Making heads or tails of the HU proteins in the planctomycete *Gemmata obscuriglobus*

Benjamin Yee, Evgeny Sagulenko and John A. Fuerst

The University of Queensland, School of Chemistry and Molecular Biosciences, St Lucia, Queensland 4072, Australia

Correspondence

Benjamin Yee

richard_st_vier@hotmail.com

Received 14 December 2010

Revised 4 April 2011

Accepted 19 April 2011

Gemmata obscuriglobus has a highly condensed nucleoid which is implicated in its resistance to radiation. However, the mechanisms by which such compaction is achieved, and the proteins responsible, are still unknown. Here we have examined the genome of *G. obscuriglobus* for the presence of proteins homologous to those that have been associated with nucleoid condensation. We found two different proteins homologous to the bacterial nucleoid-associated protein HU, one with an N-terminal and one with a C-terminal extension relative to the amino acid sequence of the HU found in *Escherichia coli*. Sequence analysis revealed that one of these HU homologues represents a novel type with a high number of prolines in its C-terminal extension, whereas the other one has motifs similar to the N terminus of the HU homologue from the radio-resistant bacterium *Deinococcus radiodurans*. The occurrence of two such HU homologue proteins with these two different terminal extensions in one organism appears to be unique among the Bacteria.

INTRODUCTION

Architectural chromosomal proteins are best known for their role in the compaction of DNA in a cell. The compaction of the nucleoid is an important cellular process in all living cells, and is achieved via a combination or the single mode of action of bending, wrapping and bridging DNA. The architectural proteins are apparently functionally conserved but share neither sequence nor structural similarity (Luijsterburg *et al.*, 2008). In eukaryotes, a massive reduction in volume of the genome is achieved predominantly via the action of having the DNA wrapped around histone complexes at specific regions called histone-fold domains (Luger *et al.*, 1997). In bacteria, several proteins work in tandem to condense the bacterial nucleoid (Azam & Ishihama, 1999; Luijsterburg *et al.*, 2006; Dillon & Dorman, 2010). Some of these proteins, such as the HU protein and histone-like nucleoid structuring protein (HN-S), have histone-like characteristics such as a similar amino acid composition and the ability to condense DNA into a densely packed nucleoid structure upon overexpression of the protein (Rouvière-Yaniv & Gros, 1975; Rouvière-Yaniv *et al.*, 1979; Spurio *et al.*, 1992).

Members of the phylum Planctomycetes within the domain Bacteria have many uncommon features, both structural and molecular. All planctomycetes examined display a compartmentalized cell plan as well as a much more condensed nucleoid compared with most other bacteria, and in the case of *Gemmata obscuriglobus*, the nucleoid is

further bounded within a double-membrane envelope (Fuerst & Webb, 1991; Lindsay *et al.*, 2001; Fuerst, 2005). A recent study of *G. obscuriglobus* has revealed resistance to UV radiation 40 times greater than that of *Escherichia coli*, and this increased tolerance is proposed to be associated with a highly effective DNA repair mechanism at least partly due to the condensed chromatin structure in this species (Lieber *et al.*, 2009). However, the identity and mechanism of the proteins responsible for the condensation of the nucleoid in *G. obscuriglobus* are still unclear. The relationships within the Planctomycetes, Verrucomicrobia, Chlamydiae (PVC) superphylum (Wagner & Horn, 2006) could be significant in the search for the nucleoid compaction protein in *G. obscuriglobus*, because chlamydiae encode a functional histone-like protein analogous to histone H1 of eukaryotes (Barry *et al.*, 1992; Remacha *et al.*, 1996).

An *in silico* study of the histone-like proteins identified from the genome of *G. obscuriglobus* was performed and the functional implications are discussed.

METHODS

Homology search. Query amino acid sequences of histones and histone-like proteins from eukaryotes, bacteria and archaea were used in similarity searches via the TBLASTN version of BLAST against the TBLASTN translations of the *G. obscuriglobus* genome within the Whole-Genome Shotgun Reads (WGS) database at the National Center for Biotechnology Information (NCBI) website.

Sequence analysis. Alignment of HU protein sequences was performed using CLUSTAL_X v2.0 (Thompson *et al.*, 1997). Secondary structure prediction was done with the Jpred prediction

Abbreviations: HN-S, histone-like nucleoid structuring protein; LGT, lateral gene transfer; PVC superphylum, Planctomycetes, Verrucomicrobia, Chlamydiae superphylum.

server (<http://www.compbio.dundee.ac.uk/jpred>) (Cuff *et al.*, 1998; Cole *et al.*, 2008). Similarity matrices were created using the PHYLIP package (Felsenstein, 1989). Codon usage was tabulated using the Codon Usage Tool at the Gene Infinity website (http://www.geneinfinity.org/sms/sms_codonusage.html).

