)               | TCCCTTAATACCTCCATCCTTA | 56 | 1   |
| +1u (A-8uT)               | TCCCTTAATACCTCCATCCTTA | 47 | 46 – 49 | 3  |
| -8d (C-1uG)               | ATACTCGATCCTCAATAC1C | 42 | 40 – 43 | 2  |
| -8u (C-16,-15uA/T-9uC)*   | CATCTTAATACCTCCATCCTTA | 41 | 1   |
| -8d (C-2,-1uG)            | ATACTCGATCCTCAATAC1C | 32 | 28 – 35 | 2  |
| -8d (C-2uG)               | ATACTCGATCCTCAATAC1C | 30 | 1   |
| -8d (WT)                  | ATACTCCATCCTCAATAC1C | 27 | 14 – 31 | 11 |
| -8d (C-1uT)               | ATACTCTATCCTCAATAC1C | 25 | 1   |
| -8d (A+1uT)               | ATACTCCATCCTCAATAC1C | 24 | 18 – 30 | 2  |
| +1u (C-1uA)               | TCCCTTAATACCTCAATCCTTA | 21 | 1   |
| -8d (A-5uT)               | ATACTCAATCCTCAATAC1C | 21 | 1   |
| -8u (C-16,-15uA)*         | ATACTCAATCCTCAATAC1C | 19 | 1   |
| +1u (A-8,-7uT)            | TCCCTTAATACCTCCATCCTTA | 17 | 1   |
| +1u (C-2,-1uG)            | TCCCTTAATACCTCAGATCCTTA | 16 | 12 – 18 | 3  |
| +1u (C-1uG)               | TCCCTTAATACCTCAGATCCTTA | 13 | 2   |
| -8u (T-9uC)*              | TCCCTTAATACCTCAGATCCTTA | 10 | 1   |
| +1u (A+1uT)               | TCCCTTAATACCTCAGATCCTTA | 7  | 1   |
| -8u (WT)*                 | TCCCTTAATACCTCAGATCCTTA | 4  | 1 – 6 | 14 |

\(^a\) Asterisks indicate alleles analyzed in Fig. 5c. Reverse text indicates positions divergent from the A\(+1\) WT start site.

\(^b\) Efficiencies were calculated as described in Fig. 4 and are shown as an average in cases where \(n > 1\).
substitutions in residues 63, 64, 66, and 78 of TFIIB all caused a similarly modest shift in transcription initiation from upstream to downstream sites, reducing +1u start site efficiency by about 1.5-fold (Fig. 6A). The effect of the W63R substitution was less severe than W63P and the other TFIIB substitutions, consistent with what has been observed at the effect of the W63R substitution was less severe than W63P and the other TFIIB substitutions in residues 63, 64, 66, and 78 of TFIIB all caused a similarly modest shift in transcription initiation from upstream to downstream sites, reducing +1u start site efficiency by about 1.5-fold (Fig. 6A). The effect of the W63R substitution was less severe than W63P and the other TFIIB substitutions, consistent with what has been observed at the W63R position.

Other potential functions for Abf1 and its homologues include nucleosome partitioning, nucleosome positioning, and genome partitioning, nucleosome positioning, and genome partitioning. For example, the initiation of yeast snRNA genes, especially snRNAs and snoRNAs, is dependent on the sequence of the TFIIB TFIIH homologue that is adjacent to the transcription start site. Abf1 sites have been shown to be targets for the RNA modification machinery.

