Mini Review

Role of Nucleoid Associated Protein HU in the Intercellular and Extracellular Milieu

Debayan Dey

Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30329 USA

Abstract: Nucleoid-associated proteins (NAPs) play an architectural role by bending, bridging, and wrapping the DNA along with a regulatory role of controlling various transcriptional units in the cell. Previews reviews have highlighted the role of HU and its paralog IHF plays in intracellular function as a transcriptional regulator, nucleoid bending protein and sometimes also moonlights in other functions. This review highlights along with the canonical functions of HU and IHF which affects genes responsible for translational machineries, cell wall biosynthesis, aerobic respiration and virulence ; other non-canonical roles which HU plays outside the cellular milieu, notably in acting as an adhesin and playing role in host-cell adhesion, its role in biofilm architecture and its association with cationic low complexity region, resembling histone like H1 proteins. HU and IHF thus has evolved as a hub protein performing a vast type of functions which makes it a important drug target for antibacterial therapy.

Keywords: Nucleoid-associated proteins (NAPs), moonlighting proteins, drug target, biofilm, specificity determination, phylogenetic analysis

Nucleoid-associated proteins (NAPs) in bacteria

In all the three domains of life an absolute requirement is to package and store DNA in such a manner that it is compatible with replication and transcriptional machinery. In eukaryotes a sophisticated mechanism of histone mediated DNA packaging and other associated proteins which unwrap the DNA is present, which is absent in eubacteria (Isenberg 1979). In bacteria similar job is performed by “histone like proteins” which can bend, bridge or wrap DNA (Dillon and Dorman 2010). Due to the different structural and functional reasons, the older term is replaced by ‘Nucleoid associated proteins’ (NAPs). They influence the transcriptional process in both positive and negative manner and usually exhibit multi-specificity. They also take part in formation of multi-protein complexes at origin of replication, transcriptional start points and recombination/repair complexes (Mott and Berger 2007). There are also evidences of regulatory interplay among the NAPs influencing complex genetic switches. Major NAPs include HU, IHF, FIS, LRP, H-NS, Lsr2, YbaB etc

There is a good correlation between transcriptional activity and the number of looped DNA domains in a prokaryotic genome. In active state the number and stability of looped domain is found to be more than when it is in stressed condition or in stationary stage. Many of the looped domains are formed or stabilized by NAPs. They bind to the intergenic AT rich regions and bend DNA or stabilize curved DNA segments. HU, IHF, Fis plays a crucial role in such DNA looping processes (Luijsterburg 2006). Another process by which transcription is silenced is by DNA bridging in which the bridge has the potential to trap RNA polymerase or remove it from promoters, thereby acting as a transcriptional silencer (Navarre et al. 2006). H-NS and its homolog StpA, Ler are DNA bridging proteins found in Proteobacteria and in many cases are essential for the organisms’ survival (Dorman 2004). Atomic force microscopy indicates that it constrains supercoils in DNA and forms protein-DNA-protein bridges. H-NS is a homodimer with an N-terminal oligomerization domain and C-terminal DNA binding domain characterized by AT hook motif which binds to the minor groove of DNA. It is characterized as a sequence non specific DNA binding

protein with a preference towards AT rich DNA sequences. H-NS also forms heterodimer with StpA and plasmid encoded homolog Sfh (Johansson et al. 2001). It is believed that the homo and hetero-dimerization imparts different transcriptional responses in the cell. It also interacts with Hha/YdgT family of proteins which can interact with oligomerization domain and influence the bridging capability in a negative manner (Madrid et al. 2001). H-NS protein is a wonderful example of modular protein architecture, in which the C-terminal DNA binding module remains conserved, while the N-terminal oligomerization domain is varied. Locus of Enterocyte Effacement-Encoded Regulator (Ler) is an H-NS like protein found in Proteobacteria with non-homologous N-terminal oligomerization domain (Laaberki et al. 2006). It is found in enteropathogenic and enterohaemorrhagic Escherichia coli (EPEC and EHEC) and functions to activate transcription of virulence genes silenced by the H-NS. In lpf operon of Enterohemorrhagic Escherichia coli (EHEC) serotype O157:H7, H-NS binds to the regulatory sequence of lpfA and silences the transcription of lpf, and Ler binds to a nearby sequence and counteract the H-NS binding (Torres et al. 2007). Thus, these two similar proteins act against each other in a well orchestrated fashion to regulate virulence genes.

Lsr2 is a small, basic protein present in Mycobacterium and related actinomycetes. It is functionally analogous to H-NS like proteins in Enterobacteria, and Rok protein in Firmicutes (Gordon et al. 2011). Atom force microscopy indicates that Lsr2 has the ability to bridge distant DNA segments, playing role in the overall compactness of the nucleoid. Recent developments showed that Mycobacterium tuberculosis Lsr2 is involved in a variety of crucial cellular processes like lipid biosynthesis, response to hypoxia and antibiotic resistance. Lsr2 is involved in the biosynthesis of mycolyl-diacylglycerols in the cell wall of M. smegmatis and negatively regulates the iniBAC operon, which is a multidrug efflux system in *M.tuberculosis* (Colangeli et al. 2007). It also participates in the negative regulation of the mps operon, the biosynthetic locus of glycopeptidolipids in M. smegmatis.

