Complete nucleotide sequence of an *E.coli* IS5 insertion element containing an internal 88 base pair direct repeat (IS5-D)

David M.Reinitz, Jill A.Inverso and John M.Mansfield

Department of Veterinary Science, University of Wisconsin, Madison, WI 53706, USA

Submitted April 3, 1989

EMBL accession no. X13668

An unusual *E. coli* insertion sequence which transposed into plasmid pGEM-3Z (Promega Corp.) during propagation in strain DH5 α was analyzed¹, found to possess an 88 Bp direct repeat (bases 147 to 235, underlined below) and to be otherwise homologous to initial isolates of IS5². This 1291 Bp sequence transposed into a CTAG site (bases 34 to 37 in the plasmid polylinker) which was duplicated to flank the element. Duplicated regions included the promoter (IS5:157-203) and transcription initiation site (GTG at IS5:209) for the IS5C gene³ and on the complement of the strand shown below, the last 27 codons and the terminator stem and loop structure of the IS5A gene⁴. The duplication was predicted¹ to form a stem and loop structure with 23 of 88 bases paired and an overall free energy of -17.8 Kcals/mole. Similar analyses of the entire RNA sequence, with and without the direct repeat, yielded grossly different folding patterns, however, with similar free energies of -438.6 and -414.7 Kcals/mole, respectively. In addition, the 88 Bp repeat was flanked by a run of 5 thymine residues. These bases or secondary structural elements may have been involved in the duplication process and represent functional elements in IS5 transposition.

```

0001 CTAGGGAAGGTGCGAATAAGCGGGGAAATTCTTCTCGGCTGACTCAGTCATTTTCATTTCT
0061 TCATGTTTGAGCCGATTTTTTCTCCGTAATGCCTTGAATCAGCCTATTTAGACCGTTT
0121 CTTCCGCCATTTAAGGCGTTATCCCCAGTTTTTGTAGTGAATCTCTCCCACTGACGTATCAT
0181 TTGGTCCGCCCCGAAACAGGTTGGCCAGCGTGAATAACATCGCCAGTTGGTTATCGTTTTT
0241 AGTGAGATCTCTCCCACTGACGTATCATTGGTCCGCCGAAACAGGTTGGCCAGCGTGA
0301 ATAACATCGCCAGTTGGTTATCGTTTTTTCAGCAACCCCTGTATCTGGCTTTTCACGAAGC
0361 CGAACTGTCGCTTGATGATGCGAAATGGGTGCTCCACCCTGGCCCGGATGCTGGCTTCA
0421 TGTATTCGATGTTGATGGCCGTTTGTTCCTTGCGTGGATGCTGTTTCAAGGTTCTTACCT
0481 TGCCGGGGCGCTCGGCGATCAGCCAGTCCACATCCACCCTCGGCCAGCTCCTCGCGCTGTG
0541 GCGCCCTTGGTAGCCGGCATCGGCTGAGACAAATTGCTCCTCTCCATGCAGCAGATTAC
0601 CCAGCTGATTGAGGTCATGCTCGTTGGCCGCGGTGGTGACCAGGCTGTGGGTCAGGCCAC
0661 TCTTGGCATCGACACCAATGTGGGCCTTCATGCCAAAGTGCCACTGATTGCCCTTCTTGG
0721 TCTGATGCATCTCCGGATCGCGTTGCTGCTCTTTGTTCTTGGTCGAGCTGGGTGCCCTCAA
0781 TGATGGTGGCATCGACCAAGGTCCTTGAGTCATCATGACGCCTGCTTCGGCCAGCCAGC
0841 GATTGATGGTCTTGAACAATTGGCGGGCCAGTTGATGCTGCCAGCAGGTGGCGGAAAT
0901 TCATGATGGTGGTGCAGTCCGGCAAGGCGCTATCCAGGGATAACCGGGCAACAGACGCA
0961 TGGAGGCGATTTCTGACAGAGCATCTTCCATCGCGCCATCGCTCAGGTTGTACCAATGCT
1021 GCATGCAGTGAATGCGTAGCATGGTTTCCAGCGGATAAGGTGCGCCGCCATTACCAGCCT
1081 TGGGGTAAACCGGCTCGATGACTTCCACCATTGTTTGCCATGGCAGAATCTGCTCCATGC
1141 GGGACAAGAAAATCTCTTTTCTGGTCTGACGGCGCTTACTGCTGAATCACTGTGCGCGA
1201 AGGTAAGTTGATGACTCATGATGAACCCTGTTCTATGGCTCCAGATGACAAACATGATCT
1261 CATATCAGGGACTTGTTCGCACCTTCCCTAG

```

Figure: Complete nucleotide sequence deduced from the *E. coli* IS5-D element with an 88Bp direct repeat (underlined).

Acknowledgements

Supported by grants from NIH to JMM (AI 22441-04) and DMR (1 F32 AI07876-01). We acknowledge the expert assistance of Brian Aizenstein in DNA sequencing.

References

1. Devereux, J., Haeberli, P., & Smithies, O., *Nucleic Acids Res.* 12:387 (1984).
2. Schoner, B., & Kahn, M., *Gene* 14:165 (1981).
3. Rak, B., & von Reutern, M., *EMBO J.* 3:807 (1984).
4. Kroger, M., & Hobom, G., *Nature* 297:159 (1982).