Gene HI1472 of *Haemophilus influenzae* Rd is a novel gene involved in DNA repair

Pradnya K. Kowtal & Vasudha P. Joshi

Molecular Biology and Agriculture Division, Bhabha Atomic Research Centre, Trombay, Mumbai 400 085, India.

Received 21 December 1998; revised 22 March 1999

A chimeric plasmid, pJPuvr4, consists of a 16.7 kbp *Haemophilus influenzae* Rd chromosomal DNA insert at the EcoRI site of vector pJ1-8. This plasmid complements the UV and gamma ray sensitivity of the mutant strain MBH4. This plasmid carries the wild type allele of gene *uvr4* which was localised to a 3.8 kbp *DraI* fragment, with an internal *EcoRI* site. Partial sequencing of the gene and its alignment with the published genome sequence of *H. influenzae* Rd revealed *uvr4* to be HI1472. HI1472 is a putatively identified open reading frame (ORF), which has been assigned no function so far. The partial sequence did show nt database match with 3D exon of N cadherin gene of homosepians and moaA gene of *H. influenzae*. Cadherins are involved in cell adhesion, cell to cell contact and morphogenesis in homosepians and moaA gene codes for molybdenum biosynthesis subunit A. This report implicates HI1472 of *Haemophilus influenzae* Rd in DNA repair. Nucleotide sequence obtained for the gene *uvr4* was compared with the published sequence of gene HI1472. A wild type strain variation was observed at the 592nd nucleotide position corresponding to a change from aspartic acid to threonine.

*Haemophilus influenzae* is a naturally transformable Gram negative bacterium. The ability of *H. influenzae* to attain competence naturally has been utilised to study its DNA repair systems. The DNA repair systems studied in *H. influenzae* Rd are the nucleotide excision repair and post replication repair. The photoreactivation repair system is absent in *Haemophilus influenzae*.

Whole genome sequencing of *H. influenzae* Rd has putatively identified 1,743 ORFs (genes). In keeping with the earlier studies on the DNA repair systems of *H. influenzae*, genes functioning in the nucleotide excision repair and post replication repair pathways have been identified by the genome sequencing data. The genes involved in other DNA repair systems like the mismatch repair, base excision repair and oxidative repair systems have also been identified, based on amino acid sequence homology of their translation products with various deposits in the databases for other bacteria.

Four UV sensitive mutants of *H. influenzae* Rd have been isolated in our laboratory. These mutants were isolated by N-methyl N-nitro N-nitrosoguanidine treatment of the wild type strain Rd. The mutant strains earlier named as MB01, MBO2, MBO3 and MBO4 in addition to being UV-sensitive, were found to be deficient in host cell reactivation of phage HP1c1 to varying degrees. These strains are now designated as MBH1, MBH2, MBH3 and MBH4. One of the UV-sensitive mutant strain MBH-4 has been used in this study. This mutant strain is about 12 times more sensitive to UV radiation and about 1.3 times more sensitive to gamma radiation compared to the wild type strain Rd. This report describes the cloning, sequencing and thereby identification of the wild type *uvr4* of *H. influenzae* Rd.

Materials and Methods

**Bacterial strains and plasmids**—The bacterial strains used in this study are wild type and UV-sensitive mutant strain MBH4 of *Haemophilus influenzae* Rd. DH5α and XL1 Blue-MRF7 strains of *Escherichia coli*. The plasmids used were pJ1-8 as *H. influenzae* cloning vector, pBluescriptII SK+ for subcloning and M13K07 helper phage for rescuing single stranded DNA for sequencing. Both the plasmids carry ampR (ampicillin resistance) marker. The UV-sensitive mutant strain MBH4 was isolated by N-methyl N-nitro N-nitrosoguanidine treatment of the wild type strain Rd and out crossed for any secondary mutations. Alleles of the gene governing UV resistance or UV sensitivity, in wild type Rd and mutant strain MBH-4 respectively is referred as *uvr4*.
Media, plasmid DNA isolation and transformation—*H. influenzae* cells were grown in brain heart infusion broth/agar (Difco) supplemented with NAD (2 μg/ml) and hemin (10 μg/ml). *Escherichia coli* cells were grown in Luria Bertani broth/agar 11. Plasmid DNA was isolated by standard alkali lysis method 11. *H. influenzae* cells were made competent by the aerobic-anaerobic-aerobic method 12, while competence was induced in *E. coli* cells by calcium chloride treatment and transformed 13. Restriction endonucleases, T4 DNA ligase and Taq DNA polymerase were purchased from M/s Bangalore Genei Pvt. Ltd. and used as per their instructions.