Nucleoid associated proteins in archaea and eubacteria determined the overall architecture of genome, which in turn determined the regulatory landscape of the cell. Other than HU, IHF, HNS like proteins (and Lsr2), other crucial NAPs include CRP, LRP,

Fis, Alba Ybab etc (Dillon and Dorman 2010). Fis (Factor of inversion stimulation) is identified as a protein which stimulates the site specific recombination of DNA catalyzed by DNA invertases. Later its role as a global regulatory protein, modulating DNA architecture was established. Fis is dimeric protein of 22 kDa, consisting two intertwined α helical subunits forming a globular structure, constituting a helix turn helix, which is involved in DNA binding (Finkel and Johnson 1992). Fis binding has been demonstrated to lead to bending of DNA between 50° to 90°. Its biological role during the exponential phase of E.coli gained significance as it binds to nearly 68000 sites in chromosome. LRP (Leucine responsive protein) constitutes a family of NAPs which plays a global regulatory role in regulating cellular adaptation to environmental stimuli. It is involved in amino acid metabolism and wraps the DNA in its octameric form. It resembles nucleosomal assembly, similar to eukaryotic histones (Calvo and Matthews 1994). It is one of the NAPs whose homologs are present in nearly all of the eubacteria. EbfC or Ybab is NAPs which is present in the exponential phase of grown in many organisms. Studies defined it as a dimeric extended alpha-helical “tweezers” as the DNA-binding domain, while the protein exists as a homodimer but can also form higher-ordered structures such as tetramers and octamer, with the potential to aggregate/bridge DNA (Jutras et al. 2010).

HU and IHF: Flexible DNA architectural proteins in bacteria

Early in 1970s Francois Gros *et al*. were trying to identify proteins which stimulate RNA synthesis and found small protein factors that stabilize the nucleic acid secondary structure and thus inhibit or favor the RNA polymerase to act. One of the proteins, which they termed as HU, stimulates the transcription of bacteriophage λ-DNA and displayed amino acid compositional similarities to eukaryotic histones (Rouviere-Yaniv and Gros 1975). Geiduschek and coworkers described a protein, TF1, which is synthesized during phage SP01 infection of *Bacillus subtilis* (Wilson and Geiduschek 1969). Its amino acid composition showed similarity with *E.coli* HU with slight differences in lysine/arginine ratio. Similarly during that period many HU homologs were characterized in different species like cyanobacteria (*Anabaena sp* and *Aphanocapsa sp*), *Mycoplasma* and *Pseudomonas aeruginosa* (Haselkorn and Rouviere-Yaniv 1976, Hawkins and Woottons 1981).

The name HU has originated from the 'histone like protein' term and U93 strain of *E.coli* in which it was first identified. It is a small obligate dimeric protein and the most abundant NAP in most of the bacterial species. Except in *Proteobacteria*, mostly HU exists as homodimer, while in Proteobacterial species like *E.coli, Salmonella, Shigella* etc it predominantly it exists as a heterodimer with each subunit being ~9.5 kDa in size. In *Proteobacteria* and *Bacteroidetes*, it can occur in both homo and heterodimeric form encoded by various paralogs of HU and each dimeric species affect the transcription in a different manner. In *E.coli* it is coded by hupA and hupB genes while homologous proteins termed as Integration Host Factor (IHF) is also coded in the same genome. IHF shares the same protein family and tertiary/quaternary fold, related by common ancestor to HU, but is characterized as moderately sequence specific DNA binding protein. *E. coli* integration host factor (IHF) participates in phage lambda site-specific recombination and regulates the expression of phage and other bacterial genes (Craig and Nash 1984).

Binding preferences of HU and IHF toward different DNA substrates

The first studies on affinity and selectivity of various nucleic acid substrates of *E.coli* HU were published by Kleppe *et al*. in 1985. They found that HU binds to ssDNA, dsDNA and RNA substrates with preference towards *E.coli* ribosomal RNA than foreign RNA species (Holck and Kleppe 1985). It was also reported that HU was isolated tightly bound to 30S ribosomal subunit, implicating a role beyond transcriptional activation/repression (Suryanarayana and Subramanian 1978). HU also stimulates the formation of DNA closure rings thus assisting T4 DNA ligase by stabilizing the bent DNA (Hodges-Garcia *et al*. 1989).

HU protein binds double stranded DNA with adjacent dimer occupying ~9bp although variations exist depending on the nature of HU dimer (homodimer or heterodimer). The ββ homodimer of *E.coli* HU binds dsDNA four times weaker than αβ and αα form, while the adjacent dimer-dimer interaction is most favored in heterodimer form. Differences in specificity, binding site length and curvature etc between HU and its homologue IHF was first reported by Rouviere-Yaniv's group in 1991(Bonnefoy and Rouviere-Yaniv 1991). As mentioned earlier HU binds to ssDNA while IHF does not exhibit any strong preference, while showing a distinctively higher preference towards curved DNA. They showed that

due to these differences in DNA binding properties IHF cannot functionally replace HU in *E.coli* hupAB deletion mutant (Bonnefoy and Rouviere-Yaniv 1991). In the last decade, many biophysical, crystallographic and biochemical studies pointed out differences in nature of DNA recognition and binding by IHF and HU homologues. Despite homology and similar tertiary structure, IHF and HU can be primarily distinguished based on its sequence preference. The HU protein mainly described in the literature is *E.coli* HU, which is studied as a model protein in this family. HU binds to random DNA sequence with a K_d of 200–2500 nM, whereas IHF binds to such sequences 100 times more weakly (K_d=20–30mM). IHF strongly binds (K_d=2-20 mM) to its cognate sites represented by the consensus WATCARXXXXTTR (W is A or T; X is A, T, C or G; R is A or G). HU and IHF binds to DNA with flexible β sheet arms with intercalating proline tips which induces sharp bends to DNA 9bp apart. Other differences between HU and IHF are related to binding site length, DNA bending angle and degree of cooperatively. Fluorescence resonance energy transfer (FRET) experiments and crystallographic studies confirm the degree of bending of IHF is ~160° with a binding site size of 35 bp. Most of the biophysical and biochemical experiments agree in the case of IHF, while they show wide variations in HU. HU binds to DNA with a wide range of cooperatively with binding sites ranging between 9 and 42 bp, under various experimental conditions. Differences in bend angle also exist in HU which depends on the type of experiment and the conditions. Other than this binding site, the distal DNA binding is influenced by the “lateral section” of HU and IHF. In IHF, this region is positively charged and tightly holds the DNA, stabilizing the sharper bends, while in HU this region is variable in charge and polarity, thus differs in its DNA stabilizing ability and is dependent on factors like salt concentration and pH etc. This phenomenon is termed as “flexible DNA bending” by Rice *et al*. and it is evidenced in various crystal structures of the same protein bound to DNA. The bend angles in each homodimer differ, as does the nonplanarity between bends and other structural parameters. The flexibility in its DNA bending is a property very crucial for a global regulator allowing it to mould itself to different regulatory and associative scenarios. It allows HU to form higher order protein-DNA complexes and gives plasticity in the cellular jigsaw. Although the differences in HU and IHF are observed from various experimental approaches, a molecular mechanism or evolutionary process behind it remained unexplored.