DISCUSSION

The synthesis of non-coding RNAs, especially snRNAs and snoRNAs, puts strong demands on the accuracy and efficiency of transcription initiation by RNAP II. The transcription start site of such RNAs usually corresponds to a unique m5'-end, and its precise placement may be required for optimal RNA function. Yeast snRNAs and snoRNAs typically have steady-state levels of hundreds of copies per cell, so their genes must be actively transcribed. Non-coding RNA gene promoters are, therefore, interesting subjects for study of the optimal sequences for directing initiation by RNAP II. Here we provide evidence that the S. cerevisiae U4 snRNA gene, SNR14, fulfills these stringent requirements by coupling a consensus TATA box with a nearly perfect initiator element. Furthermore, the DNA between the TATA box and initiator is devoid of initiator-like sequences that might divert RNAP II from the proper start site as it scans downstream from the TATA box. The differential utilization of duplicated initiator elements in artificial variants of the SNR14 promoter strongly supports the scanning model of start site selection by RNAP II in budding yeast and demonstrates that initiator element efficiency is an intrinsic property dependent primarily on the sequence at positions −8, −7, −1, and +1 relative to the start site.

Architecture of the SNR14 Promoter—In terms of their general promoter architecture, yeast snRNA genes bear a strong resemblance to mRNA genes both in the position and function of their core elements. In agreement with their observed roles in yeast mRNA genes, the conserved TATA box and initiator elements of the SNR14 promoter primarily influence RNAP II transcription efficiency and accuracy, respectively. A more distinctive feature of the SNR14 promoter is the presence of a conserved T-stretch and putative Abf1 binding site. Abf1 sites have been shown to be targets for the RNA modification machinery.

In one orientation of a plasmid-borne SNR14 allele, deletion of the putative Abf1 site and T-stretch led to a 2-fold reduction in U4 RNA level, suggesting a potential role for one or both of these elements in transcription efficiency. In the context of the ribosomal protein-coding gene RPS28A, a mutation that destroys Abf1 binding in vitro reduced transcription by 10-fold, whereas substitutions in the T-rich element reduced transcription by 2-fold (39). Other potential functions for Abf1 within the SNR14 promoter (e.g. genome partitioning, nucleosome

Initiator Selection by Yeast RNAP II

![Diagram of TFIIB and SNR14 promoter](image)

**FIGURE 6.** Substitutions within the TFIIB B-finger exacerbate the effect of mutations at −8 and −1. Primer extension of U4 and scR1 RNAs in total cellular RNA isolated from strains containing plasmid-borne alleles of wild-type (A) or mutant (B) SNR14-StDup and wild-type or mutant SUA7 (TFIIH). Start site efficiencies were calculated as in Fig. 4, except that +7d and a site further downstream (indicated with asterisks) were also included in the total signal intensity.
preferences extend beyond sensus YA, the latest evidence suggests that yeast initiator sequence reported for higher organisms like Drosophila, maize and yeast initiator consensus that influence start site selection. Maicas and Friesen (44) identified a region centered at −9 of the TCM1 gene and 95 other mRNA genes that they termed the "locator." The locator was defined as a region where the base composition of the non-template strand sharply switched from a preponderance of thymine residues to predominately adenine residues. Rathjen and Mellor (45) identified a region from −10 to −4 (ACAGATC) of the major PGK1 start site as a "determinator" element. Deletion of the determinator resulted in a loss of initiation from the normal start site and increased use of more downstream sites. Healy and Zitomer (46) were able to show that insertion of CAAG upstream of the CYC7 gene could direct initiation at a site at which it did not normally occur, and it is interesting to note that their insertion also introduced an adenine at the −8 position on the non-template strand. Our genetic evidence supporting a preference for adenine at positions −8 and −7 of the SNR14 initiator expands the older yeast initiator consensus and can account for the earlier observations described above for the TRP4, TCM1, PGK1, and CYC7 genes.

The general importance of the −8, −7, −1, and +1 positions for initiator efficiency across all RNAP II-transcribed yeast genes is supported by a recent bioinformatics study that compared 4637 yeast transcription start sites. Sequence alignment produced the consensus A(\text{rich})_3NYA(T)/N(N(\text{rich})_2, where the underlined A is the initiation site (15). This yeast initiator consensus is more expansive than that provided in our study (42). The potential role for Abf1 in yeasts may be specific to the upstream and downstream. Zhang and Dietrich (15) could not conclude whether the A-richness of the yeast initiator consensus sequence was important for transcription initiation or a consequence of some other aspect of genome structure. Here, we present direct evidence indicating that the adenine at position −8 and to a lesser extent the adenine at −7 contributes to the functionality of the yeast initiator as a start site determinant. The fact that substitution of any base besides adenine at −8 significantly decreases start site efficiency suggests that the functional impairment is not merely related to the melting potential of an A-T base pair. Rather, it suggests that the −8 position is recognized in a sequence-specific manner.

