Nucleic Acids Research, 1995, Vol. 23, No. 4 595–598

Cloning and characterization of the *hrpA* gene in the *terC* region of *Escherichia coli* that is highly similar to the DEAH family RNA helicase genes of *Saccharomyces cerevisiae*

Hisao Moriya, Hiroaki Kasai+ and Katsumi Isono*

Department of Biology, Faculty of Science, Kobe University, Rokkodai, Kobe 657, Japan

Received November 22, 1994; Revised and Accepted January 9, 1995

DDBJ accession no. D42105

ABSTRACT

During the course of systematic nucleotide sequence analysis of the *terC* region of *E.coli* K-12 by using the ordered lambda phage clones, we found the presence of a gene, termed *hrpA*, that showed a high degree of sequence similarity to the *PRP2*, *PRP16* and *PRP22* genes of *Saccharomyces cerevisiae*. The products of these yeast genes are known to play their roles in mRNA splicing, and belong to a group of proteins collectively called the DEAH family. The *hrpA* gene is the first example of a DEAH family gene in prokaryotes. The N-terminal region of the protein it encodes contains conserved sequence stretches characteristic of an RNA helicase. Its molecular mass is calculated to be 146 kDa. Previously, a 135 kDa protein was identified by Moir *et al.* [*J. Bacteriol.* (1992) 174, 2102–2110] in this region which is most likely identical to that encoded by *hrpA*. The C-terminal region of the *hrpA* gene product seems to contain an RNA binding motif weakly resembling that of ribosomal protein S1 of *E.coli*. Disruption of the *hrpA* gene suggested that it is not essential for the growth of *E.coli*.

INTRODUCTION

RNA metabolism is an integral part of cellular activity which includes transcription, translation, ribosome assembly, mRNA-splicing, and so on. In some of these processes, a group of proteins called DEAD-box proteins has been found to play an important role. These proteins have a stretch of tetra amino acid sequence, D–E–A–D (Asp–Glu–Ala–Asp), in common (hence their collective name), and are presumed to be RNA helicases (1). The first DEAD-box protein that was shown to have RNA helicase activity was a eukaryotic translational initiation factor eIF4A (2). The eIF4A protein, along with another factor named eIF4B, unwinds double stranded RNA in an ATP-dependent manner. Many of the DEAD-box proteins reported afterwards were so named because they contained a DEAD-box within their sequences. Some of them have been shown to have ATP-dependent RNA helicase activity, while others have not been experimentally confirmed yet as to whether they are also RNA-helicases or not (reviewed in 3 and 4).

More recently, the products of three *S.cerevisiae* genes, *PRP2*, *PRP16* and *PRP22*, have been identified to be involved in mRNA splicing (5–8). Their gene products are presumed to function in two *trans*-esterification steps of mRNA splicing as well as in spliceosome release. They are similar to DEAD-box family proteins, but are relatively larger, and part of the conserved region is different from the DEAD-box proteins. In particular, the conserved sequence that corresponds to the DEAD-box is not D–E–A–D in these proteins: instead, it is D–E–A–H. The other conserved stretches are, however, similar to those of DEAD-box RNA helicases. For these reasons, these proteins were grouped separately into the DEAH family (7).

In the *terC* (terminus of chromosome replication) region of *E.coli* which we have been systematically sequencing, we found the presence of a likely gene at the kilobase coordinate of 1493. Its product has a high degree of sequence similarity to the DEAH family proteins mentioned above. The gene is capable of encoding a protein with a calculated molecular mass of 145 kDa, one of the largest proteins in *E.coli*. Its product is likely to be identical to one of the proteins identified by Moir *et al.* (9). Until now, five DEAD-box family proteins have been identified in *E.coli* (10) some of which were shown to be involved in ribosome assembly (11–13). However, no DEAH family protein has been reported yet to be present in *E.coli* or any other bacteria. The *hrpA* gene is, therefore, the first example of a gene that encodes a DEAH-family RNA helicase in prokaryotes. Since no splicing is known in *E.coli*, the function of this gene appears to be interesting.

