Volume 11 Number 6 1983 Nucleic Acids Research

Nucleotide sequence of the yeast *SUC2* gene for invertase

Ronald Taussig and Marian Carlson

Department of Human Genetics and Development, College of Physicians and Surgeons of Columbia University, 701 West 168th Street, New York, NY 10032, USA

Received 18 October 1982; Revised and Accepted 3 January 1983

ABSTRACT

The yeast SUC2 gene is a structural gene for both the secreted and intracellular forms of invertase. We have determined the nucleotide sequence of the coding region and the 5' and 3' flanking regions. The coding regions for the signal peptide-containing precursor to secreted invertase and for the intracellular invertase begin at different initiation codons within the SUC2 gene but share the same reading frame. The amino acid sequences predicted for the two forms of invertase from the nucleotide sequence are consistent with the properties of the purified enzymes. Potential sites for glycosylation of the secreted invertase are identified.

INTRODUCTION

Yeasts (Saccharomyces) carrying a SUC^+ gene produce two forms of the sucrose-cleaving enzyme invertase: a secreted, glycosylated enzyme and an intracellular, nonglycosylated enzyme. The SUC genes of yeast are a dispersed multigene family, including SUC1-SUC5 and SUC7 (1,2). We previously cloned one member of this family, the SUC2 gene, and showed that it is a structural gene for both forms of invertase (3,4). The SUC2 gene encodes two mRNAs, 1.9 kb and 1.8 kb in size, which differ at their 5' ends (3). Nucleotide sequence analysis of the 5' end of the gene showed that the 1.9 kb mRNA encodes a signal peptide-containing precursor to secreted invertase and that the 1.8 kb mRNA does not include the complete signal sequence and thus is translated to yield intracellular invertase (4). The two mRNAs are differently regulated in response to glucose concentration, as is synthesis of the two forms of invertase (3).

Yeast invertases have been extensively studied for over a century [for review, see (5)], and the secreted and intracellular invertases produced from the SUC2 gene have been purified and

characterized (6,7,8). The protein moiety of each consists of two identical subunits of approximately 60,000 daltons (6,7,9). The secreted enzyme is a glycoprotein containing about 50% carbohydrate (mannose and 3% glucosamine; 6) which is attached as approximately nine asparagine-linked polysaccharide chains per protein subunit (9,10). Secreted invertase has served as a model glycoprotein for studies of carbohydrate synthesis and structure (11,12) and for analysis of the secretory pathway in yeast (13).

We report the nucleotide sequence of the SUC2 structural gene for invertase, including the entire coding region and both 5' and 3' flanking sequences. The amino acid sequences of the two forms of invertase are predicted from the DNA sequence, and potential glycosylation sites on the secreted enzyme are identified.

MATERIALS AND METHODS

Preparation of end-labeled restriction fragments

Plasmid DNAs were prepared as described previously (3). Restriction sites were mapped by standard methods and by analysis of purified 5' end-labeled fragments. Restriction enzymes were purchased from New England BioLabs or Bethesda Research Laboratory and used under the conditions specified by the supplier. The 5' ends of restriction fragments were dephosphorylated using calf intestinal alkaline phosphatase (Boehringer-Mannheim) and then radioactively labeled using T4 polynucleotide kinase (Bethesda Research Laboratory) and γ-^{32}P-ATP (3000 Ci/mmol; Amersham), as described by Maxam and Gilbert (14).

Fragments labeled at a unique end were usually obtained by cleaving with a second restriction enzyme and separating the resulting two end-labeled fragments. In a few cases, the two labeled strands were separated. End-labeled fragments were purified by electrophoresis in 4% to 6% polyacrylamide gels prepared in 89 mM Tris-OH, 89 mM boric acid, 2 mM EDTA, pH 8.3, from a stock of 29% acrylamide, 1% N,N'-methylene-bis-acrylamide. For strand separation, the DNA was heated to 90^{0}C for 1 minute in 30% dimethylsulfoxide, 1 mM EDTA, prior to electrophoresis on a

5% polyacrylamide gel. DNA fragments were recovered from gels by crushing the gel and eluting the DNA (14).

