Sequence of a DNA injection gene from *Salmonella typhimurium* phage P22

Pratima Adhikari and Peter B. Berget
Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA

Received December 7, 1992; Accepted December 11, 1992

GenBank accession no. L07556

In the *Salmonella* phage P22, the presence of three structural proteins in the phage particle, gp7, gp16, and gp20 is required for successful phage DNA injection (1, 2). The genes coding for these proteins are contiguous, occupy about 4 kilobases in the late operon of P22 and are located between the *imml* region and genes which specify other capsid proteins. The specific function of any one of the DNA injection gene products remains obscure. The sequence of these genes will prove useful toward the study of DNA injection in this virus. The sequences of two of these genes, gene 7 and gene 16 have recently been reported (3, 4). Here we report the sequence of gene 20.

Gene 20 is flanked on its 5' side by gene 7 and by gene 16 on its 3' side. Sequencing of the gene was performed off of both strands, employing the dideoxy chain termination method (5). The sequence reveals a single 1416 base pair ORF which could code for a polypeptide of 472 amino acids with a predicted molecular weight of 50,069 daltons. The reading frame of this ORF was confirmed through the sequencing of two amber alleles, H1025 and H1032 which were identified as CAG (Gin) to TAG (amber) transition mutations (underlined in Figure 1) at codons 315 and 407 respectively. P22 particles prepared by growing the H1025 mutant under non-permissive conditions were analyzed by SDS/PAGE (not shown). These particles contain the normal complement of P22 structural proteins except for gp20 which is replaced by a polypeptide of the correct size to be the predicted 33,111 dalton amber fragment generated by this mutation. Thus as many as 158 amino acids can be missing from the carboxyl terminus of gp20 and result in a polypeptide that is still recognized for assembly into phage particles although it is not functional in DNA injection.

In the sequence that is shown in Figure 1, nucleotides 1 to 306 were previously published (4) and represent the end of gene 7 extending to nucleotide 276 of gene 20. Nucleotides 1067 to 1454 have been previously published (3) and extend from nucleotide 1037 in gene 20 (nucleotide 1067 in Figure 1) to nucleotide 9 in gene 16. There are three conflicts between our sequence and that of Umlauf and Dreiseikelmann (3) in the 3' region of gene 20. In Figure 1 the sequence GC at nucleotides 1199 and 1200 is reversed compared to their sequence; and the G residues at positions 1272 and 1402 are missing from their sequence. These differences are indicated in bold type face. We believe our sequence to be correct because this portion of the previously reported sequence was determined from only one strand (Figure 2, reference 3). There is a gap of 10 nucleotides between the stop codon of gene 7 and the start codon of gene 20. This gap contains a purine rich sequence which may be utilized as a ribosome binding site (6). The region where gene 20 ends and gene 16 begins, contains the more common stop and start codon overlap.

No significant homology was found between the gene 20 sequence and any sequence in the GenBank non-animal database. The hydropathy profile of the putative gene 20 protein was equally devoid of any notable features.

REFERENCES

Figure 1. Sequence and calculated amino acid coding capacity of P22 gene 20. The locations of the two amber alleles which were sequenced are underlined. Changes associated with these mutations are indicated next to their labels. Positions of conflicts between our sequence and previously published sequence (3) are indicated by bold type face. Nucleotides for all stop and start codons in gene 20 and surrounding genes are indicated by carets below the DNA sequence and are labeled.