Phylogenetic analysis. A suitable substitution model based on the protein alignment was determined by the program ModelGenerator (Keane *et al.*, 2006). A likelihood tree was created using the TreeFinder program (Jobb *et al.*, 2004).

RESULTS

Homology search

A homology search was performed on the draft genome of *G. obscuriglobus* to search for homology matches using histones and histone-like proteins from eukaryotes, bacteria and archaea as query sequences, and the results are presented in Table 1. Query sequences of the HU proteins from the thermophilic archaeon *Thermoplasma acidophilum*, the chloroplast of the red alga *Cyanidioschyzon merolae* from the eukaryotes, as well as one of two HU proteins from the spore-producing bacterium *Streptomyces coelicolor*, all yielded matches with significant E-values to an annotated 'probable DNA-binding protein HU' (ZP_02732218), herein termed GoC. When the *G. obscuriglobus* genome was queried with HU proteins from bacteria other than *E. coli*, there were also matches of lower significance (expected value of e^{-3} or above) for query HU proteins from *Deinococcus* species with

the *G. obscuriglobus* hypothetical protein GobsU_35800 (ZP_02737229) (hevein termed GoN), and for the histone-like protein (Hlp) from *Mycobacterium tuberculosis* and an HU protein from *S. coelicolor* with a *G. obscuriglobus* protein annotated as a probable DNA-binding HU protein distinct from GobsU_35800. In addition to HU protein homologues, we also identified a homologue to the DNA starvation/stationary phase protection protein (Dps). However, there were no significant homology search matches with any of the eukaryote or archaeal histones, or HN-S.

We noticed that the sequence lengths of the two HU proteins from *G. obscuriglobus* were much longer than that found in *E. coli* and were more similar to the length of the HU proteins from *Deinococcus* and *Streptomyces*. The sequences of GoN and GoC were amplified from genomic DNA of *G. obscuriglobus* using specific primers and confirmed by sequencing to determine that the length difference was not an artefact of genome sequencing.

Sequence analysis

Based on the alignment of HU proteins (Fig. 1), there are three different types, as recently reviewed (Grove, 2010): the type occurring in *Deinococcus*, with an N-terminal extension; the type in actinobacteria, with a C-terminal extension; and lastly, the typical type found in *E. coli*, with just the core HU domain. The core HU region of both HU proteins from *G. obscuriglobus* is similar to that of *E. coli* HU in terms of amino acid composition, with 12 % lysine

Table 1. Results of a homology search for histones and histone-like proteins in the draft genome of *G. obscuriglobus*

Organism	Query protein	E-value	Homologue in <i>G. obscuriglobus</i>
<i>Cyanidioschyzon merolae</i>	CmHupA	6e-8	Probable DNA-binding protein HU
<i>Cryptocodinium cohnii</i>	HCC3	2.8	WD-40 repeat protein
	HCC4	2.9	Geranylgeranyl pyrophosphate synthase
	HCC4	2.9	Geranylgeranyl pyrophosphate synthase
<i>Saccharomyces cerevisiae</i>	Histone H3	0.49	PAS/PAC sensor hybrid histidine kinase
	Histone H2A	6.5	No match to any protein
	Histone H2B	0.15	No match to any protein
	Histone H4	2	No match to any protein
	Histone H1	—	No match to any protein
<i>Chlamydia trachomatis</i>	Hc1	—	No match to any protein
	Hc2	5.1	No match to any protein
<i>E. coli</i>	HN-S	0.4	No match to any protein
	Dps	7e-28	DNA starvation/stationary phase protection
	HU-A	1.5	No match to any protein
	HU-B	1	No match to any protein
<i>D. radiodurans</i>	DrHU	0.005	Hypothetical protein GobsU_35800
<i>Deinococcus geothermalis</i>	DgHU	0.002	Hypothetical protein GobsU_35800
<i>M. tuberculosis</i>	Hlp	0.017	Probable DNA-binding protein HU
<i>S. coelicolor</i>	HupS	0.015	Probable DNA-binding protein HU
	HS1	3e-4	Probable DNA-binding protein HU
<i>Methanothermus fervidus</i>	HmfA	6.2	—
	HmfB	1.4	Excinuclease ABC subunit C
<i>T. acidophilum</i>	Hta	1e-6	Probable DNA-binding protein HU
<i>Cenarchaeum symbiosum</i>	HisA	2.7	—

Fig. 1. Sequence alignment of bacterial HU proteins. Secondary structure based on that of *E. coli* is indicated as solid bars above the alignment: helical segments are indicated by red bars, β -sheets by blue bars. The conserved DNA-intercalating proline is identified by an arrow. Asterisks highlight the three thermotability-associated residues in *Bacillus subtilis*, *Geobacillus stearothermophilus* and *Thermotoga maritima* (Christodoulou & Vorgias, 2002).

