Mini Review

Endogenous and Foreign Nucleoid-Associated Proteins of Bacteria: Occurrence, Interactions and Effects on Mobile Genetic Elements and Host’s Biology

Rodrigo Flores-Ríos a, Raquel Quatrini a,b,⁎, Alejandra Loyola a,*

a Fundación Ciencia y Vida, Avenida Zañartu 1482, Ñuñoa, Santiago, Chile
b Millennium Nucleus in the Biology of Intestinal Microbiota, Santiago, Chile

ABSTRACT

Mobile Genetic Elements (MGEs) are mosaics of functional gene modules of diverse evolutionary origin and are generally divergent from the hosts’ genetic background. Existing biases in base composition and codon usage of these elements’ genes impose transcription and translation limitations that may affect the physical and regulatory integration of MGEs in new hosts. Stable appropriation of the foreign DNA depends on a number of host factors among which are the Nucleoid-Associated Proteins (NAPs). These small, basic, highly abundant proteins bind and bend DNA, altering its topology and folding, thereby affecting all known essential DNA metabolism related processes. Both chromosomally- (endogenous) and MGE- (foreign) encoded NAPs have been shown to exist in bacteria. While the role of host-encoded NAPs in xenogeneic silencing of both episomal (plasmids) and integrative MGEs (pathogenicity islands and prophages) is well acknowledged, less is known about the role of MGE-encoded NAPs in the foreign elements biology or their influence on the host’s chromosome expression dynamics. Here we review existing literature on the topic, present examples on the positive and negative effects that endogenous and foreign NAPs exert on global transcriptional gene expression, MGE integrative and excisive recombination dynamics, persistence and transfer to suitable hosts and discuss the nature and relevance of synergistic and antagonizing higher order interactions between diverse types of NAPs.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords:
NAP
Bacterial nucleoid
MGE
HGT
Xenogeneic silencing
Regulatory hierarchies

Contents

1. Introduction ............................................. 747
2. Endogenous Nucleoid Associated Protein Types and Functions ............................................. 748
  2.1. H-NS .................................................. 748
  2.2. HU .................................................... 748
  2.3. IHF .................................................... 749
  2.4. Fis ..................................................... 749
  2.5. Multiple Interacting NAPs .......................................... 750
3. NAPs Occurrence and Distribution in Bacterial Genomes ............................................. 751
  3.1. Occurrence of NAPs in Chromosomes ............................................. 751
  3.2. Occurrence of NAPs in Episomal MGEs ............................................. 751
  3.3. Occurrence of NAPs in Integrated MGEs ............................................. 751
4. Functional Interactions between Endogenous and Xenogenic NAPs ............................................. 752
  4.1. Prevention of Endogenous NAPs Depletions ............................................. 752
  4.2. Modulation of NAPs Target Selectivity ............................................. 752
  4.3. Establishment of MGE-host Regulatory Circuits ............................................. 752
  4.4. Anti and Counter Silencing .................................................. 752

⁎ Corresponding authors at: Fundación Ciencia y Vida, Avenida Zañartu 1482, Ñuñoa, Santiago, Chile.
E-mail addresses: rquatrini@cienciavida.org (R. Quatrini), aloyola@cienciavida.org (A. Loyola).

https://doi.org/10.1016/j.csbj.2019.06.010
2001-0370/© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

Prokaryotes have the ability to acquire DNA from other microorganisms or their environment and incorporate it into their genomes through a process collectively referred to as Horizontal Gene Transfer (HGT) [145]. This phenomenon contributes to bacterial adaptation to changing environments and plays an important role in prokaryotic evolution. HGT is frequently mediated by mobile genetic elements (MGEs), i.e. by segments of DNA that encode proteins that mediate or facilitate self-movement within (intracellular mobility) or between cells (intercellular mobility). The MGEs are highly diverse and broadly include plages, plasmids and genomic islands (GI) [56,77]. While some MGEs are episomal and replicative in nature (e.g. plasmids), others are translocate or integrative and exist mostly in an intrachromosomal state (e.g. transposons and GIs). Others, such as Integrative Conjugal Elements (ICEs) fluctuate between the integrated and episomal state [173]. Regardless of the type of MGE all tend to be mosaics of functional gene modules of diverse evolutionary origins and genetically divergent from their hosts (e.g. [164]). Well known MGEs carry along pathogenicity [20], symbiosis [148], antibiotic and/or metal resistance functions [10,14], that convey increased fitness to their hosts under a diverse set of adaptive conditions.

Stable appropriation of MGEs and their gene cargo is determined by a number of host factors (e.g. replication compatibility factors) and mechanisms (e.g. restriction modification surveillance and/or CRISPR-Cas mediated degradation) that limit dispersal of MGEs [120,158]. However, when an MGE is acquired by HGT from distantly related species and enters a new host cell, the element has to face the problem of physical integration to a structured nucleoid and its gene cargo has to face the challenge of regulatory integration to its host regulatory network [38,39]. The bacterial nucleoids are highly compacted, folded and dynamic macromolecules, organized into large (30–300 kb) domains defined by long-range contacts (e.g. [87,92]) and smaller (15–30 kb) domains [73] grouping co-regulated genes [161]. In a multilayered structure compatible with all needed DNA transactions (replication, segregation, transcription, recombination, repair) [32]. These domains are shaped and modulated by different physical (e.g. supercoiling), chemical (e.g. hydration level) and biological (e.g. proteins) factors [85]. Among the later, Nucleoid-Associated Proteins (NAPs) act as central regulators of chromosome organization and play a relevant role in the facilitation or limitation of the physical integration of MGEs [92].