*UV* irradiation of *H. influenzae* cells—*H. influenzae* cells were grown to exponential phase in 10 ml of supplemented BHI broth. These cells were washed and resuspended in 2 ml of saline buffer (0.5 g NaCl, 2.7 g Na₂HPO₄, and 4.4 g KH₂PO₄ in 1000 ml distilled water). Two ml of the bacterial culture at an approximate cell density of 2 x 10⁸ cells/ml was exposed to UV light in a 4.5 cm diameter petridish. Source of UV radiation was a 15 W germicidal type, violet UV-irradiation of *H. influenzae* cells—*H. influenzae* cells were grown to exponential phase in 10 ml of supplemented BHI broth. These cells were washed and resuspended in 2 ml of saline buffer (0.5 g NaCl, 2.7 g Na₂HPO₄, and 4.4 g KH₂PO₄ in 1000 ml distilled water). Two ml of the bacterial culture at an approximate cell density of 2 x 10⁸ cells / ml was exposed to UV light in a 4.5 cm diameter petridish. Source of UV radiation was a 15 W Phillips TUV germicidal lamp (wavelength 254 nm) with a fluence of 0.3 J/m²/sec as measured by a Black Ray UV meter (Ultra Violet Products Inc.). Aliquots of the cell culture were removed after various doses of UV irradiation, diluted suitably, plated with supplemented BHI agar and incubated at 37°C for 16 h. A graph of percent survivors versus the UV dose was plotted.

Preparation of genomic DNA library—The wild type, *H. influenzae* Rd, chromosomal DNA (10 μg in 20 μl volume) was partially digested with 0.5U/μg EcoRI enzyme for 20 minutes at 37°C. The vector, pJl-8 (2 μg in 10 μl volume) was linearised with 1U/μg EcoRI for 1 hour at 37°C. A typical ligation reaction contained 6 μl digested chromosomal DNA, 4 μl linearised plasmid DNA, 1X ligase buffer, 10mM ATP and 5U T4 DNA ligase. The total volume of the reaction mixture was 20 μl and ligation was carried out at 14°C over night.

Subcloning and sequencing—The 3.8 kbp DraI fragment was digested with EcoRI. The 1.3 kbp and 2.5 kbp DraI/EcoRI fragments were recovered from low melting temperature agarose gel after electrophoresis and subcloned into the EcoRI/SmaI digested phagemid vector SK 11. The plasmids were "stabilised" by transforming into the *E. coli* strain DH5α and later transformed into the *E. coli* strain XL1-Blue MRF. Single stranded DNA was isolated after infecting XL1-Blue MRF with helper phage M13K07. The gene *uvr4* was partially sequenced using the dideoxynucleotide chain termination method 14. The sequencing kit used was the Sequenase Version 2.0 (United States Biochemicals Inc.). Using the universal primer, *uvr4* was sequenced from the EcoRI site present within the gene. The partial sequence obtained for *uvr4* was aligned with the published genomic sequence of *H. influenzae* Rd, at the National Center for Biotechnology Information, Maryland, USA.

Amplification of gene *uvr4* using PCR-Amplification of the gene *uvr4* was performed by using genomic DNA as template. Taq DNA polymerase used in the amplification reaction was purchased from M/s Bangalore Genei and the amplification reactions were carried out as per their recommendations. The thermocycling profile consisted of 30 cycles, with 1 cycle consisting of denaturation at 94°C for 1 min, annealing at 62°C for 1 min and elongation at 72°C for 2 min. A final extension was carried out at 72°C for 5 min, in a Perkin Elmer Gene Amp PCR system 2400.

Results and Discussion

Cloning and complementation—The wild type gene *uvr4* was cloned by its ability to restore UV-resistance of the mutant strain, MBH4, to the wild type level. Competent MBH4 cells were transformed with the genomic DNA library. Ampicillin resistant transformants were screened for UV resistance at a dose of 4.5 J/m². A UV dose of 4.5 J/m² was selected because the survival of the mutant strain at this dose is about 0.002%, while the survival of the wild type strain is about 60%. Eighteen AmpR UVK transformants were scored. The plasmid DNA was isolated and used to transform competent MBH4 cells. One of the recombinant plasmids complemented the defect in MBH4, by conferring UV-resistance on MBH4, comparable to the wild type strain levels (Fig. 1). This chimeric plasmid carrying the wild type allele of *uvr4* gene was designated as pJPuvr4.

Gamma ray inactivation of colony formation was studied to find out if (a) the mutant strain MBH4 is more sensitive to gamma radiation, as compared with the wild type strain, and (b) the chimeric plasmid pJPuvr4 was able to complement the gamma ray
sensitivity of the mutant. These studies showed that
the mutant strain MBH4 is about 1.3 times more
sensitive to gamma radiation as compared to the wild
type strain. Chimeric plasmid pJPuvr4 complements
wild type levels (Fig. 2).

