The abundant and essential HU proteins in *Deinococcus deserti* and *Deinococcus radiodurans* are translated from leaderless mRNA

Claire Bouthier de la Tour, Laurence Blanchard, Rémi Dulermo, Monika Ludanyi, Alice Devigne, Jean Armengaud, Suzanne Sommer and Arjan de Groot

HU proteins have an important architectural role in nucleoid organization in bacteria. Compared with HU of many bacteria, HU proteins from *Deinococcus* species possess an N-terminal lysine-rich extension similar to the eukaryotic histone H1 C-terminal domain involved in DNA compaction. The single HU gene in *Deinococcus radiodurans*, encoding DrHU, is required for nucleoid compaction and cell viability. *Deinococcus deserti* contains three expressed HU genes, encoding DdHU1, DdHU2 and DdHU3. Here, we show that either DdHU1 or DdHU2 is essential in *D. deserti*. DdHU1 and DdHU2, but not DdHU3, can substitute for DrHU in *D. radiodurans*, indicating that DdHU3 may have a non-essential function different from DdHU1, DdHU2 and DrHU. Interestingly, the highly abundant DrHU and DdHU1 proteins, and also the less expressed DdHU2, are translated in *Deinococcus* from leaderless mRNAs, which lack a 5′-untranslated region and, hence, the Shine–Dalgarno sequence. Unexpectedly, cloning the DrHU or DdHU1 gene under control of a strong promoter in an expression plasmid, which results in leadered transcripts, strongly reduced the DrHU and DdHU1 protein level in *D. radiodurans* compared with that obtained from the natural leaderless gene. We also show that the start codon position for DrHU and DdHU1 should be reannotated, resulting in proteins that are 15 and 4 aa residues shorter than initially reported. The expression level and start codon correction were crucial for functional characterization of HU in *Deinococcus*.

INTRODUCTION

Genomic DNA of bacteria is folded and compacted into a structure called the nucleoid. Various DNA-binding architectural proteins, such as HU, H-NS and Fis in *Escherichia coli*, play a role in nucleoid compaction and also participate in DNA-dependent functions including replication, recombination, repair and gene regulation (Dillon & Dorman, 2010; Grove, 2011; Luijsterburg et al., 2006). Microscopy images revealed that radiation-resistant *Deinococcus* species and other radiation-resistant bacteria show a higher degree of nucleoid compaction compared with radiation-sensitive bacteria, which suggested that a more condensed nucleoid might contribute to radiation resistance (Levin-Zaidman et al., 2003; Zimmerman & Battista, 2005). Shotgun proteomics of isolated nucleoids
showed that histone-like HU is the most abundant nucleoid-associated protein in *Deinococcus radiodurans* and *Deinococcus deserti* among the very few candidate nucleoid-associated proteins encoded by these species (Toueille et al., 2012).

HU proteins are present in all eubacteria (Grove, 2011). HU proteins of *E. coli*, *Bacillus subtilis* and most other characterized bacteria are composed of about 90 aa residues. Compared with this core HU domain, a low-complexity region rich in lysine, alanine and proline residues is present as a C-terminal extension in HU proteins from *Mycobacterium* species, *Streptomyces coelicolor* and *Kineococcus radiotolerans* (Grove, 2011; Yee et al., 2011). Members of the genera *Deinococcus*, *Xanthomonas*, *Xylella*, *Gemmatia* and a few others encode HU proteins with a similar Lys- and Ala-rich extension at the N terminus. The N- and C-terminal extensions in HU resemble the (S/T)PKK repeats in the C-terminal domain of eukaryotic histone H1, which are involved in DNA compaction (Bharath et al., 2002). In vitro experiments have shown that the *Mycobacterium* HU extensions have an important contribution to DNA binding, modulate the DNA binding specificities compared with the corresponding truncated proteins consisting of the conserved core HU domain, and may have a role in DNA protection and compaction (Grove, 2011; Kumar et al., 2010; Mukherjee et al., 2008; Sharadamma et al., 2011). In vitro DNA-binding experiments have also been performed with recombinant *D. radiodurans* HU proteins (i.e. full-length protein and HU lacking the 47 aa residue N-terminal extension) (Ghosh & Grove, 2004, 2006). The truncated protein was found to bind short duplex DNA and to show preferred binding to DNA with nicks or gaps, similar to the binding properties of *E. coli* HU. The presence of the N-terminal extension prevented binding to short duplex DNA, attenuated the preference for DNA with nicks or gaps, and resulted in binding preferentially only to four-way junction DNA structures. In addition, *D. radiodurans* HU (DrHU) did not bend duplex DNA. These results suggested an *in vivo* role of DrHU in stabilizing homologous recombination intermediates rather than an architectural role (Ghosh & Grove, 2004, 2006; Grove, 2011).