On the other hand, base composition (e.g., AT-rich sequences) and codon usage biases define horizontally acquired genes, gene clusters or genomic islands as xenogenic (or foreign), and impose transcription and translation limitations that may affect their persistence in the new host. Most if not all bacterial species have proteins that downregulate gene expression from these xenogenic sequences [118]. Xenogenic silencing enables bacteria to distinguishing their own DNA from foreign DNA and buy time for genetic amelioration (i.e. adjustment to the DNA composition of the new genome [86]) and for regulatory integration to occur (e.g. [40,96,108]). NAPs are relevant representatives of the group of regulators that perform these tasks as well, offsetting the fitness costs of foreign genes acquisition.

NAPs are small, basic proteins that bind DNA [32]. They are among the most abundant proteins in bacteria, reaching up to 60,000–80,000 copies/cell in well-studied model microbes such as Escherichia coli [2] or Pseudomonas putida [149]. Depending on the case, NAPs exert their action as either homo or heterodimers. Thanks to their sequence-independent affinity for DNA, NAPs show high promiscuity in DNA binding and can target a wide range of endogenous and xenogeneic sequences [39,136]. The binding of NAPs to the DNA alters its topology by wrapping, bending or bridging the nucleic acid [32] and also alters its folding into higher order structures (also known as supercoiling) by influencing both nano and meso-scale interactions [38,67,92]. The emerging models of NAP contribution to nucleoid structuring are transforming our general understanding of how the bacterial chromosomes are organized and function. Not only do NAPs compact the DNA, but they constrain negative supercoils and generate diffusion barriers for the formation of topological domains, while preserving a degree of supercoiling compatible with all DNA-related processes [159,160]. Exhaustive and profound reviews on these topics have been published in recent years [32,125], and are not further covered in this manuscript.

Due to their high protein titers and their DNA binding and bending properties, NAPs impact not only genome architecture [5,32], but affect all known essential DNA metabolism related processes such as replication [25,81], recombination and repair [80], and transcription [12,79]. NAP effects on gene expression, may be either repressive or stimulative, and may result from either global chromosomal DNA topology modification or specific and local effects on gene promoters. Because of the functional resemblance of these proteins to eukaryotic histones, NAPs were initially named “histone-like” proteins, yet they lack sequence homology to support the contention [45].

NAP encoding genes in bacteria are diverse and ubiquitous [39]; most if not all sequenced bacterial species encode at least one NAP per genome, being thus considered as housekeeping genes. Yet, the specific NAP cellular pools vary with the species or even with the strain considered (e.g. [114,149]). In E. coli, at least 12 NAPs have been found to be associated with the nucleoid [114]. In turn, the Lyme disease spirochete Borrelia burgdorferi, which possesses at least one of the most complex bacterial genomes known, with as many as 25 distinct replicons per cell, encodes only 4 putative NAPs [78]. Chromosomally-encoded NAPs (endogenous NAPs) bind DNA with different specificities and affinities at diverse locations along the bacterial chromosomes [64] and in apparent concentration gradients, with a coverage of roughly 1 NAP per 100 to 200 base pairs (reviewed in [125]). Also, their genes show distinct positioning along the bacterial chromosome [144], and distinct relative expression patterns according to the bacterial growth stage (e.g. [7]). Therefore, at any given time, the overall composition of the pool of NAPs of a bacterium, and the interactions these proteins establish, can be quite different and their effects on the gene expression programs dynamics highly complex [159].

The role of endogenous NAPs in silencing the transcriptional expression of foreign genes acquired by HGT is well established [38,107], as is the mechanism by which silencing is exerted and antagonized in several well studied microbial models and/or in several different growth conditions [40]. The relation between protein structure and function has also been dissected for some NAPs (e.g. [63]) and the global DNA-binding profiles have been determined for others (e.g. [121]). Recent studies have also uncovered extensive post-translational modifications of several NAPs [34], the functional significance of which still await investigation. In contrast, much less evidence on the occurrence of NAPs in MGEs (exogenous or foreign NAPs) has been presented to date and considerably little is known about the role of MGE-encoded NAPs in foreign elements biology and interactions, or on their influence on the host’s chromosome expression dynamics. Here, we review published evidence on the roles and mechanisms used by best-studied endogenous NAPs in the regulation of MGEs and different aspects of their biology (Fig. 1), along with the roles of MGE-encoded NAPs in the regulation of MGE.
and host biology (Fig. 2). Main challenges in dissecting the complex regulatory interactions these NAPs establish are also pinpointed.

2. Endogenous Nucleoid Associated Protein Types and Functions

Currently known NAPs belong to several distinct protein families (Table 1), including the “histone-like nucleoid structuring protein” or H-NS (Pfam PF00816), the “histone-like protein from Escherichia coli strain U93” or HU (Pfam PF14848), the closely related “integration host factor” or IHF (cd13832), and the “factor for inversion stimulation” or Fis (Pfam PF02954). Yet, other alternative NAPs and/or NAPs-like proteins have been described more recently, that also play important roles in cellular and DNA metabolism [32,39,41]. Most of these are small proteins, sharing little identity at the amino acid sequence level, able DNA-binding and oligomerization domains.

