

Nucleotide sequence of the *lsp-dapB* interval in *Escherichia coli*

J. Bouvier and P. Stragier

Institut de Microbiologie, Bâtiment 409, Université Paris-Sud, 91405 Orsay Cedex, France

Submitted October 31, 1990

EMBL accession no. X54925

The *Escherichia coli* chromosomal region extending from the end of the *lsp* gene to the beginning of the *dapB* gene was sequenced. The *lsp* gene is followed by two inverted repeats (shown as arrows) which belong to the REP/PU family (1) and by two open reading frames, *orf149* and *orf316*, which are part of the same operon as *lsp* (2). A putative transcription termination signal is immediately followed by a third open reading frame, *orf304*, encoding a potentially membrane-bound protein. These three open reading frames are preceded by good ribosome binding sites (underlined twice) and could encode proteins of 17K, 36K and 32k that have been previously described (3). Transcription of the *dapB* gene (single arrow) starts 136 bp downstream of *orf304* (4).

REFERENCES

1. Gilson, E., Clement, J.M., Brutlag, D. and Hofnung, M. (1984) *EMBO J.* **3**, 1417-1421.
2. Miller, K.W., Bouvier, J., Stragier, P. and Wu, H.C. (1987) *J. Biol. Chem.* **262**, 7391-7397.
3. Mackie, G.A. (1980) *J. Biol. Chem.* **255**, 8928-8935.
4. Bouvier, J., Richaud, C., Richaud, F., Patte, J.C. and Stragier, P. (1984) *J. Biol. Chem.* **259**, 14829-14834.

```

TCTAGAGCGAAAAACAATAATAAACCCGCGGATCGCGCTGACCGCATCTTATCCGGCTACAGATTGCTCGCAATCGTAGCCGGATAAGCGCTTACCGCGCATCCGGCAAAAATCCTTAATATATAGCGCAAACTGCATGTC 150
S R A K K Q ***
TGAATCTGTACAGAGCAATAGCCCGCTCTGGTGCACCTTCACCGCTAAACCTCGACGATGCGACCCACCGCGGAGTCTACCCGCAACACCGTAACCGCGCGCTGTTCCGCCCTGGGTGATGCTTCTCTTTCTGAGGGCTGGAGCAACACCT 300
E S V Q S N S A V L V H F T L K L D D G T T A E S T R N N G K P A L F R L G D A S L S E G L E Q H L
GTTGGGGCTGAAAGTGGCGATAAAACCACTTCTCGTTGGAGCCAGATGCGCGCTTGGCGTCCCGTACCGGACCTGATTCAGTACTTCTCCCGCGTGAATTTATGGATGCAGGCGAGCCAGAAAATGGCGCAATCATGCTTTTTAC 450
L G L K V G D K T T F S L E P D A A F G V P S P D L I Q Y F S R R E F M D A G E P E I G A I N L F T
CGCAATGGATGGCAGTGAATGCTTGGCGTATCCGCGAAATTAACGGCGACTCCATTACCGTTGATTTCAMCCATCCGCTGGCCGGGACAGCGTTCAATTTGATATTGAAGTCTGGAAATCGATCCGGCAGTGGAGCGTAACATGC 600
A M D G S E M P G V I R E I N G D S I T V D F N H P L A G Q T V H F D I E V L E I D P A L E A *** M Q
AGATCCTGTGGCCAAACCGCGTGGTTTTTGTGCCGGGTAGACCGGCTATCAGCATGTTGAAAACCGCTGGCCATTTACGGCCGACCGGATATATGTCGCTACGAAAGTGTACATAACCGCTATGTGGTGCATAGCTTGGTGGAC 750
I L L A N P R G F C A G V D R A I S I V E N A L A I Y G A P I Y V R H E V V H N R Y V V D S L R E R
GTGGGCTATCTTTATTAGCAGATTAGCGAGTACCGGACCGCGGATCCTGATTTCTCCGACACCGGTTTCTACGGCGGTACGTACGAGCAAAAAGTCCGGATTTGACGGTGTTTGATGCCACCTGTCCCGTGGTGCACCAAG 900
G A I F I E Q I S E V P D G A I L I F S A H G V S Q A V R N E A K S R D L T V F D A T C P L V T K V