A more detailed examination of the sequences of the two *Gemmata* HU-like proteins revealed a (S/P/K)AAK motif repeat in the lysine- and alanine-rich region of the N terminus of GoN, which was highly similar to the extensions found in DrHU and Hlp. The sequence of the 70 aa C terminal of GoC does not bear any motif structure homologous to that of any of the HU proteins, but was observed to contain the slightly higher number of nine proline residues when compared with the seven present in the 130 aa C terminal of HupS. The lack of similarities to the N-terminal versions, the core-only versions, and the lysine- and alanine-rich C-terminal versions in *Streptomyces* means that GoC represents a novel type of HU protein.

The fully sequenced planctomycete *Rhodopirellula baltica* was also found to possess two types of HU protein, which indicated that such a feature might be conserved within the phylum Planctomycetes. Therefore, sequences of HU proteins from members of the Planctomycetes and the related phylum Verrucomicrobia were retrieved from the NCBI database and used in the alignment with the two HU proteins from *G. obscuriglobus*, and the results are presented in Fig. 2. Based on the alignment, all members of the Planctomycetes have two types of HU-like protein, with either an N-terminal extension or a C-terminal extension of varying length. Similarly, two representatives of the closely related phylum Verrucomicrobia, *Verrucomicrobium spinosum* and the acidophilic *Methylococcus thermophilus*, also appear to have two types of HU protein, although with a shorter N-terminal extension. It is interesting that none of the other planctomycetes has the discernible motif repeat (S/P/K)AAK mentioned above within the sequence of the N-terminal extensions, with the exception of *Gemmata* sp. strain Wal-1, suggesting a possible characteristic specific to the *Gemmata* group.

As observed above for the C-terminal region of GoC, the C-terminal extensions of all planctomycetes also appear to be rich in proline (up to 12 in *Planctomyces maris*), whereas *V. spinosum* does not contain such a high number of

Fig. 2. Alignment of HU-like proteins of planctomycetes and verrucomicrobia in comparison with deinococcal and actinobacterial HU proteins. Red bars to the left of the figure represent the planctomycetes N-terminus type of HU, whereas the blue bars denote the C-terminus type of HU. The black box marks the (S/P/K)AAK motif in the N terminus of *Gemmata* sp. The red box shows the high number of prolines concentrated at the end of the C-terminus-type planctomycetes HU.

prolines and has a much shorter C-terminal region. Therefore, it would appear that all budding peptidoglycan-less species of the Planctomycetes have such a proline-rich C-terminal region in this version of the HU protein, unlike *V. spinosum*, related to the Planctomycetes within the PVC superphylum but possessing peptidoglycan cell walls and an FtsZ-based cell division mechanism. This suggests the possible correlation of such proline-rich

regions with the unusual compartmentalization and cell division cycle of planctomycetes.

Phylogenetic analysis and lateral gene transfer (LGT)

A phylogenetic tree constructed from an alignment of the core regions only of HU proteins from members of

different phyla is presented in Fig. 3. It is clear that the overall topology of an HU protein tree does not conform to typical 16S gene trees, likely due to the above-mentioned fact that architectural proteins such as HU and histones have limited sequence conservation, and are often conserved at the functional level only (Luijsterburg *et al.*, 2008). Planctomycete HU sequences of the same terminus type cluster together with significant bootstrap support but otherwise do not display any homology between different terminus types, as shown by the clear separation of the two

clades containing the planctomycete HU protein sequences.

Interestingly, the HU proteins from the proteobacterial species *Xylella fastidiosa*, *Xanthomonas campestris*, *Methylosinus trichosporium* and *Teredinibacter turnerae* form a clade together with the N-terminal HU proteins of the Planctomycetes with significant bootstrap support. These proteobacterial HU proteins have a weak signal for the (S/P/K) AAK motif in their N-terminal region and form a cluster

Fig. 3. Phylogenetics of HU proteins from representatives of different phyla. Planctomycetes HU proteins with an N-terminus domain are coloured blue, whereas those with a C-terminus domain are coloured red.

