Computers in Biology and Medicine 42 (2012) 1133–1140

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Interplay between Fur and HNS in controlling virulence gene expression in *Salmonella typhimurium*

Mahendra Kumar Prajapat [a], Supreet Saini [b,*]

[a] Chemical Engineering, Indian Institute of Technology Gandhinagar, VGEC Campus, Chandkheda, Ahmedabad, Gujarat 382424, India
[b] Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400067, India

ARTICLE INFO

Article history:
Received 7 June 2012
Accepted 11 September 2012

Keywords:
Salmonella typhimurium
Salmonella Pathogenicity Island
Fur
HNS
T3SS
AT%
Transcription factors
Gene expression
Fimbriae
Flagella

ABSTRACT

Salmonella enterica is responsible for a large number of diseases in a wide-range of hosts. Two of the global regulators involved in controlling gene expression during the infection cycle of the bacterium are Fur and HNS. In this paper, we demonstrate computationally that Fur and HNS have disproportionately high density of binding sites in the Pathogenicity Islands on the *Salmonella* chromosome. Moreover, the frequency of binding sites for the two proteins is correlated throughout the genome of the organism. These results indicate a complex interplay between Fur and HNS in regulating cellular global behavior.

1. Introduction

Salmonella is a gram-negative bacterium which causes a variety of diseases, ranging from the self-limiting gastroenteritis to the life-threatening systemic infection in a broad range of hosts [1–3]. The ability to cause infection is due to presence of a number of virulence factors (like SPI1, SPI2, fim, SPI4, SPI5) encoded on the *Salmonella* chromosome [4–12]. These virulence factors are in turn controlled by an assortment of virulence specific gene expression regulators as well as global regulators [13–16]. Two of the global regulators involved in control of gene expression are Fur and HNS [17,18].

Ferric uptake regulator (Fur) is a Fe^{2+} dependent transcriptional regulator in bacteria, which can act as an activator or repressor of gene expression [17,19–24]. In its most common mode of regulation, in the presence of Fe^{2+}, Fur binds to the Fe^{2+} ion, and represses expression of its target genes [20,25]. Fur has also been found to control virulence gene expression in *Salmonella*. In fact, in a fur mutant, the expression of critical invasion regulators, like *hilA*, is reduced to near-background levels [17]. The precise mechanistic details of this Fur-mediated positive regulation are not clearly known. It is also not well understood whether Fur controls the expression of pathogenicity-related genes in *Salmonella* directly, or by its action on another global/local regulator [17,19,22,26–28].

HNS belongs to a family of small nucleoid-associated proteins in gram-negative bacteria [29,30]. The proteins in this family have the ability to bind DNA and are known to negatively regulate gene expression by shielding DNA from transcription factors and RNA polymerase. It is thought that the shielding of DNA by HNS begins at a preferred nucleation site (~9 bases), and thereafter, HNS molecules polymerize to shield large regions of DNA [31]. HNS is known to preferentially bind AT-rich regions in the chromosome and silence gene expression. In terms of pathogenesis in *Salmonella*, HNS is known to negatively regulate expression of the genes involved in invasion [31–34]. Since these genes are horizontally acquired, they often have an AT% different from the rest of the chromosome, and an effective target for HNS-mediated silencing [31,35–37].

How Fur and HNS control virulence gene expression in *Salmonella* has been the focus of a number of studies recently [17,19,20,22]. The precise mechanistic details of this regulation are not yet understood. In this report, we do a genome scan of the Fur-binding sites and HNS-nucleation sites in the *Salmonella* chromosome. Our results indicate that, as expected, Fur binding and HNS nucleation sites are well spread throughout the chromosome. However, there is an uncharacteristic high density of HNS and Fur binding sites in the various *Salmonella* Pathogenicity Islands (SPIs).

* Corresponding author. Tel.: +91 22 2576 7216.
E-mail addresses: supreet.saini@gmail.com, saini@che.iitb.ac.in (S. Saini).

0010-4825/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2012.09.005

The SPIs are horizontally acquired pieces of DNA and are responsible for the bacterium gaining entry into the mammalian cells and for survival inside host cells [38–40]. Moreover, even within SPIs, the Fur and HNS binding sites seem to be concentrated around the transcriptional regulators in the two Pathogenicity Islands [12,41]. These include *hilD* and *ssrAB*, the two key regulators which control SPI1 and SPI2 gene expression [17,42–45]. In addition, we also located a number of Fur and HNS binding sites in close proximity to accessory regulators *hilC*, *sprB*, and *hilA* [17,19,33,46,47]. Overall, there seems to be a strong correlation in the frequency of Fur binding and HNS nucleation sites in the *Salmonella* chromosome. These results point to a complex regulatory interplay between HNS and Fur at a global level.

2. Materials and methods

The genome of *Salmonella typhimurium* LT2 14028 [48] was used as a template to perform scanning for locating Fur binding sites, HNS-nucleation sites, and computing AT base percentage in the chromosome. To locate putative Fur boxes, we constructed a Position Specific Scoring Matrix (PSSM) from experimentally verified Fur boxes in the Enterobacteriaceae family [49].