How the yeast initiator sequence determines start site usage is as yet unknown. It seems likely that the initiator is recognized by a protein component of a scanning pre-initiation complex. Mutations that alter yeast transcription start site selection have been identified in numerous protein components of the PIC, including RNAP II (Rpb1, Rpb2, Rpb9), TFIIB, and TFIIF. Of these proteins, Rpb1, Rpb2, TFIIB, and TFIIF (Tfg1, Tfg2) have also been cross-linked to DNA at or near a transcription start site (53).

Faitar et al. (23) determined that mutations in yeast initiator sequences genetically interact with substitutions in the TFIIB B-finger, making certain start sites more or less sensitive to downstream shifts in transcription start site selection. Specifically, they showed that among a set of mutations made from positions −6 to +5 of an ADH1 initiator, changes at −2, +1, and +2 significantly increased or decreased the sensitivity of the +1 transcription start site to TFIIB-V79L. Here, we present evidence supporting a genetic interaction between the −8 position of the SNR14 initiator with the TFIIB B-finger, expanding our view of what constitutes an initiator element and where potential protein-DNA interactions may occur. Furthermore, the fact that the −8d start site is highly sensitive to the TFIIB-R64A substitution in both the StDup-A-8uT and -C-1uG alleles whereas the −5d start site is essentially insensitive in both contexts indicates that sensitivity correlates with sequence and not simply initiator strength.

**Implications for a Scanning Model of RNAP II Start Site Selection**—A scanning model is currently the best supported explanation for how yeast start sites are selected, but experiments directly testing the basic implications of the model have been scarce. We constructed an allele of SNR14 with tandemly duplicated start sites as a means to test the yeast scanning model in both a qualitative and quantitative fashion. We observed that although the upstream start site had a higher relative yield than the downstream start site, a reduced level of RNAP II flux fully accounted for the lower relative yield from the downstream site. RNAP II flux is an inherent property of a unidirectional scanning model for transcription initiation. Thus, the fact that RNAP II flux can be used to resolve the observed differences in relative utilization of two identical start sites is in itself strong support for the model. The fact that our estimations of RNAP II flux so closely agree with shifts in utilization of start sites 14−22 bp apart in response to mutations suggests that RNAP II scanning is reasonably processive. A scanning mechanism of start site selection requires RNAP II to be processive to accommodate the large and variable distances between yeast TATA boxes and initiator elements. It will be interesting to test the limits of processive scanning by RNAP II. The identification of initiators with a range of efficiencies (Table 1) should aid such studies.

The yeast initiator sequence consensus is readily apparent among RNAP II-transcribed snRNA genes (SNR14, SNR19, SNR20, SNR7) and genes. A Weblogo alignment (54) using a pool of 22 yeast snRNA and snoRNA transcription start sites results in an initiator consensus very similar to that reported for mRNA genes, A(\text{rich})_3NYY(A/T)N(N(\text{rich})_2. Given that yeast RNAP II transcribes all mRNA genes and most snRNA and snoRNA genes, it is expected that the cis-acting sequence requirements for their start site selection would be similar. However, the identification of a strong match to the initiator consensus from such a small sample size of snRNA/snoRNA genes suggests a basis for why yeast RNAP II scanning model is currently the best supported explanation for how yeast start sites are selected, but experimental testing the basic implications of the model have been scarce. We constructed an allele of SNR14 with tandemly duplicated start sites as a means to test the yeast scanning model in both a qualitative and quantitative fashion. We observed that although the upstream start site had a higher relative yield than the downstream start site, a reduced level of RNAP II flux fully accounted for the lower relative yield from the downstream site. RNAP II flux is an inherent property of a unidirectional scanning model for transcription initiation. Thus, the fact that RNAP II flux can be used to resolve the observed differences in relative utilization of two identical start sites is in itself strong support for the model.