separate from the other proteobacterial HU proteins such as those of *E. coli* and *Pseudomonas aeruginosa*. Of great interest is the sequence from the endosymbiotic marine proteobacterium *Teredinibacter turnerae* (Distel *et al.*, 2002; Yang *et al.*, 2009), which contains an N-terminal domain which is lysine- and alanine-rich and consists of PAKK repeats, a feature very similar to that of GoN. The C-terminal types of Planctomycetes form a significant cluster away from GoN, superficially forming a clade with other bacteria from unique environments, such as the deep-branching *T. maritima* and *Aquifex aeolicus*, and also the green sulphur bacteria group. More importantly, they form a distinct clade away from the C terminus-containing actinobacterial histone-like proteins of streptomycetes and mycobacteria, supporting the earlier suggestion that the C-terminus types of planctomycetes are a novel variety of HU protein.

The presence of two distinct HU-like proteins within an organism raises questions about the origin of these genes. Our study has shown that such a feature is preserved within the phylum Planctomycetes, although the distinct separation between the two types of HU protein hints at the possibility of LGT. The GC content and codon usage of GoN and GoC were compared with those of five housekeeping genes (RpoB, GyrA, RnpB, DnaK and EF-TU) as well as with values for the overall genome of *G. obscuriglobus* obtained from Santarella-Mellwig *et al.* (2010). Examination of the codon usage of GoN and GoC does not reveal any obvious difference in codon preferences compared with other *G. obscuriglobus* genes. When comparing GC content, the only notable difference was that of GoC (58 %), which exhibited a 9 % difference when compared with the genome (67 %) and most of the housekeeping genes. There are three copies of EF-TU in the genome of *G. obscuriglobus*, one of which has a GC content similar to that of GoC (59 %), so there is a possibility that GoC and this copy of EF-TU are remnants of a gene transfer event; however, more extensive tests would be required to make a better judgment on the validity of this putative gene transfer.

DISCUSSION

Cells of all planctomycete species have a nucleoid with a condensed structure, and in the member species *G. obscuriglobus*, the compaction of the nucleoid has been correlated with an inherent resistance to high levels of radiation. Therefore, it is of interest to identify the proteins responsible for the highly compacted nucleoid. A homology search of the genome of *G. obscuriglobus* with histone-like proteins as queries revealed two different HU proteins of different types. The HU protein belongs to a class of DNA-bending proteins associated with the folding of the bacterial nucleoid (White *et al.*, 1989; Swinger & Rice, 2004). Known to exist in a monodimeric form, this protein in most bacteria is usually encoded by a single gene, whereas in others, such as the enterobacteria, it exists in a heterodimeric form encoded by two different genes (HU- β and HU- α) (Rouvière-Yaniv & Kjeldgaard, 1979) with different preferential binding

affinities, and thus interacts differently with DNA, with effects on DNA replication and thermostability (Shindo *et al.*, 1992; Tanaka *et al.*, 1993; Bahloul *et al.*, 2001).

Sequence analysis of the two HU proteins from *G. obscuriglobus* showed that the proteins contain few or no cysteine, tryptophan or tyrosine residues, which is important, since HU proteins and other nucleoid-associated proteins are known to be homologous in terms more of structure and the underlying properties of the amino acid residues rather than sequence (Laine *et al.*, 1980; Grove & Lim, 2001; Luijsterburg *et al.*, 2006). Other than the above-mentioned general characteristics of HU proteins, both *G. obscuriglobus* proteins are unlike the heterodimeric form found in enteric bacteria. One of the proteins, GoN, contains an N-terminal extension rich in lysine and alanine, a feature similar to the HU protein from *Deinococcus radiodurans* (Ghosh & Grove, 2006), whereas the other HU, GoC, has a C-terminal extension after the core HU domain, which is a prominent feature of a specific group of HU proteins from the Actinobacteria (Prabhakar *et al.*, 1998; Mukherjee *et al.*, 2009; Salerno *et al.*, 2009; Kumar *et al.*, 2010).

The C terminal of eukaryote histone H1 can have up to 42 % lysine residues in addition to other basic and hydrophilic amino acids (except for proline), which accounts for its intrinsically disordered conformation (Maeder & Bohm, 1991; Hansen *et al.*, 2006). Where they occur among bacterial HU proteins, extensions at either the N or the C terminal are of exceptional interest, as these extensions are similar in amino acid composition to that of the C terminal of eukaryotic histone H1, even though in some cases a homology search at the primary sequence level does not reveal significant matches with histones. Not only is the amino acid composition of the extensions in the HU protein from *Deinococcus* (DrHU), the mycobacterial histone-like protein (Hlp), the spore maturation-associated HupS of *S. coelicolor*, and both HU-like proteins from *G. obscuriglobus* rich in basic amino acids, but also some of these extensions contain regions similar to the (S/T)PKK repeats of eukaryotic histone H1, involved in DNA compaction (Bharath *et al.*, 2002).

