MOLECULAR AND CELLULAR BIOLOGY, Mar. 1983, p. 439–447
0270-7306/83/030439-09$02.00/0
Copyright © 1983, American Society for Microbiology
Vol. 3, No. 3

The Secreted Form of Invertase in *Saccharomyces cerevisiae* Is Synthesized from mRNA Encoding a Signal Sequence

MARIAN CARLSON,[1,2]* RONALD TAUSSIG,[1] SYDNEY KUSTU,[2]† AND DAVID BOTSTEIN[2]

Department of Human Genetics and Development, College of Physicians and Surgeons of Columbia University, New York, New York 10032,[1] and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[2]

Received 8 July 1982/Accepted 8 November 1982

The *SUC2* gene of *Saccharomyces cerevisiae* encodes two differently regulated mRNAs (1.8 and 1.9 kilobases) that differ at their 5′ ends. The larger RNA encodes a secreted, glycosylated form of invertase and the smaller RNA encodes an intracellular, nonglycosylated form. We have determined the nucleotide sequence of the amino-terminal coding region of the *SUC2* gene and its upstream flanking region and have mapped the 5′ ends of the *SUC2* mRNAs relative to the DNA sequence. The 1.9-kilobase RNA contains a signal peptide coding sequence and presumably encodes a precursor to secreted invertase. The 1.8-kilobase RNA does not include the complete coding sequence for the signal peptide. The nucleotide sequence data prove that *SUC2* is a structural gene for invertase, and translation of the coding information provides the complete amino acid sequence of an *S. cerevisiae* signal peptide.

Saccharomyces cerevisiae cells carrying a *SUC*$^+$ gene produce two forms of the sucrose-hydrolyzing enzyme invertase: a secreted, glycosylated form and an intracellular, nonglycosylated form. Synthesis of the secreted invertase is regulated by glucose repression. The intracellular invertase is synthesized at low levels constitutively. A variety of evidence has suggested that each of the *SUC* genes, a family including *SUC1-SUC5* and *SUC7* (5, 19), is a structural gene for both forms of invertase (6, 12, 13, 21, 24). We have previously reported cloning the *SUC2* gene by complementation of a *suc2* mutation in *S. cerevisiae* (4). We used the cloned DNA to investigate the synthesis of two forms of invertase from one gene. Two mRNA products of the *SUC2* gene were identified: a 1.9-kilobase (kb) RNA and a 1.8-kb RNA. The steady-state level of the 1.9-kb RNA is regulated by glucose repression and that of the 1.8-kb RNA is not regulated. Studies of wild-type cells and *snf1* mutant cells, which are defective in derepression of secreted invertase, showed that both RNAs are present in cells synthesizing both forms of invertase and that only the 1.8-kb RNA is present in cells synthesizing only intracellular invertase (4). The two RNAs were shown by S1 mapping to differ at their 5′ ends. No intervening sequences were detected in either RNA, although the data could not exclude a small exon very close to an end. We proposed that the larger RNA includes the initiation codon for a signal peptide constituting the amino-terminal end of a precursor to secreted invertase and that the smaller RNA does not include the complete coding sequence for the signal peptide. Translation of the smaller RNA would begin at a methionine downstream from the signal peptide initiator methionine and lead to synthesis of an intracellular enzyme.

In related studies, Perlman and Halvorson (21) found that in vitro translation of mRNA from a *SUC2* strain generates three invertase polypeptides, two of which can be processed and "secreted" by dog pancreas membranes. Partial amino acid sequences of the amino termini of these in vitro translation products suggested the presence of a signal peptide on the major polypeptide secreted by dog pancreas membranes (22). These authors proposed that the three polypeptides they observed are encoded by three distinct mRNAs (21), and that the mRNA for the major secreted polypeptide contains an exon encoding the signal sequence and the mRNA for cytoplasmic invertase lacks this exon (22).

To test our model for the synthesis of two forms of invertase, we determined the nucleotide sequence of the amino-terminal coding region of the *SUC2* structural gene and its upstream flanking sequence. We mapped the 5′ ends of the *SUC2* mRNAs by the methods of primer extension with reverse transcriptase (3)

† Permanent address: Department of Bacteriology, University of California, Davis, CA 95616.

and S1 nuclease protection (2, 28). The results substantiate our model. In addition, we showed that RNA splicing is not involved in glucose regulation of the stable levels of the 1.9-kb RNA, and we proved definitively that *SUC2* is a structural gene for invertase.

MATERIALS AND METHODS

DNA preparation, labeling, and sequencing. Plasmid DNAs were prepared as described previously (4). DNAs were digested with restriction enzymes (New England BioLabs), using conditions recommended by the supplier. The 5′ ends of restriction fragments were dephosphorylated with bacterial alkaline phosphatase or calf intestinal alkaline phosphatase (Boehringer Mannheim) and then radioactively labeled by using [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase (Bethesda Research Laboratories) as described by Maxam and Gilbert (18). The 3′ ends of fragments were radioactively labeled by using [α-^{32}P]dGTP (Amersham) and the Klenow fragment of DNA polymerase I (Boehringer Mannheim). Fragments were purified and sequenced by the methods of Maxam and Gilbert (18). The G, G+A or A>C, C, and C+T cleavage reactions were carried out. The final lyophilizations were sometimes replaced by precipitations with ethanol (25).