Remarkably, members of the genus *Deinococcus* possess different numbers of HU genes. The genome of *Deinococcus geothermus* contains two HU genes (Dgeo_0175 and Dgeo_2501). We are characterizing *D. radiodurans* and *D. deserti*, which produce one and three HU proteins, respectively (Toueille et al., 2012). DrHU shares, respectively, 89, 76 and 76 % identical residues with the *D. deserti* HU proteins Deide_2p01940 (DdHU1), Deide_3p00060 (DdHU2) and Deide_00200 (DdHU3) over 89 aa residues corresponding to the core HU domain. The HU proteins are among the most highly expressed proteins in *D. radiodurans* and *D. deserti* (de Groot et al., 2009; Toueille et al., 2012). The single HU gene (*DR_A0065*) in *D. radiodurans* is essential for viability, and its gene product DrHU is located all over the nucleoid (Nguyen et al., 2009). When DrHU is expressed from a thermosensitive plasmid in *D. radiodurans*, its progressive depletion at the non-thermopermissive temperature generates DNA decondensation, nucleoid fractionation and finally cell lysis, strongly suggesting that DrHU does have a major role in nucleoid organization and DNA compaction (Nguyen et al., 2009).

Here we investigated if HU is also essential for viability in *D. deserti*, and if so, which of the three HU genes provides this essential function. We also investigated if one or more of the *D. deserti* HU proteins could substitute for the essential DrHU protein in *D. radiodurans*. Each of the annotated HU proteins of *D. deserti*, *D. geothermus* and *D. radiodurans* possesses an N-terminal extension, but these are of different lengths, varying from 26 aa residues for DdHU3 to 50 residues for Dgeo_2501. However, it appeared that the translation initiation codon position for DdHU1 and DrHU (and Dgeo_2501), and hence the actual length and composition of the N-terminal extension, were unclear. Therefore, experiments were performed to identify the true start of DdHU1 and DrHU. We showed that the presence of either DdHU1 or DdHU2 is required for viability of *D. deserti*, and that both DdHU1 and DdHU2 can substitute for DrHU in *D. radiodurans*. The results allowed a correction of the DrHU and DdHU1 start codon positions, resulting in shorter proteins than initially annotated, which was important for the complementation and expression experiments performed. We also found that DrHU and DdHU1 proteins are highly expressed from leaderless mRNA, and that recloning of the corresponding genes in a classical expression plasmid reduced expression in *Deinococcus* to levels that, for DrHU, were insufficient for cell viability.

**METHODS**

**Bacterial strains, growth conditions and transformation.** The strains used are listed in Table S1 (available in the online Supplementary Material). All *D. deserti* and *D. radiodurans* strains were derivatives of strain RD19 and strain R1, respectively. *D. deserti* was grown at 30 °C in 10-fold diluted trypticase soy broth supplemented with trace elements (Vujicˇic´-Zagar et al., 2009), or on plates with the same medium and 1.5 % agar. *D. radiodurans* was grown at 30 °C in TGY2X (1 % tryptone, 0.2 % glucose, 0.6 % yeast extract) or plated on TGY containing 1.5 % agar. *D. radiodurans* was grown at 30 °C in TGYX2 (1 % tryptone, 0.2 % glucose, 0.6 % yeast extract) or plated on TGY containing 1.5 % agar. *E. coli* was grown in Luria–Bertani medium (LB) at 37 °C. Antibiotics were used at the following concentrations for *D. deserti*: streptomycin, 10 μg ml⁻¹; kanamycin, 10 μg ml⁻¹; chloramphenicol, 2 μg ml⁻¹. For *D. radiodurans*: chloramphenicol, 3.5 μg ml⁻¹; spectinomycin, 75 μg ml⁻¹. For *E. coli*: kanamycin, 50 μg ml⁻¹; ampicillin, 100 μg ml⁻¹; spectinomycin, 40 μg ml⁻¹. Transformation of *D. deserti* (Dulermo et al., 2009) and *D. radiodurans* (Bonacossa de Almeida et al., 2002) was performed as described.

**Plasmids and DNA manipulations.** The plasmids used are listed in Table S2. Genomic DNA was extracted from *D. deserti* (Dulermo et al., 2009) and *D. radiodurans* (Noraies et al., 2013) as described. To construct plasmids, PCR products were generally obtained using primers including restriction sites, and cloned in pCRBlunt-TOPO prior to recloning in the desired vectors. All cloned PCR fragments were analysed by DNA sequencing to verify absence of potential PCR errors. The sequences of primers used for amplification of genes or
other DNA fragments, diagnostic PCR or rapid amplification of cDNA ends (5′-RACE) are listed in Table S3. Three different start codons for DrHU and DdHU1 were considered, and M1, M2 and M3 indicate the start of the DNA fragment cloned in plasmid p11559 or p14001 (start codon overlapping the NdeI site CATATG). For the M1 construct of DrHU fused to the SPA-tag, the DR_A0065::spa fragment was amplified from genomic DNA of strain GY13320 and cloned as an NdeI–DraI fragment in p11559. In this fusion protein, the DrHU sequence is followed by the HU DNA was cloned as an NdeI–DraI fragment in p14001, resulting in proteins in which the HU sequence is followed by phenylalanine and lysine (corresponding to the DraI site) and then the SPA-tag.