Others, such as Rok [143], NdpA [134] and Lrp [88] are larger in size (up to ~40 KDa) and bare additional domains that direct protein-protein interactions or alternative functions. This is the case of Rok, which interacts with the DNA bacterial replication initiator protein [134] and of NdpA/YejK, which interacts with the ParE ATPase subunit of the Topoisomerase IV [88]. In turn, the N-terminal domain of Lrp mediates leucine-dependent allosteric regulation of aminoacid metabolism genes (e.g. RAM domain [48]). Several NAP families remain less-well characterized (e.g. Efc, from Deinococcus radiodurans involved in protection from DNA damage [78,167,180]) or GapR, from Caulobacter crescentus involved in initiation of chromosome replication and partitioning [157]), and probably others still await discovery.

Relevant aspects of the biology of H-NS, HU/IHF and Fis family NAPs, summarized in Table 1, have been reviewed in several excellent and comprehensive publications in recent years [39,141,147], and only those that have been acknowledged to contribute to host-MGE or MGE-MGE biological interactions understanding are further covered in the sub/sections bellow.

2.1. H-NS

H-NS is a ~15 kDa, dimeric DNA binding protein encoded by the hns gene. It is very abundant in E. coli, which expresses up to 20,000 copies of the protein monomer per cell at the exponential growth phase [2]. Orthologs of hns are ubiquitous in bacteria, including several well-known pathogens (e.g. [22,62,103]). H-NS binds non-specifically to DNA, with a preference for AT-rich regions and curved DNA [175] typically found at promoters, insertion sequences and horizontally acquired genetic elements, exerting effects on host gene regulation (e.g. [124]), xenogeneic silencing (e.g. [96,181]), transposome stabilization (e.g. [153,168]) and nucleoid architecture (e.g. [126]) (Supplementary Table 1). H-NS acts as a repressor of its own transcription and that of over 200 other genes in E. coli [36,162]. Depending on the intracellular magnesium concentration H-NS shows two binding modes on DNA; oligomerizing along AT-rich regions or forming cross-bridges [39,182]. As a result, H-NS forms repressive nucleoprotein complexes that block and trap RNA polymerase at gene promoters [96], stiffen the DNA [94] or stabilize hairpins [63]. Recent evidence has demonstrated that local binding of H-NS restricts short-range interactions between discrete DNA regions and neighboring loci, supporting the role of this NAP in nucleoid organization [59,92]. The binding mode, the location and the repressive mechanism of H-NS varies in response to diverse types of perturbations in different model microorganisms (e.g. [3,63,183]). For example, Rafiei and colleagues analyzed the spatial reorganization of H-NS in response to osmotic stress, and uncovered the growth phase dependent detachment of H-NS and exclusion from the nucleoid volumes in response the presence of potassium or sodium ions [122]. H-NS and its orthologs have been found to play a crucial role in protecting the cell from detrimental effects of HGT by recognizing and silencing diverse low GC content MGEs in Salmonella [108], Bacillus [143] and Pseudomonas spp. [35,96,108]. Smaller versions of H-NS-like NAPs, showing structural mimicry to the H-NS oligimerization domain have also been identified in the chromosmes and MGEs of several microorganisms [97]. These proteins (Hha-like) form complexes with full-length H-NS and appear to comodulate the expression of several genes with this latter protein, many of which appear to be foreign in origin (see section 4).

2.2. HU

HU proteins are also low molecular weight, basic, heat-stable proteins that bind DNA as dimers in a sequence independent manner with a preference for AT-rich sequences [121] and at high frequency in most, if not all, bacteria [66]. In many bacteria, HU occur exclusively as homodimers (e.g. [113]), whereas in E. coli, two highly homologous subunits α and β (70% amino acidic identity), encoded by hupA and hupB respectively [128], can conform homo- or heterodimers depending on the growth cycle stage [26]. In microorganisms in which HU is the only NAP, its absence has proven to be lethal (e.g. [101]). In E. coli, the absence of HU is not lethal unless IHF and H-NS are deleted as well [176]. The interaction between HU and DNA seems to be nonspecific, but HU has a binding preference for distorted regions containing nicks or gaps [21]. Recently, a role for HU in maintaining DNA contacts in the megabase range, with a consequence in DNA folding of the ter macrodomain, has been presented [92]. Whilst individual HU dimers bend the DNA, the cooperative binding of HU can lead to higher-order complexes by dimer-dimer interaction [30], exerting not only DNA packaging functions that affect replication [75] and repair [21], but aiding in several DNA recombination-dependent events such as transposition [102] and/or inversion [76]. For instance, for transposition of phage Mu to occur optimally in E. coli, supercoiled Mu donor DNA, along with the bacterial protein HU bound at one of the att ends of the element, and E. coli’s HU protein bound at an enhancer site within the Mu genome, are required [70]. A comprehensive list of endogenous and foreign HU-family NAPs and their effects on MGEs can be found in...
2.3. IHF

IHF is closely related to HU and shares a high degree of amino acid sequence identity [1]. As in the case of HU, two genes, *ihfA* and *ihfB*, encode the α and β subunits. In *E. coli*, IHF exists as a heterodimer [106, 184], and its cellular concentration reaches its maximum at the end of exponential growth [2, 17]. In contrast to HU, IHF binds DNA with significant sequence specificity [15, 152, 153], yet it recognizes its binding site through indirect readout on the basis of structural or topological parameters. Nearly 1000 specific IHF binding sites have been identified in the *E. coli* chromosome, most of which occur in close vicinity of promoters [61, 163], having an extensive and global effect on the transcriptome [24, 84]. The mechanism of IHF DNA binding has been described in great detail, and key aspects have been summarized by Dillon and Dorman [32]. First described as a host cofactor with a role in integration and excision of bacteriophage lambda [105], IHF is now acknowledged to be involved in several other processes, such as transposition [68, 129], recombination [28, 47], and plasmid replication [50], among others. An extensive list of examples on the effects and influence of IHF on MGEs biology are listed in Supplementary Table 1. MGE-encoded IHF family proteins are covered in further detail below (Sections 3 and 4).