RESULTS AND DISCUSSION

Identification of the *hrpA* gene

We have been performing systematic nucleotide sequence determination of the *terC* region of *E.coli* by using the transposon- and PCR-based method (14,15). In this project, we sequenced *E.coli* ordered λ phage clones #255 through #275 (16)

* To whom correspondence should be addressed

+Present address: Division of Cell Fusion, National Institute for Basic Biology, Myodaiji, Okazaki 444, Japan

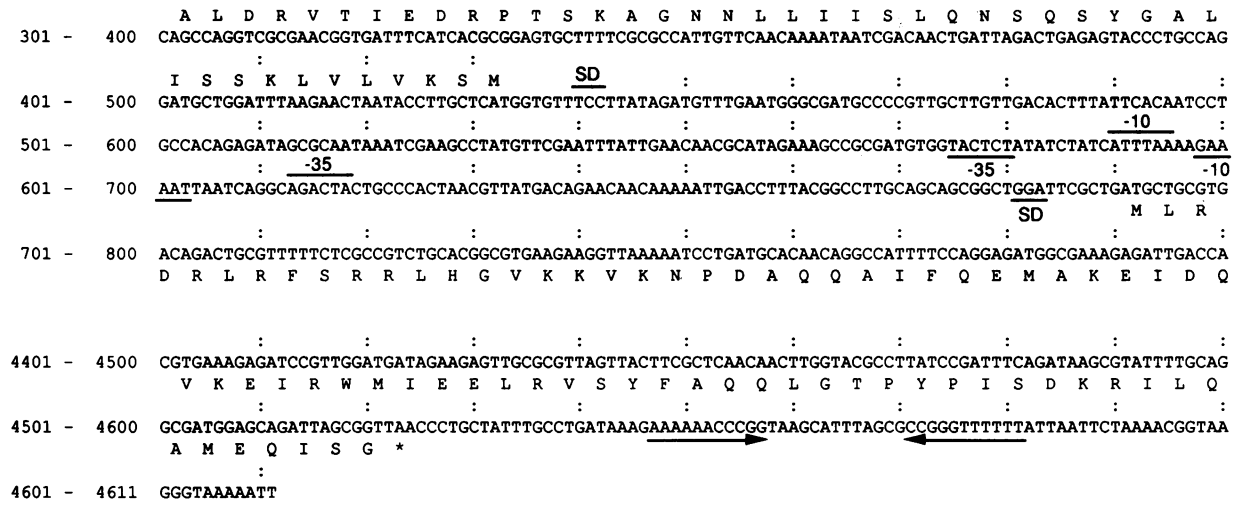

Figure 1. Nucleotide sequence flanking *hrpA*. The chromosomal regions containing the *hrpA* gene and its 5′- and 3′-untranslated regions are shown. –35 and –10 indicate, respectively, the conserved stretches of a likely promoter, and SD indicates a probable ribosome-binding site. A possible stem–loop structure observed downstream of the coding region of *hrpA* is indicated by thin horizontal arrows. The translation of an ORF situated upstream of *hrpA* is given above the nucleotide sequence and in the opposite direction. Numbers correspond to those in the data deposited in the DDBJ database (accession number D42105).

and subsequently analyzed the nucleotide sequence data by computer programs fasta (17) and GeneMark (18) after pre-processing the data. The two computer programs were found to be very useful in identifying likely genes in the sequence data. During the course of this analysis, we found the presence of a large and highly probable gene in the chromosomal segment contained in clone #269. The nucleotide sequence data of the putative gene along with its 5′- and 3′-untranslated regions have been deposited to DDBJ (accession number D42105). Its 5′- and 3′-untranslated regions are presented in Figure 1. There is a promoter-like sequence (–35 and –10 conserved regions) and a probable Shine–Dalgarno (SD) sequence in the 5′-untranslated region, and a terminator-like sequence capable of forming a stem–loop structure in the 3′-untranslated region. We have tentatively assigned the translational initiator situated at position 691 as indicated, because (i) neither the first in-frame ATG codon at position 634 nor the one located further downstream, i.e. at position 781, is preceded by a likely SD sequence, and because (ii) the prediction of a likely gene in this region by program GeneMark matches the one starting with the ATG at position 691. The coding region of the putative gene showed a considerably high degree of sequence similarity to the *S.cerevisiae* DEAH family genes as shown in Figure 2. Consequently, the gene was named *hrpA* (DEAH-family RNA helicase-like protein). Flanking the *hrpA* gene, there are two previously published sequence data, ECIS2IS30 (19; accession no. X62680) and ECOORFAA (20; accession number L09068), both of which are contained in the GenBank nucleotide sequence database (release 85.0). By comparing the restriction enzyme cleavage sites generated from the nucleotide sequence data and the sites in the physical map of *E.coli* (21), the *hrpA* gene has been located at kilobase coordinates of 1493. The putative gene is capable of encoding a protein with 1277 amino acid residues and a predicted molecular mass of 146 130. This is one of the largest proteins in *E.coli*. The *hrpA* gene is likely to be a lowly expressed gene from its prediction pattern by GeneMark (not shown) as well as from its codon usage comparison (22). It should be noted that there is another likely gene upstream of *hrpA* in the opposite direction, starting at position 433 (Fig. 1). No sequence similarity was found with this open reading frame to any known genes and their protein products, however.

