Letter to the Editor

The *GCR1* Gene of *Saccharomyces cerevisiae* Is a Split Gene With an Unusually Long Intron

In *Saccharomyces cerevisiae*, *GCR1* 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 *GCR1*, 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 *GCR1* is not spliced.

***GCR1* is a split gene:** Random point mutations were introduced throughout the 5'-half of the coding region of *GCR1* by using error-prone polymerase chain reaction (PCR) (LEUNG *et al.* 1989). The mutagenized genes were assayed for complementation by transforming *S. cerevisiae* cells containing a *gcr1* null mutation [generated by replacing the wild-type *GCR1* 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 *GCR1* genes indicated that the beginning of the predicted coding region is not translated, contrary to the conclusions of earlier published sequence analyses of the *GCR1* 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 branch-point sequence (TACTAAC) is found at nucleotide positions 798–804.

The *GCR1* intron is unusually long: To confirm that the *GCR1* mRNA is spliced, and to map the 5' and 3' intron/exon junctions precisely, we did DNA sequence analysis of cDNA copies of the *GCR1* mRNA. An *S. cer-*

visiae 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 *GCR1*. 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 *BclI* 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 *GCR1* 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 *BclI* 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 *GCR1* 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 *GCR1* 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 *GCR1* mRNA (data not shown), a result that suggests that the *GCR1* intron may be inefficiently spliced. This leaves

FIGURE 1.—The *GCR1* gene contains an intron. A cDNA copy of the *GCR1* mRNA was isolated from a λ gt11 library, subcloned and subjected to DNA sequence analysis. Along the right border of the photograph is the DNA sequence of the region, showing both complementary and sense strands. The data in the autoradiogram represent the complementary strand of the cloned cDNA. The loop represents the sequence that is present in the genomic *GCR1* sequence, but absent in the cDNA. The intron sequence contains a TACTAAC box (as indicated), and is bordered by canonical splice junction signals. The predicted start codon is boxed. The numbering of the nucleotides is relative to the upstream *Hind*III site (HOLLAND *et al.* 1987).

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 *gcr1* 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.

JOANNE TORNOW and GEORGE SANTANGELO
Department of Biological Sciences
Center for Molecular and Cellular Biosciences
University of Southern Mississippi
Hattiesburg, Mississippi 39406-5018

LITERATURE CITED

- BAKER, H. V., 1986 Glycolytic gene expression in *Saccharomyces cerevisiae*: nucleotide sequence of *GCR1*, null mutants, and evidence for expression. *Mol. Cell. Biol.* **6**: 3774–3784.
- CLIFTON, D., and D. G. FRAENKEL, 1981 The *gcr* (glycolysis regulation) mutation of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **256**: 13074–13078.
- CLIFTON, D., S. B. WEINSTOCK and D. G. FRAENKEL, 1978 Glycolysis mutants in *Saccharomyces cerevisiae*. *Genetics* **88**: 1–11.
- HOLLAND, B. J., T. YOKOI, J. P. HOLLAND, K. MYAMBO and M. A. INNIS, 1987 The *GCR1* gene encodes a positive transcriptional regulator of the enolase and glyceraldehyde-3-phosphate dehydrogenase gene families in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **7**: 813–820.
- LEUNG, D. W., E. CHEN and D. V. GOEDDEL, 1989 A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**: 11–15.
- RYMOND, B. C., and M. ROSBASH, 1992 Yeast pre-mRNA splicing, pp. 143–192 in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, edited by E. W. JONES, J. R. PRINGLE and J. R. BROACH. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- SANTANGELO, G. M., and J. TORNOW, 1990 Efficient transcription of the glycolytic gene *ADH1* and three translational component genes require the *GCR1* product, which can act through TUF/GRF/RAP binding sites. *Mol. Cell. Biol.* **10**: 859–862.
- TORNOW, J., and G. M. SANTANGELO, 1990 Efficient expression of the *Saccharomyces cerevisiae* glycolytic gene *ADH1* is dependent upon a *cis*-acting regulatory element (*UAS_{RPC}*) found initially in genes encoding ribosomal proteins. *Gene* **90**: 79–85.
- TORNOW, J., X. ZENG, W. GAO and G. M. SANTANGELO, 1993 *GCR1*, a transcriptional activator in *Saccharomyces cerevisiae*, complexes with RAP1 and can function without its DNA binding domain. *EMBO J.* **12**: 2431–2437.

Communicating editor: S. JINKS-ROBERTSON