Understanding of the molecular and evolutionary mechanism giving rise to specificity or promiscuity in substrate recognition is thus a fundamental question in biological systems with implications for protein engineering and drug development. Our study found, the molecular mechanism giving rise to specificity or multi-specificity depends on a combination effect of the amino acid composition of the binding site, its flexibility, and steric constraints (Dey et. al. 2017)

To understand the sequence determinants, which influence the degree of DNA binding specificity, we undertook a phylogenetic study in conjunction with analysis of three-dimensional structures. Our phylogenetic analysis using evolutionary trace method, found HU/IHF proteins can be majorly categorized into HU like proteins, b) IHFα and c) IHFβ like proteins (with reference to E. coli). Evolutionary trace method was previously used as a technique to understand the specificity determining residues in a larger protein family or superfamily (Dey 2018, Shrilakshmi et al. 2019, Nosrati et al. 2019, Kuiper et al. 2019, Dey et al. 2020).

These clades differ significantly in their amino acid compositions of charged, small and aromatic residues. IHF is significantly enriched in both positive and negatively charged residues in the DNA binding region, which contributes to the salt bridge formation, affecting its conformational flexibility. HU is enriched in small amino acids, particularly alanine, which does not impede the side chain conformations of the neighboring DNA binding residues. Thus, the interplay between electrostatic and steric restrictions, affect the binding specificity of HU and IHF clade proteins. We observed an enrichment of Arg in both DNA binding and alpha-helical region of IHF clade protein while Lys is favored in HU. Lys is mainly involved in interaction with DNA phosphate whereas Arg can also interact with DNA bases. Thus, the nature of DNA binding residues in IHF promotes specific interactions to DNA as opposed to HU which promotes non-specific one. Based on comparative analysis of 3D structures, we also found the differential dimer stabilization strategies in HU and IHFs, which might influence DNA binding and bending. HU clade proteins have a single aromatic cluster formed by three aromatic residues from each protomer, while in IHF clade proteins; an additional aromatic cluster is present.

The present study provides a model system to understand the differences in specificity of proteins of the same fold towards their target, thus, the three factors mentioned above can be utilized to engineer proteins for changing their substrate specificity and also design selective inhibitors for HU and IHF from pathogenic organisms. In a recent study, by Doudna and coworkers, E. coli IHF was shown to plays an important role in the functioning of CRISPR-Cas mediated adaptive immunity by inducing sharp DNA bend at the AT-rich leader sequence in the CRISPR loci, allowing the Cas1-Cas2 integrase to catalyze the integration (Nuñez et al. 2016). The guidelines from our present study can help custom designed HU and IHF like proteins with different binding and bending ability, for studying the role of bending angle in Cas-mediated integration.

Recent studies to probe the nucleoid compaction of bacteria using super-resolution microscopy used HU tagged with mCherry, which is a fluorophore (a fluorescent protein), used to track the protein attached with it (Fisher et al. 2013, Gahlmann et al. 2014, Stracy et al. 2015). These studies use HU (mainly from E. coli) to probe bacterial nucleoid morphogenesis and dynamics. Our structural and evolutionary analysis presented in the thesis can be used as a mutagenesis guide to vary the specificity and the bending ability of HU or IHF protein, thus along with mCherry assisted four-dimensional imaging, different properties of the nucleoid compaction can be probed.

Thus, as a future possibility in this project, different mutants can be generated and their DNA binding, recognition can be probed with a combination of site biochemical experiments like electrophoretic mobility shift assay along with crystal structures of the mutants to gain structural insights complemented by molecular dynamics studies to understand the role of structural constraints like salt-bridges and steric hindrances in promiscuity.

Diverse functions of HU

HU and IHF family proteins are involved in a multitude of DNA transaction processes and regulatory complexes. HU and IHF proteins are required in the formation of pre-replication complex at OriC (Ryan et al. 2002). It interacts with DnaA and facilitates the opening of replication origin (Hwang and Kornberg 1992). HU binds to DNA containing nicks and

gaps more preferentially than dsDNA. It is also upregulated during SOS response or γ irradiation related stress events. It binds to DNA repair and recombination intermediates like DNA overhangs, forks or invasions with an affinity 1000 times higher than dsDNA (Kamashev and Rouviere-Yaniv 2000). It also interacts with other DNA repair and recombination machineries like RecA and plays a critical role in the recA-recB dependent recombination, DNA repair and SOS induction pathways in UV-irradiated *E. coli* (Miyabe et al 2000). HU and IHF is also involved in many plasmid replication and recombination events by specifically binding to operators of plasmid regulatory proteins or formation of nucleoprotein complexes with replication associated factors (Petit et al. 1995, Pérez-Martín and De Lorenzo 1997, Yasukawa et al. 1997).