S1 mapping. Polyadenylate [poly(A)]-containing RNA was prepared from glucose-repressed and derepressed cells as described previously (4), and 5′ end-labeled restriction fragments (approximately 1×10^6 to 2×10^6 cpm/pmol) were prepared as described above. S1 mapping was carried out essentially according to Berk and Sharp (2) except with end-labeled DNA (28). Hybridizations were performed at 42°C in 80% formamide (deionized)–0.4 M NaCl–0.04 M PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (pH 7.8)]–0.001

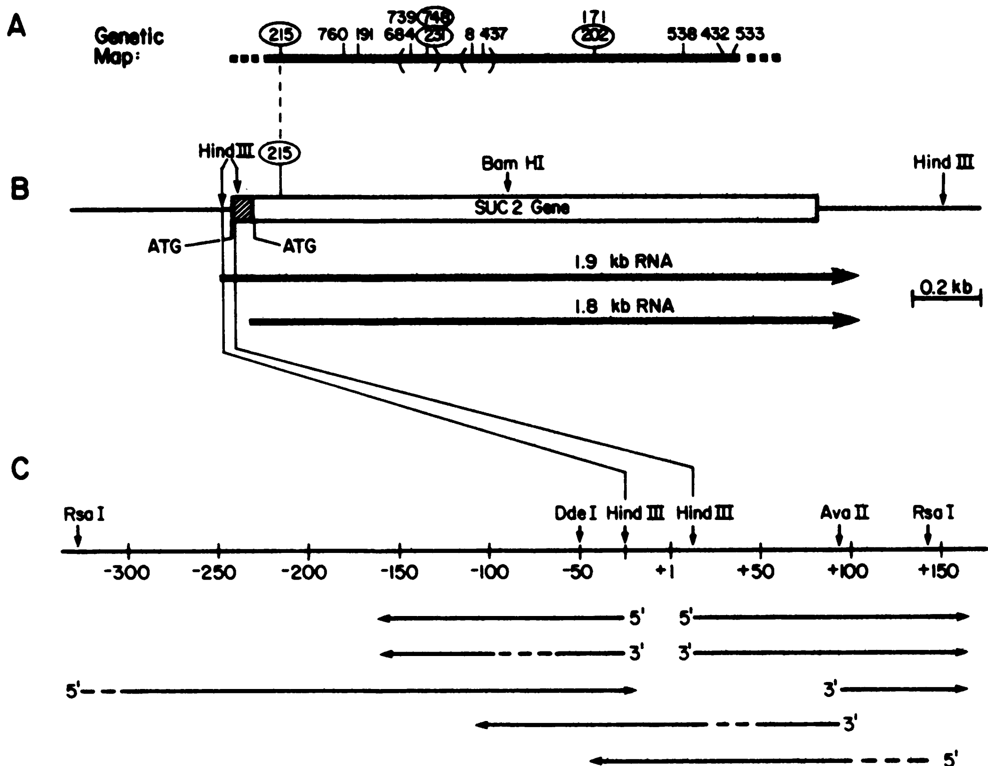

FIG. 1. Maps of the *SUC2* gene and sequencing strategy. (A) Fine-structure genetic map of *SUC2* taken from Carlson et al. (6). Amber alleles are circled. Map is based on sunlamp radiation-induced recombination frequencies in heteroallelic diploids. (B) Restriction map of the *SUC2* gene and map of its RNA transcripts taken from Carlson and Botstein (4). The model proposed earlier (4) for the structure of the *SUC2* gene is redrawn according to the data reported in this paper. The bar represents the structural gene and the hatched region encodes the signal sequence. The positions of initiation codons and the sequenced *suc2-215* amber mutation are indicated. The arrows represent RNA transcripts, pointing in the direction of the 3′ end. (C) The region near the 5′ end of the *SUC2* gene is expanded and nucleotide positions are numbered as in Fig. 2. Restriction sites used to determine the sequence are shown. Each arrow starts at the radioactively labeled 5′ or 3′ end, as indicated, and the unbroken portion of the arrow up to the head indicates the extent of the nucleotide sequence determined. The sequences from −162 to −27 and +12 to +162 were determined from both pRB118 and pRB58, and the remainder were determined from pRB118.

M EDTA. Poly(A)-containing RNA derived from 100 μg of total RNA was used for each sample applied to one gel slot. When no *S. cerevisiae* RNA was added, *Escherichia coli* tRNA (Boehringer Mannheim) was substituted. Digestion with S1 nuclease (Miles Laboratories) was carried out as described previously (2), and sonicated salmon sperm DNA was used as carrier in the final ethanol precipitation. Protected DNA fragments were analyzed on 0.4-mm-thick, 8% polyacrylamide–7 M urea sequencing gels.