The architectural role of the HU protein *in vivo* is still unclear, with *in vitro* studies suggesting that HU proteins might actually counteract condensation, have dual functions dependent on concentration (Dame & Goosen, 2002; van Noort *et al.*, 2004), or that they facilitate DNA condensation (Sarkar *et al.*, 2007). The effect of terminal extensions on the condensation of DNA is even more unclear, considering the rarity of such extensions among bacterial HU proteins. Functions associated with extensions in HU proteins include a developmental role similar to that of actinobacterial HupS (Salerno *et al.*, 2009), and protection of the DNA from strand breakages from either radiation (Ghosh & Grove, 2006; Whiteford *et al.*, 2011) or enzymes (Kumar *et al.*, 2010), as in actinobacterial Hlp and deinococcal DrHU. More importantly, suggestions of an architectural role for HU proteins with extensions have come from studies of the *D. radiodurans* HU protein with

an N-terminal extension, which is proposed to stabilize four-way junctions and thus promote condensed toroidal DNA (Englander *et al.*, 2004; Ghosh & Grove, 2004, 2006). In addition, deletion studies on the C-terminal region of mycobacterial Hlp hint at the protein having a passive architectural role as well. Thus, the extensions of HU proteins at either the N or the C terminal appear not only to confer protection upon the DNA to which they bind but also to act to compact DNA to some degree. However, the functionality of the two *Gemmata* HU proteins will have to be determined in future studies.

LGT

The phylogenetic tree of HU proteins including the C-terminal and N-terminal types suggests that in the case of the N-terminal type, there may have been gene transfer from planctomycetes or their ancestors to an ancestor of some members of the phylum Proteobacteria, as the N-terminal types of HU protein of several proteobacteria cluster at a shallow level within the clade formed by the N-terminal HU protein of planctomycetes. Since planctomycete sequences form the deepest branches of this clade, the LGT is likely to have been from a planctomycete to a proteobacterium. The similarity of N-terminal HU proteins of planctomycetes to those of at least one of the proteobacterial sequences at a sequence feature level, e.g. the PAKK repeats in the lysine- and alanine-rich N-terminal domain in *Teredinibacter turnerae*, is consistent with the occurrence of planctomycete and proteobacteria sequences in the same clade and the transfer of an ancestral sequence with these shared features.

LGT could conceivably be an explanation for the occurrence of two extension types of HU protein in *G. obscuriglobus*, N-terminal and C-terminal, e.g. the acquisition of a second type from a non-planctomycete. In the case of the C-terminal type, there is some indication of possible LGT based on the aberrant GC content of the GoC-type HU protein. There was no indication of such a disparity in GC content in the case of the N-terminal type of HU protein in *G. obscuriglobus*. However, an alternative explanation for the co-occurrence of two extension types in the same phylum is that there were two or more types of HU protein in a common ancestor of the Bacteria, and that there has been extensive loss of one or more types in most Bacteria. There may be good reasons for selective retention of one or more extension types in particular species, e.g. DNA repair or protection conferring radiation resistance, known to be a property of both *G. obscuriglobus*, in which both extension types occur, and *D. radiodurans*, in which the N-terminal extension type occurs.

Members of the phyla *Deinococcus* and *Actinobacteria* are the only other bacteria known to have HU protein types with extensions, with either N- or C-terminal extensions but not with both in the same organism. On the assumption that the extensions at the terminals of HU proteins such as those that we have observed are rare (Dillon & Dorman, 2010; Grove, 2010), any LGT that occurred into *G. obscuriglobus* or an ancestor of planctomycetes would likely be a transfer from a

donor organism with an already evolved two-domain HU protein with either extension type, such as those of *Deinococcus* or the *Actinobacteria* or their ancestors. On the other hand, if such two-domain HU proteins with extensions are widely distributed across the domain Bacteria, it is more than likely that the ancient form of HU was a two-domain protein and that the Planctomycetes belong to a conserved group that retains the ancient form, whereas most other bacteria have lost the extension domain. The deviant GC content of GoC implies that it is the result of a transfer, but the strong bootstrap support for the clustering together of the C-terminal HU proteins of planctomycetes in the HU protein tree suggests that *G. obscuriglobus* as well as other planctomycete genera could have retained an ancient version of this HU from an ancestral planctomycete.

