Letter to the Editor

The GCRI Gene of Saccharomyces cerevisiae Is a Split Gene
With an Unusually Long Intron

In Saccharomyces cerevisiae, GCRI encodes a transcriptional activator that contacts at least two other proteins [Rap1p and an unidentified component of the transcriptional machinery (TORNOW et al. 1993)]. Cells that lack Gcr1p grow much more slowly than wild-type cells and exhibit a reduction in transcription of glycolytic and translational component genes (CLIFTON and FRAENKEL 1981; CLIFTON et al. 1978; SANTANGELO and TORNOW 1990; TORNOW and SANTANGELO 1990). Activation of these genes is accomplished interdependently with Rap1p (TORNOW et al. 1993), an essential multifunctional protein. Our laboratory is interested in determining the role of Gcr1p in transcriptional activation of its many target genes. During an analysis of the effects of point mutagenesis on GCRI, we discovered that it contains an intron, contrary to earlier published reports (HOLLAND et al. 1987; BAKER 1986). Given that this gene is of interest to many laboratories, we report here the evidence indicating the presence and location of the intron to correct the misconception that GCRI is not spliced.

GCRI is a split gene: Random point mutations were introduced throughout the 5'-half of the coding region of GCRI by using error-prone polymerase chain reaction (PCR) (LEUNG et al. 1989). The mutagenized genes were assayed by complementation by transforming S. cerevisiae cells containing a gcr1 null mutation [generated by replacing the wild-type GCRI locus with one in which the region between codon 66 and 797 has been replaced with the URA3 gene (HOLLAND et al. 1987)]. DNA sequence analysis of the mutated genes produced a striking result, namely that three different fully functional variants were isolated that had stop codons near the beginning of the coding region (codons 37, 42, and 54). The presence of stop codons in functional GCRI genes indicated that the beginning of the predicted coding region is not translated, contrary to the conclusions of earlier published sequence analyses of the GCRI gene (BAKER 1986; HOLLAND et al. 1987), for which two independent laboratories predicted an identical, uninterrupted coding region. A second look at the published DNA sequence suggested the presence of an intron, since the conserved branchpoint sequence (TACTAAC) is found at nucleotide positions 798–804.

The GCRI intron is unusually long: To confirm that the GCRI mRNA is spliced, and to map the 5' and 3' intron/exon junctions precisely, we did DNA sequence analysis of cDNA copies of the GCRI mRNA. An S. cerevisiae cDNA library carried in lambda gt11 was purchased from Clontech Laboratories (Palo Alto, California) and screened with the HindIII to NcoI fragment that contains the 5' end of GCRI. The insert DNA from several independent clones was isolated and subcloned into pGEM11zf+ (Promega, Madison, Wisconsin). DNA sequence analysis of the cDNA clones confirmed the existence of an intron (Figure 1). The 5' splice junction is at nucleotide position 103/104 (G/gtatg . . . ) just upstream from the BglII site, and the 3' junction is at position 854/855 ( . . . tag/T), thus bordering an intron that is 751 bases long (the first base in the upstream HindIII site is base 1). This is unusually long for an S. cerevisiae intron; only the DBP2 gene has a longer one [1001 bases (RYMOND and ROSBASH 1992)]. The GCRI intron includes 62 N-terminal codons previously predicted to be part of the open reading frame, beginning at the ATG at position 670–672. Although it was reported that a gcr1 null mutant could be complemented with a BglII fragment beginning at position 107 (BAKER 1986), which is within the intron, we were unable to repeat that result. We suggest that the true start codon is at position 96–98, and that the gene is interrupted between the second and third base of the third codon. The actual coding region of GCRI thus predicts a 785-amino acid protein, with a molecular mass of 87.7 kilodaltons, and methionine, valine and cysteine as the first three residues. We propose that the numbering of GCRI codons be changed to reflect this new information, i.e., Thr63 is now Thr4.

It seems likely that Gcr1p is necessary for efficient expression of most translational component genes (SANTANGELO and TORNOW 1990; TORNOW et al. 1993); it is interesting that this regulatory gene now joins the intron-containing class of genes in S. cerevisiae, a majority of which are involved in production of the translational machinery. As a general rule, most S. cerevisiae genes are not interrupted; when introns do exist, they are found almost exclusively at the 5' end of the gene (RYMOND and ROSBASH 1992). It is therefore quite possible to overlook an intron when open reading frames are deduced by relying on DNA sequence data alone, without careful analysis of the corresponding mRNA products (by S1 analysis, for example). Other eukaryotic genes in the database may therefore similarly contain overlooked introns. Finally, several cDNA clones were isolated that contained copies of unspliced GCRI mRNA (data not shown), a result that suggests that the GCRI intron may be inefficiently spliced. This leaves
open the possibility that there may be alternative forms of Gcr1p. However, the alleles of GCR1 containing stop codons within the intron fully complemented the ger1 null mutant, which strongly suggests that the unspliced GCR1 mRNAs do not contribute to expression of functional Gcr1p. Discovery of the intron and unspliced cDNA clones raises the possibility that Gcr1p levels may be regulated at the level of splicing.

The random point mutagenesis was done by Stephen Deminoff, a graduate student in the Santangelo laboratory. This work was supported by National Institutes of Health grant GM45828 to G.M.S.

**LITERATURE CITED**


Communicating editor: S. Jinks-Robertson