Until now, five RNA helicase-like proteins have been identified in *E.coli* all of which belong to the DEAD RNA helicase family (10). None of them shows a high degree of sequence similarity to the *hrpA* gene product. Furthermore, no protein has been found that shows a significant degree of sequence similarity to the C-terminal half of HrpA protein except for a short stretch of sequence that shows weak similarity (not shown) to the RNA-binding motif of ribosomal protein S1 (23). This stretch is identical to the one discussed by Company *et al.* (7), although its similarity to *E.coli* S1 protein is much weaker than the similarity between yeast PRP22 and *E.coli* S1. Moreover, unlike yeast PRP22, the region showing this similarity resides in the C-terminal region. In this connection, it may be interesting to note that the region showing a high degree of similarity to the conserved stretches in DEAH helicases (Fig. 2) is located more to the N-terminus in HrpA protein than in yeast PRP22; in other words, HrpA protein has a much longer C-terminal half.

The *hrpA* gene is not essential for the growth of *E.coli*

To estimate the level of expression of the *hrpA* gene, we performed Northern hybridization analysis using an *hrpA* gene probe. However, no hybridization was detected in this manner. An RNA band of the expected size was detected only when it was amplified with reverse transcriptase and PCR (data not shown). Therefore, we concluded that the transcription of *hrpA* must be very weak. To analyze the *in vivo* function of the *hrpA* gene, therefore, we attempted to construct an *hrpA* disruption mutant. For this purpose a plasmid, termed pKH-Δ*hrpA*::*Kan*, was introduced into *E.coli* strain MC1061 and Kan^r Sm^r transformants were isolated according to the methods of Ohmori *et al.* (24). From one of such isolates, derivatives were isolated that had become Sm^r again to substitute the wild-type allele of *hrpA* on the

```
                                        I                                   Ia
HrpA   55 DNLPVSQKKQDILEAIR.DHQVVIVAGETGSGKTTQLPKICMELGRGI.KG.L..IGHTQPRRLAARTVANRIAEELKTEPGVCIGYKVRFSDHS..DNTM
PRP22 480 QTLPVYAMRSELIQAVR.DNQFLVIVGETGSGKTTQITQYLDEEGFSN....YGMIGCTQPRRVAAVSVAKRVAEEVGCKVGHDVGYTIRFEDVTGPD.TR
PRP16 347 EQLPVFRCRSQLLSLIR.ENQVVVIIGETGSGKTTQLAQYLYEEGYANDRGK.S.IVVTQPRRVAAISVAKRVAMEMQVPLGKEVGYSIRFEDVTDSECTK
PRP2  220 KLLPVHQYKDELLQEIK.KNQVLIIMGETGSGKTTQLPQYLVEDGFTD.QGKL.QIAITQPRRVAATSVAARVADEMNVVLGKEVGYQIRFEDKTTPNKTV
HrpB    1 SSLPVAAVLPELLTALDCAPQVLL.SAPTGAGKSTWLP.LQLLAHPGI.NGKI..I.LLEPRRLAARNVAQRLAELLNEKPGDTVGYRMRAQNCVGPN.TR

                                    II                                  III
HrpA  VKLMTDGILLAEIQQDRLLMQYDTIIIDEAHEHSLNIDFLLGYLKELLP.RRPGLKIIITSATIDPERFSRHFNNAPIIEVSGRTYPVEVRYRPIVEEADDTSAI
PRP22 IKYMTDGMLQREALLDPEMSKYSVIMLDEAHERTVATDVLFALLKKAAI.KRPELKVIVTSATLNSAKFSEYFLNCPIINIPGKTFPVEVLYSQTPQMDYIEAAL
PRP16 LKFVTDGILLRETLLDDTLDKYSCVIIDEAHERSLNTDILLGFFKILLA.RRRDLKLIITSATMNAKKFSAFFGNAPQFTIPGRTFPVQTIYTSNPVQDYVEAAV
PRP2  LKYMTDGMLLREFLTDSKLSKYSCIMIDEAHERTLATDILIGLLKDILP.QRPTLKLLISSATMNAKKFSEFFDNCPIFNVPGRRYPVDIHYTLQPDGNYIHAAI
HrpB  LEVVTEGVLTRMIQRDPELSGVGLVILDEFHERSLQADLALALLLDVQQGLRDDLKLLIMSATLDNDRLQQMLPEAPVVISEGRSFPVERRYLPLPAHQRFDDAV

                                   IV                                                                    V
HrpA  SCRAIFDAVDELSQESHGDILIFMSG.EEIPDTADALNKLNLRHTE..............ILPLYARLSNSEQNRVFQS.HSG.RRIVLATNVAETSLTVPGIKYV
PRP22 DC..VIDIHINEGP...GDILVFLTGQEEIDSCCEILYDRVKTLGDSIGELL.........ILPVYSALPSEIQSKIFEPTPKGSRKVVFATNIAETSITIDGIYYV
PRP16 SQ..AVKIHLANDCSS.GDILIFMTGQEDIETTFDTLQEKFLQVYSKKFGTANFEEINDIEILPIYSALPADLQFKIFQDLHGTKRKIIIATNIAETSLTIKGIRYV
PRP2  TT..IFQIHTTQSLP..GDILVFLTGQEEIERTKTKLEEIMSKLGSRTKQMI.........ITPIYANLPQEQQLKIFQPTPENCRKVVLATNIAETSLTIDGIRYV
HrpB  ....AVATAEMLRQES.GSLLLFLPGVGEIQRVQEQLASRIGSDVL..............LCPLYGALSLNDQRKAILPAPQGMRKVVLATNIAETSLTIEGIRLV

                                   VI
HrpA  IDPGTARISRYSYRTKVQRLPIEPISQASANQRKGRCGRVSEGICIRLYSEDDFLS.RPEFTDPEILRTNLASVILQMTALGLGDIAAFPFVAEPDKRNI
PRP22 VDPGFAKINIYNARAGIEQLIVSPISQAQANQRKGRAGRTGPGKCYRLYTESAFYNEMLENTVPEIQRQNLSHTILMLKAMGINDLLKFDFMDPPPKNLM
PRP16 IDCGYSKLKVYNPKIGLDSLVITPISKANADQRSGRAGRTAPGTRYRLYTEDTFKEDMYLQTIPEIQRTNLSNTLLLLKSLDVTDFLSKFPFIDKPPLQT
PRP2  IDPGFVKENSYVPSTGMTQLLTVPCSRASVDQRAGRAGRVGPGKCFRIFTKWSYLHELELMPKPEITRTNLSNTVLLLLSLGVTDLIKFPLMDKPSIPTL
HrpB  VDCAQERVARFDPRTGLTRLITQRVSQASMTQRAGRAGRLEPGISLHLIAKEQAER.AAAQSEPEILQSDLSGLLMELLQWGCSDPAQMSWLDQPPVVNL
```