**Deletion of HU genes in *D. deserti* and *D. radiodurans*. DNA fragments upstream and downstream of genes to be deleted were cloned in the correct orientation respectively upstream and downstream of the antibiotic resistance cassette in pRD0, pSF0kan or pSF0cat, and the resulting plasmids were used to transform *D. deserti* to construct gene deletion mutants by allelic replacement as described by Dulermo et al. (2009). Like *D. radiodurans* (Hansen, 1978), *D. deserti* is multigenomic (Dulermo et al., 2009; Ludanyi et al., 2014). Therefore, single colonies of transformants were streaked three successive times on plates with appropriate antibiotic, which is generally sufficient to obtain a strain with the gene replacement in each copy of the genome unless the gene is essential for viability. If the wild-type allele was still present, strains were restreaked four additional times on selective medium. Complete absence of the gene of interest was verified by diagnostic PCR. Assays to delete native *DR_A0065* in *D. radiodurans* containing p11559 derivatives with HU genes were performed after transformation of these strains with genomic DNA of strain GY13372 (non-homogenized ΔDR_A0065::cat). Transformants were grown and restreaked at least seven times on plates supplemented with chloramphenicol, spectinomycin and 1 mM IPTG. Deletion of DR_A0065 from each genome copy in the resulting strains was analysed by diagnostic PCR.

**5′-RACE.** RNA from *D. deserti* and *D. radiodurans* was isolated from 2 ml of mid-exponential phase cultures using RNAProtect Bacteria Reagent (Qiagen) and the RNeasy Mini kit (Qiagen) following the protocol of the manufacturer, with for *D. deserti* the protocol using enzyme lysis and proteinase K digestion, and for *D. radiodurans* using enzyme lysis, proteinase K digestion and mechanical disruption (Qiagen). 5′-RACE was performed as described by Tillett et al. (2000).

**Western blot analysis of SPA-tagged proteins.** Cell pellets from 20 ml of exponentially growing *D. radiodurans* cultures (OD650 of 0.3–0.4) were obtained by centrifugation and resuspended in 150 μl of SSC1X (150 mM NaCl, 15 mM trisodium citrate, buffered at pH 7). The protein extract was prepared as previously described (Bouthier de la Tour et al., 2013). Samples of 5 μg of proteins were submitted to electrophoresis in a 12% SDS-polyacrylamide gel. Proteins were then transferred to a PVDF membrane (Amersham). The membrane was successively incubated with monoclonal mouse anti-FLAG antibodies (Sigma-Aldrich) and secondary alkaline phosphatase-labelled anti-mouse antibody, and revealed using NBT (nitro blue tetrazolium) and BCIP (5-bromo-4-chloro-3-indolyl phosphate) as a substrate for alkaline phosphatase.

**Fluorescence microscopy.** Cells were grown to an OD650 of 0.3. Aliquots (1 ml) were removed and cells were treated as previously described (Nguyen et al., 2009). DNA and membranes were stained with DAPI (2 μg ml⁻¹) and FM4-64 [N-[3-(triethylammonium-propyl)]-4-(6-(4-(diethylamino)phenyl)hexatrienyl]pyridinium dibromide] (10 μg ml⁻¹), respectively. Stained cells were observed using a Leica DM RXA microscope and images were analysed using ImageJ software.

**RESULTS**

Either DdHU1 or DdHU2, but not DdHU3, is required for viability of *D. deserti*

*D. deserti* produces three different HU proteins (Touille et al., 2012), here called DdHU1, DdHU2 and DdHU3, encoded by genes located on plasmid P2, plasmid P3 and the main chromosome, respectively. Fig. 1 shows a multiple alignment of the HU proteins from *D. deserti*, *D. radiodurans* and *D. geothermalis*. The DNA-intercalating proline, conserved among all HU homologues (Grove, 2011), is present in each of these six HU proteins. According to the initial gene annotation, DdHU1, DdHU2 and DdHU3 contain 125, 121 and 115 aa residues, respectively. Of these, DdHU1 (Deide_2p01940) is most similar to the single and essential HU of *D. radiodurans* (DrHU), encoded by *DR_A0065* on chromosome 2. Shotgun proteomics of isolated nucleoids indicated that DdHU1 is also most abundant, with about five and ten times more spectral counts than obtained for DdHU2 and DdHU3, respectively (Touille et al., 2012). To investigate if HU is also essential for viability in *D. deserti*, we attempted to construct various HU deletion mutants. It was possible to completely delete Deide_2p01940 from the genome (Fig. S1), showing that DdHU1 is not required for viability in an otherwise wild-type background. Similarly, either Deide_3p00060 (DdHU2) or Deide_00200 (DdHU3) could be entirely deleted in *D. deserti* (Fig. S1). Two different antibiotic-resistance gene cassettes are available to obtain gene deletion mutants in *D. deserti* (Dulermo et al., 2009). We therefore tried to construct double mutants for the HU genes. Strains with deletions of Deide_2p01940 and Deide_00200 or Deide_3p00060 and Deide_00200 were obtained, but generating a strain with a deletion of both Deide_2p01940 and Deide_3p00060 failed (Fig. S1), indicating that HU is required for viability in *D. deserti* and that the essential function is provided by either DdHU1 or DdHU2 but not by DdHU3.