Primer extension. RNA and excess 5′ end-labeled DNA primer were hybridized under conditions identical to those used for the S1 mapping experiments. Poly(A)-containing RNA prepared from 15 to 30 μg of total RNA was used for each gel sample. The hybrids were recovered by ethanol precipitation, and the DNA primers were elongated with avian myeloblastosis virus reverse transcriptase (500 U/ml; Life Sciences Inc.) as described by Lamb and Lai (17). Except where otherwise indicated, the extension reactions were carried out in 50 mM Tris-hydrochloride (pH 8.3), 10 mM $MgCl_2$, 50 mM KCl, 5 mM dithiothreitol, 40 μg of actinomycin D per ml, and 500 μM each of dATP, dGTP, dCTP, and dTTP for 1 h at 41°C. After precipitation with ethanol, the products were analyzed on 8% polyacrylamide–7 M urea sequencing gels.

RESULTS

Figure 1 shows the restriction map of the *SUC2* gene and the map of its two RNA transcripts (4). Nucleotide sequence analysis of the region near the 5′ end of the gene was carried out on restriction fragments prepared from two plasmids, pRB58 and pRB118, according to the strategy indicated in Fig. 1. Plasmid pRB58 carries a *SUC2*$^+$ gene and pRB118 contains a subcloned segment of the *suc2-215* amber allele (4). Figure 2 shows the DNA sequence and the amino acid translation of the open reading frame within the region where the *SUC2* mRNAs map.

Structural gene for invertase. To confirm that *SUC2* is a structural gene for invertase, the amino acid sequence encoded by the *SUC2* gene was compared with the partial amino-terminal sequence of invertase reported by Perlman et al. (22). By analysis of invertase polypeptides translated in vitro from mRNA from a *SUC2* strain, Perlman et al. identified 17 of the first 42 amino acids. Our inferred amino acid sequence agrees with their data at 16 residues (see Fig. 2). Our

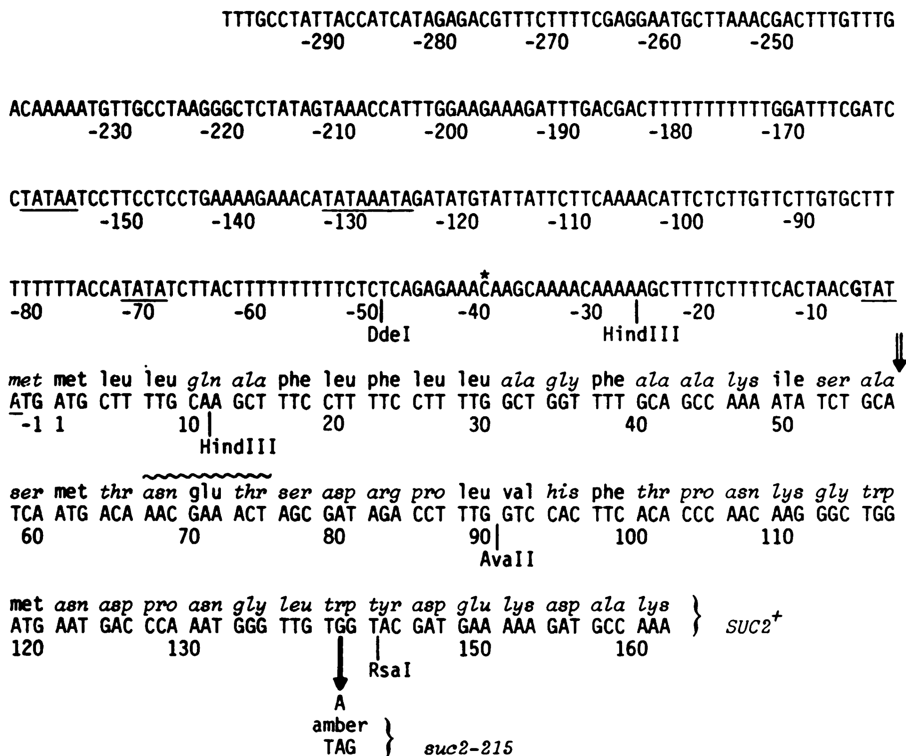

FIG. 2. Nucleotide sequence of the amino-terminal coding region of the *SUC2* gene and 5′ flanking region. At position +140, the sequence of the *suc2-215* amber allele differs from that of the *SUC2*$^+$ allele as indicated. The amino acid translation of the open reading frame within the region where the *SUC2* mRNAs map is shown. The 16 amino acids predicted from the partial amino acid sequence of Perlman et al. (22) are in roman type. The signal peptide cleavage site reported by Perlman et al. (22) is marked with an arrow. The position of the 5′ end of the 1.9-kb *SUC2* RNA (nucleotide −40 ± 1) is marked with an asterisk. TATA-like sequences are underlined. The putative glycosylation site is marked above with a wavy line.

results differ in that the DNA sequence encodes serine at nucleotides +58 to +60, whereas they tentatively found leucine. Our sequence shows two methionine codons at the beginning of the coding region. We have assigned position 1 to the second methionine codon because the in vitro translational product begins with a single methionine residue and because comparisons with other *S. cerevisiae* gene sequences suggest that the second ATG is the initiation codon (see below) (8). The agreement between the DNA sequence of our cloned *SUC2* gene and the amino acid sequence of invertase establishes without question that *SUC2* is a structural gene for invertase.