Sequence analysis

DNA sequence was determined by the methods of Maxam and Gilbert (14). The G, G+A, C+T, and C reactions were carried out with only minor modifications, except that piperidinium-formate, pH 4.0, was used in the G+A reaction. The final lyophilizations were usually replaced by precipitations with ethanol (15). Cleavage products were analyzed by electrophoresis in 6% and 8% polyacrylamide/7 M urea gels (80 cm x 38 cm x 0.4 mm). Autoradiography was carried out at -70°C with Kodak XR-5 film and DuPont Lightning Plus intensifying screens. Sequences were analyzed on a PDP11/45 computer using programs obtained from R. Staden (MRC laboratory, Cambridge, U.K.) (16).

The nucleotide sequence was determined on both strands for greater than 55% of the sequence (66% of the coding region). The sequence near each restriction site was determined by analysis of at least one fragment spanning the site.

RESULTS

Nucleotide sequence analysis

Figure 1 shows a restriction map of the *SUC2* gene indicating the sites and strategy used for sequence analysis. The position of the coding region and the structures of the 1.9 kb and 1.8 kb mRNAs are also shown. Nucleotide sequence analysis was performed on restriction fragments prepared from plasmids pRB58, pRB59, pRB117 and pRB118, which have been described elsewhere (3). Plasmid pRB58 carries a *SUC2*$^+$ gene and pRB59, pRB117 and pRB118 contain subcloned fragments of the *suc2-215* amber allele (17).

Figure 2 presents the nucleotide sequence of the *SUC2*$^+$ gene, including the complete sequence of the coding region, 300 nucleotides of the 5' noncoding region and 493 nucleotides of 3' flanking sequence. The only open reading frame encoded by the *SUC2* mRNAs is translated into amino acids. The A residue in the ATG codon which initiates translation of the precursor to secreted invertase (see Discussion) has been assigned position +1. The reading frame is open through position +1596.

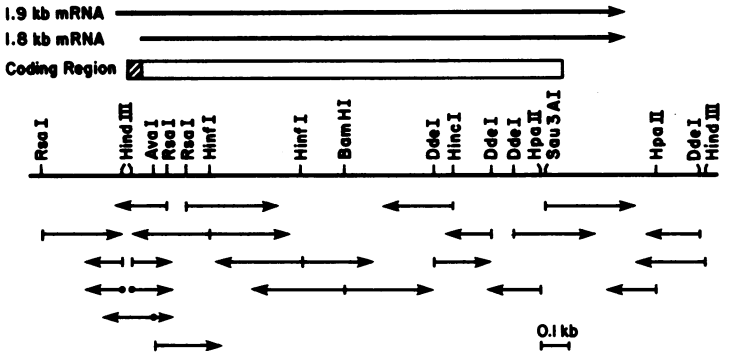

Figure 1. Sequencing strategy. The map of the SUC2 gene shows only those restriction sites used for sequence analysis (other sites are not shown). Arrows indicate the direction and extent of the nucleotide sequence read from each 5' end-labeled site. Those arrows with a circular head indicate 3' end-labeled sites analysed in a previous study (4). The open bar represents the coding region; the signal sequence is hatched. The 1.9 kb and 1.8 kb mRNAs are indicated by bold arrows pointing in the direction of the 3' end.

DISCUSSION

Coding region

From the nucleotide sequence of the SUC2 gene we can infer the amino acid sequences of intracellular invertase, secreted invertase, and the signal peptide-containing precursor to secreted invertase. We showed in a previous study that the coding regions for the precursor to secreted invertase and for the intracellular invertase begin at different points within the SUC2 gene but share the same reading frame (4). The initiator methionine codon for translation of the precursor to secreted invertase from the 1.9 kb mRNA was identified as the ATG codon at position +1. The initiation codon for translation of the intracellular enzyme from the 1.8 kb mRNA was identified as the ATG at position +61.