The yeast initiator sequence consensus is readily apparent among RNAP II-transcribed snRNA genes (SNR14, SNR19, SNR20, SNR7) and snoRNA genes. A Weblogo alignment (54) using a pool of 22 yeast snRNA and snoRNA transcription start sites results in a consensus very similar to the one reported for mRNA genes, A(\text{rich})_3NYY(A/T)N(N(\text{rich})_2. Given that yeast RNAP II transcribes all mRNA genes and most snRNA and snoRNA genes, it is expected that the cis-acting sequence requirements for their start site selection would be similar. However, the identification of a strong match to the initiator consensus from such small sample size of snRNA/snoRNA genes suggests a basis for why yeast RNAP II scanning model is currently the best supported explanation for how yeast start sites are selected, but experimental testing the basic implications of the model have been scarce. We constructed an allele of SNR14 with tandemly duplicated start sites as a means to test the yeast scanning model in both a qualitative and quantitative fashion. We observed that although the upstream start site had a higher relative yield than the downstream start site, a reduced level of RNAP II flux fully accounted for the lower relative yield from the downstream site. RNAP II flux is an inherent property of a unidirectional scanning model for transcription initiation. Thus, the fact that RNAP II flux can be used to resolve the observed differences in relative utilization of two identical start sites is in itself strong support for the model. The fact that our estimations of RNAP II flux so closely agree with shifts in utilization of start sites 14−22 bp apart in response to mutations suggests that RNAP II scanning is reasonably processive. A scanning mechanism of start site selection requires RNAP II to be processive to accommodate the large and variable distances between yeast TATA boxes and initiator elements. It will be interesting to test the limits of processive scanning by RNAP II. The identification of initiators with a range of efficiencies (Table 1) should aid such studies.

The yeast initiator sequence consensus is readily apparent among RNAP II-transcribed snRNA genes (SNR14, SNR19, SNR20, SNR7) and snoRNA genes. A Weblogo alignment (54) using a pool of 22 yeast snRNA and snoRNA transcription start sites results in an initiator consensus very similar to the one reported for mRNA genes, A(\text{rich})_3NYY(A/T)N(N(\text{rich})_2. Given that yeast RNAP II transcribes all mRNA genes and most snRNA and snoRNA genes, it is expected that the cis-acting sequence requirements for their start site selection would be similar. However, the identification of a strong match to the initiator consensus from such a small sample size of snRNA/snoRNA genes suggests a basis for why yeast RNAP II scanning model is currently the best supported explanation for how yeast start sites are selected, but experimental testing the basic implications of the model have been scarce. We constructed an allele of SNR14 with tandemly duplicated start sites as a means to test the yeast scanning model in both a qualitative and quantitative fashion. We observed that although the upstream start site had a higher relative yield than the downstream start site, a reduced level of RNAP II flux fully accounted for the lower relative yield from the downstream site. RNAP II flux is an inherent property of a unidirectional scanning model for transcription initiation. Thus, the fact that RNAP II flux can be used to resolve the observed differences in relative utilization of two identical start sites is in itself strong support for the model. The fact that our estimations of RNAP II flux so closely agree with shifts in utilization of start sites 14−22 bp apart in response to mutations suggests that RNAP II scanning is reasonably processive. A scanning mechanism of start site selection requires RNAP II to be processive to accommodate the large and variable distances between yeast TATA boxes and initiator elements. It will be interesting to test the limits of processive scanning by RNAP II. The identification of initiators with a range of efficiencies (Table 1) should aid such studies.