Nevertheless, we are aware that even if an LGT event had occurred, the transfer could have been a very ancient event, and that since the gene reflects the host genome, the signal may have drifted towards neutrality and be undetectable via GC content (Lawrence & Ochman, 1997). Alternatively, a gene could have been originally transferred in a 'ready-to-use' form, and therefore displayed compatible and thus similar codon usage (Medrano-Soto *et al.*, 2004).

Possible role in DNA repair?

The GoN protein of *G. obscuriglobus* and the HU protein of *Deinococcus* display a similar N terminus with respect to properties and motifs. Although there is still a lack of *in vivo* evidence linking the HU protein of *D. radiodurans* to radiation resistance, the unique preferential binding of the N terminus to four-way DNA junctions coupled to an effective DNA repair system during recombination events in *D. radiodurans* are highly suggestive of the role of the N-terminus HU protein in radiation survival (Ghosh & Grove, 2006; Blasius *et al.*, 2008). Chromatin organization and radiation survival in *G. obscuriglobus* have been linked to an effective DNA repair mechanism (Lieber *et al.*, 2009). Analogous to the example in *D. radiodurans*, the GoN HU protein of *G. obscuriglobus* could be an important accessory protein that stabilizes DNA at the four-way junction via N-terminus binding and thus contributes to the radiation survival of *G. obscuriglobus* via promotion of homologous recombination. Furthermore, there are other examples which might support this hypothesis. For example, *Coxiella burnetii*, an intracellular parasitic proteobacterium, has an SOS system with a highly effective DNA repair mechanism, even using a reduced set of DNA repair genes (Mertens *et al.*, 2005), and coincidentally has a 31 aa-long N terminus rich in hydrophobic residues in its HU protein. Similarly, in the HU protein of a *Xanthomonas* species, our analysis indicates that an N-terminus region of 41 aa comprising more than 60 % of the lysine and alanine residues might have some relationship to an SOS system highly responsive to DNA-damaging agents. Also in *Xanthomonas*, it has been reported that deletion of a nucleotide excision repair gene renders cells four orders of magnitude more susceptible to UV radiation

(Yang *et al.*, 2001; Shen *et al.*, 2007). GoN might thus play a role in radiation survival, and determining this would first require robust *in vitro* experimentation to investigate its DNA-binding properties as well as to determine the radiation sensitivity of gene knockouts (dependent on a yet-to-be-developed genetic system for *G. obscuriglobus*) and experiments to establish its *in vivo* association with the *G. obscuriglobus* nucleoid.

Conclusion

G. obscuriglobus has an intriguing compartmentalized cell structure possessing a condensed nucleoid with a double-membrane envelope. The ability to recover from an otherwise lethal dose of radiation has been attributed to an efficient DNA repair system operating on a tightly packed nucleoid. In this study, we have attempted to identify possible DNA-compacting proteins of *G. obscuriglobus* and have found two bacterial HU-like proteins with extensions at different termini of the protein, one of which represents a novel type of HU protein with a concentration of prolines in the C-terminus region. The distribution of the proteins suggests that members of the phylum Planctomycetes and perhaps some related PVC superphylum members in the phylum Verrucomicrobia are unique in the domain Bacteria in possessing versions of these histone-like HU proteins with both N- and C-terminal extensions. Based on deviant GC content there is a possibility that one of these HU proteins is the product of an LGT event. Nevertheless, it would be interesting to find out whether *G. obscuriglobus* has managed to integrate the function of this protein into its cellular processes. Hence, future research will have to focus on determining the DNA-binding properties of these two proteins with respect to their association with the nucleoid, as well as secondary functions such as involvement in radiation survival and cell development.

ACKNOWLEDGEMENTS

Research in J.A.F.'s laboratory on planctomycetes is funded by the Australian Research Council; B.Y. thanks Dow-Agroscience, the University of Washington and The University of Queensland for scholarship support.