**DdHU1 can functionally replace DrHU in *D. radiodurans***

The viability or non-viability of the different *D. deserti* HU mutants may be related to differences in expression levels of the three HU proteins. To investigate this, each HU gene of *D. deserti* was cloned under control of the strong IPTG-inducible Psac promoter on plasmid p11559, which can be introduced in *D. radiodurans*. It has been shown previously that DrHU production from a similar expression plasmid in *D. radiodurans* allows deletion of the native chromosomal DrHU gene (*DR_A0065*), which is otherwise essential for viability (Nguyen et al., 2009). The possibility to delete native *DR_A0065* in a strain that
contains an HU gene on p11559 thus indicates that the plasmid-derived HU is functionally expressed.

The various HU genes are cloned in the NdeI–DraI sites located downstream of the Pspac promoter in p11559 in such a way that the translation initiation codon overlaps with ATG of the NdeI site (CATATG). A Shine–Dalgarno sequence is present upstream of the NdeI site (Lecointe et al., 2004; Mennecier et al., 2004). We first performed the experiment with p11559 carrying the annotated gene for the 125-residue-long DdHU1 protein (start corresponding to M2 in Fig. 1). In transformed D. radiodurans cells containing this construct we observed a failure to delete DR_A0065 from chromosome 2 (Fig. 2c). In the control experiment with the p11559 derivative encoding the annotated 137-residue-long DrHU protein (starting with M1 in Fig. 1), DR_A0065 could be deleted from the chromosome (Fig. 2b). These results indicate that DdHU1 may not be functional in D. radiodurans.

However, after inspection of the sequence upstream of the annotated start codon of DdHU1, we found an in-frame CTG codon. With this start codon, DdHU1 would be 13 residues longer and have increased similarity to the annotated DrHU and Dgeo_2501 proteins (Fig. 1). A CTG translation initiation codon is rare, but a CTG start codon was previously found and confirmed in D. deserti for rpsL encoding a ribosomal protein (Baudet et al., 2010). Also for DrHU and Dgeo_2501 the annotated translation initiation codon ATG is not the most frequently used, but TTG. We therefore cloned a DNA fragment for a 138-residue-long DdHU1 in p11559 (starting with M1). With this construct in D. radiodurans it was possible to delete DR_A0065 (Fig. 2b), showing that DdHU1 can substitute for DrHU to support cell viability. Moreover, the morphology of the nucleoids in the resulting strain was indistinguishable from that of the control strains (Fig. 3). As observed for the annotated DdHU1 gene (start at M2), deletion of native DR_A0065 failed when p11559 carried a fragment for an 11-residue-shorter DrHU protein (starting with M2) (Fig. 2c). Therefore, the data also suggest that the additional residues at the N terminus are important for the function of HU, and that DdHU1 should be longer than initially annotated. However, as described below, both assumptions appeared to be incorrect.

---

Fig. 1. Multiple sequence alignment of HU proteins of D. deserti, D. radiodurans and D. geothermalis. Initially annotated N-terminal residues are indicated with a grey box. Other potential starts are shown with a black box. M1, M2 and M3 indicate the three potential starts of DrHU, DdHU1 and Dgeo_2501. Triangles point to the DNA-intercalating proline conserved among all HU homologues, and to a glycine associated with thermostability in HU from thermophiles. Theoretical pI values are given at the end of each sequence. The core HU domain is overlined (thick line). Identified tryptic peptides closest to the N terminus are underlined and in bold.
What is the correct start codon position for DrHU and DdHU1?

Although HU proteins from *D. radiodurans* and *D. deserti* have been very well detected in various proteomic studies, tryptic peptides corresponding to their N terminus have not been identified, probably because of the many lysine residues in the N-terminal region (Baudet *et al.*, 2010; de Groot *et al.*, 2009; Toueille *et al.*, 2012). The identified tryptic peptides closest to the N terminus were TQLVEMVADQTGLTK for DrHU and GAVAAESNKVAK for DdHU1 (and IGKTQVYMDIADR and TQLIEMVADR for DdHU2 and DdHU3, respectively) (Fig. 1). Furthermore, attempts to purify native DrHU or a C-terminally tagged DrHU fusion protein expressed from its native locus were not successful. We therefore performed other experiments to obtain support that translation of DrHU and DdHU1 starts at M1, as suggested by the results in the previous section.