The yeast initiator sequence consensus is readily apparent among RNAP II-transcribed snRNA genes (SNR14, SNR19, SNR20, SNR7) and snoRNA genes. A Weblogo alignment (54) using a pool of 22 yeast snRNA and snoRNA transcription start sites results in an initiator consensus very similar to the one reported for mRNA genes, A(\text{rich})_3NYY(A/T)N(N(\text{rich})_2. Given that yeast RNAP II transcribes all mRNA genes and most snRNA and snoRNA genes, it is expected that the cis-acting sequence requirements for their start site selection would be similar. However, the identification of a strong match to the initiator consensus from such a small sample size of snRNA/snoRNA genes suggests a basis for why yeast RNAP II scanning model is currently the best supported explanation for how yeast start sites are selected, but experimental testing the basic implications of the model have been scarce. We constructed an allele of SNR14 with tandemly duplicated start sites as a means to test the yeast scanning model in both a qualitative and quantitative fashion. We observed that although the upstream start site had a higher relative yield than the downstream start site, a reduced level of RNAP II flux fully accounted for the lower relative yield from the downstream site. RNAP II flux is an inherent property of a unidirectional scanning model for transcription initiation. Thus, the fact that RNAP II flux can be used to resolve the observed differences in relative utilization of two identical start sites is in itself strong support for the model. The fact that our estimations of RNAP II flux so closely agree with shifts in utilization of start sites 14−22 bp apart in response to mutations suggests that RNAP II scanning is reasonably processive. A scanning mechanism of start site selection requires RNAP II to be processive to accommodate the large and variable distances between yeast TATA boxes and initiator elements. It will be interesting to test the limits of processive scanning by RNAP II. The identification of initiators with a range of efficiencies (Table 1) should aid such studies.
window between the SNR14 TATA box and the +1 start site revealed that none contained the preferred adenine at the –8 or –7 positions. The driving force behind this proposed kind of snRNA promoter evolution would be to increase the functional capacity of the RNA gene products, which are essential for yeast viability. snRNA genes encode structural RNAs that require precise 5′ ends for their function. For example, the 5′ end of yeast U4 RNA engages in base-pairing interactions with the yeast U6 RNA during the splicing cycle. In contrast, mRNA transcripts contain 5′-untranslated regions that typically have no precise length requirement for proper expression. Consequently, the 5′ end of mRNAs typically need not be formed in as precise a fashion as those of snRNAs, and their gene promoters would likely not have undergone the same type of evolutionary selection.

The study of transcription initiation on yeast non-coding RNA genes has provided useful insight into the fundamental process by which RNAP II initiates RNA synthesis, particularly with regard to its accuracy. Additional genetic, biochemical, and structural studies are necessary to elucidate the underlying mechanism by which both initiator DNA and PIC proteins function in the process of RNAP II transcription start site selection.

Acknowledgments—We thank E. Craig, P. James, C. Guthrie, K. Shannon, and A. Ponticelli for providing plasmids and/or yeast strains and R. Landick, V. Gerlach, and current members of the Brow laboratory for discussions.