TGCATATGGANGTCCCGCCGCGAGTCCCGTGGCGGANGAATCTATTCTCATCGGTACCGCGGACCCCGAAGTGAAGGGCAATGGCCAGTACAGTAACCCGGAAGGGGGAATGATCTGGTGCATCCCGGACGATGTGTGA 1050
H H E V A R A S R R G E E S I L I G H A G H P E V E G T N G Q Y S N P E G G M Y L V E S P D D V W K
AACTGACGGTCAAAAACGAGAGACTCTCCCTTATGACCCGAGACCGCTGTCCGTTGGATGACAGCTGTGATGATCGACGGCTCCGTAACCGCTTCCCGAAAATGTCCGTCGCCGCAAGATGCATCTGCTACGCCAGACTA 1200
L T V K N E E K L S F M T Q T T L S V D D T S D V I D A L R K R F P K I V G P R K D D I C Y A T T N
ACCGTCAGGAGCGGTACCGCCCTGGCAGAACGGCGGANGTGTGTGGTGGTGGTTCGAAAACCTCCCACTCCAACCGCTTGGCGGAGCTGGCCAGCGTATGGCAACCGCGCTTTTGGATTGACGATCGCAAGACATCC 1350
R Q E A V R A L A E Q A E V V L V V G S K N S S N S N R L A E L A Q R M G K R A F L I D D A K D I Q
AGGAGAGTGGGTGAAGAGGTTAAATCGCTCGCGGTGACTCGCGGCGCATCCGCTCCGGATATCTGGTCCAGAAATGTGTGGCAGCTTTGCGAGCTGGCGGCGTGGCCAGCGTATGGCAACCGCGCTTTTGGATTGACGATCGCAAGACATCC 1500
E E W V K E V K C V G V T A G A S A P D I L V Q N V A R L Q Q L G G G E A I P L E G R E E N I V F
TCGANGTCCGAAAGAGCTGCGTGTGATATTCTGANGTGCATTAGTCAATAGCAGCCTAAGTTATCGGAAAATCGCGCTTGTTCACCGGCAATTTTATGGAGAAAACATCGGTTTACCTATCTCTCCGATCTGACCCCGCAT 1650
E V P K E L R V D I R E V D ***
M R L P I F L D T D P G I
TGACGATCGCTCGCCATTCGCCCGCGGATTTTTCACCCGCACTCGACCTGCACTGATGACCACCGCTCGCGGGTAAATGTCTCGGTGGAGAAAAC TACCCGCAATGCCCTGCAACTGCATCTCTGGAATGCGGAGATTCGCTCGC 1800
D D A V A I A A A I F A P E L D L Q L M T T V A G N V S V E K T T R N A L Q L L H F W N A E I P L A
CCAAAGGGCGCTGTGCCACTGGTACCGGACCGCGGTGATCGCGCATCTGTCCAGCGGAAATCGGGATGGCTGACTACTTTGTTGAGCACAACGAAAAGCGCTCGGGATACCGCGGTTTCTGGCGATTCCGGATGCCCTGATGCG 1950
Q G A A V P L V R A P R D A A S V H G E S G M A G Y D F V E E N R K P L G I P A F L A I R D A L M R
TGCACAGAGCTGTTACCCGTTGGCCATCCGCGGTTAACCAATATTCGCGTGTACTTTCACAAATGCCCGAATGCAAGCGGTATATTCGCGCTGTTGGTATCATGGGTGGTTCCTCGGACCGCGCACTGTACGCCAAACCGCGA 2100
A P E V T L V A I G P L T N I A L L S Q C P E C K P Y I R R L V I M G G S A G R G N C T P N A E
GTTAATATTGCTGCCGATCCAGAGCTGCTGCCGTGTCTTCCGAGTGGTATTGAAATCGTCAATGTCGCGTTTGGATGTCACCAATCAGGCAATATTAACCTGACTATCTCTACACTGCCCGAGTTAAACCGTACCGGGAAAAT 2250
F N I A A D P E A A A C V F R S G I E I V M C G L D V T N Q A I L T P D Y L S T L P Q L N R T G K M
GCTTACCGCCGTTTACCCACTACCGTACCGGATGATGCAAGCGGCTTGGCAATGACGATCTTCCGCAATCCGCTGCTGCGCGGACCTGTTCACTCTCAAAACCTGTTTGTGGCGAGTGGCACTACGGCGAATTTAC 2400
L H A L F S H Y R S G S M Q S G L R M H D L C A I A N L V R P D L F T L K P C F V A V E T Q G E F T
CTCAGGCACGAGCGTGGTGTATATCGACGGTTGCCCTGGGCAAGCCAGCAATGTCAGGTTGGATGATGATGTAAGGCTCCAGCAGTGGGTGGCTGAGGTGCTGGCTTGGCGTCTGAACTGTCACATGTTATTGGCATGC 2550
S G T T V D I D G C L G K P A N V Q V A L D L D V K G F Q Q W V A E V L A L A S ***
AGTCATTCAGACTCATGCTTTCAGTATATCCCTCCCTGTTTATCATTAATTTCTAATATACCGCTTTTTGGCTGGCGCGTAGCGATGCGCTGGTACTCTGAAAACGGTCTATGCAAAATACAAAAGGAATAGTATGCAT 2699
M H
    
```