We first made a pair of constructs in p11559 that again contained the HU-encoding DNA fragment in which the NdeI site overlapped with the M1 start codon for DrHU and DdHU1, but this ATG was now immediately followed by an in-frame stop codon. Plasmids were introduced in *D. radiodurans*, and attempts to delete native DR_A0065 were performed. Unexpectedly, with both these constructs it was possible to delete DR_A0065 (Fig. 2d), showing that HU was still functionally expressed and, importantly, that M1 is not the start of DrHU and DdHU1, which is in

---

**Fig. 2.** Assays to delete native DR_A0065 in *D. radiodurans* containing plasmids with various DrHU or DdHU1 constructs. (a) Schematic drawing of the PCR products to verify DR_A0065 deletion. If DR_A0065 can be entirely deleted, the 593 bp PCR fragment (wild-type) will be absent and replaced by a 1254 bp PCR product. (b–f) Schematic drawings of different constructs and results of deletion assays. Both for DrHU and for DdHU1, five constructs were made after cloning in expression vector p11559. Only 5′-ends of cloned genes (thick lines) are indicated, corresponding to the region with the three potential start codons of DdHU1 and DrHU. PCR results for the strains with the plasmids containing DrHU genes are shown in the gels on the left, those with DdHU1 constructs on the right. DR_A0065 deletion was analysed in six different transformants. The lane on the right of each gel shows the 593 bp PCR product obtained with the wild-type strain R1. Marker DNA fragments are at the left of each gel.
contrast to the indications obtained in the previous section. To analyse this further, a second pair of constructs was made, which was the same as the first one except that the DNA fragments also contained a mutation at the M2 codon of DdHU1 and DrHU, with the ATG changed into TAG. Obtaining a deletion of native DR_A0065 failed with both these constructs (Fig. 2e), indicating that this mutation abolished functional expression of HU. This could mean that M2 is the start for DrHU and DdHU1, but this is not supported by the results described above because a DR_A0065 deletion was not obtained when the cloned fragments started with the ATG for M2 (Fig. 2c).

To gain more insight into these contradictory results, we analysed the expression of the HU proteins. For this, a p11559 derivative was used to express DrHU or DdHU1 fused at the C terminus with an SPA-tag, which can be detected by Western blotting. The cloned DNA fragments started with the ATG corresponding to either M1 or M2. As a control we used a previously constructed D. radiodurans strain that expresses the DrHU-SPA fusion from its native locus. No expression of the M2 DdHU1-SPA and DrHU-SPA proteins was detected in D. radiodurans (Fig. 4a, lane 3 and 6), indicating that the AUG for M2 is not used for translation initiation with these constructs and explaining why the deletion of native DR_A0065 was not obtained when the cloned fragments started with the ATG for M2 (Fig. 2c).

Remarkably, two proteins were detected in the strain containing the M1 construct for DrHU-SPA (Fig. 4a, lane 2), with the lower band corresponding to the size of the fusion protein expressed from the native DrHU locus (Fig. 4a, lane 8). We assume that the upper band corresponds to the protein starting with M1, and that the lower band results from translation initiation at the true start codon. Only one band for the M1 construct encoding DdHU1-SPA was clearly observed (Fig. 4a, lane 5). As the DrHU and DdHU1 sequences are very similar, it is possible that translation initiation of this DdHU1-SPA protein occurred at the position equivalent to the true start for DrHU. As the results show that the true start for DrHU and DdHU1 is not M1 (the annotated start for DrHU) and do not support that it corresponds to M2 (the annotated start for DdHU1), the true start could be M3, which corresponds to another ATG conserved in the genes encoding DrHU, DdHU1 and Dgeo_2501 (Fig. 1). To test this, p11559 derivatives containing DNA fragments starting with the ATG corresponding to M3 were made and introduced in D. radiodurans. The plasmid carrying the gene for DdHU1 allowed deletion of DR_A0065 (Fig. 2f). Surprisingly, however, deletion of DR_A0065 failed with the equivalent construct for DrHU (Fig. 2f). After cloning the fragments starting with the ATG for M3 in the vector for generating SPA-tag fusions, Western blotting (Fig. 4a, lane 7) showed expression of DdHU1-SPA with a size indistinguishable from that observed with the construct that started with ATG for M1 (Fig. 4a, lane 5). With the equivalent construct for DrHU, expression of DrHU-SPA was also observed (Fig. 4a, lane 4), but in lower amounts compared with DdHU1-SPA, which is a likely explanation for the failure to delete native DR_A0065 in this case. Taken together, the results suggest that M3 is the true start of DdHU1 and DrHU.