Signal peptide sequence. At the amino terminus of invertase, the amino acid sequence translated from the DNA sequence includes a core of hydrophobic residues and strongly resembles procaryotic and eucaryotic signal peptides (for review, see references 7 and 16). Perlman et al. (22) reported that an invertase polypeptide translational product synthesized in vitro contains an amino-terminal signal sequence of 19 amino acids that is cleaved during secretion in a cell-free system. The position at which they observed cleavage corresponds to the bond between the alanine and serine residues predicted by nucleotides +55 to +60 in the DNA sequence. Alanine frequently constitutes the residue on the amino-terminal side of signal peptide cleavage sites. The DNA sequence from +1 to +57 predicts the amino acid sequence of the invertase signal peptide.

Glycosylation site. The amino acid translation of nucleotides +67 to +75 is asparagine-glutamate-threonine. The sequences asparagine-X-threonine and asparagine-X-serine are sites for glycosylation in eucaryotic proteins (14), and secreted invertase is glycosylated at an average of nine sites on each of its two identical subunits (27). Thus, it is likely that we have identified one of the glycosylation sites.

Incomplete signal sequence in smaller *SUC2* mRNA. The sequence in Fig. 2 shows that the *Hin*dIII site at position +11 to +16 lies within the signal peptide coding region. The position of this site is significant because previous work indicated that the 5′ end of the glucose-regulated 1.9-kb *SUC2* RNA (encoding secreted invertase) extends to the left of this *Hin*dIII site (see Fig. 1) (4). These studies also showed that the 5′ end (or, possibly, a splice junction) of the 1.8-kb *SUC2* RNA (encoding intracellular invertase) begins to the right of this *Hin*dIII site. In either case, the smaller RNA could not encode the complete signal sequence.

Mapping the 5′ ends of *SUC2* mRNAs. We undertook to define more precisely the 5′ ends of the *SUC2* mRNAs and also to investigate the possibility that splicing is involved in their synthesis or regulation. We used both the methods of S1 protection mapping (2, 28) and primer extension with reverse transcriptase (3). In order that the mapping data would be directly comparable, both procedures were carried out by using restriction fragments that were 5′ end-labeled at the *Rsa*I site at position +141 to +144 (see Fig. 1). For S1 mapping, an end-labeled fragment extending from this *Rsa*I site past the 5′ ends of the *SUC2* RNAs to the next *Rsa*I site was hybridized to *SUC2* RNA and the portion of the fragment protected from S1 digestion was determined. For primer extension analysis, the DNA fragment between the end-labeled *Rsa*I site and the *Ava*II site at position +90 to +94 was hybridized to *SUC2* RNA and used as a primer for elongation by reverse transcriptase. If a *SUC2* RNA species has no splice located upstream from the labeled *Rsa*I site, the DNA fragment protected from S1 digestion and the primer extension product should be the same size. If a splice is present, the S1-protected fragment should be smaller, as it should extend from the *Rsa*I site only to the position of the first upstream splice.

Poly(A)-containing RNA from both glucose-repressed and derepressed yeast cells was used in these experiments. Glucose-derepressed cells contain both the 1.9 and 1.8-kb *SUC2* RNAs, and glucose-repressed cells contain only the 1.8-kb *SUC2* RNA (4). Figure 3 shows that RNA from glucose-derepressed cells protects from S1 digestion a DNA fragment of the same size (approximately 180 nucleotides) as a primer extension product synthesized by using this RNA as a template for reverse transcriptase. Neither of these DNA species was observed when RNA from glucose-repressed cells was used (Fig. 3 and 4). These data indicate that the 1.9-kb RNA encoding secreted invertase has no splice upstream from the *Rsa*I site. The absence of any splice enabled us to map the 5′ end of this RNA simply by determining the size of either product. The size of the primer extension product was determined by electrophoresis on a sequencing gel adjacent to sequencing reactions carried out on a restriction fragment end-labeled at the same *Rsa*I site (Fig. 4). The primer extension product migrates between the sequencing reaction products resulting from cleavages at the T residues at positions −41 and −42. Thus, the primer extension product probably ends with the guanine nucleotide at position −40 (see the legend to Fig. 4) (26). This result means that the mRNA probably begins within one nucleotide of the cytosine at position −40 in the DNA sequence (Fig. 2). The 5′ end of the 1.9-kb *SUC2* RNA thus maps ahead of the initiation codon for the signal sequence, and therefore this RNA contains cod-