In this study we have shown that the reading frame for the invertase gene is open through position +1596, where it is terminated by a nonsense codon (TAG) at position +1597. A second nonsense codon (TAA) in the same frame follows at position +1606. A contiguous coding region was expected because no intervening sequence was detected in the SUC2 gene by S1 mapping and primer

extension analysis of the mRNAs (3,4).

From the translation of the nucleotide sequence, we infer that the precursor to secreted invertase is a polypeptide of 532 amino acids. Its calculated molecular weight is 60,572. Cleavage of the 19 amino acid signal peptide (4,18) from the precursor (see Figure 2) is then predicted to yield a mature protein of 513 amino acids and a calculated molecular weight of 58,567. The coding region for the intracellular invertase includes the 512 codons between nucleotides +61 and +1596 and would be translated into a polypeptide of calculated molecular weight 58,480. These values agree well with studies of the intracellular and secreted enzymes indicating that the protein in each subunit has a molecular weight of about 60,000 (6,7,9).

Amino acid sequence data derived from direct analysis of the protein is available only for the amino and carboxyl termini of invertase. The amino acid sequence predicted by nucleotides +1 to +120 has been compared previously to partial amino acid sequences of the amino termini of invertase in vitro translation products (18) and shows good agreement (4). Trimble and Maley used carboxypeptidases A and B to determine the following sequence for the carboxyl terminus of secreted invertase: Tyr-Val-Arg-Phe-(Ser or Gln)-Glu-Val-Lys-COOH (9). The translation of the 3' end of the coding region of our nucleotide sequence is:

Lys-Phe-Gln-Val-Arg-Glu-Val-Lys-COOH.

These two sequences are identical for the three carboxyl terminal residues and they also agree with respect to the amino acid composition, although not the order, of the next four residues. This result confirms that we have correctly identified the end of the coding region at position +1596.

The amino acid composition calculated for the mature secreted enzyme is shown in Table 1. The most striking feature is the high levels of aspartic and glutamic acids, which constitute 62 of the 513 amino acid residues. The predicted composition is in good accord with the empirically determined composition of the purified enzyme (6). The major discrepancy is that the previously reported levels of aspartic and glutamic acids are

TTTGCCTATTACCAT

-300

CATAGAGACGTTTCTTTTCGAGGAATGCTTAAACGACTTTGTTTGACAAAAATGTTGCCTAAGGGCTCTATAGTAAACCATTTGGAAGAAAGATT

-200

TGACGACTTTTTTTTTTTGGATTTCGATCCTATAATCCTTCCTCCTGAAAAGAAACATATAAATAGATATGTATTATTCTTCAAAACATTCTCTT

-100

*

GTTCTTGTGCTTTTTTTTTACCATATATCTTACTTTTTTTTTTCTCTCAGAGAAACAAGCAAAACAAAAAGCTTTTCTTTTCACTAACGTATATG

-1

met leu leu gln ala phe leu phe leu leu ala gly phe ala ala lys ile ser ala ser met thr asn glu

ATG CTT TTG CAA GCT TTC CTT TTC CTT TTG GCT GGT TTT GCA GCC AAA ATA TCT GCA TCA ATG ACA AAC GAA

1

thr ser asp arg pro leu val his phe thr pro asn lys gly trp met asn asp pro asn gly leu trp tyr

ACT AGC GAT AGA CCT TTG GTC CAC TTC ACA CCC AAC AAG GGC TGG ATG AAT GAC CCA AAT GGG TTG TGG TAC

100

asp glu lys asp ala lys trp his leu tyr phe gln tyr asn pro asn asp thr val trp gly thr pro leu

GAT GAA AAA GAT GCC AAA TGG CAT CTG TAC TTT CAA TAC AAC CCA AAT GAC ACC GTA TGG GGT ACG CCA TTG

200

phe trp gly his ala thr ser asp asp leu thr asn trp glu asp gln pro ile ala ile ala pro lys arg

TTT TGG GGC CAT GCT ACT TCC GAT GAT TTG ACT AAT TGG GAA GAT CAA CCC ATT GCT ATC GCT CCC AAG CGT

asn asp ser gly ala phe ser gly ser met val val asp tyr asn asn thr ser gly phe phe asn asp thr