DrHU, DdHU1 and DdHU2 are translated from leaderless mRNAs

When this work was in progress, RNA sequencing data for D. deserti became available (de Groot et al., 2014). Correlating with the observed higher levels of DdHU1 protein compared with DdHU2 and DdHU3 (Toueille et al., 2012), the number of cDNA reads obtained for Deide_2p01940 was about five times higher than that for either Deide_3p00060 or Deide_00200. The RNA sequencing
included the genome-wide identification of transcription start sites (TSSs). Strikingly, 60% of the identified *D. deserti* mRNAs were found to be leaderless, i.e. lacking a 5’-untranslated region (UTR) and, hence, a Shine–Dalgarno sequence. The AUG or GUG at the 5’-end of leaderless mRNAs is the start codon for translation (Benelli & Londei, 2009; Brock et al., 2008; de Groot et al., 2014; Hering et al., 2009; O’Donnell & Janssen, 2002). The TSS for DdHU3 was found at 41 nt upstream of the annotated start codon. Interestingly, the mRNAs for DdHU1 and DdHU2 were found to be leaderless. The TSS for DdHU2 is located at the first nucleotide of the annotated ATG start codon, and that for DdHU1 at the first nucleotide of the third potential start codon (M3 in Fig. 1).

Independent 5’-RACE confirmed the TSS for DdHU1. A 5’-RACE experiment was also performed to identify the TSS for DrHU. As for DdHU1, the 5’-end of the mRNA was found to correspond to the first nucleotide of the third potential start codon (M3). From these data we conclude that the start codon position for DdHU1 and DrHU has to be reannotated, with the new translation initiation codon corresponding to M3 in Fig. 1. A similar start codon correction can be predicted for Dgeo_2501, giving a protein that is 16 residues shorter than initially annotated (Figs 1 and 5).

With the identified TSSs, the promoter-containing regions immediately upstream of the TSSs are also known. A conserved motif resembling the −10 promoter element

---

**Fig. 4.** Western blots detecting DrHU-SPA and DdHU1-SPA fusion proteins in *D. radiodurans*. Strains and constructs are indicated at the bottom of the blot. Corresponding samples in a stained SDS-PAGE gel are shown in the top panels. M1, M2 and M3 indicate the start of the cloned fragment in the expression plasmid and correspond to the three potential starts of DrHU or DdHU1. V, strain with ‘empty’ p14001. Chrom DrHU-SPA, strain expressing DrHU-SPA from the native DrHU locus. R1, wild-type. (a) Strains were grown in the presence of 1 mM IPTG. (b) Strains were grown with (+) or without (−) 1 mM IPTG.
TATAAT was found upstream of 94% of the TSSs in *D. deserti* (de Groot et al., 2014). The identified TSSs explain some of the results that were obtained with DrHU and DdHU1 constructs in plasmid p11559. It is now clear that cloned fragments starting with ATG for M1 in fact contain the native promoter of the DrHU- and DdHU1-encoding genes (Fig. 5). Indeed, when IPTG was omitted to repress the Pspac promoter in cells containing these plasmids, HU was still expressed, with only the smaller protein for DrHU (Fig. 4b, lanes 2 and 6). Furthermore, the constructs containing the mutation of the second potential start codon (M2), with which deletion of native DR_A0065 was unsuccessful (Fig. 2e), have in fact a mutation in the predicted −10 element that would abolish transcription of the leaderless mRNAs encoding DrHU and DdHU1 (Fig. 5).

Interestingly, the Western blots also indicate that DrHU and DdHU1 proteins are more efficiently expressed when the genes are under control of their own promoter compared with the IPTG-induced Pspac promoter (Fig. 4b, compare lane 2 with 3, and lane 6 with 7). This may be related to stronger promoters of the genes encoding DrHU and DdHU1, and also to the fact that DrHU and DdHU1 are translated from leaderless mRNA when under control of their own promoter, whereas the mRNA for HU contains a 5′-UTR when the DNA fragment cloned in p11559 starts with the ATG corresponding to M3. An even higher amount of DrHU was observed when expressed from the native locus on chromosome 2, suggesting that DdHU3, in contrast to DdHU1 and DdHU2, is too different from DrHU to be able to replace the latter in *D. radiodurans*. Located over the entire length of HU, there are various residues that are conserved in DrHU, DdHU1 and DdHU2 but different in DdHU3 (Fig. S2). One of these differences concerns a glycine residue in a region corresponding to a loop between two alpha helices in the HU structure. This glycine has been associated with loop flexibility, optimal helix packing and increased thermostability of HU homologues from thermophiles (Grove, 2011; Kawamura et al., 1996). A corresponding glycine is present in DrHU, DdHU1, DdHU2 and Dgeo_2501, but absent from DdHU3 and Dgeo_0175 (Fig. 1). However, no enhanced thermal stability was observed for the DrHU core domain compared with HU from mesophilic organisms (Ghosh & Grove, 2006).