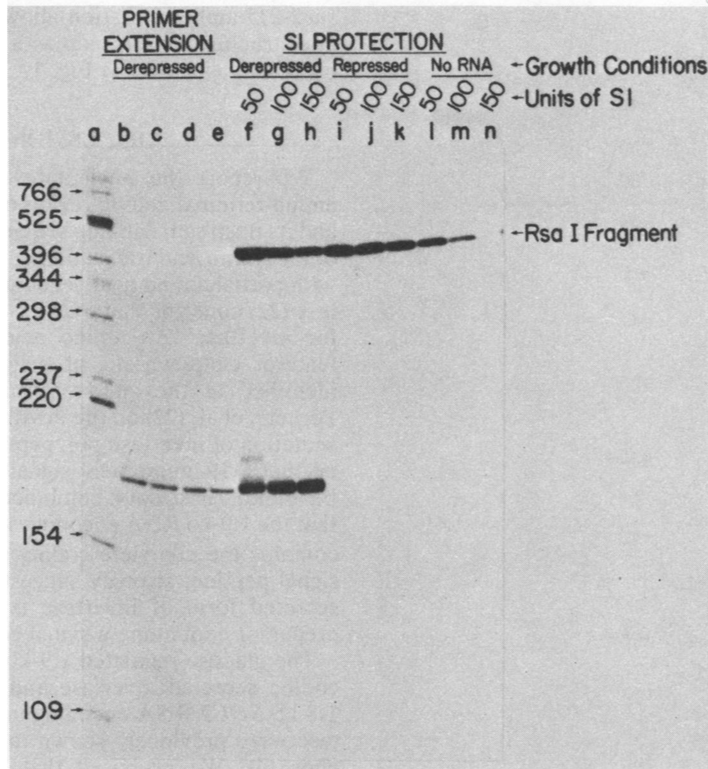

FIG. 3. mRNA encoding secreted invertase is unspliced. The radioactively labeled DNA products of the primer extension experiment and the S1 nuclease protection experiment described in the text were electrophoresed on an 8% polyacrylamide sequencing gel for size comparison. An autoradiogram of the gel is shown. (a) Marker *Hin*fI fragments of pBR322 and simian virus 40 DNA. The numbers indicate sizes in nucleotides. Lanes b through e are primer extension products synthesized from RNA from glucose-derepressed cells, using the *Rsa*I-*Ava*II fragment described in the text as primer. Reaction conditions were varied as follows: (b) as described in the text without modification; (c) with the addition of 0.01% Triton X-100; (d) with 750 μM deoxynucleoside triphosphates, 15 mM KCl, 1,000 U of reverse transcriptase per ml, incubated at 47°C; (e) as in (d) but with the addition of 0.01% Triton X-100. Lanes f through n are S1 protection products. The *Rsa*I fragment was protected from S1 digestion by the following RNA samples: (f through h) RNA from glucose-derepressed cells; (i through k) RNA from glucose-repressed cells; (l through n) no yeast RNA. RNA-DNA hybrid molecules were digested with 50, 100, or 150 U of S1 nuclease, as indicated.

ing sequences for an invertase polypeptide with an amino-terminal signal sequence.

Primer extension experiments carried out with RNA from glucose-repressed cells resulted in several extension products (Fig. 4). If the three major extension products are synthesized from unspliced RNA templates, then their 5' ends map at positions +50, +54, and +58. These same products were detected in the experiment using RNA from glucose-derepressed cells, which contain both the 1.8- and 1.9-kb *SUC2* RNAs (Fig. 4). In addition, several other products were detected, which may correspond to RNA species specific to derepressed cells or may result from premature termination by reverse transcriptase on the 1.9-kb RNA template. We were unable, for technical reasons, to characterize the 1.8-kb *SUC2* RNA species by S1 protection mapping.

Alignment of physical and genetic maps. We previously constructed a fine-structure genetic map of *SUC2* and showed that the *suc2-215* mutation lies at one extreme of the map (Fig. 1A) (6); however, we did not know which end of the genetic map corresponded to the beginning of the gene. Most of the coding sequence shown in Fig. 2 was determined from two plasmids: pRB58, which carries a $SUC2^+$ gene, and pRB118, which contains a segment of the *suc2-215* amber allele. Comparison of the sequences showed one difference at nucleotide position +140. The $SUC2^+$ sequence encodes tryptophan, and the *suc2-215* sequence is an amber codon. Knowing the physical location of the