2.4. Fis

The Fis protein is a small and versatile NAP [41, 76], known to contribute to many different DNA-metabolism related processes and to affect global transcriptional patterns, with effects ranging from rather conventional positive or negative control (e.g. [64, 65, 79, 132]) and DNA-supercoiling-dependent preservation of transcriptionally open promoter configurations [4], to bacterial chromosome organization in independent looped domains of negatively supercoiled DNA [73]. Fis functions as a homodimer that binds to a 17 bp-long AT-rich DNA consensus sequence [24], recognizing the shape of the minor groove [146]. The dimer bends the DNA at its binding site [115], having a role in the genome architecture maintenance and remodeling [69], with both local and global effects [131]. By promoting contacts beyond 100 kb along the genome, this NAP is a global player of chromosome folding [92]. In *E. coli* there are up to ~1200 Fis-binding sites, one site every ~200 bases [72]. Its expression is maximal at the beginning of the...
exponential growth phase, having an important role in boosting the expression of genes involved in supplying components of the translation machinery [111]. During exponential growth Fis represses its own transcription and that of the stationary phase sigma factor rpoS, thus stimulating and maintaining the activity of relevant growth stage-dependent genes [24,79,98]. Importantly, Fis is also a transcriptional regulator of the DNA topoisomerase I (topA) and the DNA gyrase (gyrA/gyrB) encoding genes, mediating control of opposed DNA topology effects (relaxed vs negative supercoiled). The former is transcriptionally activated when DNA becomes more negatively supercoiled, while the latter is activated when DNA is relaxed [83,170]. In this way, Fis is thought to act as a superhelicity monitor and topological buffer controller, with relevant effects in all physiology [131,159,160]. Through various positive vs. negative effects, direct vs. indirect mechanisms, as well as through the interaction with other NAPs and transcriptional regulators, Fis efficiently controls basic cellular processes and specific genetic programs including virulence, e.g. in pathogenic Salmonella carrying pathogenicity islands SPI-1 and SPI-2 [19,166] and several other bacteria (reviewed by Duprey et al. [46]). The role for endogenous Fis family proteins as a regulators of site-specific DNA recombination of MGEs was uncovered early on. In E. coli Fis plays a role in lambda prophage lysogeny maintenance and dynamics, stimulating both excision and integration, depending on the presence or absence of the directionality factor Xis [6]. Also, Fis improves the efficiency of serine-invertase-driven recombination systems responsible for Gin-gix and Cin-cix phage tail fibre switching in bacteriophages Mu and P1, respectively [16,74], stimulates the frequency of transposition Tn5 and insertion sequence IS50 transposition events [169] and participates in site-specific recombination of class 1 integrons [18]. Additional effects of Fis-type NAPs on MGEs biology are listed in Supplementary Table 1.

2.5. Multiple Interacting NAPs

The individual roles of host-encoded NAPs in xenogeneic silencing of integrative MGEs such as pathogenicity islands and prophages was alluded-to above. However, frequently several bacterial NAPs are required to interact in order to fulfill silencing of whole MGEs as reported for the Mu transposable phase [59,70] or MGE-encoded genes, including several virulence factors of pathogenic bacteria (e.g. [130]). Interacting NAPs, also perform additional roles related with the integration and dispersal capacities of the MGEs. Requirement and role of host-encoded NAPs IHF and Fis, together with viral proteins Int (site-specific recombinase) and Xis (excisionase), in the assembly of the nucleoprotein complexes that carry out either integrative or excisive recombination of the Lambda bacteriophage are supported by a vast array of genetic, biochemical, and structural data (reviewed in Seah et al. [133]). This effect of host NAPs on the directionality of viral-host genome recombination extends to other lysogenic viral models (e.g. P2 [57] and T4 [179]). In addition to this role, HU and IHF proteins have been shown to influence lambda DNA packaging in E. coli. In mutants lacking these NAPs, the interaction of the Lambda terminase with the cos site is affected and the generation of the cohesive ends on the mature viral chromosome is impaired [100,174].

Evidence on the role of host NAPs on MGE biology has also begun to build up for integrative conjugal type of elements. In 2002, Connolly et al. [27] tested the role of IHF, Fis and HU on ICE_{Tn916} excision in E. coli MG1655 using single and double mutant strains. While Fis and IHF had no major impact on excision and growth rate, the absence of HU showed a ~90% attenuation of the excision efficiency. The authors postulated that HU may act by binding to and bending the ICE ends or stabilizing looped structures, similar to IHF in Lambda phage recombination. IHF is also required for efficient transfer of ICE_{STb} in Vibrio cholera [99], while Fis had no measurable effect [99]. In 2015, the Gardner’s group demonstrated that a protein named BHFα (Bacteroides host factor A), a member of the IHF/HU family, participated in the integrative recombination of the ICE_{CTODOT}, being the first host factor identified for a site-specific recombination reaction in this ICE family [127]. The authors speculated that BHFα participates in the integration reaction of ICE_{CTODOT} by binding and bending the DNA. Other in vitro studies have shown integration/excision of ICE elements and chromosomal target sequences without requirement of endogenous NAPs (e.g. [23]). Most studies though, have disregarded the influence of endogenous NAPs in MGE recombination. Host-encoded NAPs, Fis, IHF and HU also play a role in modulating the rates of transposition of transposable phase Mu [70] and several Insertion Sequences and transposons, including Tn10/J510 [135,142], Tn5/550 [171] and Tn903 [153]. Target sequences (attB) need to be located within a negatively supercoiled DNA substrate for recombination to proceed, which frequently occurs within other larger horizontally acquired mobile elements as reviewed by Dorman and Bogue [185].