Localisation of gene uvr4 on pJPuvr4 - Size of the
pJPuvr4 insert was estimated to be 16.7 kbp with an
internal EcoRI site dividing the insert into 9.6 kbp and
7.1 kbp fragments. Both the fragments were
subcloned at the EcoRI cloning site of the vector pJ1-
8 individually. The chimeric plasmids were
designated as pJP13.4 and pJP10.9. Competent
MBH4 cells were transformed with these plasmids
and scored for Amp^ UV^ transformants. No Amp^ UV^ transformants were obtained, indicating that
neither of the two plasmids was able to confer
UV-resistance on MBH4. This meant that the intact insert
is essential for complementation. It was speculated
that the EcoRI site is located within the gene uvr4.
The restriction map of pJPuvr4 showed that a 3.8 kbp
Dra I fragment carries the internal EcoRI site. The
presence of uvr4 on this fragment was therefore
tested. Attempts at subcloning the 3.8 kbp Dra I
fragment into suitable vectors proved to be
unsuccessful. Hence, a possible presence of repetitive
sequences/palindromes on this fragment was
speculated.

Competent MBH4 cells were transformed using the
3.8 kbp DraI fragment. The transformants were
scored for UV-resistance at 4.5J/m^2 (Table I), TE
buffer without DNA served as a negative control. A
number of UV^ transformants scored using either 3.8
kb Dra I fragment or pJPuvr4 plasmid DNA is
comparable, indicating presence of uvr4 gene on 3.8
kb Dra fragment. It was also gathered that at least
one copy of the 9 bp sequence (AAG TGC GGT)
essential for the efficient uptake of linear fragments in
H. influenzae was present on the 3.8 kbp Dra I
fragment.

Nucleotide sequence of gene uvr4 - The EcoRI/
DraI fragments of the 3.8 kbp DraI fragment were
subcloned in the phagemid vector and used for
sequencing. The partial sequence (240 nucleotides)
obtained for uvr4 was found to tally 100% with a part
of a putatively identified open reading frame (ORF)

---

Fig. 1—Survival curves of wild type Rd and mutant strains
(MBH-4 and MBH-4 containing plasmid pJPuvr4) of Haemophi-
lus influenzae Rd upon exposure to varying doses of UV radia-
tions.

Fig. 2—Survival curves of wild type Rd and mutant strains
(MBH-4 and MBH-4 containing plasmid pJPuvr4) upon exposure
to varying doses of gamma radiation.
HI1472, carried on the locus HIU 32825 of the whole genome. Location of ORF 1472/uvr4 on *Haemophilus influenzae* Rd genome and restriction sites for some of the restriction endonucleases is shown in Fig.3. On obtaining the sequence information for the flanking loci HIU 32824 and HIU 32826, the restriction map deduced for the insert of pJPuvr4 was also found to tally. Only a 96 bp *Pst* I fragment was missed out in this study, probably because of its small size. A total of four ORFs HI1471 (hemU), HI1472 (no assigned function), HI1473 (modD) and HI1474 (sfrC) are present on 3.8 kb *Dra*I fragment. ORF HI1472 is 1053 bp in length and it did show partial nucleotide sequence homology with a part of *H. influenzae* moaA gene and 3D exon of N cadherin gene of homosapians. This ORF, HI1472, has not been assigned any function. As per our observation HI1472 is involved in DNA repair of *H. influenzae* Rd.

PC/Genome analysis of the sequence information of HIU32825 (10.8kbp) confirmed the speculations made earlier on, with respect to the uptake specific sites and repetitive sequences on the 3.8 kbp *Dra*I fragment. The *H. influenzae* genome sequence shows that a 3976 bp *Dra*I fragment on the HIU 32825 locus, carries 5 uptake specific sequences. A potential Shine-Dalgarno sequence "AGGA" was found 6 bp upstream of the translation start site of HI1472. A probable -10 sequence (TAT AA T) and a probable -35 sequence (TTA TCA) of the promoter elements were located at 31 bases and 52 bases upstream of the translation start site. The putative protein coded for by HI 1472/uvr4 has a molecular size of 39.2 kDa, with 351 amino acid residues and an isoelectric point at 7.3.

Based on the published nucleotide sequence of gene HI1472, appropriate primers were designed and used for amplification of the gene HI1472/uvr4 by the polymerase chain reaction. Internal primers were designed, synthesised, purified and used to sequence the 1053 bp open reading frame. DNA sequencing of the PCR product (Fig. 4) showed a variation at the nucleotide position 592 of the open reading frame of wild type allele of uvr4. A GAC codon is replaced by ACG. In effect, the 198th amino acid residue of the putative protein showed a change from aspartic acid (GAC) to a threonine residue (ACG).