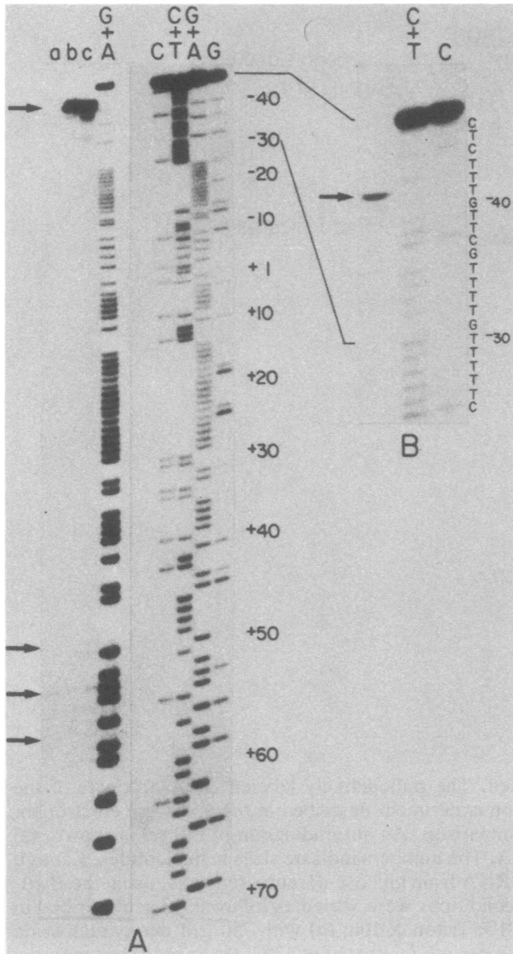

FIG. 4. Position of the 5′ end of the 1.9-kb *SUC2* RNA. (A) Extension products of the *Rsa*I-*Ava*II primer were synthesized from RNA prepared from (a) glucose-repressed cells and (b and c) glucose-derepressed cells. Reaction conditions were as described in the text except that 0.01% Triton X-100 was added in (b). Products were electrophoresed on an 8% polyacrylamide sequencing gel along with sequencing reaction products from the *Rsa*I-*Dde*I fragment (see Fig. 1), which was 5′ end-labeled at the *Rsa*I site. The sequence is that of the anti-sense strand, and the nucleotides are numbered as in Fig. 2. The primer extension products are also the anti-sense strand. Arrows indicate the major extension products. (B) Autoradiograph of an 8% polyacrylamide sequencing gel used to determine the size of the major primer extension product synthesized from the 1.9-kb *SUC2* mRNA. A sample identical to sample (c) described above was electrophoresed along with sequencing reaction products, also as described for (A). The extension product (indicated by the arrow) appears to terminate with the G residue at −40. The corresponding sequencing reaction product migrates faster because the sequencing reaction eliminates the 3′-terminal nucleoside and leaves a phosphate group (see Sollner-Webb and Reeder [26] for discussion).

suc2-215 amber mutation allows us to align the fine-structure genetic map with the physical map of *SUC2*, as shown in Fig. 1.

DISCUSSION

We report the nucleotide sequence of the amino-terminal coding region of the *SUC2* gene and its upstream flanking sequence. Comparison of the amino acid translation of the coding region with partial amino acid sequence data for invertase (22) confirms that *SUC2* is a structural gene for invertase. An amino acid sequence with features characteristic of a signal peptide was identified at the amino terminus; studies by Perlman et al. (22) on the in vitro translation and secretion of invertase polypeptides had suggested that a 19-amino acid signal sequence would be found. These data, combined with our finding that the 1.9-kb RNA encoding secreted invertase contains the complete coding sequence for the signal peptide, strongly suggest that in vivo the secreted form of invertase is derived from a precursor containing a signal sequence.

The glucose-regulated 1.9-kb *SUC2* RNA encoding secreted invertase and the unregulated 1.8-kb *SUC2* RNA encoding intracellular invertase were previously shown to differ at their 5′ ends (4). We proposed that the larger RNA encodes a signal sequence-containing precursor to secreted invertase and that the smaller RNA does not encode the complete signal sequence and is translated to yield intracellular invertase. The data presented here support this idea. The 5′ end of the 1.9-kb *SUC2* RNA was mapped to within one nucleotide of position −40, showing that this RNA contains the complete coding sequence for a signal peptide-containing precursor of invertase. The conclusion that the 5′ end of the 1.9-kb RNA is at position −40, and not further upstream, is subject to the reservation that the S1 protection and primer extension mapping techniques might be sensitive to secondary structure in DNA and RNA. It should also be noted that these methods do not prove that this 5′ end corresponds to a transcriptional start; it is possible that the 1.9-kb RNA is produced from a larger precursor by removal of 5′ terminal nucleotides.

Previous S1 mapping studies (4) showed that either the 5′ end of the 1.8-kb *SUC2* RNA or possibly a splice site lies downstream from the *Hin*dIII site at position +11 to +16, which is shown here to map within the signal sequence. These data suggest, in accord with our model, that the 1.8-kb RNA cannot contain the complete signal peptide coding sequence. An in-phase methionine codon which could serve to initiate translation of the intracellular invertase is found at position +61 to +63, just three

nucleotides downstream from the putative signal peptide cleavage site. Comparison of the amino acid translation of the DNA sequence with the partial amino acid sequence of one of the invertase polypeptides translated in vitro (22) indicates that this methionine codon does indeed serve to initiate translation; this polypeptide contained glutamate, leucine, valine, phenylalanine, and methionine residues (shown in Fig. 2) at positions 3, 9, 10, 12, and 19, respectively (the initiator methionine was presumably removed). Our data are consistent with the idea that the 1.8-kb RNA includes this methionine codon and the downstream sequence.

Our efforts to map precisely the 5′ end(s) of the 1.8-kb RNA were inconclusive. Several primer extension products were synthesized from the 1.8-kb RNA. If the RNA templates are unspliced, and all evidence indicates that the RNA encoding intracellular invertase is colinear with the *SUC2* gene downstream from nucleotide +61, then the three major products terminate at nucleotides +50, +54, and +58. This result is consistent with the size difference between the 1.8- and 1.9-kb RNAs and the position of the 5′ end of the 1.9-kb RNA at −40. We were, however, unable to confirm by S1 mapping that these are the positions of the 5′ ends. In our previous S1 mapping studies and in the current efforts, the fragments apparently protected by the 1.8-kb RNA were heterogeneous in size under conditions in which a discrete fragment was protected by the 1.9-kb RNA. We still do not know whether this result reflects real heterogeneity in the 1.8-kb size class of RNA or merely experimental difficulties. We are inclined to take this heterogeneity seriously, particularly since similar heterogeneity has been reported for other *S. cerevisiae* RNAs (10, 20). Although we can draw no definite conclusions regarding the exact 5′ ends of the 1.8-kb RNA species, it is in any case clear that this RNA does not encode a complete signal sequence.