REFERENCES

- Azam, T. A. & Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* **274**, 33105–33113.
- Bahloul, A., Boubrik, F. & Rouviere-Yaniv, J. (2001). Roles of *Escherichia coli* histone-like protein HU in DNA replication: HU-beta suppresses the thermosensitivity of *dnaA46ts*. *Biochimie* **83**, 219–229.
- Barry, C. E., III, Hayes, S. F. & Hackstadt, T. (1992). Nucleoid condensation in *Escherichia coli* that express a chlamydial histone homolog. *Science* **256**, 377–379.
- Bharath, M. M., Ramesh, S., Chandra, N. R. & Rao, M. R. (2002). Identification of a 34 amino acid stretch within the C-terminus of histone H1 as the DNA-condensing domain by site-directed mutagenesis. *Biochemistry* **41**, 7617–7627.
- Blasius, M., Sommer, S. & Hübscher, U. (2008). *Deinococcus radiodurans*: what belongs to the survival kit? *Crit Rev Biochem Mol Biol* **43**, 221–238.
- Christodoulou, E. & Vorgias, C. E. (2002). The thermostability of DNA-binding protein HU from mesophilic, thermophilic, and extreme thermophilic bacteria. *Extremophiles* **6**, 21–31.
- Cole, C., Barber, J. D. & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36** (Web Server issue), W197–W201.
- Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892–893.
- Dame, R. T. & Goosen, N. (2002). HU: promoting or counteracting DNA compaction? *FEBS Lett* **529**, 151–156.
- Dillon, S. C. & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* **8**, 185–195.
- Distel, D. L., Morrill, W., MacLaren-Toussaint, N., Franks, D. & Waterbury, J. (2002). *Teredinibacter turnerae* gen. nov., sp. nov., a dinitrogen-fixing, cellulolytic, endosymbiotic gamma-proteobacterium isolated from the gills of wood-boring molluscs (Bivalvia: Teredinidae). *Int J Syst Evol Microbiol* **52**, 2261–2269.
- Englander, J., Klein, E., Brumfeld, V., Sharma, A. K., Doherty, A. J. & Minsky, A. (2004). DNA toroids: framework for DNA repair in *Deinococcus radiodurans* and in germinating bacterial spores. *J Bacteriol* **186**, 5973–5977.
- Felsenstein, J. (1989). PHYLIP – phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.
- Fuerst, J. A. (2005). Intracellular compartmentation in planctomycetes. *Annu Rev Microbiol* **59**, 299–328.
- Fuerst, J. A. & Webb, R. I. (1991). Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A* **88**, 8184–8188.
- Ghosh, S. & Grove, A. (2004). Histone-like protein HU from *Deinococcus radiodurans* binds preferentially to four-way DNA junctions. *J Mol Biol* **337**, 561–571.
- Ghosh, S. & Grove, A. (2006). The *Deinococcus radiodurans*-encoded HU protein has two DNA-binding domains. *Biochemistry* **45**, 1723–1733.
- Grove, A. (2010). Functional evolution of bacterial histone-like HU proteins. *Curr Issues Mol Biol* **13**, 1–12.
- Grove, A. & Lim, L. (2001). High-affinity DNA binding of HU protein from the hyperthermophile *Thermotoga maritima*. *J Mol Biol* **311**, 491–502.
- Hansen, J. C., Lu, X., Ross, E. D. & Woody, R. W. (2006). Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J Biol Chem* **281**, 1853–1856.
- Jobb, G., von Haeseler, A. & Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* **4**, 18.
- Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McInerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* **6**, 29.
- Kumar, S., Sardesai, A. A., Basu, D., Muniyappa, K. & Hasnain, S. E. (2010). DNA clasping by mycobacterial HU: the C-terminal region of HupB mediates increased specificity of DNA binding. *PLoS ONE* **5**, e12551.
- Laine, B., Kmiecik, D., Sautiere, P., Biserte, G. & Cohen-Solal, M. (1980). Complete amino-acid sequences of DNA-binding proteins HU-1 and HU-2 from *Escherichia coli*. *Eur J Biochem* **103**, 447–461.