The whole genome sequencing of *H. influenzae* Rd has identified 1,743 genes. Of these, the function of
about 736 genes is unknown. The reasons for genes not being assigned a function may include (a) the genes, whose function in regulatory circuits, are not yet known (b) the genes, may be essential for the cell viability or (c) the genes, may function in processes where there are multiple regulatory mechanisms. Also, genes affecting the cell viability or functioning in the regulatory mechanisms may be difficult to identify by the classical genetic methods.

The exact role of the gene product of HI1472 in the repair of UV-damaged DNA is not yet known. But, as the gene is also involved in the repair of gamma-ray induced damage, it appears to be involved in some common step in the DNA repair mechanism. The partial nucleotide sequence of the gene uvr4 shares some homology with 3D exon of N cadherin of homosapians and mao A gene of H. influenzae Rd. Cadherins are a family of cell adhesion receptors.

SQ SEQUENCE 1400 bp, 421 A, 258 C, 317 G, 404 T, 0 OTHERS.

-241 AGGCTTTAGA TCTACTTCAA TGAAAACAGA AAGATATTAT GCTTTCTTGTC
-191 GCTGGAGGAG TAAATAAAAA TAATGTATCA GATTACGCCA AATTGGGTAT
-141 TGCTTTATTAT ATCACTTCTG CACCCTTAAT TGTTACGCCA GAAATATATA
-91 AGGTCGTTGAT TGAAAAAGATA TGAGTACATA GTTTTATTAT ATATAATAT
-41 GTGTATATAG ATATTTTCTTT ACTTAAAATA AGGATGTGCC A'TGAAACTTA
 11 AATCTTGTGT AATTGCCGTC TTGTCGAGTT CCCCTTCTTT TCGAGCATT
 61 GCCTGATCGCA TTTATTACGA CGAACTCGAT CGAAAGTGA CAATTCCCGA
111 TCACATTAAAC CGAGCCCGTG GCCATTACAAG GAAATCTCCTTG CGACCCGGAAC
161 TGCAATTGGA TGCCACAAAA CAGATTGTGG GAGTCCCTTC CAACTGGAAG
211 AACAAACTAG GCAAAAACCTA GTGTCGCTCTT GCAAGCAGGC TGGAAAAACAT
261 GGCATGGCCG GGCAGTTTG AAATTCCCTCG ATGGTTGCCTT TTGTGGCCGT
311 TAAAACCGGA TGTGGTTTGT GTCGACTAAC ACGCCGGTTC TGAAATGATC
361 AAGCAATCGA TGACCGGAA TACCCCTCGG TTAGCCATT TCTTCCGCTAC
411 TGCGTGGATGG GCCGAAAAAGG TGAACCTTAA TCAACCTTAA ACTGACGAAG
461 ATAAAAGCTTA TGAATCGGCG TAAAACAGGT GCAAGAGCTAC AGTGGGCAGG
511 GTGTTTGGAG AAAAAAACAA AGGGTATGAG TTAGTGAAGG AGCGCCCTTC
561 TAATCCTGAA TGTTGCGGAG ATGCGGCTAGG CGACGGCTTCT GCCGATTAAC
611 GTTGCGCGCA ATATATGCGA AATCGCGATT TAGGTCCTTTA TTGTGGCCGT
661 AAATACGGG GCCTTTATGT GGAACATCGG GCGCTTATAA ACCGCGGCG
711 GCGAACCATG AAAGGTGTCCA AACAGGTTCG TCTGAAATAT TGTTGCGGAT
761 GGAATCTCGG GTGGATTGTT GTGCAAGATC GTTATCTGTG TGTAGTACAC
811 AAAATTGGGA AGCTAAGAGG TGCCGGAATT ATCGGGTATA GAAATATAGA
861 AAAAGGTTTTT TGATGGCGCG AAATATGCTAA AGGCTGGGGA TACCGATGCG
911 CGGAAGGCGT AGCGTCGCGGG GAAATGATGT GTGCTAAAGC CTTCGTTGG
961 CAACGTGTGAG AAGCGGCGGG TTGGATAAAA ATGGTGAAAC ATATTACAC
1101 AAAATTCTAC CGCACTCTCT AAAACCCCGA CAATCGGGGC AGATAGTTAC

Fig. 4—Nucleotide sequence of the gene uvr4 / HI1472. The probable -35, -10 and S.D. regions, start and stop codon are indicated as bold letters and underlined. The primers used for sequencing are underlined, while the bold italic underlined letters show the bases of strain variation.
involved in cell morphology and cell to cell contact\textsuperscript{18}. A mutation in cadherin gene is responsible for the formation of cancerous tissue. At this stage, it is not possible to hypothesise any function to the HI1472 coded protein, further work is essential.

Acknowledgement
The authors are grateful to Dr. S. K. Mahajan for comments and advise, and to Mrs. M. Mathur for help in gene sequence alignment and computational analysis. PKK was supported by a fellowship of the Department of Atomic Energy, Government of India.

References