The 1.8- and 1.9-kb RNAs are regulated differently. The stable level of the 1.9-kb RNA is regulated in response to glucose concentration in the medium, as is the synthesis of secreted invertase (4). In contrast, the stable level of the 1.8-kb RNA and the synthesis of intracellular invertase remain relatively constant. A variety of regulatory mechanisms can be suggested to account for these findings. The simplest is that the two RNAs are transcribed from different promoters and only the promoter for the 1.9-kb RNA is regulated in response to glucose. Our results are consistent with this idea, and we favor it as the most likely mechanism. The possibility that the stability of the 1.9-kb RNA (rather than synthesis) is regulated cannot be excluded; however, any mechanism for differential stability must be specific for the 1.9-kb RNA and not the 1.8-kb RNA. A class of models involving splicing is ruled out by our data showing that the 1.9-kb RNA has no splice near its 5′ end. Another possibility is that transcription occurs from a single promoter and that the two RNAs are produced by differential processing of the 5′ terminus of the transcript (perhaps by removal of nucleotides); any such model would, however, have to provide specially for both regulation of the level of the 1.9-kb RNA in response to glucose and production of a constant level of the 1.8-kb RNA.

Several interesting features of the nucleotide sequence can be noted. There are two ATG codons at the beginning of the structural gene (Fig. 2). We have assigned the ATG at +1 as the initiation codon on the basis of in vitro translation data (22) and its sequence context. The sequence surrounding this ATG codon has features characteristic of a translation initiation region in *S. cerevisiae* (8), which differs somewhat from those of other eucaryotes (15). The adenine residue at −3 has been found in all *S. cerevisiae* genes examined, and the thymine at +6 is common to most *S. cerevisiae* genes. These observations support the assignment of the ATG at +1 (rather than the ATG at −3) as the initiation codon.

TATA-like sequences are located 120, 93, and 32 nucleotides upstream from the 5′ end of the 1.9-kb RNA, beginning at positions −160, −133, and −72, respectively. The sequence TATA at position −6 is also upstream from the apparent 5′ ends of the 1.8-kb RNA species. We do not know which, if any, of these TATA-like sequences is important for transcription of the *SUC2* gene; however, computer-assisted comparison of the 5′ flanking regions of *SUC2* and the *S. cerevisiae PHO5* gene, which encodes secreted acid phosphatase, revealed that the sequence surrounding the TATAAATA element 93 nucleotides upstream from the 5′ end of the 1.9-kb *SUC2* RNA is homologous to the sequence surrounding the TATATAA element about 65 nucleotides upstream from the 5′ end of the major *PHO5* RNAs (R. Kramer and G. Thill, personal communication). Comparison with the yeast alcohol dehydrogenase I gene shows that this region is also homologous to the sequence surrounding the TATAATA element 91 and 101 nucleotides upstream from the two 5′ ends of alcohol dehydrogenase I mRNAs (1).

Another feature of the 5′ flanking region of the *SUC2* gene is a pyrimidine-rich block between positions −68 and −48, followed by the sequence CAAG at −40. This structure (and its relative location) is common to a variety of *S. cerevisiae* genes (8). Two other pyrimidine-rich blocks are found in the region upstream of the

1.8-kb RNA (at nucleotides −24 to −9 and +14 to +29), and the sequence CAAG is present at position +10.

It is also of interest to note that the sequence from −300 to −1 has an open reading frame in the same frame as the invertase gene. An ATG codon is located at position −234. No RNA transcript containing this region was detected (4).

A likely glycosylation site for the secreted invertase was identified from the amino acid translation of the DNA sequence. Nucleotides +67 to +75 encode asparagine-glutamate-threonine, which is known to serve as a site for glycosylation of eucaryotic proteins (14). The secreted invertase is glycosylated at approximately nine sites per subunit (27). The intracellular invertase is not glycosylated (11), although these same sites are present in the protein. This finding is consistent with evidence that glycosylation is carried out during the secretory process (9, 14).

The nucleotide sequence of the *suc2-215* allele confirms directly that it is an amber mutation; previous identification of *suc2-215* as an amber mutation was based on analysis of suppression patterns (6). Comparison of the nucleotide sequence in the $SUC2^+$ and *suc2-215* alleles shows that the mutation is a G · C to A · T transition, which is a change commonly induced in *S. cerevisiae* by the mutagen used, ethyl methane sulfonate (23). This interpretation is, however, subject to the following reservation. The $SUC2^+$ gene in pRB58 was derived from a *suc2-215* amber allele, probably by recombination between *suc2-215* and another mutant allele (4). It is thus likely that this $SUC2^+$ gene is a wild-type allele, but we cannot exclude the possibility that the TGG codon is the result of reversion, rather than recombination, and therefore is not the codon at this position in a wild-type $SUC2^+$ gene.