AAC GAT TCA GGT GCT TTC TCT GGC TCC ATG GTG GTT GAT TAC AAC AAC ACG AGT GGG TTT TTC AAT GAT ACT

300

ile asp pro arg gln arg cys val ala ile trp thr tyr asn thr pro glu ser glu glu gln tyr ile ser

ATT GAT CCA AGA CAA AGA TGC GTT GCG ATT TGG ACT TAT AAC ACT CCT GAA AGT GAA GAG CAA TAC ATT AGC

400

tyr ser leu asp gly gly tyr thr phe thr glu tyr gln lys asn pro val leu ala ala asn ser thr gln

TAT TCT CTT GAT GGT GGT TAC ACT TTT ACT GAA TAC CAA AAG AAC CCT GTT TTA GCT GCC AAC TCC ACT CAA

500

phe arg asp pro lys val phe trp tyr glu pro ser gln lys trp ile met thr ala ala lys ser gln asp

TTC AGA GAT CCA AAG GTG TTC TGG TAT GAA CCT TCT CAA AAA TGG ATT ATG ACG GCT GCC AAA TCA CAA GAC

tyr lys ile glu ile tyr ser ser asp asp leu lys ser trp lys leu glu ser ala phe ala asn glu gly

TAC AAA ATT GAA ATT TAC TCC TCT GAT GAC TTG AAG TCC TGG AAG CTA GAA TCT GCA TTT GCC AAC GAA GGT

600

phe leu gly tyr gln tyr glu cys pro gly leu ile glu val pro thr glu gln asp pro ser lys ser tyr

TTC TTA GGC TAC CAA TAC GAA TGT CCA GGT TTG ATT GAA GTC CCA ACT GAG CAA GAT CCT TCC AAA TCT TAT

700

trp val met phe ile ser ile asn pro gly ala pro ala gly gly ser phe asn gln tyr phe val gly ser

TGG GTC ATG TTT ATT TCT ATC AAC CCA GGT GCA CCT GCT GGC GGT TCC TTC AAC CAA TAT TTT GTT GGA TCC

phe asn gly thr his phe glu ala phe asp asn gln ser arg val val asp phe gly lys asp tyr tyr ala

TTC AAT GGT ACT CAT TTT GAA GCG TTT GAC AAT CAA TCT AGA GTG GTA GAT TTT GGT AAG GAC TAC TAT GCC

800

leu gln thr phe phe asn thr asp pro thr tyr gly ser ala leu gly ile ala trp ala ser asn trp glu

TTG CAA ACT TTC TTC AAC ACT GAC CCA ACC TAC GGT TCA GCA TTA GGT ATT GCC TGG GCT TCA AAC TGG GAG

900

tyr ser ala phe val pro thr asn pro trp arg ser ser met ser leu val arg lys phe ser leu asn thr

TAC AGT GCC TTT GTC CCA ACT AAC CCA TGG AGA TCA TCC ATG TCT TTG GTC CGC AAG TTT TCT TTG AAC ACT

1000

glu tyr gln ala asn pro glu thr glu leu ile asn leu lys ala glu pro ile leu asn ile ser asn ala