In a given environmental condition, supercoiling is maintained by the combined action of gyrases, the major source of negative supercoiling (Drlica and Snyder, 1978) and by topoisomerase I, which relaxes DNA (Wang 1985, 2002) and prevents excess supercoiling (Pruss et al., 1982). Homeostatic control is limited in this situation and in topoisomerase I mutants, to prevent the supercoiling reaching unacceptably higher limit, compensatory mutations occur in gyrase. Functional coupling exists between the levels of topoisomerase I and HU (Bensaid et al. 1996). Decreasing HU increases the relaxing activity, while increasing intracellular HU levels decreases the topoisomerase I activity (Bensaid et al. 1996). They also proposed a mechanism in which HU in involved in Gyrase mediated topology maintenance. Structural studies on HU complexed with DNA revealed that HU dimer introduces a positive twist into the DNA (Swinger et al. 2003) and an octamer of HU constrains left handed toroidal coils (Guo and Adhya 2007).

HU and IHF family proteins are known to bind to specific sites of genome and impart local architectural role in different DNA based transactions. Examples of such regulatory control can be found in the repressosome loops maintained by HU at lac, gal and araBAD operons in *E.coli* (Kramer et al. 1987, Bellomy et al. 1988, Lee and Schleif 1989, Aki et al. 1996). In the well studied example of gal operon, HU binds in the inter operator region, which is between the two distal binding sites of galR repressor. HU facilitate in bending the DNA to bring the two distal GalR binding sites near which helps in formation of tetrameric GalR complex (Kar and Adhya 2001). This looping of DNA assisted by HU and

IHF plays crucial role in formation of many different nucleoprotein complexes. Similarly, in site specific DNA inversion by Hin recombinase, HU plays a crucial role in assembly of invertasome (Haykinson and Johnson 1993). Serine recombinase (Sin) collaborates with HU for activity at its recombination site resH. HU and Sin subunits bound at regulatory site of resH, together regulate recombination by specifying assembly of an intertwined synapse (Rowland et al. 2006). HU binds to Nodulation genes (nod) of rhizobia which are essential for establishment of its symbiosis. It stimulates or inhibits the transcription of nodD protein at its promoter and might have sequence specific binding sites (Liu et al. 1998). IHF, a homologous moderately specific DNA binding protein also stimulate many different DNA transactions processes like transposition excision and modulation of tn10 transposition process (Chalmers et al. 1998). It was inferred that IHF acts as an “architectural catalyst” promoting the initial transpososome assembly and acts as a sensory transducer of chromosomal supercoiling status. In absence of supercoiling, it acts as a "supercoiling relief factor (Chalmers et al. 1998). IHF binding sites are also present in many viral and plasmid insertion segments. IHF binds specifically to the ends of the transposable element IS1, as well as many other sites within a short segment of the plasmid pBR322 (Prentki et al. 1987).

The role of HU is also exemplified by knock out and mutant studies in the *E.coli K-12* strain and others. Different studies showed although HU is not essential in *E.coli*, but the knock out strain significantly changes the morphology, gene expression landscape and motility of the cells (Huisman at al. 1989, Dri et al. 1991). Cells lacking both hupA and hupB (genes of both HU paralogs) accumulate secondary mutations which include mutations in the minCDE operon, inactivating the system of septation control and hence forming minicells (Jaffe et al. 1997). It also plays substantial role in the motility of cells, which is demonstrated by Nishida et al. it was reported that a double deletion mutant of hupA and hupB results in immotile phenotype and reduced flagella (Nishida et al. 1997).

It is well known that many DNA binding transcription factors also interacts with other proteins during the formation of transcription initiation complex or other multi-protein transcriptional complexes (Merika and Orkin 1995). Thus, other than its DNA binding interface, protein-protein interaction interfaces are also very crucial for such proteins.

E.coli GalR piggybacks HU, which in turn facilitates the formation of nucleoprotein complex at the Gal operon in *E.coli* (Kar and Adhya 2001). Recently, in *Mycobacterium tuberculosis*, HU was reported to be associated with Topoisomerase I using the same site as that of *E.coli* HU (Ghosh et al. 2015). IHF, a homologous protein also was found to interact with C-terminal domain RNA polymerase (Giladi et al. 1998).

DNA protection and association with low complexity regions

Hyperthermophilic *Thermotoga maritime* HU (TmHU) exhibits DNA protection activity and shows remarkable resistant to DNase I-mediated degradation. It also binds to a ~35 bp stretch of DNA unlike *E.coli* HU with shorter binding site size (Mukherjee et al. 2008). HU proteins in many organisms are also associated with a short low complexity region rich in proline, alanine and lysine (Grove 2010). In *Deinococcus radiodurans*, it precedes the HU/IHF fold and forms a 47-amino acid long extension which binds to DNA. The truncated *Deinococcus radiodurans* HU (DrHU), comprising only the conserved DNA-binding fold, has a site size of approximately 11 bp in contrast to full-length DrHU which engages 50 bp DNA. The truncated and wild type DrHU also differs in DNA binding specificity and affinity. DrHU cannot distinguish between dsDNA and DNA with nicks and gaps, while the truncated HU which is similar to *E.coli* HU can distinguish them (Ghosh and Grove 2006). Also it's binding nature towards four way junctions differ, with DrHU binding to the arms of the junction, and the truncated DrHU primarily protects the junction crossover (Ghosh and Grove 2006). The HU homologs of *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* along with other *Actinobacteria* is also associated with a similar lysine rich "PAKKA" repeat in the C-terminal. It is implicated in protection of DNA from adverse condition by binding strongly to it. The HU associated with C-terminal tail (full length HU) binds to DNA strongly than *E.coli* HU. Its affinity for 76 bp linear DNA is higher (Kd = 0.037 ± 0.001 nM) compared to an HU variant without the C-terminal repeats(K_d = 2.5 ± 0.1 nM) (Mukherjee et al. 2008). It also favors the DNA end-joining in the presence of T4 DNA ligase which is not facilitated by the HU variant without the C-terminal repeats (Mukherjee et al. 2008). Our study to determined the possible folding of this PAKKA repeat low complexity regions associated with HU and other proteins show 'flexible disorder' and consists of characteristic pentapeptide repeats.