ACKNOWLEDGMENTS

We thank Paula Grisafi for excellent technical assistance.

This work was supported by American Cancer Society grants NP-358 to M.C. and MV-90A to D.B.

LITERATURE CITED

1. **Bennetzen, J. L., and B. D. Hall.** 1982. The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. J. Biol. Chem. **257**:3018–3025.
2. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
3. **Bina-Stein, M., M. Thoren, N. Salzman, and J. Thompson.** 1979. Rapid sequence determination of late SV40 16S mRNA leader by using inhibitors of reverse transcriptase. Proc. Natl. Acad. Sci. U.S.A. **76**:731–735.
4. **Carlson, M., and D. Botstein.** 1982. Two differentially regulated mRNAs with different 5′ ends encode secreted and intracellular forms of yeast invertase. Cell **28**:145–154.
5. **Carlson, M., B. C. Osmond, and D. Botstein.** 1980. *SUC* genes of yeast: a dispersed gene family. Cold Spring Harbor Symp. Quant. Biol. **45**:799–803.
6. **Carlson, M., B. C. Osmond, and D. Botstein.** 1981. Mutants of yeast defective in sucrose utilization. Genetics **98**:25–40.
7. **Davis, B. D., and P.-C. Tai.** 1980. The mechanism of protein secretion across membranes. Nature (London) **283**:433–438.
8. **Dobson, M. J., M. F. Tuite, N. A. Roberts, A. J. Kingsman, and S. M. Kingsman.** 1982. Conservation of high efficiency promoter sequences in *Saccharomyces cerevisiae*. Nucleic Acids Res. **10**:2625–2637.
9. **Esmon, B., P. Novick, and R. Schekman.** 1981. Compartmentalized assembly of oligosaccharides on exported glycoproteins in yeast. Cell **25**:451–460.
10. **Faye, G., D. W. Leung, K. Tatchell, B. D. Hall, and M. Smith.** 1981. Deletion mapping of sequences essential for *in vivo* transcription of the iso-1-cytochrome *c* gene. Proc. Natl. Acad. Sci. U.S.A. **78**:2258–2262.
11. **Gascon, S., N. P. Neumann, and J. O. Lampen.** 1968. Comparative study of the properties of the purified internal and external invertases from yeast. J. Biol. Chem. **243**:1573–1577.
12. **Grossmann, M. K., and F. K. Zimmermann.** 1979. The structural genes of internal invertases in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **175**:223–229.
13. **Hackel, R. A.** 1975. Genetic control of invertase formation in *Saccharomyces cerevisiae*. I. Isolation and characterization of mutants affecting sucrose utilization. Mol. Gen. Genet. **140**:361–370.
14. **Hubbard, S. C., and R. J. Ivatt.** 1981. Synthesis and processing of asparagine-linked oligosaccharides. Annu. Rev. Biochem. **50**:555–583.
15. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5252.
16. **Kreil, G.** 1981. Transfer of proteins across membranes. Annu. Rev. Biochem. **50**:317–348.
17. **Lamb, R. A., and C.-J. Lai.** 1980. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. Cell **21**:475–485.
18. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
19. **Mortimer, R. K., and D. C. Hawthorne.** 1969. Yeast genetics, p. 385–460. *In* A. H. Rose and J. S. Harrison (ed.), The yeasts, vol. 1. Academic Press, Inc., New York.
20. **Nasmyth, K. A., K. Tatchell, B. D. Hall, C. Astell, and M. Smith.** 1981. A position effect in the control of transcription at yeast mating type loci. Nature (London) **289**:244–250.
21. **Perlman, D., and H. O. Halvorson.** 1981. Distinct repressible mRNAs for cytoplasmic and secreted yeast invertase are encoded by a single gene. Cell **25**:525–536.
22. **Perlman, D., H. O. Halvorson, and L. E. Cannon.** 1982. Presecretory and cytoplasmic invertase polypeptides encoded by distinct mRNAs derived from the same structural gene differ by a signal sequence. Proc. Natl. Acad. Sci. U.S.A. **79**:781–785.
23. **Prakash, L., and F. Sherman.** 1973. Mutagenic specificity: reversion of iso-1-cytochrome *c* mutants of yeast. J. Mol. Biol. **79**:65–82.
24. **Rodriguez, L., J. O. Lampen, and V. L. MacKay.** 1981. *SUC1* gene of *Saccharomyces*: a structural gene for the large (glycoprotein) and small (carbohydrate-free) forms of invertase. Mol. Cell. Biol. **1**:469–474.
25. **Smith, D. R., and J. M. Calvo.** 1980. Nucleotide sequence of the *E. coli* gene coding for dihydrofolate reductase.

Nucleic Acids Res. 8:2255–2273.
26. **Sollner-Webb, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in X. laevis. Cell 18:485–499.
27. **Trimble, R. B., and F. Maley.** 1977. Subunit structure of external invertase from *Saccharomyces cerevisiae*. J. Biol. Chem. 252:4409–4412.
28. **Weaver, R. F., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. 7:1175–1193.