Table 1

| Characteristics of the main nucleoid-associated proteins in bacteria. |
|------------------|------------------|------------------|------------------|
| **NAPs**         | **Size (kDa)**   | **Accession ID** | **Representative organism** | **Presence in MGEs** | **Acknowledged function** |
| H-NS family proteins | Y                | ND               | Y                | ND               | Y                   |
| H-NSe            | 15               | P0ACF8           | E. coli          | Y                | Xenogeneic silencing, Nucleoid structuring |
| H-SmT            | 15               | P0ACG1           | E. coli          | ND               | Functional analog of H-NS |
| MvlT             | 14               | Q9HW86           | Pseudomonas aeruginosa | Y                | Functional analog of H-NS |
| LrpF             | 14               | A0A0H0PFT0       | E. coli          | Y                | Homologue and antagonist of H-NS. Activator of LEE |
| Hfp              | 14               | O07507           | Bordetella pertussis | Y                | Functional analog of H-NS |
| BpE三十        | 14               | P0ACF8           | E. coli          | Y                | Functional analog of H-NS. Essential for B. pertussis |
| BVf3             | 13               | A4JS72           | Burkholderia viemanniensis | ND               | Functional analog of H-NS |
| HvrA             | 11               | P42505           | Rhodobacter capsulatus | ND               | Functional analog of H-NS |
| Ltr2             | 12               | P9WFIP7          | Mycobacterium tuberculosis | ND               | Functional analog of H-NS |
| XiVA             | 15               | Q56835           | Xanthomonas oryzae | ND               | Functional analog of H-NS |
| Rok              | 22               | O34857           | Bacillus subtilis | ND               | Functional analog of H-NS. Repressor of ComK |
| HU/HIF family proteins | Y                | ND               | Y                | ND               | Y                   |
| HU               | 9                | P0ACF0           | E. coli          | Y                | DNA replication, repair, recombination packaging |
| IHF              | 11               | P0A687/0A6Y1     | E. coli          | Y                | DNA transposition, recombination, plasmid replication |
| Fis family proteins | ND               | ND               | ND               | ND               | Gene regulation, nucleoid architecture, DNA remodeling |
| Fis              | 11               | P0A6R3           | E. coli          | ND               | Gene regulation |
| Other            |                  |                  |                  |                  | Gene regulation |
| Lrp              | 19               | P0ACJ0           | E. coli          | ND               | Gene regulation |
| Ebc              | 11               | G51418           | B. burgdorferi    | ND               | Gene regulation |
| Nse              | 37               | A0A02441X9       | E. coli          | Y                | Gene regulation |
3. NAPs Occurrence and Distribution in Bacterial Genomes

3.1. Occurrence of NAPs in Chromosomes

NAPs are recognized as ubiquitous chromosomal genes in bacteria, ranging from one in small genome-sized microbes such as Mycoplasma (e.g. [37]) to more than 10 diverse NAP gene variants in bacteria such as E. coli (e.g. [2]). Even if the specific literature on the topic acknowledges that each species has its own unique set of NAP variants, systematic assessment of the species-specific NAP pools or the phyletic patterns of the NAP protein families are still lacking. The few genome-wide searches performed to date have discrepant results and are not entirely comparable. Using a sequence similarity-based search approach (BLASTn), and a rather confined set of query proteins, Takeda and colleagues [155] determined the occurrence of NAP-encoding genes among 588 proteobacterial genomes. Few years later the same group performed a follow-up study and extended the analysis to 3056 closed bacterial genomic sequences spanning 3 taxonomic phyla (Proteobacteria, Firmicutes and Actinobacteria) [136]. Relevant results from these studies revealed that nearly 30% of the genomes encoded at least one NAP, with Beta- and Gammaproteobacteria carrying the majority of the NAPs identified and Hu/IFH being the most highly recovered NAPs (~63%). A study by Perez-Rueda and Ibarra [118] using 2265 bacterial genomes (non-redundant selection) and Hidden Markov Model based searches of candidate NAPs uncovered a remarkably different figure of ~94% NAP occurrence in the microbial genomes analyzed, with H-NS (82%) being the most frequent type of NAP in the set. The differences in types and percentages recovered by both studies arise fundamentally from the search methodology used by each, although annotation issues cannot be discarded.

Likewise, little evidence has been gathered so far on the distribution of the NAPs between genomic compartments (including chromosomes vs. plasmids), and subcompartments (including integrated vs. episomal MGEs). A notable exception is E. coli, where NAP-encoding genes have been mapped along the chromosome [226]. Most of these NAP genes’ position (with the exception of hns) correlate with the temporal order of expression during bacterial growth, being the exponential phase induced genes hupA and fis located in the Ori-region, while late exponential or stationary phase genes ihfA and ihfB, are located towards the Ter-region (reviewed in Rimsky and Travers [125]). Much room for exploration remains in this respect for non-model microbes, from both bioinformatics and experimental standpoints. Some of the relevant questions still awaiting an answer are: What is the frequency of NAP genes in MGEs? Are MGE-encoded NAP variants different from chromosomally encoded ones? Do MGE-encoded NAPs exert a role on host biology or are their functions related exclusively to MGE biology? In the following sections we review pioneering and recent literature that have begun to address these questions.