GAA TAT CAA GCT AAT CCA GAG ACT GAA TTG ATC AAT TTG AAA GCC GAA CCA ATA TTG AAC ATT AGT AAT GCT

```
gly pro trp ser arg phe ala thr asn thr thr leu thr lys ala asn ser tyr asn val asp leu ser asn
GGT CCC TGG TCT CGT TTT GCT ACT AAC ACA ACT CTA ACT AAG GCC AAT TCT TAC AAT GTC GAT TTG AGC AAC
                      1100

ser thr gly thr leu glu phe glu leu val tyr ala val asn thr thr gln thr ile ser lys ser val phe
TCG ACT GGT ACC CTA GAG TTT GAG TTG GTT TAC GCT GTT AAC ACC ACA CAA ACC ATA TCC AAA TCC GTC TTT
                                                  1200

ala asp leu ser leu trp phe lys gly leu glu asp pro glu glu tyr leu arg met gly phe glu val ser
GCC GAC TTA TCA CTT TGG TTC AAG GGT TTA GAA GAT CCT GAA GAA TAT TTG AGA ATG GGT TTT GAA GTC AGT

ala ser ser phe phe leu asp arg gly asn ser lys val lys phe val lys glu asn pro tyr phe thr asn
GCT TCT TCC TTC TTT TTG GAC CGT GGT AAC TCT AAG GTC AAG TTT GTC AAG GAG AAC CCA TAT TTC ACA AAC
 1300

arg met ser val asn asn gln pro phe lys ser glu asn asp leu ser tyr tyr lys val tyr gly leu leu
AGA ATG TCT GTC AAC AAC CAA CCA TTC AAG TCT GAG AAC GAC CTA AGT TAC TAT AAA GTG TAC GGC CTA CTG
                                    1400

asp gln asn ile leu glu leu tyr phe asn asp gly asp val val ser thr asn thr tyr phe met thr thr
GAT CAA AAC ATC TTG GAA TTG TAC TTC AAC GAT GGA GAT GTG GTT TCT ACA AAT ACC TAC TTC ATG ACC ACC
                                                      1500

gly asn ala leu gly ser val asn met thr thr gly val asp asn leu phe tyr ile asp lys phe gln val
GGT AAC GCT CTA GGA TCT GTG AAC ATG ACC ACT GGT GTC GAT AAT TTG TTC TAC ATT GAC AAG TTC CAA GTA

arg glu val lys *** ***
AGG GAA GTA AAA TAGAGGTTATAAAACTTATTGTCTTTTTTATTTTTTTCAAAAGCCATTCTAAAGGGCTTTAGCAACGAGTGACGAATGT
            1600

AAAACTTTATGATTTCAAAGAATAACCTCCAAACCATTGAAAATGTATTTTTATTTTTATTTTCTCCCGACCCCAGTTACCTGGAATTTGTTCTT
                    1700

TATGTACTTTATATAAGTATAATTCTCTTAAAAATTTTTACTACTTTGCAATAGACATCATTTTTTCACGTAATAAACCCACAATCGTAATGTAG
                          1800

TTGCCTTACACTACTAGGATGGACCTTTTTGCCTTTATCTGTTTTGTTACTGACACAATGAAACCGGGTAAAGTATTAGTTATGTGAAAATTTAA
                                1900

AAGCATTAAGTAGAAGTATACCATATTGTAAAAAAAAAAAAAGCGTTGTCTTCTACGTAAAAGTGTTCTCAAAAAGAAGTAGTGAGGGAAATGGAT
                                      2000

ACCAAGCTATCTGTAACAGGAGCTAAAAAATCTC
```

Figure 2. Nucleotide sequence and amino acid translation of the SUC2 gene. The complete sequence of a contiguous 2389 nucleotide region encompassing the SUC2 structural gene is shown. The sequence of the mRNA-identical strand of the $SUC2^{+}$ gene is presented; the suc2-215 amber allele has a mutation at position +140 which produces an amber codon (4). The open reading frame encoded by the 1.9 kb mRNA is translated into amino acids. TATA-like sequences preceding the coding region are underlined and the approximate position of the 5' end of the 1.9 kb mRNA is marked with an asterisk (4). The two termination codons immediately following the coding region are also marked with asterisks. The site for cleavage of the signal peptide is indicated by a vertical arrow (4,18). Tripeptide sequences which may serve as glycosylation sites are boxed. The putative recognition sequence for poly(A) addition, TATATAAG, is underlined twice. The sequence contains the 6 bp recognition sites of the following enzymes: HindIII (at positions -27, +11), BamHI (+787), XbaI (+829), BclI (+1037), KpnI (+1159) and HpaI (+1189).