- Lawrence, J. G. & Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**, 383–397.
- Lieber, A., Leis, A., Kushmaro, A., Minsky, A. & Medalia, O. (2009). Chromatin organization and radio resistance in the bacterium *Gemmata obscuriglobus*. *J Bacteriol* **191**, 1439–1445.
- Lindsay, M. R., Webb, R. I., Strous, M., Jetten, M. S., Butler, M. K., Forde, R. J. & Fuerst, J. A. (2001). Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Arch Microbiol* **175**, 413–429.
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.
- Luijsterburg, M. S., Noom, M. C., Wuite, G. J. & Dame, R. T. (2006). The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J Struct Biol* **156**, 262–272.
- Luijsterburg, M. S., White, M. F., van Driel, R. & Dame, R. T. (2008). The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* **43**, 393–418.
- Maeder, D. L. & Bohm, L. (1991). The C-domain in the H1 histone is structurally conserved. *Biochim Biophys Acta* **1076**, 233–238.
- Medrano-Soto, A., Moreno-Hagelsieb, G., Vinuesa, P., Christen, J. A. & Collado-Vides, J. (2004). Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol Biol Evol* **21**, 1884–1894.
- Mertens, K., Lantsheer, L. & Samuel, J. E. (2005). A minimal set of DNA repair genes is sufficient for survival of *Coxiella burnetii* under oxidative stress. *Ann N Y Acad Sci* **1063**, 73–75.
- Mukherjee, A., DiMario, P. J. & Grove, A. (2009). *Mycobacterium smegmatis* histone-like protein Hlp is nucleoid associated. *FEMS Microbiol Lett* **291**, 232–240.
- Prabhakar, S., Annapurna, P. S., Jain, N. K., Dey, A. B., Tyagi, J. S. & Prasad, H. K. (1998). Identification of an immunogenic histone-like protein (HLPMT) of *Mycobacterium tuberculosis*. *Tuber Lung Dis* **79**, 43–53.
- Remacha, M., Kaul, R., Sherburne, R. & Wenman, W. M. (1996). Functional domains of chlamydial histone H1-like protein. *Biochem J* **315**, 481–486.
- Rouvière-Yaniv, J. & Gros, F. (1975). Characterization of a novel, low-molecular-weight DNA-binding protein from *Escherichia coli*. *Proc Natl Acad Sci U S A* **72**, 3428–3432.
- Rouvière-Yaniv, J. & Kjeldgaard, N. O. (1979). Native *Escherichia coli* HU protein is a heterotypic dimer. *FEBS Lett* **106**, 297–300.
- Rouvière-Yaniv, J., Yaniv, M. & Germond, J. E. (1979). *E. coli* DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA. *Cell* **17**, 265–274.
- Salerno, P., Larsson, J., Bucca, G., Laing, E., Smith, C. P. & Flärdh, K. (2009). One of the two genes encoding nucleoid-associated HU proteins in *Streptomyces coelicolor* is developmentally regulated and specifically involved in spore maturation. *J Bacteriol* **191**, 6489–6500.
- Santarella-Mellwig, R., Franke, J., Jaedicke, A., Gorjanacz, M., Bauer, U., Budd, A., Mattaj, I. W. & Devos, D. P. (2010). The compartmentalized bacteria of the Planctomycetes-Verrucomicrobia-Chlamydiae superphylum have membrane coat-like proteins. *PLoS Biol* **8**, e1000281.
- Sarkar, T., Vitoc, I., Mukerji, I. & Hud, N. V. (2007). Bacterial protein HU dictates the morphology of DNA condensates produced by crowding agents and polyamines. *Nucleic Acids Res* **35**, 951–961.
- Shen, C. H., Chiang, Y. C., Hsu, C. H. & Yang, M. K. (2007). Identification and characterization of two *uvrA* genes of *Xanthomonas axonopodis* pathovar *citri*. *Mol Genet Genomics* **277**, 149–160.
- Shindo, H., Furubayashi, A., Shimizu, M., Miyake, M. & Imamoto, F. (1992). Preferential binding of *E. coli* histone-like protein HU α to negatively supercoiled DNA. *Nucleic Acids Res* **20**, 1553–1558.
- Spurio, R., Dürrenberger, M., Falconi, M., La Teana, A., Pon, C. L. & Gualerzi, C. O. (1992). Lethal overproduction of the *Escherichia coli* nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *Mol Gen Genet* **231**, 201–211.
- Swinger, K. K. & Rice, P. A. (2004). IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* **14**, 28–35.
- Tanaka, H., Goshima, N., Kohno, K., Kano, Y. & Imamoto, F. (1993). Properties of DNA-binding of HU heterotypic and homotypic dimers from *Escherichia coli*. *J Biochem* **113**, 568–572.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- van Noort, J., Verbrugge, S., Goosen, N., Dekker, C. & Dame, R. T. (2004). Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci U S A* **101**, 6969–6974.
- Wagner, M. & Horn, M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* **17**, 241–249.
- White, S. W., Appelt, K., Wilson, K. S. & Tanaka, I. (1989). A protein structural motif that bends DNA. *Proteins* **5**, 281–288.
- Whiteford, D. C., Klingelhoets, J. J., Bambenek, M. H. & Dahl, J. L. (2011). Deletion of the histone-like protein (Hlp) from *Mycobacterium smegmatis* results in increased sensitivity to UV exposure, freezing and isoniazid. *Microbiology* **157**, 327–335.
- Yang, M. K., Chou, M. E. & Yang, Y. C. (2001). Molecular characterization and expression of the *recX* gene of *Xanthomonas campestris* pv. *citri*. *Curr Microbiol* **42**, 257–263.
- Yang, J. C., Madupu, R., Durkin, A. S., Ekborg, N. A., Pedamallu, C. S., Hostetler, J. B., Radune, D., Toms, B. S., Henrissat, B. & other authors (2009). The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS ONE* **4**, e6085.

Edited by: D. W. Ussery