Nucleotide sequence of the galactose gene cluster of *Kluyveromyces lactis*

Thomas D. Webster⁺ and Robert C. Dickson^{*}

Department of Biochemistry, College of Medicine, University of Kentucky, Lexington, KY 40536-0084, USA
Submitted March 2, 1988

Accession no. X07039

The GAL gene cluster of *Kluyveromyces lactis* was isolated by complementation of a mutant strain (1). The sequence presented corresponds to the coding strand of GAL10 and GAL7 and the non-coding strand of GAL1 and Gene X. The GAL1 coding region (kinase) begins at nucleotide 552 and extends leftward to nucleotide 1 (not all of the gene sequence was determined). Transcription start sites were mapped (1) and are indicated by asterisks; for major sites the asterisk is underlined. Transcript termination sites are indicated by dots below the sequence. Potential TATA sequences are underlined. Upstream activator sequences (UAS) are overlined

```

10      20      30      40      50
ATT CAT CCT TGA GGA TTG GCT TCC TGC AGG AAC ATT AGA GTA TAT AGT TGC CAA
Asn Met Arg Ser Ser Gln Ser Gly Ala Pro Val Asn Ser Tyr Ile Thr Ala Leu

60      70      80      90      100
AGA AAC CGC ACA GAT GAA GGC AGC ACT TGA AGA CAA ACC ACC GCT AGG AAT
Ser Val Ala Cys Ile Phe Ala Ile Ser Ser Leu Gly Gly Gly Ser Pro Ile

110     120     130     140     150     160
GTC ACC CTT GAC GTA GAT TTC CAT ACC ATG AAC TGG ACC CTT GAA GTT GTA TTT
Asp Gly Lys Val Tyr Ile Glu Met Gly His Val Pro Gly Lys Phe Asn Tyr Lys

170     180     190     200     210
CTC TTG TAA GAA TTG TTG AAC GAC TAG GAG GCC ACA CTT GAA ATA ATT GGA CCA
Glu Gln Leu Phe Gln Gln Val Val Leu Leu Gly Cys Lys Phe Tyr Asn Ser Val

220     230     240     250     260     270
GTC TGA CAC TGA TGG GTC AAT TTC AAT CAA GGA TCC ATC CAA TGG TAA GTC GAA
Asp Ser Lys Val Tyr Pro Asp Ile Glu Ile Leu Ser Gly Asp Leu Pro Leu Asp Phe

280     290     300     310     320
CTT ACG CTG AGC AAA ATT GCT ATC AGT ATT AGT CAA AGT AAT GGA TGG ATT CTC
Lys Arg Gln Ala Phe Asn Ser Asp His Asn Thr Leu Thr Ile Ser Pro Asn Glu

330     340     350     360     370
GCT TTC AGA AGT TAA GCG GCA CGC TAG CAT CAA ATC GTT TTC ACT AGC CAT TGG
Ser Gly Ser Thr Leu Arg Cys Ala Leu Met Leu Asp Asn Glu Ser Ala Met Pro

380     390     400     410     420     430
CAC ATG AAA CTG ACA ATA ATC GAT ATG TTC ACC AAT CAA ATT AAC TCT ACC AGG
Val His Phe Gln Cys Tyr Ile His Glu Gly Ile Leu Asn Val Arg Gly Pro

440     450     460     470     480
AGA TCT GGT TAT AAA AAA CTT CCT GGA ATC ATC TGT ATT ATA ACA TTC CAC AAA
Ser Arg Thr Ile Phe Phe Lys Arg Ser Asp Asp Thr Asp Tyr Cys Glu Val Phe

490     500     510     520     530     540
TTC ATT CTT CAA TTG TTC CAA TTT TGG AGC GTG AGC AAA ATT TGT TGG CAC AAT
Glu Asn Lys Leu Gln Glu Leu Lys Pro Ala His Ala Phe Asp Thr Pro Val Ile

550     562     572     582     592     602
AGG AAC GGA CAT TATTATTCT TATATTTCAG CCTGTGTTCA AACCTGTGT CATTTAAAG
Pro Val Ser Met

612     622     632     642     652     662     672
GATTTCAGG TCAAAATTTCC AACTTTATAG GTTGAATAT ATTCTCACCC GTCTATTGAA AGATCAACC
**

682     692     702     712     722     732     742
TGGTATCAAG ATATGTTCTCG AACTGTCAAA CTTTITTAAT GTAATTTCAC TTCAATAGAA CGGTTCTTT
*

752     762     772     782     792     802     812
ATAGACCTTT GGAAGCTCG TCATCATTTT AACAGACGG CTACCGAGTT TCTTTCGAT CACTTCAATC
UAS

822     832     842     852     862     872     882
CGATACAGT GATACATACG GATTGACGA ATCCGCTCTA CGGAATACAC CAATCCGAA TGAGCGGAG
UAS

892     902     912     922     932     942     952
ACTATGCCCC GGATGCCCCA GATTITTTGT TTCTGACGAG GCAAGCTACC GACCGTACTA TTGTAACGA

962     972     982     992     1002    1012    1022
GAGATGAGTT AGGAATTAAA CACCGTGAAA GAAGGAAGA AACAGAAAT ACTAATAGT TGTCTATCA

1032    1042    1052    1062    1072    1082    1092