Temporally orchestrated regulation of HU and IHF homologs

One of the most efficient ways for a transcription factor to control the global regulatory landscape is to form various homo and hetero-oligomer regulated by differential conditions. In *E.coli* three forms of HU proteins are produced (HUαα, HUββ and HUαβ) dependent on growth stage. Transcription of the hupA gene is activated early in logarithmic phase while at mid to late exponential phase, RNA originating at the hupB-P2 promoter is detected. The hupB-P3 promoter is activated later when the cells enter stationary phase (Claret and Rouviere-Yaniv 1997). In the exponential phase, both HUαα and HUαβ are present while in stationary phase, predominantly HUαβ is present. Another example is found in *Streptomyces coilicolor* HU homologs, in which two different HU varients are expressed at different phases (Salerno et al. 2009). HupA is the homolog which is similar in length with *E.coli* HU which is expressed during the exponential growth phase, while a longer HU (with C-terminal PAKKA like repeat region), termed as HupS is expressed during the sporulation phase. HupS is found to be associated with the nucleoid and the HupS mutant was defective in heat resistance and spore pigmentation (Salerno et al. 2009). A differential gene expression of HU and IHF is best illustrated in *Legionella pneumophila*, an intracellular parasite of protozoa, which expresses both HU and IHF but in a reciprocal manner (Morash et al. 2009). HU is expressed during exponential phase while IHF is expressed in the stationary phase.

HU family proteins in eukaryotes

HU homologs are not only present in eubacteria but also in archaea and eukaryotes. In eukaryotes, it plays crucial architectural role in compacting and regulating chloroplast or mitochondrial DNA. In chloroplast it is organized into discrete structures called chloroplast nucleoids (cp-nucleoids) (Kobayashi et al. 2002). Pathogens like *Plasmodium* and *Toxoplasma*, carry a nonphotosynthetic plastid of secondary endosymbiotic origin called the apicoplast, which is organized by a HU homolog. The apicoplast contains a 35 kb, circular DNA genome which is compacted and coordinately regulated by HU. It was reported to be essential for the parasites life cycle. In contrast to bacterial HUs that bends DNA; Plasmodium HU promotes concatenation of linear DNA and inhibits DNA

circularization (Ram et al. 2008, Sasaki et al. 2009). HU homologs were also found in *Neospora caninum, Theileria parva and Theileria annulata* (Arenas et al. 2008).

Extracellular role of HU homologs

HU, other than playing DNA compaction and regulatory role, has been shown to other roles in extracellular milieu. MtbHU (also known as MDP1) interacts with lung epithelia by binding to hyaluronic acid (Aoki et al. 2004, Katsube et al. 2007). It also interacts with mycolic acid intermediates like trehalose-6-monomycolate (TMM) and trehalose-6,6′-dimycolate (TDM) (also called cord factor) (Katsube et al. 2007). It also interacts with other proteins like antigen 85 complex in Mtb, which is a crucial mycolyltransferase (Katsube et al. 2007). HU controls the K-capsule formation genes in oral pathogen *Porphyromonas gingivalis* (Alberti-Segui et al. 2010, Tjokro et al. 2014). The decrease of HU transcript is correlated with decrease in the amount of polysaccharide produced, which influence the pathogens virulence (Alberti-Segui et al. 2010). Similarly, in *Streptococcus intermediusm* histone like protein HU (SiHU) controls many genes involved in virulence and quorum sensing (Liu et al. 2008). Investigation of extracellular role of HU homologs led to discovery of streptococcal histone-like HU protein which binds to cellular epithelia. It binding to cell surface polysaccharides like glycosaminoglycans (GAGs), such as heparin, dextran sulfate, heparan sulfate and chondroitin sulfate (Liu et al. 2008). It induces the production of cytokines in murine peritoneal macrophages (Zhang et al 1999). HU stimulation induces the activation of cell signal transduction pathways, extracellular signal-regulated kinase 1/2 (ERK1/2) and c-Jun N-terminal kinase (JNK). It can also trigger a response which induces pro-inflammatory signaling via ERK1/2 and JNK signal pathways, suggesting that it may contribute to the activation of host innate immunity during bacterial infection (Liu et al. 2008).

In recent years, National Institutes of Health (USA) estimate that >80% of all bacterial infections require a necessary biofilm phase during the infection (http://grants.nih.gov/grants/guide/pa-files/PA-03-047.html). The biofilm is required for interbacterial communication, and transportation network, and encompasses an extensive extracellular polymeric substance (EPS) that covers the bacterial colonies. Moreover, the EPS limit access to the effectors of both the innate and the adaptive

immune systems to and also prevents the effectiveness of antibacterial therapeutics. The biofilms formed in vivo are likely to be composed of eDNA which is found to be organized by HU and IHF like proteins. Previous experiments have shown that a cocktail of monoclonal antibodies directed against specific epitopes of an HU/IHF proteins are highly effective to disrupt diverse biofilms.