Figure 2. Similarity of the *hrpA* gene product and RNA helicases. The N-terminal half of the deduced amino acid sequence of the *hrpA* gene product is aligned with the amino acid sequences of the proteins encoded by the yeast DEAH family genes, *PRP2*, *PRP16* and *PRP22*, as indicated. Identical or highly similar amino acid residues in all four sequences are indicated in bold face. In addition, the amino acid sequence of the deduced product of the *hrpB* gene is aligned. Residues in bold face in this case indicate its similarity to the above-mentioned four gene products. Dots were introduced to obtain maximal sequence alignment. Numbers at the top indicate the first amino acid residue numbers aligned for comparison.

chromosome with the introduced segment containing a disrupted allele. Southern hybridization analysis of several of such derivatives showed that indeed they contained only the disrupted allele of the *hrpA* gene. They grew normally and no aberrant features were noticed such as increased UV-sensitivity, high- or low-temperature sensitivity, sensitivity to RNA (f2) and DNA (Δ) phages, or reduced viability by prolonged incubation, etc. Thus, we concluded that the *hrpA* gene is not essential for the growth of *E.coli* cells as well as for the proliferation of *E.coli* phages under the conditions employed.

We thought it possible that there is a gene that is not only functionally but also structurally similar to *hrpA* and can compensate the disruption of *hrpA*. To test this possibility, we performed genomic Southern hybridization under low-stringency conditions and by using *hrpA* gene probes. However, no hybridization other than the *hrpA* gene itself was observed (data not shown). One of the *E.coli* DEAD family genes, *rhlB*, was reported to show a strain-dependent phenotype (10), although the nature of the dependency has not been clarified yet. Among the four DEAD family genes reported so far, *srmB* (or, *rba*) and *dbpA* null mutants showed cold-sensitive growth (13), but the disruption of *deaD* and *rhlE* did not result in any noticeable phenotypic alteration (13,25). Also, many DNA helicase genes in *E.coli* are not essential. It could be that some helicases may be functionally compensating each other.