Table 1. Amino acid composition of secreted invertase

Amino Acid	Number of residues	
	Predicted	Reported
Aspartic acid	32	77
Asparagine	44	--
Threonine	39	36
Serine	47	49
Glutamic acid	30	50
Glutamine	18	--
Proline	26	28
Glycine	30	31
Alanine	29	30
Cysteine	2	2
Valine	29	30
Methionine	10	9
Isoleucine	18	17
Leucine	34	36
Tyrosine	31	28
Phenylalanine	36	35
Tryptophan	16	14
Lysine	25	26
Histidine	4	7
Arginine	13	12

The values predicted from the nucleotide sequence are compared with those reported for the purified enzyme by Neumann and Lampen (6). The latter values have been recalculated for a molecular weight of 58,567.

higher than predicted by the nucleotide sequence. This could be attributed to breakdown of asparagine and glutamine.

Glycosylation sites on secreted invertase

Secreted invertase is a glycoprotein carrying approximately nine asparagine-linked polysaccharide chains per subunit (9,10). Glycosylation of eukaryotic proteins occurs at the tripeptide sequences asparagine-X-threonine and asparagine-X-serine, where X may be any amino acid residue except possibly aspartate (19,20). The amino acid sequence of invertase includes 13 such tripeptide sequences, three of which contain aspartate as the middle residue

Table 2. Codon usage in the SUC2 gene

	U	C	A	G	
U	Phe 19	Ser 19	Tyr 10	Cys 1	U
	20	12	21	1	C
	Leu 5	7			A
	21	1		Trp 16	G
C	Leu 5	Pro 7	His 3	Arg 3	U
	0	4	1	1	C
	6	15	Gln 19	0	A
	2	0	0	0	G
A	Ile 12	Thr 21	Asn 14	Ser 6	U
	4	9	30	3	C
	3	6	Lys 11	Arg 8	A
	Met 11	3	15	1	G
G	Val 7	Ala 16	Asp 21	Gly 20	U
	12	11	11	6	C
	4	5	Glu 22	3	A
	6	2	8	2	G

The first position of the codon is indicated at the left, the second position at the top, and the third position at the right. The number indicates the utilization of each codon in the coding region for the precursor to secreted invertase.

(boxed in Figure 2). In other glycoproteins, not all asparagine-X-threonine/serine acceptor sites are glycosylated (19,20), and it remains to be determined which of the asparagine residues in invertase are actually glycosylated.

Codon usage

Codon usage in the SUC2 gene is summarized in Table 2. A bias is evident in the codon representation for several amino acids; for example, the codon CAA for glutamine is used 19 times and the codon CAG is not used. The codon bias in the SUC2 gene follows the same pattern observed for other yeast genes by

Bennetzen and Hall (21), but the degree of bias is relatively low. Of the 61 possible codons, 56 are used in the SUC2 gene; in contrast only 33 codons are used in the gene for alcohol dehydrogenase I (22), and 29 codons in the two genes for glyceraldehyde-3-phosphate dehydrogenase (23). Bennetzen and Hall note a correlation between the degree of codon bias and the level of gene expression (22); with respect to degree of bias, the SUC2 gene resembles the less highly expressed yeast genes.

Sequences preceding and following the coding region

The sequence reported includes 300 nucleotides preceding the coding region and 493 nucleotides following the coding region. The base composition of these noncoding sequences is more AT-rich (70%) than the coding region (58%); such AT-rich noncoding sequences have been found adjacent to most of the yeast genes which have been sequenced (23-28). The 5' ends of the 1.9 kb and 1.8 kb mRNAs have been mapped and TATA-like sequences and other features of the 5' noncoding region have been described previously (4; see Figure 2). The 3' ends of the 1.9 kb and 1.8 kb mRNAs were mapped by S1 nuclease protection analysis to a position 1.03 kb downstream from the BamHI site (3). The sequence reported here shows that the relevant BamHI cleavage site is between nucleotides +787 and +788. The 3' ends therefore map approximately at position +1817 and the mRNAs contain about 220 nucleotides of 3' noncoding sequence. The sequence TATATAAG starts at position +1780 (see Figure 2). This sequence closely resembles the consensus sequence TAAATAAA/G, which is often found 28 to 33 nucleotides before the poly(A) addition site in yeast genes (22). The sequence TATGT is found at position +1771; Zaret and Sherman suggested that the sequence TATGT or TAGT, located 10 to 40 bp before the poly(A) site, may have a role in transcription termination and polyadenylation in yeast (29).