References

Alberti-Segui, C., Arndt, A., Cugini, C., Priyadarshini, R., & Davey, M. E. (2010). HU protein affects transcription of surface polysaccharide synthesis genes in Porphyromonas gingivalis. Journal of bacteriology, 192(23), 6217-6229.

Aki, T., Choy, H. E., & Adhya, S. (1996). Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by Gal repressor. Genes to Cells, 1(2), 179-188.

Arenas, A. F., Escobar, A. J. G., & Gómez-Marin, J. E. (2008). Evolutionary origin of the protozoan parasites histone-like proteins (HU). In silico biology,8(1), 15-20.

Bellomy, G. R., Mossing, M. C., & Record Jr, M. T. (1988). Physical properties of DNA in vivo as probed by the length dependence of the lac operator looping process. Biochemistry, 27(11), 3900-3906.

Bensaid A, Almeida A, Drlica K, Rouviere-Yaniv J. 1996. Cross-talk Between Topoisomerase I and HU in Escherichia coli. Journal of molecular biology, 256(2): 292-300.

Bertoni, G., Fujita, N., Ishihama, A., & de Lorenzo, V. (1998). Active recruitment of σ54-RNA polymerase to the Pu promoter of Pseudomonas putida: role of IHF and αCTD. The EMBO journal, 17(17), 5120-5128.

Bhowmick T, Ghosh S, Dixit K, Ganesan V, Ramagopal UA, Dey D, Sarma SP, Ramakumar S, Nagaraja V. 2014. Targeting Mycobacterium tuberculosis nucleoid-associated protein HU with structure-based inhibitors. Nature communications 5.

Bonnefoy E, Rouviere-Yaniv J. 1991. HU and IHF, two homologous histone-like proteins of Escherichia coli, form different protein-DNA complexes with short DNA fragments. The EMBO journal, 10(3): 687

Browning, D. F., Grainger, D. C., & Busby, S. J. (2010). Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. Current opinion in microbiology, 13(6), 773-780.

Calvo, J. M., & Matthews, R. G. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in Escherichia coli. Microbiological reviews, 58(3), 466-490.

Chalmers, R., Guhathakurta, A., Benjamin, H., & Kleckner, N. (1998). IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. Cell, 93(5), 897-908.

Claret L, Rouviere-Yaniv J. 1997. Variation in HU composition during growth of Escherichia coli: the heterodimer is required for long term survival. Journal of molecular biology, 273(1): 93-104.Craig, N. L., & Nash, H. A. (1984). E. coli integration host factor binds to specific sites in DNA. Cell, 39(3), 707-716.

Colangeli, R., Helb, D., Vilchèze, C., Hazbón, M. H., Lee, C. G., Safi, H., ... & Peterson, S. N. (2007). Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in M. tuberculosis. PLoS Pathog, 3(6), e87.

Daumann, L. J., McCarthy, B. Y., Hadler, K. S., Murray, T. P., Gahan, L. R., Larrabee, J. A., ... & Schenk, G. (2013). Promiscuity comes at a price: Catalytic versatility vs efficiency in different

metal ion derivatives of the potential bioremediator GpdQ. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1834(1), 425-432.

Dey, D., Nagaraja, V., & Ramakumar, S. (2017). Structural and evolutionary analyses reveal determinants of DNA binding specificities of nucleoid-associated proteins HU and IHF. Molecular phylogenetics and evolution, 107, 356-366.

Dey, Debayan. "Crystal Structures of Native and AdoMet Bound rRNA Methyltransferase from Sinorhizobium meliloti: Structural Insights into rRNA Recognition. Evolutionary, Structural and Functional Studies on Nucleoid-Associated Proteins HU and IHF." PhD diss., 2018.

Dey, D., Kavanaugh, L. G., & Conn, G. L. (2020). Antibiotic substrate selectivity of Pseudomonas aeruginosa MexY and MexB efflux systems is determined by a Goldilocks affinity. Antimicrobial Agents and Chemotherapy.

Dillon SC, Dorman CJ. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nature Reviews Microbiology, 8(3): 185-195.

Dorman, C. J. (2004). H-NS: a universal regulator for a dynamic genome. Nature Reviews Microbiology, 2(5), 391-400.

Dri, A. M., Rouviere-Yaniv, J., & Moreau, P. L. (1991). Inhibition of cell division in hupA hupB mutant bacteria lacking HU protein. Journal of bacteriology, 173(9), 2852-2863.

Drlica, K., & Snyder, M. (1978). Superhelical Escherichia coli DNA: relaxation by coumermycin. Journal of molecular biology, 120(2), 145-154.

Finkel, S. E., & Johnson, R. C. (1992). The Fis protein: it's not just for DNA inversion anymore. Molecular microbiology, 6(22), 3257-3265.

Haselkorn, R., & Rouvière-Yaniv, J. (1976). Cyanobacterial DNA-binding protein related to Escherichia coli HU. Proceedings of the National Academy of Sciences, 73(6), 1917-1920.

Hawkins, A. R., & Wootton, J. C. (1981). A single DNA-binding protein from Pseudomonas aeruginosa homologous to proteins NS1 and NS2 (HU proteins) of Escherichia coli and other bacteria. FEBS letters, 130(2), 275-278.

Heidi Ledford (2009). Promiscuous antibody targets cancer .Nature doi:10.1038/news.2009.175

Holck, A., & Kleppe, K. (1985). Affinity of protein HU for different nucleic acids. FEBS letters, 185(1), 121-124.

Huisman, O, Faelen, M, Girard, D., Jaffe, A., Toussaint, A., & Rouviere-Yaniv, J. (1989). Multiple defects in Escherichia coli mutants lacking HU protein. Journal of bacteriology, 171(7), 3704-3712.