3.2. Occurrence of NAPs in Episomal MGEs

Over the past decade, a number of plasmids carrying NAP gene orthologs have been described and the role of these proteins in the regulation of transcriptional networks between the host’s chromosome and the episomal replicons have begun to be elucidated. This is the case of the H-NS paralogues encoded in several large conjugative plasmids including the IncP-7 plasmid pCAR1 from Pseudomonas putida KT2440 [177], the IncH1 group R27 family plasmids of Shigella and Salmonella spp. [9,11], and p0908 from Vibrio spp. [71]. Many H-NS-like proteins (designated Hha/RmoA/Hmo/YdfA), showing structural similarity to the oligomerization domain of these NAPs but having only about half of their molecular mass, have also been identified in the plasmids of several enterobacteria [87,109,116].

Nojiri’s group recently analyzed more than 4600 plasmid sequences spanning diverse replication and mobilization incompatibility groups [136,155]. Results of these studies revealed that ~10% of the plasmids analyzed carried identifiable NAP orthologs, and that little less than 3% of them encoded more than one NAP gene per replicon. One remarkable observation of these studies was the higher frequency of NAP genes in plasmids (1 per 236 kb) relative to proteobacterial chromosomes (1 per 1.8 Mb), suggesting that episomal NAPs play relevant roles in plasmid adaptive biology (some examples are shown in Section 4). The authors also uncovered a clear positive correlation between NAP gene occurrence and plasmids size [155], possibly indicating that large plasmids have higher costs on host’s fitness than small plasmids, and would thus require more resources to ensure xenogeneic silencing for successful appropriation. However, to support this assertion NAP binding sites frequency and distribution should also be analyzed in this set of plasmids and the cognate host’s genomes. To our knowledge, few studies of the kind have been published to date [96,108,177]. The plasmid-encoded NAPs identified so far include H-NS, Hu, IFH and Fis representatives, in decreasing order of abundance. The majority of the H-NS family proteins were found in plasmids of the Gammaproteobacteria, while Hu and IFH family representatives were distributed all over Proteobacteria. Remarkably, no Fis ortholog was found in the queried plasmid dataset in 2011 and only one was retrieved in the 2015 search [136,155]. Reasons for the distinct distribution of fis orthologs remain to be explored, but could be related to their role in site-specific recombination which facilitates integration, excision and/or inversion of integrative MGEs rather than plasmid-related processes. It will be interesting to assess the presence of Fis-like NAPs in integrative elements more systematically to put this interpretation to test.

3.3. Occurrence of NAPs in Integrated MGEs

The occurrence of NAPs in integrative mobile elements has been much less explored than both host and plasmid-encoded NAPs. A relatively small number of NAP-encoding genes occurring in integrated MGEs, such as prophages, transposons and ICEs are presently acknowledged (Supplementary Table 1). Phage-encoded NAPs include the Lss2-type Cgp3 protein (H-NS-like) of the CGP3 prophage from Corynebacterium glutamicum ATCC 13032 [119], the 5.5 protein of the E. coli bacteriophage T7 [95] and the TFI protein (HU-like) of the E. coli bacteriophage SPO1 [60], as well as the MuB protein of the transposable phage Mu [59]. An ortholog of the IFH/HU family of NAPs, BFIa, has been reported to occur in the CtnDOT conjugative transposon of Bacteroides species [127]. Full-length H-NS-like proteins have been described in a number of integrative MGEs, including the H-NSB and HfP proteins from the serU-island of uprphlogenic E. coli strains [104,172] and the Ler protein from the Locus of Enterocyte Effacement (LEE) pathogenicity island of enteropathogenic E. coli (EPEC) [43,89]. Also, truncated H-NS variants (H-NST) lacking the DNA-binding carboxyl terminal domain and acting as interaction partners for H-NS [13], have been reported to occur in the LEE pathogenicity island [90] and other integrated MGEs [90,172].

Many additional, as of yet uncharacterized, NAPs and NAP-like genes can be traced to reported annotations of integrated genomic islands (e.g. [51,156]) or available sequences in MGE data repositories (e.g. ICE-Berg database [93]). Given the similar challenges experienced by integrative MGEs to persist and adapt to their host’s biology, and the accumulating evidence on the high abundance of NAPs in episomal MGEs, occurrence of NAP-encoding genes in integrative elements is also expectable. The functional interactions established by the best studied MGE-encoded NAPs are further discussed bellow.

4. Functional Interactions between Endogenous and Xenogenic NAPs

Even if most functional studies on NAPs to date have focused on endogenous host-encoded proteins acting as silencers of foreign DNA, a couple of thorough studies have provided relevant clues on the interaction mechanisms between endogenous and xenogenic NAPs and the
functional outputs of those interactions. Four aspects are further considered herein.