More recently, the nucleotide sequence of a region of the *E.coli* genome spanning from 2.4–4.1 min of the *E.coli* genetic map has been reported (26; accession number D26562) which contains a putative gene the product of which is significantly similar (33% identity) to that of the *hrpA* gene. We tentatively named this gene *hrpB*. The two genes show a considerable degree of nucleotide sequence similarity, but the predicted *hrpB* gene product does not contain some of the amino acid residues highly conserved among the DEAH family proteins as well as in the *hrpA* gene product as shown in Figure 2. A more detailed analysis of the two genes and their protein products is now in progress. Our preliminary results suggest that an *hrpA* and *hprB* double disruptant can grow without any appreciable phenotypic anomaly.

ACKNOWLEDGEMENTS

This work was supported in part by grants-in-aid for scientific research Nos 01880029, 02237105, 03221105 and 04305005 from the Ministry of Education of Japan. We are grateful to H. Ohmori of Kyoto University for providing the *E.coli* strains and plasmid for gene disruption analysis, M. Borodovsky of Georgia Technical Institute for useful suggestions in analyzing data with the computer program GeneMark, and to Dynal-Japan (Tokyo) and Showa-Hokokai (Osaka) for their financial support.

REFERENCES

1 Linder P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., Slonimski, P.P. (1989) *Nature* **337**, 121–122.
2 Ray, B.K., Lawson, T.G., Kramer, J.C., Cladaras, M.H., Grifo, J.A., Abramson, R.D., Merrick, W.C., Thach, R.E., (1985) *J. Biol. Chem.* **260**, 7651–7658.
3 Schmid, S.R., Linder, P. (1992) *Mol. Microbiol.* **6**, 283–292.
4 Fuller-Pace F.V., Lane D.P. (1992) *Nucleic Acids Mol. Biol.* **6**, 159–173.
5 Chen, J.H., Lin, R.J. (1990) *Nucleic Acids Res.* **18**, 6447.
6 Burgess, S., Couto, J.R., Guthrie, C. (1990) *Cell* **60**, 705–717.
7 Company, M., Arenas, J., Abelson, J. (1991) *Nature* **349**, 487–493.
8 Ruby, S.W., Abelson, J. (1991) *Trends Genet.* **7**, 79–85.
9 Moir, P.D., Spiegelberg, R., Oliver, L.R., Pringle, J.H., Masters, M., (1992) *J. Bacteriol.* **174**, 2102–2110.
10 Kalman M., Murphy H., Cashel M.l (1991) *New Biologist* **3**, 886–895.
11 Nish, K., Morel-Deville, F., Hershey, J.W.B., Leighton, T., Schnier, J. (1988) *Nature* **336**, 496–498
12 Toone, W.M., Rudd, K.E., Friesen, J.D., (1991) *J. Bacteriol.* **173**, 3291–3302.

13 Nashimoto H. (1993) In *The Translational Apparatus* (Nierhaus, *et al.* eds) Plenum Publishing, New York, pp. 185–195.
14 Kasai, H., Isono, S., Kitakawa, M., Mineno, J., Akiyama, H., Kurnit, D.M, Berg, D.E., Isono, K. (1992) *Nucleic Acids Res.* **20**, 6509–6515.
15 Kasai, H., Kim, S.-O., Isono, S., Borodovsky, M., Isono, K. (1994) In *Advances in Biomagnetic Separation* (Uhlén, M. *et al.* eds.) Eaton Publishing, pp. 113–125.
16 Kohara, Y., Akiyama, K., Isono, K. (1987) *Cell* **50**, 495–508.
17 Pearson, W.R., Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
18 Borodovsky M., McIninch J. (1993) *Comput. Chem.* **17**, 123–133.
19 Moszer, I., Glaser, P., Danchin, A. (1991) *Biochimie* **73**, 1361–1374.
20 Hidalgo, E., Limon, A. and Aguilar, J. (unpublished) L09068.
21 Rudd, K.E. (1992) In *A Short Course in Bacterial Genetics* (Miller, J.H.) Cold Spring Harbor Laboratory, pp. 2.3–2.43.
22 Aota, S., Gojobori, T., Ishibashi, F., Maruyama, T., Ikemura, T. (1988) *Nucleic Acids Res.* **16**, r315–402.
23 Subramanian, A.R. (1984) *Trends Biochem. Sci.* **9**, 491–494.
24 Ohmori, H., Saito, M., Yasuda, T., Nagata, T., Fujii, T., Wachi, M., Nagai, K. (1995) *J. Bacteriol.* **177**, 156–165.
25 Ohmori, H. (1994) *Jpn J. Genet.* **69**, 1–12.
26 Fujita, N., Mori, H., Yura, T., Ishihama, A. (1994) *Nucleic Acids Res.* **22**, 1637–1639.