Hwang, D. S., & Kornberg, A. (1992). Opening of the replication origin of Escherichia coli by DnaA protein with protein HU or IHF. Journal of Biological Chemistry, 267(32), 23083-23086.

Ghosh S, Grove A. 2006. The Deinococcus radiodurans-encoded HU protein has two DNA-binding domains. Biochemistry, 45(6): 1723-1733

Gordon, B. R., Li, Y., Cote, A., Weirauch, M. T., Ding, P., Hughes, T. R., ... & Liu, J. (2011). Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. Proceedings of the National Academy of Sciences, 108(26), 10690-10695.

Guo, F., & Adhya, S. (2007). Spiral structure of Escherichia coli HUαβ provides foundation for DNA supercoiling. Proceedings of the National Academy of Sciences, 104(11), 4309-4314.

Haykinson, M. J., & Johnson, R. C. (1993). DNA looping and the helical repeat in vitro and in vivo: effect of HU protein and enhancer location on Hin invertasome assembly. The EMBO journal, 12(6), 2503.

Hodges-Garcia, Y., Hagerman, P. J., & Pettijohn, D. E. (1989). DNA ring closure mediated by protein HU. Journal of Biological Chemistry, 264(25), 14621-14623.

Isenberg, I. (1979). Histones. Annual review of biochemistry, 48(1), 159-191.

Jaffe, A., Vinella, D., & D'Ari, R. (1997). The Escherichia coli histone-like protein HU affects DNA initiation, chromosome partitioning via MukB, and cell division via MinCDE. Journal of bacteriology, 179(11), 3494-3499.

Johansson, J., Eriksson, S., Sondén, B., Wai, S. N., & Uhlin, B. E. (2001). Heteromeric interactions among nucleoid-associated bacterial proteins: localization of StpA-stabilizing regions in H-NS of Escherichia coli. Journal of bacteriology, 183(7), 2343-2347.

Jutras, B. L., Bowman, A., Brissette, C. A., Adams, C. A., Verma, A., Chenail, A. M., & Stevenson, B. (2012). EbfC (YbaB) is a new type of bacterial nucleoid-associated protein and a global regulator of gene expression in the Lyme disease spirochete. Journal of bacteriology, 194(13), 3395-3406.

Kamashev D, Rouviere-Yaniv J. 2000. The histone-like protein HU binds specifically to DNA recombination and repair intermediates. The EMBO journal, 19(23): 6527-6535.

Kar, S., & Adhya, S. (2001). Recruitment of HU by piggyback: a special role of GalR in repressosome assembly. Genes & development, 15(17), 2273-2281.

Katsube T, Matsumoto S, Takatsuka M, Okuyama M, Ozeki Y, Naito M, Kobayashi K. 2007. Control of cell wall assembly by a histone-like protein in mycobacteria. Journal of bacteriology, 189(22):8241-8249.

Khare, H., Dey, D., Madhu, C., Senapati, D., Raghothama, S., Govindaraju, T., & Ramakumar, S. (2017). Conformational heterogeneity in tails of DNA-binding proteins is augmented by proline containing repeats. Molecular BioSystems, 13(12), 2531-2544.

Kobayashi T, Takahara M, Miyagishima SY, Kuroiwa H, Sasaki N, Ohta N, Motomichi M, Kuroiwa T. 2002. Detection and localization of a chloroplast-encoded HU-like protein that organizes chloroplast nucleoids. The Plant Cell Online, 14(7): 1579-1589.

Kuiper, E. G., Dey, D., LaMore, P. A., Owings, J. P., Prezioso, S. M., Goldberg, J. B., & Conn, G. L. (2019). Substrate recognition by the Pseudomonas aeruginosa EF-Tu–modifying methyltransferase EftM. Journal of Biological Chemistry, 294(52), 20109-20121.

Laaberki, M. H., Janabi, N., Oswald, E., & Repoila, F. (2006). Concert of regulators to switch on LEE expression in enterohemorrhagic Escherichia coli O157: H7: interplay between Ler, GrlA, HNS and RpoS. International journal of medical microbiology, 296(4), 197-210.

Lee, D. H., & Schleif, R. F. (1989). In vivo DNA loops in araCBAD: size limits and helical repeat. Proceedings of the National Academy of Sciences,86(2), 476-480.

Link, T. M., Valentin-Hansen, P., & Brennan, R. G. (2009). Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proceedings of the National Academy of Sciences, 106(46), 19292-19297.

Liu, D., Yumoto, H., Murakami, K., Hirota, K., Ono, T., Nagamune, H., ... & Miyake, Y. (2008). The essentiality and involvement of Streptococcus intermedius histone-like DNA-binding protein in bacterial viability and normal growth. Molecular microbiology, 68(5), 1268-1282.

Liu, S. T., Chang, W. Z., Cao, H. M., Hu, H. L., Chen, Z. H., Ni, F. D., ... & Hong, G. F. (1998). A HU-like Protein Binds to Specific Sites within nodPromoters of Rhizobium leguminosarum. Journal of Biological Chemistry,273(32), 20568-20574.

Louw, G. E., Warren, R. M., Van Pittius, N. G., McEvoy, C. R. E., Van Helden, P. D., & Victor, T. C. (2009). A balancing act: efflux/influx in mycobacterial drug resistance. Antimicrobial agents and chemotherapy, 53(8), 3181-3189.

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. Journal of structural biology,156(2), 262-272.

Madrid, C., Nieto, J. M., & Juárez, A. (2001). Role of the Hha/YmoA family of proteins in the thermoregulation of the expression of virulence factors.International journal of medical microbiology, 291(6), 425-432.

Miyabe, Q.-M. Zhang, Y. Kano, S. Yonei, I. (2000). Histone-like protein HU is required for recA gene-dependent DNA repair and SOS induction pathways in UV-irradiated Escherichia coli. International journal of radiation biology,76(1), 43-49.