2.1. Constructing the Position Specific Scoring Matrix (PSSM)

The calculation of the PSSM for Fur boxes was done using fifteen previously experimentally identified Fur boxes [49]. Background frequencies B1 and B2 were obtained from Multiple Expectation Maximum for Motif Elicitation (MEME) algorithm [50,51] and GIBBS sampling algorithm [51,52], respectively for getting relative frequencies (Eqs. (1) and (2)). In case, a count and probability of nucleotide at particular position in Fur box encountered as zero, which is quite common, might not be converted to logarithms, so the background frequency at that position was taken under consideration and penalized by multiplying of intergenic fraction or overall frequency [53]. Generally it is considered as 0.25 for each nucleotide in case of bacteria because of the almost equal *A*, *T*, *G*, and *C* frequencies [51,54]. The averaged relative frequency (Eq. (3)) was then normalized (Eq. (4)) with minimum individual frequency (0.25) and then converted it into likelihood score (Eq. (5)) by taking logarithm of each for better observation. By having logarithm value of each score (positive or negative) at every position on Fur box, we were able to determine the possibility of occurrence a nucleotide which had maximum score at same position and constructed PSSM. The resulting PSSM has four columns representing score of occurring possibility of bases *A*, *T*, *C* and *G* and the number of the rows represents the length of the Fur box (Table 1).

$$F_m = [n+(B_1 \times 0.25)]/(N+B_1) \quad (1)$$

$$F_g = [n+(B_2 \times 0.25)]/(N+B_2) \quad (2)$$

$$F = (F_m+F_g)/2 \quad (3)$$

$$F_n = F/0.25 \quad (4)$$

$$Score = \log_2(F_n) \quad (5)$$

$$Score_{(FUR)} = \sum_{i=1}^{N} Score(i) \quad (6)$$

$$Threshold/Cutoff = \min\left(\sum_{j=1}^{19} Score_{(FUR)}\right) \quad (7)$$

Table 1
The Position Specific Scoring Matrix (PSSM) for Fur box. The table lists the relative scores a base at all 19 bases in a Fur box. The cumulative scores for a sequence of 19 bases were used to compute the Fur box score.

Position	A	T	G	C
1	−1.538	0.376	0.828	−0.779
2	0.084	−0.284	−0.779	0.620
3	0.828	1.010	−3.239	−3.239
4	1.881	−3.239	−3.239	−3.239
5	1.784	−1.538	−3.239	−3.239
6	−1.538	1.681	−3.239	−1.538
7	−0.779	−1.538	1.316	−0.779
8	1.681	−1.538	−3.239	−1.538
9	−1.538	1.681	−1.538	−3.239
10	1.569	−0.284	−3.239	−3.239
11	1.569	−0.284	−3.239	−3.239
12	−1.538	1.316	−0.284	−1.538
13	−0.284	0.376	−3.239	0.828
14	1.681	−3.239	−0.779	−3.239
15	−3.239	1.569	−3.239	−0.284
16	−3.239	1.881	−3.239	−3.239
17	0.828	1.010	−3.239	−3.239
18	−3.239	1.784	−1.538	−3.239
19	−3.239	0.376	−3.239	1.316

Here *N* is the number of particular nucleotide in a column, $N=15$ is the total number of nucleotides in a column, 0.25 is the overall frequency (intergenic fraction), $B_1=0.1$ is the MEME algorithm, $B_2=\sqrt{N}$ is the Gibb's algorithm, F_m and F_g are relative frequencies, F is the average of relative frequencies (F_m and F_g), F_n is the normalized relative frequency, *Score* is the likelihood score, and $Score_{(Fur)}$ is the score of whole FUR box.

We took sum of each calculated likelihood score at each position in a Fur box that was $score_{(Fur)}$ and repeated this procedure for all 15 sequences and were used in constructing PSSM, the minimum value of sum of likelihood scores among all Fur boxes was set as Threshold value from Eq. (7) [55]. The *S. typhimurium* genome was scanned through both forward and reverse strand through sliding a window of 19 bases one base at a time.

The likelihood score of each position in window $Score_{(Fur)}$ through complete scanning was calculated. If the $Score_{(Fur)}$ of each window was equal or more than the threshold (12.20), the window was then considered as a hit else neglected. Every hit was located on the genome [56] and listed with its particular target, location (in a gene, upstream or/and downstream to gene) and sequence of Fur box in Table S1.

2.2. HNS nucleation sites

The *Salmonella* genome was scanned for locating HNS nucleation sites (9 bases long) [31]. The strategy was either exact fit of HNS box (fit value 9 out of 9) or a fit with freedom of having one different nucleotide (fit value 8 out 9) at any position. Both these situations were considered as a hit for H-NS boxes prediction and listed in Table S2.

From the lists of predicted Fur boxes and HNS nucleation sites, we calculated the frequency of these sites in a particular window through entire genome. Specifically, we calculated frequency of hits (Fur boxes and HNS-nucleation sites) along with AT% in bin size of 100,000 nucleotides. All codes used to perform task were programmed in the PERL, and are given in the Supplementary material (S3).

2.3. AT percent calculation

$$AT\% = \left[\frac{(\text{Total no. of } A+\text{Total no. of } T)}{\text{Total no. of nucleotides over entire genome}}\right] \times 100 \quad (8)$$