4.1. Prevention of Endogenous NAPs Depletions

Work by Dorman and colleagues using Shigella and E. coli as models (reviewed in [39]) has shown that these bacteria maintain constant ratios of H-NS to DNA, required to meet the physiological needs of the bacteria during rapid exponential growth [44,53]. Frequent concurring endogenous H-NS-like NAPs (e.g. [11]), showing growth phase-dependent transient patterns of expression, supplement the cellular NAP pools as needed (e.g. [31]). In this context, newly acquired MGEs with potential target sites for the endogenous NAPs are likely to titrate the cellular NAP pool and cause a redistribution of the endogenous proteins, with probable pleiotropic effects on the physiology and fitness of bacteria. The presence of NAPs genes in MGEs has been interpreted as source of NAPs that could act to prevent depletion of the endogenous proteins by binding to self sequences, allowing MGE maintenance and/or transmission to new hosts without reducing their fitness (e.g. [44]).

4.2. Modulation of NAPs Target Selectivity

Using a hns mutant of Salmonella enterica serovar Typhimurium SV5015, microarrays and a ChiP on chip approach, H-NS binding sites have been mapped to low GC content regions along the chromosome, including both core and foreign genes [96,108]. Studies on this strain by other authors further demonstrated that the plasmid encoded variant of H-NS (H-NS<sup>AC27</sup>) targets only a subset of the genes bound by the endogenous H-NS, selectively modulating horizontally acquired genes [9]. Similar evidence has been presented also for Shigella [33]. Although plasmid and endogenous forms of NAPs have been assumed to be functionally equivalent (e.g. [52]), evidence on the structural and functional features that differentiate them has begun to emerge (e.g. [8]). At least in the case of H-NS-like proteins, these differentiating features seem to contribute to target selectivity, and entail differential interactions with the Hha family of co-modulators of gene expression (all lacking the nucleic acid binding and linker domains of full length H-NS NAPs). While endogenous H-NS homologomers modulate core genes, silencing of foreign genes by the endogenous proteins is achieved through heteromeric H-NS-Hha complexes [9,110,117]. In contrast, the plasmid-encoded full-length H-NS<sup>AC27</sup> protein selectively modulates foreign genes [8]. The fact that many Hha-like antagonist proteins are encoded by genes within plasmids and pathogenicity islands of several microorganisms [97,147], suggests that their contribution to the hosts capacity to discriminate self from foreign DNA provides a fitness advantage and thereby facilitates the acquisition, amelioration and maintenance of these MGEs.

Other H-NS/H-NST (truncated) heterodimers that directly modulate the repressive activity of the host-encoded H-NS protein with either anti-silencing or co-repressor effects have been described in the literature (summarized by [39,147]). This is the case of the T7-phage-encoded protein 5.5, which interacts with H-NS in a similar manner to H-NS in order to derepress T7 RNA polymerase-mediated transcription [95], and also that of several horizontally acquired genomic islands of Yersinia (e.g. [29,110]), Salmonella (e.g. [140]) and pathogenic strains of E. coli [172]. In addition to H-NS, C-terminal truncated variants of H-NS can also interact with other full-length H-NS-like proteins, like the StpA paralog of H-NS in E. coli, to co-repress the H-NS target operons [55] in a dose-dependent manner [54]. Additional, non-mutually exclusive, mechanisms of action for truncated variants of H-NS (H-NST) have also been put forward by other authors. These mechanisms rely on an H-NS-independent DNA binding activity of the truncated proteins [90] and a proteolysis protection effect (e.g. [117]). The different alleviation mechanisms proposed for the H-NST variants, on H-NS, have been extensively covered by Dorman [39].

4.3. Establishment of MGE-host Regulatory Circuits

One of the most thorough follow up studies reported to date that provides proof of the interaction between plasmid-encoded NAPs and host-encoded NAPs is that performed on the carbazole-degradative plasmid of Pseudomonas spp., designated pCAR1 [137]. pCAR1 is a self-transmissible plasmid used as model to study the interactions between MGEs and their hosts at several different levels, from the molecular [138] to the ecophysiological level [186]. One of the pCAR1 bearing species, Pseudomonas putida KT2440, encodes in its chromosome five H-NS family NAP genes (turA, turb, turC, turD, turE) [139]. The pCAR1 plasmid itself carries three NAP genes including pmr, phu and pnd, coding for orthologs of H-NS, HU and NdpA, respectively [112]. All these proteins are involved in the control of the transcriptional network between the plasmid and chromosome. Early work on this system demonstrated that the presence of pCAR1 affects the transcriptional expression of the plasmid-encoded NAP gene pmr and the endogenous NAP-encoding genes turA and turb, among other chromosomally encoded genes. Comparative studies of the transcriptional profiles of the wild type strain with or without the pCAR1 plasmid, and a pmr mutant, showed that pmr disruption had far greater effects on the host transcriptome than did pCAR1 carriage [177]. Follow-up experiments using proteomic approaches and pCAR1-free versus pCAR1-harboring cells, further confirmed that plasmid carriage greatly affects many host related processes, including the utilization of several metabolites and amino acids, along with respiration [165].

