Volume 15 Number 19 1987 Nucleic Acids Research

The nucleotide sequence of the *gal* T gene of *Escherichia coli*

T.L.Cornwell, S.L.Adhya[1], W.S.Reznikoff and P.A.Frey

Institute for Enzyme Research, Graduate School, and Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin, Madison, WI 53705 and [1]Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
Submitted August 4, 1987

Lemaire and Muller-Hill recently reported the nucleotide sequence of the gal T gene of Escherichia coli and deduced the amino acid sequence of galactose 1-phosphate uridylyltransferase (EC 2.7.7.12) (1). We determined the nucleotide sequence of this gene by application of the Sanger dideoxy method to EcoRV, Sau3A, RsaI and PvuII restriction fragments. Approximately 80% of the sequence was determined in both strands, and the remaining sequences determined in single strands were confirmed by partial sequences published earlier (2-4). The sequence we obtained was similar to that reported by Lemaire and Muller-Hill, with the exception of the GC-rich region between nucleotides 79 and 114. The sequencing gels for this region were ambiguous owing to band compression caused by extensive secondary structure. We overcame the ambiguities by using dITP in place of dGTP in the sequencing reactions. This procedure clarified the sequencing gels and allowed the following nucleotide sequence to be determined in both strands. Comparison with the sequence of Lemaire and Muller-Hill (1)

```
(This work, 79--->114)  CACCGCGCTAAGCGCCCCTGGCAGGGGGCGCAGGAA
(Ref. 1, 79--->111)     CACCGC-CTAAGC--CCCTGGCAGGGGGCGCAGGAA
```

showed that an alignment could be achieved by assuming deletions of G at position 85 and GC at positions 92-93 in the Lemaire and Muller-Hill sequence. Our translated amino acid sequence also differred as follows.

```
(This work, aa's 27-38)    HisArgAlaLysArgProTrpGlnGlyAlaGlnGlu
(Ref 1, aa's 27-37)        HisArgLeuSer---ProTrpGlnGlyAlaGlnGlu
```

Our sequence analysis indicates that gal T consists of 1047 bp and codes for a 348 amino acid protein. This compares with the 1044 bp gene and 347 amino acid protein reported by Lemaire and Muller-Hill (1).

ACKNOWLEDGEMENT This work was supported by Grants GM 31343 and GM 30480 (P.A.F.) and GM 19670 (W.S.R.) from the National Institute of General Medical Sciences.

REFERENCES

1. Lemaire, H.G. and Muller-Hill, B. (1986) Nucleic Acids Res. 14, 7705-7712.
2. Grindley, N. (1978) Cell 13, 419-426.
3. Ahmed, A. (1984) J. Mol. Biol. 178, 941-948.
4. Debouck, C., Roccio, A., Schumperli, D., McKenny, L., Jeffers, J., Hughes, C. and Rosenberg, M. (1981) Nucleic Acids Res. 12, 539-549.

