

Nucleotide sequence of invasion plasmid antigen gene *ipaA* from *Shigella flexneri* 5

Malabi M.Venkatesan and Jerry M.Buysse

Department of Bacterial Immunology, Division of Communicable Diseases and Immunology, Walter Reed Army Institute of Research, Washington, DC 20307 – 5100, USA

Submitted December 21, 1989

EMBL accession no. X17628

The large 180–220 kb invasion plasmid of *Shigella* and enteroinvasive *Escherichia coli* (EIEC) encodes several major immunogens detected in convalescent sera from monkeys and humans. These include the gene products of *ipaA* (70 kDa), *ipaB* (62 kDa), *ipaC* (41 kDa) and *ipaD* (37 kDa) (1). The four *ipa* genes are ordered as *ipaBCDA* and are located contiguously on the invasion plasmid (1). We have previously reported the complete nucleotide sequence of *ipaB*, *C*, *D* and approximately 250 bp of sequence encoding the 5' end of *ipaA*, within which is located a Shine Dalgarno (GAGG) sequence (2). Here we report the complete nucleotide sequence and the predicted amino acid sequence of *S. flexneri* 5 *ipaA*, from the translation initiation

codon to the sequence for the carboxy-terminal amino acid, obtained by double-stranded DNA sequencing of plasmid clones overlapping the *ipaDA* region. The *ipaA* DNA sequence encodes a protein of 633 amino acids with a molecular weight of 70 kDa. The first four amino acids (MHNV) predicted by the *ipaA* gene sequence are similar to those seen in the *ipaB* gene sequence (2).

REFERENCES

1. Buysse, J.M., Stover, C.K., Oaks, E.V., Venkatesan, M.M. and Kopecko, D.J. (1987) *J. Bacteriol.* **169**, 2561–2569.
2. Venkatesan, M.M., Buysse, J.M. and Kopecko, D.J. (1989) *Proc. Natl. Acad. Sci. USA* **85**, 9317–9321.

```

ATGCATAATGTAATAATACTCAAGCGGCCAATTCCTTATATAAGGCAACTTCAACCATCATCAACAGAATACAGCGAGTTAAAAAGCAAATATCCGATATCCATAGTTCGGCAACTTCT 120
M H N V N N T Q A P T F L Y K A T S P S S T E Y S E L K S K I S D I H S S Q T S
CTAAAAACCCAGCATCAGTGTCTGAAAAAGAAAACCTTGCAGGCTCTTTAATCAGAAATGCTTGAATTTTTTATTTTCTTCTCAGGGAAGAGATGTGTGTAAGAGCATTTTATTC 240
L K T P A S V S E K E N F A T S F N Q K C L D F L F S S S G K E D V L R S I Y S
AACTCAATGAATGGTATGCCAAAAGGAGATTCGAAATTTCAAAATGTTTGTACTCTTGTATCATCAAAATGGTCTTAATTTTGAAGAAGAAAAGGACTTCAAAAAATTTGTCGCA 360
N S M N A Y A K S E I L E F S N V L Y S L V H Q N G L N F E N E K G L Q K I V A
CAGTATTCGGAACTAATTTATAAAGATAAATTTATCCAGATTCCTGCTTTGGACCATGGTGGCAAGATAAGAACTCCATCAATTAAGCAAAAACATTGAGCAGACTTGCAC 480
Q Y S E L I I K D K L S Q D S A F G P W S A K N K K L H Q L R Q N I E H R L A L
TTAGCAACAACACATCTGGTGAAGCTTTATCATTTGGGCAAAAACCTCTCAATCTGAGTATCATCAATTTATCAAGATAAATATTTCTGCTGAATTTAAGTTAAGTAACT 600
L A Q Q H T S G E A L S L G Q K L L N T E V S S F I K N N I L A E L K L S N E T
GTTTCATCTCTCAACTAGATGATTTAGTTGAGCAGGCAAACTTGCCTTTGATAGTTTGGCAATCAAGTAAAAATACTATTGATAGTAAAGGATTTGGTATAGGTAACCTG 720
V S S L K L D D L V D A Q A K L A F D S L R N Q R K N T I D S K G F G I G K L S
AGAGACTTAAATACAGTACGGTGTGTTCTGAGCTGTGAGAAAAGTCTTAAATGATATTTAGAGATATAAAGATTCGATCTTCCAGATGGCTCCCTACCTCCCGAAGAT 840
R D L N T V A V F P E L L R K V L N D I L E D I K D S H P I Q D G L P T P P E D
ATGCCAGTGGGGACCAACCCCGGAGCAATGCAAAACATCCCACTGTAATTCATCATATAAATAATGATAATAGACTTACGATAATAGATTTTTCACACAGAGTATAT 960
M P D G G P T P G A N E K T S Q P V I H Y H I N N D N R T Y D N R V F D N R V Y
GACAATGCTATCAAGCAACCCGAAAATGATGACAGCTCTCTCTCAGCAAAAGATCTATTTAATCCGTAAGGAACTCAATTAATCCAAAGAGCAGTGTTCAAAA 1080
D N S Y H E N P E N D A Q S P T S Q T N D L L S R N G N S L L N P Q R A L V Q K
GTAATTCCTCTCAACTATATATGATAGTCTGTCAGCAATTTGCAAAATTAATTCAGCTTTAGAAAAGTTTCAACATACTCCAGATAATTCGATGGAAATAGGTTACAGCTG 1200
V T S V L P H S I S D T V Q T F A N N S A L E K V F N H T P D N S D G I G S D L
TTAACTACCGAGTACAGAAAGATCTGCAAAATTAATCTCTTCTTGGGGCACAGGCTCTGACATACAGAACTCTTCAACCAACCCCTCTCCACCCGAGGAGTACAGCAGT 1320
L T T S S Q E R S A N N S L S R G H R P L N I Q N S S T T P P L H P E G V T S S
AATGATAACTCATGATACAACTAAAAGTAGCGCTTCTTCTCATAGAGTACCTTCGCAATCAATTAATTCACCTCAAACTGATTCAAAAGTACTTCAACTGATTTTTTATCA 1440
N D N S S D T T K S S A S L S H R V A S Q I N K F N S N T D S K V L Q T D F L S
AGAAATGGAGCACATATTTAACCGGGAAAGATATTTGAGCTTCAAAAAAGTAAACAACTCCCTAAGTAACTTATATCTCTCATTGGAACTAAAATCAGGACACAAAGCAG 1560
R N G D T Y L T R E T I F E A S K K V T N S L S N L I S L I G T K S G T Q E R E
TTACAGAAAATCAAGGACATTTACAAAATCCAAACAGACATAGATAAACAACTAAAAGTTACAGATGCAAAATATAAGAACTAAGTAACTGAAACCAAGCAGGATACAA 1680
L Q E K S K D I T K S T T E H R I N N K L K V T D A N I R N Y V T E T N A D T I
GATAAAAATCATGCGATCATGAAAAGGCAAAAGATATCTAGCGOCTCAGCAAGGTATTGTCAAAAATGAGGATACTCTGCAAAATTTACTTACAGATGATATATCTGATTTAAA 1800
D K N H A I Y E K A K E V S S A L S K V L S K I D D T S A E L L T D D I S D L K
AATAACAATGATTTACAGCTGAAAACAAATATATATAAGCAGCAAAAGATGTAACCACTTCCCTATCAAAAGTATTAAAGAAATCAATAAGGAT 1899
N N N D I T A E N N N I Y K A A K D V T T S L S K V L K N I N K D
    
```