Morash, M. G., Brassinga, A. K. C., Warthan, M., Gourabathini, P., Garduno, R. A., Goodman, S. D., & Hoffman, P. S. (2009). Reciprocal expression of integration host factor and HU in the

developmental cycle and infectivity of Legionella pneumophila. Applied and environmental microbiology, 75(7), 1826-1837.

Mott, M. L., & Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. Nature Reviews Microbiology, 5(5), 343-354.

Mukherjee A, Bhattacharyya G, Grove, A. 2008. The C-Terminal Domain of HU-Related Histone-like Protein Hlp from Mycobacterium smegmatis Mediates DNA End-Joining. Biochemistry, 47(33): 8744-8753.

Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J., & Fang, F. C. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. Science, 313(5784), 236-238.

Nishida, S., Mizushima, T., Miki, T., & Sekimizu, K. (1997). Immotile phenotype of an Escherichia coli mutant lacking the histone-like protein HU.FEMS microbiology letters, 150(2), 297-301.

Nosrati, M., Dey, D., Mehrani, A., Strassler, S. E., Zelinskaya, N., Hoffer, E. D., ... & Conn, G. L. (2019). Functionally critical residues in the aminoglycoside resistance-associated methyltransferase RmtC play distinct roles in 30S substrate recognition. Journal of Biological Chemistry, 294(46), 17642-17653.

Painbeni, E., Caroff, M., & Rouviere-Yaniv, J. (1997). Alterations of the outer membrane composition in Escherichia coli lacking the histone-like protein HU. Proceedings of the National Academy of Sciences, 94(13), 6712-6717.

Pérez-Martín, J., & de Lorenzo, V. (1997). Clues and consequences of DNA bending in transcription. Annual Reviews in Microbiology, 51(1), 593-628.

Petit, M. A., Ehrlich, D., & Jannière, L. (1995). pAMβ1 resolvase has an atypical recombination site and requires a histone-like protein HU. Molecular microbiology, 18(2), 271-282.

Prentki, P., Chandler, M., & Galas, D. J. (1987). Escherichia coli integration host factor bends the DNA at the ends of IS1 and in an insertion hotspot with multiple IHF binding sites. The EMBO journal, 6(8), 2479.

Pruss, G. J., Manes, S. H., & Drlica, K. (1982). Escherichia coli DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. Cell, 31(1), 35-42.

Ram, E. R., Naik, R., Ganguli, M., & Habib, S. (2008). DNA organization by the apicoplast-targeted bacterial histone-like protein of Plasmodium falciparum. Nucleic acids research, 36(15), 5061-5073.

Rouviere-Yaniv, J., & Gros, F. (1975). Characterization of a novel, low-molecular-weight DNA-binding protein from Escherichia coli. Proceedings of the National Academy of Sciences, 72(9), 3428-3432.

Rowland, S. J., Boocock, M. R., & Stark, W. M. (2006). DNA bending in the Sin recombination synapse: functional replacement of HU by IHF. Molecular microbiology, 59(6), 1730-1743.

Ryan VT, Grimwade JE, Nievera CJ, Leonard AC. 2002. IHF and HU stimulate assembly of pre-replication complexes at Escherichia coli oriC by two different mechanisms. Molecular microbiology, 46(1): 113-124

Salerno P, Larsson J, Bucca G, Laing E, Smith CP, 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. Journal of bacteriology, 191(21): 6489-6500.

Sasaki, N., Hirai, M., Maeda, K., Yui, R., Itoh, K., Namiki, S., ... & Kita, K. (2009). The Plasmodium HU homolog, which binds the plastid DNA sequence-independent manner, is essential for the parasite's survival. FEBS letters, 583(9), 1446-1450.

Shrilakshmi, S., Kundapura, S., Dey, D., Ramagopal, U. A., & Kulal, A. (2019). Phylogenetic, sequence and structural analysis of Insulin superfamily proteins reveals an indelible link between evolution and structure-function relationship. bioRxiv, 769497.

.

Suryanarayana, T., & Subramanian, A. R. (1978). Specific association of two homologous DNA-binding proteins to the native 30-S ribosomal subunits of Escherichia coli. Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis, 520(2), 342-357.

Swinger, K. K., Lemberg, K. M., Zhang, Y., & Rice, P. A. (2003). Flexible DNA bending in HU–DNA cocrystal structures. The EMBO Journal, 22(14), 3749-3760.

Tjokro, N. O., Rocco, C. J., Priyadarshini, R., Davey, M. E., & Goodman, S. D. (2014). A Biochemical Analysis of the Interaction of Porphyromonas gingivalis HU PG0121 Protein with DNA. PloS one, 9(3), e93266.

Torres, A. G., López-Sánchez, G. N., Milflores-Flores, L., Patel, S. D., Rojas-López, M., de la Peña, C. F. M., ... & Martínez-Laguna, Y. (2007). Ler and H-NS, regulators controlling expression of the long polar fimbriae of Escherichia coli O157: H7. Journal of bacteriology, 189(16), 5916-5928.

Wang, J. C. (1985). DNA topoisomerases. Annual review of biochemistry,54(1), 665-697.

Wang, J. C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. Nature reviews Molecular cell biology, 3(6), 430-440.

Wilson, D. L., & Geiduschek, E. P. (1969). A template-selective inhibitor of in vitro transcription. Proceedings of the National Academy of Sciences,62(2), 514-520.

Yasukawa, H., Ozaki, E., Nakahama, K., & Masamune, Y. (1997). HU protein binding to the replication origin of the rolling-circle plasmid pKYM enhances DNA replication. Molecular and General Genetics MGG, 254(5), 548-554.