ACKNOWLEDGMENTS

This work was supported by a grant (NP-358) from The American Cancer Society to M.C. We thank Lois Purcell for typing the manuscript.

REFERENCES

1. Carlson, M., Osmond, B.C., and Botstein, D. (1980) Cold Spring Harbor Symp. Quant. Biol. 45, 799-803.
2. Mortimer, R.K., and Hawthorne, D.C. (1969) in The Yeasts, Rose, A.H. and Harrison, J.S., Eds., Vol. 1, pp. 385-460, Academic Press, New York.
3. Carlson, M., and Botstein, D. (1982) Cell 28, 145-54.
4. Carlson, M., Taussig, R., Kustu, S., and Botstein, D. (1983) Mol. Cell. Biol., in press.
5. Lampen, J.O. (1971) in The Enzymes, Boyer, P.D., Ed., Vol. 5, pp. 291-305, Academic Press, New York.
6. Neumann, N.P. and Lampen, J.O. (1967) Biochemistry 6, 468-475.
7. Gascon, S, and Lampen, J.O. (1968) J. Biol. Chem. 243, 1567-1572.
8. Gascon, S., Neumann, N.P., and Lampen, J.O. (1968) J. Biol. Chem. 243, 1573-1577.
9. Trimble, R.B. and Maley, F. (1977) J. Biol. Chem. 252, 4409-4412.
10. Tarentino, A.L., Plummer, T.H., Jr., and Maley, F. (1974) J. Biol. Chem. 249, 818-824.
11. Lehle, L., Cohen, R.E., and Ballou, C.E. (1979) J. Biol. Chem. 254, 12209-12218.
12. Esmon, B., Novick, P., and Schekman, R. (1981) Cell 25, 451-460.
13. Novick, P., Ferro, S., and Schekman, R. (1981) Cell 25, 461-469.
14. Maxam, A.M. and Gilbert, W. (1980) in Methods Enzymol. Grossman, L. and Moldave, K. Eds., Vol. 65, pp. 499-560, Academic Press, New York.
15. Smith, D.R. and Calvo, J.M. (1980) Nucl. Acids Res. 8, 2255-2273.
16. Staden, R. (1977) Nucl. Acids Res. 4, 4037-4051.
17. Carlson, M., Osmond, B.C., and Botstein, D. (1981) Genetics 98, 25-40.
18. Perlman, D., Halvorson, H.O., and Cannon, L.E. (1982) Proc. Natl. Acad. Sci. 79, 781-785.
19. Struck, D.K. and Lennarz, W.J. (1980) in The Biochemistry of Glycoproteins and Proteoglycans, Lennarz, W.J. Ed., pp. 35-83, Plenum Press, New York.
20. Hubbard, S.C. and Ivatt, R.J. (1981) Ann. Rev. Biochem. 50, 555-583.
21. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3026-3031.
22. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3018-3025.
23. Holland, J.P. and Holland, M.J. (1980) J. Biol. Chem. 255, 2596-2605.
24. Smith, M., Leung, D.W., Gillam, S., Astell, C.R., Montgomery, D.L. and Hall, B.D (1979) Cell 16, 753-761.
25. Montgomery, D.L., Leung, D.W., Smith, M., Shalit, P., Faye, G. and Hall, B.D. (1980) Proc. Natl. Acad. Sci. USA. 77, 541-545.

26. Gallwitz, D. and Sures, I. (1980) Proc. Natl. Acad. Sci. USA. 77, 2546-2550.
27. Ng, R. and Abelson, J. (1980) Proc. Natl. Acad. Sci. USA. 77, 3912-3916.
28. Holland, M.J., Holland, J.P., Thill, G.P. and Jackson, K.A. (1981) J. Biol. Chem. 256, 1385-1395.
29. Zaret, K.S. and Sherman, F. (1982) Cell 28, 563-573.