**DISCUSSION**

Leaderless mRNAs are rare in *E. coli* and most other characterized bacteria. Moreover, leaderless mRNAs, which lack a 5′-UTR and, hence, a Shine–Dalgarno sequence involved in ribosome binding, are less efficiently
translated than canonical leadered mRNAs in *E. coli* (Malys & McCarthy, 2011; Moll et al., 2002; O'Donnell & Janssen, 2001; Van Etten & Janssen, 1998). Our data, however, show that the highly expressed and essential HU proteins in *D. deserti* and *D. radiodurans* are translated from leaderless mRNA, with the transcription start site exactly at the first nucleotide of the ATG translation initiation codon. One could argue that translation might start at a possible start codon downstream of the 5'-AUG. However, shotgun proteomics revealed tryptic peptides for DrHU, DdHU1 and DdHU2 that start upstream of the first possible start codon downstream of the 5'-AUG (Fig. 1). Moreover, an AUG (or GUG) at the 5'-end of an RNA molecule is the signal that is required and sufficient for ribosome binding and translation initiation at that start codon (Benelli & Londei, 2009; Brock et al., 2008; Hering et al., 2009; O'Donnell & Janssen, 2002). Indeed, an N-terminal peptide was identified for the products of 167 different leaderless mRNAs in *D. deserti*, and all these peptides corresponded to the 5'-AUG or 5'-GUG translation initiation codon (de Groot et al., 2014).

An unexpected observation of our experiments was that expression of DdHU1 and especially DrHU was strongly decreased when the genes were cloned under control of Pspac, a strong promoter in *D. radiodurans* (Lecointe et al., 2004). DrHU and DdHU1 are required in high amounts for their global role in nucleoid organization, and we assume that these proteins are optimally expressed from their leaderless mRNA and own promoter in *Deinococcus*. Changing a leaderless mRNA into a leadered mRNA, which occurs when a leaderless gene is cloned under control of Pspac in the expression plasmid, may affect translation initiation efficiency. As observed for DrHU, this may result in protein expression levels that are insufficient for its required function as a nucleoid-structuring protein. In addition to the genes encoding HU, many more leaderless genes were found or predicted in *D. deserti* and *D. radiodurans* (de Groot et al., 2014; Zheng et al., 2011). Bacterial ribosomes are not homogeneous entities always containing the same complement

---

**Fig. 6.** Assays to delete native *DR_A0065* in *D. radiodurans* containing a plasmid with the gene encoding DdHU2 (a) or DdHU3 (b). Deletion of *DR_A0065* was verified by PCR as described in Fig. 2. The lane on the right of the gel in (a) shows the 593 bp PCR product obtained with wild-type strain R1. Marker DNA fragments are indicated at the left of each gel.

**Fig. 7.** Western blots detecting DdHU2-SPA and DdHU3-SPA fusion proteins in *D. radiodurans*. Samples with DrHU-SPA expressed from p11559 (M1 construct) and from the native locus (Chrom DrHU-SPA) on chromosome 2 were included. Strains and constructs are indicated at the bottom of the blot. The corresponding samples in a stained SDS-PAGE gel are shown in the top panel. V, strain with ‘empty’ p14001. Strains were grown in the presence of 1 mM IPTG.
**Fig. 8.** Multiple sequence alignment of HU proteins from *Deinococcus*. Only N-terminal residues of the aligned proteins are shown. The N terminus of DdHU1, DrHU and Dgeo_2501 is corrected. Other *Deinococcus* HU proteins were found by BLASTP. If no locus tag was available, the species name is indicated, followed by _1, _2 or _3 if more than one HU was found. HU proteins from *E. coli* (EC_hupA and EC_hupB), *B. subtilis* (BS_hupA) and *Thermus thermophilus* (TT_C0984) are included. Gaps in the alignment were removed in the N-terminal region, except for one between the N-terminal extension and the residue corresponding to the start of *E. coli* HU. Annotated, corrected or probable start codons of *Deinococcus* HU proteins are aligned. Residues in italics for Deipr_2157m and DGo_CA0549m were found after analysis of the region upstream of the annotated start. The indicated TSS was identified for DdHU1, DdHU2 and DrHU.
of rRNAs and ribosomal proteins. Instead, evidence has been obtained for variations in the rRNA and protein components, resulting in functionally specialized ribosomes (Byrgazov et al., 2013). Therefore, Deinococcus might contain an important subpopulation of ribosomes that efficiently translate leaderless mRNAs.