REFERENCES

1. Orphanides, G., Lagrange, T., and Reinberg, D. (1996) Genes Dev. 10, 2657–2683
2. Hampsey, M. (1998) Microbiol. Mol. Biol. Rev. 62, 465–503
3. Hahn, S. (2004) Nat. Struct. Mol. Biol. 11, 394–403
4. Smale, S. T., and Kadonaga, J. T. (2003) Annu. Rev. Biochem. 72, 449–479
5. Buratowski, S., Hahn, S., Sharp, P. A., and Guarente, L. (1988) Nature 334, 37–42
6. Nakajima, N., Horikoshi, M., and Roeder, R. G. (1988) Mol. Cell. Biol. 8, 4028–4040
7. Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989) Cell 56, 549–561
8. Leuther, K. K., Bushnell, D. A., and Kornberg, R. D. (1996) Cell 85, 773–779
9. Bushnell, D. A., Westover, K. D., Davis, R. E., and Kornberg, R. D. (2004) Science 303, 983–988
10. Struhl, K. (1989) Annu. Rev. Biochem. 58, 1051–1077
11. Giardina, C., and Lis, J. T. (1993) Science 261, 759–762
12. Hahn, S., Hoar, E. T., and Guarente, L. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 8562–8566
13. Healy, A. M., Helser, T. L., and Zitomer, R. S. (1987) Mol. Cell. Biol. 7, 3785–3791
14. Furter-Graves, E. M., and Hall, B. D. (1990) Mol. Gen. Genet. 223, 407–416
15. Zhang, Z., and Dietrich, F. S. (2005) Nucleic Acids Res. 33, 2838–2851
16. Smale, S. T., and Baltimore, D. (1989) Cell 57, 103–113
17. Sakurai, H., Ohishi, T., and Fukasawa, T. (1994) Mol. Cell. Biol. 14, 6819–6828
18. Ohishi-Shofuda, T., Suzuki, Y., Yano, K., Sakurai, H., and Fukasawa, T. (1999) Biochem. Biophys. Res. Commun. 255, 157–163
19. Li, Y., Flanagan, P. M., Tschucher, H., and Kornberg, R. D. (1994) Science 263, 805–807
20. Pinto, I., Ware, D. E., and Hampsey, M. (1992) Cell 68, 977–988
21. Pinto, I., Wu, W. H., Na, J. G., and Hampsey, M. (1994) J. Biol. Chem. 269, 30569–30573
22. Pardee, T. S., Bangur, C. S., and Ponticelli, A. S. (1998) J. Biol. Chem. 273, 17859–17864
23. Faitar, S. L., Brodie, A. A., and Ponticelli, A. S. (2001) Mol. Cell. Biol. 21, 4427–4440
24. Shuman, S. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 12758–12760
25. Kornberg, R. D. (1980) Nature 283, 455–4540
26. Chen, B. S., Mandal, S. S., and Hampsey, M. (2004) J. Biol. Chem. 279, 5894–5903
27. Renfrow, M. B., Naryshkin, N., Lewis, L. M., Chen, H. T., Ebright, R. H., and Scott, R. A. (2004) J. Biol. Chem. 279, 2825–2831
28. Lascaris, R. F., Groot, E., Hoen, P. B., Mager, W. H., and Planta, R. J. (2000) Nucleic Acids Res. 28, 1390–1396
29. Qu, L. H., Henras, A., Lu, Y. J., Zhou, H., Zhou, W. X., Zhu, Y. Q., Zhao, J., Henry, Y., Caizergues-Ferrer, M., and Bachelier, J. P. (1999) Mol. Cell. Biol. 19, 1144–1158
30. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. I., Tagne, J. B., Volkert, T. L., Frisen, E., Gifford, D. K., and Young, R. A. (2002) Science 298, 799–804
31. Martin, M. P., Buratowski, L. V., and Brow, D. A. (2001) Mol. Cell. Biol. 21, 6429–6439
32. Clifton, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A., and Johnston, M. (2003) Science 301, 71–76
33. Mukherjee, S., Berger, M. F., Jona, G., Wang, X. S., Muzzey, D., Snyder, M., Young, R. A., and Bulyk, M. L. (2004) Nat. Genet. 36, 1331–1339
34. Beinoraviciute-Kellner, R., Lipps, G., and Krauss, G. (2005) FEBS Lett. 579, 4535–4540
35. Hekmatpanah, D. S., and Young, R. A. (1991) J. Biol. Chem. 266, 1859–1869
36. Shuman, S. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 12758–12760
37. Gerlach, V. L., Whitehall, S. K., Geiduschek, E. P., and Brow, D. A. (1995) Mol. Cell. Biol. 15, 1455–1466
38. Bartlett, M. S., Thomm, M., and Geiduschek, E. P. (2004) J. Biol. Chem. 279, 5894–5903
39. Faitar, S. L., Brodie, S. A., and Ponticelli, A. S. (2001) Mol. Cell. Biol. 21, 1455–1466
40. Fourel, G., Miyake, T., Defossez, P. A., Li, R., and Gilson, E. (2002) J. Biol. Chem. 277, 41736–41743
41. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. I., Tagne, J. B., Volkert, T. L., Frisen, E., Gifford, D. K., and Young, R. A. (2002) Science 298, 799–804