Studies to dissect the mechanisms behind these effects include protein-protein, protein-DNA and protein modification assays. Using in vitro pull-down assays Yun and colleagues showed that Pmr strongly interacts with itself, and with TurA, TurB and TurE, while ChiP-chip analysis showed the Pmr binding sites to locate preferentially at chromosomal intergenic regions and foreign DNA regions with low GC content [177]. Additional gene disruption studies, evaluating the effect of single or double disruption of the three pCAR1 encoded NAP genes revealed that simultaneous disruption of pmr and either pnd or phu caused decreased segregational stability and transfer frequency of pCAR1, highlighting the synergistic effects of these plasmid-encoded NAPs in pCAR1 replication, maintenance, and transfer [151]. Interestingly, double and single mutatations of the exogenous NAPs also proved to affect several aspects of the host's phenotype (e.g. energy production and conversion, biofilm formation [151]), indicating that Pmr is a key factor in optimizing gene transcription on pCAR1 and the host chromosome. Further work from this group showed also that Pmr can physically interact in vitro with the two endogenous H-NS proteins, TurA and TurB, yet the protein-protein binding affinities proved to be higher between TurA and TurB than between either of them and Pmr [150]. While Pmr intracellular protein levels remain constant (~30,000 monomers per cell) during cell growth, the relative amounts of TurB increase as TurA decreases [149], suggesting that TurA and TurB play complementary roles in different stages of cell growth. Interestingly, the binding sites of the three NAPs are almost identical and all three proteins regulate the transcriptional networks in the plasmid-harboring cells cooperatively [178]. Recently, a comprehensive analysis of the post-translational modifications (PTMs) of the Pseudomonas proteome added an additional layer of information to our current understanding of the pCAR1 plasmid impact on its host [165], being this also the first report of PTM effects in plasmid-host interactions.

4.4. Anti and Counter Silencing

NAP-mediated xenogeneic silencing of MGEs and their gene cargo prevents the expression of foreign genes acquired horizontally from other microorganisms, while these become integrated into the existing regulatory circuits of the new host. Besides the mechanisms discussed
above, involving modulation of full-length NAPs effects by truncated paralogues (e.g. Hha/H-NS), additional anti- or counter-silencing strategies which can relieve the silencing effects of NAPs have been described in fairly recent literature. Advances in our understanding of the mechanisms behind these strategies have already been reviewed at great depth by others [39,147,190] and are thus only briefly covered herein to complete the landscape of interactions affecting MGEs-encoded NAPs.

Countering silencers alludes to the regulated relief of NAP-mediated transcriptional repression and entails disruption of both simple and higher-order structure nucleoprotein complexes exerting repression of gene expression. Both physical and chemical factors such as temperature and osmolarity have been shown to undermine the ability of NAPs (e.g. H-NS) to maintain bridged structures, thus reducing the degree of DNA curvature at specific gene promoters (e.g. virF) and unmasking binding sites for further transcriptional activation [187]. Also, several specific proteins can remodel the NAPs:DNA nucleoprotein complexes locally, enabling the RNA polymerase to bind and act on the promoter sites. Such is the case of StyA which displaces H-NS at the hlyE promoter in E. coli [188], or LeuO which blocks the oligomerization of the H-NS at leuO promoter in Salmonella [189], among several others [190]. Many of these proteins appear to have been co-opted from their original functions as transcriptional repressors to serve as counter-silencers [191], playing a crucial role in facilitating adaptation and evolution of bacteria by horizontal gene transfer.

5. Concluding Remarks

Even if the evidence supporting the individual and collective roles of NAP in bacteria abounds, and their importance as DNA binding, bending and bridging factors is nowadays well established, the types and mechanistic details of the complex web of interactions they establish in diverse microbial (and physiological) contexts is far less clear (Fig. 2). To disentangle such complexity, several aspects need to be considered on a case-by-case basis: the origin of the interacting NAPs (endogenous vs. xenogenic), the nature of the target DNA (domains, regions, MGEs, genes, promoters), the type of interactions established between NAPs (e.g. regulatory, cooperative, doping) and the immediate output response of such interaction (e.g. nucleoprotein complexes remodeling, DNA stabilization, transcriptional silencing). Reports summarized above, and in several other comprehensive reviews listed in this revision, support the view that higher order interactions between NAPs occur simultaneously, of both synergistic and antagonizing nature, e.g. auto- vs. cross-regulation, formation of homo- vs. heteromers or oligo- vs. polymerization. All of these aspects are intimately linked with available NAPs titer in the cell, their relative balance and the effective (or potential) binding sites occupancy. Examples of the vast array of interactions between NAPs and other cellular- and MGE-encoded transcriptional regulators and the regulatory hierarchies they establish to exert coordinated control of gene expression, are beginning to emerge (e.g. mechanisms coordinating virulence gene expression in S. flexneri [42]).

From these few examples, it is apparent that host and MGE-encoded NAPs interact both positively and negatively to produce effects on MGE genes transcriptional expression, recombination dynamics, maintenance in the integrated state, and even dispersal to suitable hosts. Given the vast effects that NAPs exert on nucleoid architecture and compartmentation, and other DNA related processes, which exert hierarchical interactions at diverse levels in response to both physiological and environmental cues, many more influences of NAPs in MGE biology are likely to emerge in the upcoming years. Additional studies combining high throughput technologies of ever increasing resolution, are already paving the way in this direction.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2019.06.010.

Acknowledgements

This work was supported by the Comisión Nacional de Investigación Científica y Tecnológica (under Grants FONDECYT 1118251 to R.Q., FONDECYT 1160480 to A.L., Programa de Apoyo a Centros con Financiamiento Basal, Chile AFBI170004 to R.Q. and A.L.), Fundación Ciencia & Vida Hinge PostDoc Program, Chile to R.F.R. and by Millennium Science Initiative, Ministry of Economy, Development and Tourism of Chile, Chile (under Grant “Millennium Nucleus in the Biology of the Intestinal Microbiota”) to R.Q.

References


