Location of the host attachment site for phage HPI within a cluster of *Haemophilus influenzae* tRNA genes

Michael A.Hauser and John J.Scocca
Department of Biochemistry, The Johns Hopkins University School of Hygiene and Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA

Submitted July 17, 1990 EMBL accession no. X53782

The phage (*attP*) and host (*attB*) sites for integration of HPI into the *H. influenzae* chromosome contain a 182-bp common segment, consisting of two identical blocks of 93 and 62 bp, separated by a 27-bp segment which contains six mismatches. Recombination occurs within the 62-bp block (1). This 182-bp segment contains sequences of one complete and one partial tRNA gene, $tRNA^{lys}$ (anticodon: UUU) and $tRNA^{leu}$ (UAA), respectively. HPI is one of a number of phages and plasmids which integrate within tRNA gene sequences (2, 3). To characterize this region more fully, the sequence of the 636-bp *Xmn*I-*Hin*dIII segment surrounding the HPI *attB* site was determined. This sequence extends 160 bp to the left and 294 bp to the right of the 182-bp core, and incorporates additions and corrections provided in Ref. 3. The sequence, shown in Fig. 1, contains a ca. 350-bp stretch organized into an apparent operon of three tRNA genes. A presumptive promoter, containing nearly perfect consensus −35 and −10 sequences (4), is followed by the three tRNA sequences, and the region ends with a presumptive terminator containing a stem-loop sequence and an oligo-T segment (5). The *H. influenzae* sequence encodes $tRNA^{gly}$ (GCC), $tRNA^{leu}$ (UAA), and $tRNA^{lys}$ (UUU). The conserved and semi-conserved residues diagnostic for tRNAs, the appropriate anticodons, and secondary structure are all present (6). Transcription from this region in either normal or lysogenic *H. influenzae* remains to be established.

The three *H. influenzae* tRNA gene sequences reflect different extents of evolutionary divergence from other bacterial tRNA sequences (7). *H. influenzae* $tRNA^{gly}$ (GCC) differs from the *E. coli* equivalent by a single residue (60 T → C) in the TψC loop. The *H. influenzae* and *E. coli* $tRNA^{lys}$ (UUU) sequences differ at eight positions, six of which are complementary changes which maintain hydrogen bonding in stem regions. No sequence for a $tRNA^{leu}$ (UAA) from *E. coli* has been reported. The *H. influenzae* sequence differs from the *E. coli* (CAA) species at 25 positions, one within the anticodon and nine in the variable loop. The $tRNA^{leu}$ (UAA) sequences of *H. influenzae* and *Bacillus subtilis* differ at 24 positions, nine of which are in the variable loop.

ACKNOWLEDGEMENT

Supported by American Cancer Society grant MV385A

REFERENCES

1. Waldman,A.S., Goodman,S.D. and Scocca,J.J. (1989) *J. Bacteriol.* **169**, 238–246.
2. Pierson III,L.S. and Kahn,M.L. (1987) *J. Mol. Biol.* **196**, 487–496.
3. Reiter,W.-D., Palm,P. and Yeats,S. (1989) *Nucl. Acids Res.* **17**, 1907–1914.
4. Hawley,D.K. and McClure,W.R. (1983) *Nucl. Acids Res.* **11**, 2237–2255.
5. Brendel,V., Hamm,G.H. and Trifonov,E.N. (1986) *J. Biomol. Struct. Dynam.* **3**, 705–723.
6. Rich,A. and RajBhandary,U.L. (1976) *Annu. Rev. Biochem.* **45**, 805–860.
7. Sprinzl,M., Hartmann,T., Meissner,F., Moll,J. and Vorderwuelbecke,T. (1987) *Nucl. Acids Res.* **15**, r53–r188.

```
GAAGTGATTCGTGAGTCTATTAAACGTTGGAATGAGCGTTAATTAATTAAAGACGTTTGATTGTAAAAGTCCCTGATTCT    80

TTGAGTCGGGGATTTTTTATATGAGTTTCAGAGATTTAGATAATAAAAAAACACCTTTCGGTGCTTACTTTATTTTTAAG   160
                                                TTTTTTTGTGGAAAGCCACgaatgaaataaaaattc

TGGTGGGTCGTGAAGGATTCGAACCTTCGACCAACGGATTAAAAGTCCGCTGCTCTACCGACTGAGCTAACGACCCAATT   240
ACCACCCAGCACTTCCTAAGCTTGGAAGCTGGTTGCCTAATTTTCAGGCGACGAGATGGCAGACTCGATTGCTGGGttaa
          . .          .    ..     .

AGATGATTTTAAAATAAATTTTAAAATCTTTGAATTGGTGTTTAAATGGTGCCCGAAGCCAGACTTGAACTGGCACGCCT   320
tctactaaaattttatttaaaatttagaaacttaaccacaaatttACCACGGGCTTCGGTCTGAACTTGACCGTGCGCA

CGAAAGGCGAGGGATTTTAAATCCCTTGTGTCTACCGATTCCACCACTCGGGCAAATTGGAGCGGGAAACGAGGCTCGAA   400
GCTTTCCGCTCCCTAAAATTTAGGGAACACAGATGGCTAAGGTGGTGAGCCCGtttaACCTCGCCCTTTGCTCCGAGCTT

CTCGCGACCCCGACCTTGGCAAGGTCGTGCTCTACCAACTGAGCTATTCCCGCATTTATCACTAATCAGCACTACTGCGT   480
GAGCGCTGGGGCTGGAACCGTTCCAGCACGAGATGGTTGACTCGATAAGGGCGtaaatagtgattagtcgtcatgacgca

GACAACGGGGCGTATTTTACGGATTTATATATTGCTGTCAAACGAAAAAAATAAAAAATATTTTATTTGATTAAAAATAA   560
ctgttG.......TAAAAT..............ACAGTT

CTCAAATATAGATGGTTATCATAAAAAATAACTGTGTTTTATTTAATTTTATAATTGAATATTATCTAGCAAGCTT   636
```

Figure 1. Nucleotide sequence of the *H. influenzae* HPI *attB* region. The numbering begins with the first residue of the *Xmn*I site, and the sequence is oriented as in Ref. 1, and is anticoding for the tRNA genes. The complementary strand of the proposed transcript is also shown. The promoter, probable transcriptional start, the tRNA sequences, and the terminator, are all in upper case. The anticodon triplets are underlined. The common core begins at residue 161 and ends at 342. Those residues which differ in the *attB* and *attP* cores are marked with superior dots; all these changes are transitions.