Our experimental data revealed that the start codon position for DrHU and DdHU1 has to be reannotated, resulting in proteins that are 15 and 4 aa residues shorter than initially reported. Recently, (partial) genome sequence data of many more Deinococcus species have been released in databases, resulting in an increased number of known deinococcal HU protein sequences. The N terminus of several of these new HU proteins probably also requires reannotation resulting in shorter proteins (Fig. 8). Two other HU proteins (i.e. Deipr_2157 and DGo_CA0549) may be longer than initially annotated (Fig. 8). Although obvious and already reviewed (Hartmann & Armengaud, 2014), we further document that correct annotation of start codon positions is crucial for appropriate genetic and biochemical studies. Studying genes or proteins with incorrect start codon position may result in protein expression problems and, if expressed anyway, may change the properties of the protein, leading to incorrect conclusions. Moreover, with the identification of the correct start codon position the actual region upstream of the coding sequence is also delineated, which is important for recognizing sequences involved in translation, transcription or regulation. The corrected DrHU protein sequence still contains a Lys- and Ala-rich N-terminal extension, but we cannot exclude that the 15 additional residues that were present in recombinant DrHU purified from E. coli did not influence the results of previously published in vitro experiments that showed modulation of the DNA-binding properties by the N-terminal extension (Ghosh & Grove, 2004, 2006). As the presence of a Lys- and Ala-rich region is highly conserved at the N terminus of most deinococcal HU proteins (Fig. 8) and at the C terminus of HU in some other bacterial species, and as this region is similar to the eukaryotic histone H1 C-terminal domain essential for DNA condensation, the N-terminal extension is probably important for the in vivo function of HU in Deinococcus. We wanted to investigate this, but only a very low level of expression of truncated DrHU lacking the extension was observed from the cloned fragment under control of the Pspac promoter in D. radiodurans, and obtaining a deletion of the native HU gene was unsuccessful in this strain (data not shown).

The annotated or probable N-terminal residue of many HU proteins is followed by a threonine residue (Fig. 8), which will become the first residue of the polypeptide after removal of the N-terminal methionine by methionine aminopeptidase. This is in line with a recent protein N-termini survey that showed that many deinococcal proteins (23 % in D. deserti) have Thr at position 2 of the protein sequence (A. Dedieu, A. de Groot & J. Armengaud, unpublished data). Strikingly, many of these N-terminal Thr residues were found to be acetylated in D. deserti (A. Dedieu, A. de Groot & J. Armengaud, unpublished data). It is thus possible that acetylation could also occur at the N-terminal Thr residue of matured HU proteins, a post-translational modification that might be important for their function. Many other HU protein sequences have an alanine residue at position 2 (Fig. 8). N-terminal acetylation of Ala was also observed in D. deserti, but less frequently compared with Thr (A. Dedieu, A. de Groot & J. Armengaud, unpublished data). N-terminal acetylation is found in a majority of yeast and mammalian proteins, including histones. Recently, it was found that the N-terminal acetylation of yeast histone H4 inhibits the occurrence of a neighbouring modification, cooperates with other H4 modifications and is involved in gene regulation (Schiza et al., 2013). Therefore, further studies aimed at analysing the dynamics of post-translational modifications of HU proteins should be conducted.

Why some Deinococcus species possess more HU homologues than other species is still an open question. For the entirely sequenced Deinococcus genomes, we did not find any correlation between genome size and number of HU genes. For example, the genome size/number of HU genes is 3.28 Mb/1 for D. radiodurans, 3.86/3 for D. deserti, 3.22 Mb/2 for D. geothermalis, 2.89 Mb/3 for Deinococcus proteolyticus and 4.51/2 for Deinococcus peraridilitoris. D. radiodurans seems to be the exception with only one HU gene and absence of an HU gene on the main chromosome, while the others have more than one HU gene with at least one of these located on the main chromosome. Of the three HU proteins from D. deserti, DdHU1 and DdHU2 can functionally replace DrHU in D. radiodurans. As D. deserti with only DdHU3 (i.e. lacking DdHU1 and DdHU2) is not viable, and because DdHU3 cannot substitute for DrHU in D. radiodurans, DdHU3 may have a more specialized but non-essential function compared with DdHU1 and DdHU2. HU is not an exception in having a variable number of homologues in different Deinococcus species, as this is also found for several other important genes, including recA, ddrO, csp, dps and ligA (de Groot et al., 2009; Dulerimo et al., 2009; Ludanyi et al., 2014), revealing interesting differences among closely related Deinococcus species.

ACKNOWLEDGEMENTS

We thank S. Fochesato for construction of plasmids pSFOkan and pSF0cat, D. Laloue for advice about 5′-RACE and D. Pignol for critical reading of the manuscript. This work was supported by the Commissariat à l’Energie Atomique et aux Energies Alternatives (CEA), the Université Paris-Sud, the Centre National de la Recherche Scientifique, the Agence Nationale de la Recherche (ANR-07-BLAN-0106-02 to A. d. G., J. A. and S. S., ANR 11BSV3-01701 to S. S.) and Electricité de France (RB2011-17 to S. S.).

REFERENCES

Baudet, M., Ortet, P., Gaillard, J. C., Fernandez, B., Guérin, P., Enjalbal, C., Subra, G., de Groot, A., Barakat, M. & other authors...