ACACGACAA CTGCTTGCTT TGCACCTTGA AGTGACATTG AGGATCAAT CCCTTTACT TAAACACGT

1102    1112    1122    1132    1142    1152    1162
ATGATTGCT GCTCAATAG TATATAACA CTGAGTCCAG GCAATTGAAT CCAATAAGA ACAGAAGGCA

1172    1182    1192    1202    1212    1222    1232
AAACAGTAA ACTGCTTCC TTCAATATA TTGATATAAT AGACTACAT TTGAAGATA TCTCTATAG
*

1242    1252    1262    1272    1282    1292    1302
AAGCTGTCA TAAGAAGAC CAATAATCTG AATCAGGTTA AAGGAACCT CTGAGAGCA AAAAAA
*

1312
AAATCAACG CTAAG ATG TCT
*

1332    1347    1362    1377
GAA GAT AAA TAC TGT TTG GTG ACT GGT GGT GCT GGT TAC ATT GGG TCG CAT ACT
Glu Asp Lys Tyr Cys Leu Val Thr Gly Gly Ala Gly Tyr Ile Gly Ser His Thr

1392    1407    1422
GTT GTT GAA CTT TGT GAA GCT GGT TAC TGT ATT GTT GTT GAC AAC TTG TCT
Val Val Glu Leu Cys Glu Ala Gly Tyr Lys Cys Ile Val Val Asp Asn Leu Ser

1437    1452    1467    1482
AAT TCC AGT TAT GAA TCT GTC GCT CGC ATG GAA TTG TTG ACT GGC CAG GAG ATC
Asn Ser Ser Tyr Glu Ser Val Ala Arg MET Glu Leu Leu Thr Gly Gln Glu Ile

1497    1512    1527
AAG TTC GCA ARG ATT GAT TTA TGC GAA CTG GAA CCA TTG AAC AAG TTG TTC GAT
Lys Phe Ala Lys Ile Asp Leu Cys Glu Leu Glu Pro Leu Asn Lys Leu Phe Asp

1542    1557    1572    1587
GAT TAC AAG ATC GAT TCC GTT CTA CAT TTT GCC GGT TTG AAA GCG GTT GGT GAA
Asp Tyr Lys Ile Asp Ser Val Leu His Phe Ala Gly Leu Lys Ala Val Gly Glu

1602    1617    1632    1647
AGT ACT CAG ATT CCA TTG ACT TAT TAT TTC AAC AAT ACT GTT GGT ACT ATC AAC
Ser Thr Gln Ile Pro Leu Thr Tyr Tyr Phe Asn Asn Ile Val Gly Thr Ile Asn

1662    1677    1692
CTT CTA GAA TGT ATG AAG AGT CAC GAT GTA AAG AAG TTA GTC TTT TCC TCT TCT
Leu Leu Glu Cys MET Lys Ser His Asp Val Lys Lys Leu Val Phe Ser Ser Ser

1707    1722    1737    1752
GCC ACT GTT TAC GGT GAT GCC ACT AGA TTC GAA AAT ATG ATT CCA ATT CCA GAG
Ala Thr Val Tyr Gly Asp Ala Thr Arg Phe Glu Asn MET Ile Pro Ile Pro Glu

1767    1782    1797
ACC TGT CCA ACA GGT CCA ACT AAC CCA TAT GGT AAG ACC AAG TTG ACC ATT GAA
Thr Cys Pro Thr Gly Pro Thr Asn Pro Tyr Gly Lys Thr Lys Leu Thr Ile Glu

1812    1827    1842    1857
GAT ATG ATG AGA GAT CTA CAT TTC TCT GAT AAG TCT TCC TTT GCT ACT TTA
Asp MET MET Arg Asp Leu His Phe Ser Asp Lys Ser Phe Phe Ala Ile Leu

1872    1887    1902    1917
AGA TAT TTC AAT CCT ATT GGT GCT CAT CCA TCT GGT GTC ATC GGT GAA GCC TTT
Arg Tyr Phe Asn Pro Ile Gly Ala His Pro Ser Gly Val Ile Gly Glu Ala Phe

1932    1947    1962
GGT ATT CCA AAT AAC TTG TTA CCA TTC ATG GCT CCA GTT GCT ATT GGA AGA AGA
Gly Ile Pro Asn Asn Leu Leu Pro Phe MET Ala Gln Val Ala Ile Gly Arg Arg
```

[illegible][illegible]

5303	5318										5333										5348										5711										5721										5731										5741										5751										5761										5771																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
CAA	GCA	GCA	GAT	AGA	CTC	AAA	GCC	CTA	AGT	GGT	GAA	GTA	CAT	TAT	TTA	GCA	AAA											TACCC	TTAAT	GCGCA	CTT	TTGGCTATGA	CATCAT	GTGG	ATAAGABTCT	TTTGTGTTCT	CTAATACCC																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
Gln	Ala	Ala	Asp	Arg	Leu	Lys	Ala	Leu	Ser	Gly	Glu	Val	His	Tyr	Leu	Ala	Lys																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										</

Acknowledgements

This work was supported by Public Health Service grant GM-22749 from the National Institutes of Health and by Biomedical Research Support Grant No. RR05374 from the Biomedical Research Support Branch, Division of Research Facilities and Resources, NIH. T.D.W. was supported in part by a fellowship from the Mason and Hanger Corporation.

*To whom correspondence should be addressed

+Present address: Institute for Molecular Genetics, T813, Baylor College of Medicine, Texas Medical Center, Houston, TX 77030, USA

References

I) Webster, T. D. and Dickson, R. C. (1988) Nucl. Acids Res. 16,8011-8028