6762–6777 *Nucleic Acids Research, 2007, Vol. 35, No. 20* *Published online 5 October 2007*
doi:10.1093/nar/gkm631

Discovery of Fur binding site clusters in *Escherichia coli* by information theory models

Zehua Chen[1], Karen A. Lewis[1], Ryan K. Shultzaberger[1], Ilya G. Lyakhov[1,2], Ming Zheng[3], Bernard Doan[3,4], Gisela Storz[3] and Thomas D. Schneider[1,*]

[1]National Cancer Institute at Frederick, Center for Cancer Research Nanobiology Program, [2]Basic Research Program, SAIC-Frederick, Inc., National Cancer Institute at Frederick, Frederick, MD 21702-1201, [3]National Institute of Child Health and Human Development, Cell Biology and Metabolism Branch and [4]Division of Extramural Activities, Referral and Program Analysis Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA

Received June 5, 2007; Revised July 28, 2007; Accepted July 31, 2007

ABSTRACT

Fur is a DNA binding protein that represses bacterial iron uptake systems. Eleven footprinted *Escherichia coli* Fur binding sites were used to create an initial information theory model of Fur binding, which was then refined by adding 13 experimentally confirmed sites. When the refined model was scanned across all available footprinted sequences, sequence walkers, which are visual depictions of predicted binding sites, frequently appeared in clusters that fit the footprints (~83% coverage). This indicated that the model can accurately predict Fur binding. Within the clusters, individual walkers were separated from their neighbors by exactly 3 or 6 bases, consistent with models in which Fur dimers bind on different faces of the DNA helix. When the *E. coli* genome was scanned, we found 363 unique clusters, which includes all known Fur-repressed genes that are involved in iron metabolism. In contrast, only a few of the known Fur-activated genes have predicted Fur binding sites at their promoters. These observations suggest that Fur is either a direct repressor or an indirect activator. The *Pseudomonas aeruginosa* and *Bacillus subtilis* Fur models are highly similar to the *E. coli* Fur model, suggesting that the Fur–DNA recognition mechanism may be conserved for even distantly related bacteria.

INTRODUCTION

The protein Fur is the 16.8 kDa product of the *fur* (*ferric uptake regulation*) gene in *Escherichia coli* (1), so named because it was first observed to repress the transcription of genes that code for components of ferric (Fe^{+3}) uptake systems found in the cell membrane. Since then, Fur also has been found to regulate other genes that are not directly related to iron transport, such as those encoding hemolysin, Shiga-like toxin and manganese superoxide dismutase (2–5).

Fur binds to DNA and represses transcription in the presence of divalent metal ions. The ion is thought to be Fe^{+2} *in vivo* (6), however, DNase I footprinting experiments have shown that Fur also binds to DNA in the presence of Mn^{+2}, Co^{+2}, Cu^{+2}, Cd^{+2}, and Zn^{+2} (7). Recent studies have suggested that purified Fur contains at least one Zn^{+2} ion as a structural stabilizer (8). Fur has been observed to bind to DNA as a dimer and in higher order polymers (7,9), and electron microscopy has shown polymerization of Fur on DNA under high concentrations of protein and metal ions (2).

Numerous strategies have been employed to find new Fur binding sites. Various consensus sequences have been derived from both footprinted and non-footprinted Fur binding sites (3,7,10) and these have been compared to sequences in the promoter region of suspected iron-regulated genes. Putative Fur targets were then investigated further through genetic and biochemical experiments. Stojiljkovic *et al.* created a successful

*To whom correspondence should be addressed. Tel: +1 301 846 5581; Fax: +1 301 846 5598; Email: toms@ncifcrf.gov

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Present address:
Karen A. Lewis, Department of Physiology, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390-9040, USA.
Ryan K. Shultzaberger, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3202, USA.
Ming Zheng, Dupont Central Research and Development, Experimental Station E328-B31, Wilmington, DE 19880-0328, USA.
Zehua Chen, Department of Biology, Boston College, MA 02467, USA.

'Fur titration assay' to locate new Fur binding sites using an *fhuF*:*lacZ* fusion and Fur consensus sequence-containing plasmid titrant on MacConkey plates (1). Several new iron-regulated genes in *E. coli* were discovered using this consensus sequence-based technique. In addition to the above, studies have also been carried out using *E. coli* Fur for DNase I footprinting with non-*E. coli* DNA (11,12). Recently, transcriptional profiles of *E. coli* genes have been used to determine those that are regulated by iron and Fur by evaluating mRNA levels in the absence of iron or Fur protein (13).

Another method for finding Fur-regulated genes is to use molecular information theory to locate new binding sites. Using this approach, classical information theory (14,15) is applied to molecular biology (16). First, a set of binding sites is aligned by maximizing the information content (17), and then the average pattern at the sites is represented by a computer graphic called a sequence logo (18). Next, the conservation of bases in the aligned set is used to create a weight matrix model that assigns a weight in bits to each base at each position according to its frequency in the data set (19). This can be displayed using the sequence walker graphic (20).

In addition to displaying details of binding sites, sequence logos can be used to understand the mechanism of binding. In instances where factors bind in overlapping clusters, it is difficult to assign the relative contribution of a base in an overlapping region to the appropriate binder or to determine the range of the binding site. Here, we tested several Fur binding site models that were obtained by multiply aligning Fur binding sequences using different window sizes, and identified the model that best represents binding by a single Fur dimer.

Information theory has previously been used to build two models to evaluate and predict Fur binding sites (13,21). Both models used *ad hoc* variations of information theory to assign scores to the predicted binding sites, rather than classical information content in bits. In one case the model was built using some sites that had not been footprinted by Fur and were probably not aligned to maximize the information content (21).

The most rigorous approach to model building is to use a data set comprised of only footprinted binding sites from one species. By restricting the data set to experimentally proven sites, one is certain that the model will reflect the binding characteristics of the protein; the use of a single species ensures that the protein and DNA binding sequences evolved together and therefore correspond to one another (22). Many biases from previous models are thereby avoided, but not all (23). The resulting experimentally supported model is then scanned across the entire genome of the species, looking for sequences that contain a positive amount of information as evaluated by the weight matrix (19). Sequence walkers, which are graphical representations of individual binding sites, then display probable binding sites on the genome based on the model of proven sites (20). This method was successfully used to discover that the OxyR transcription factor controls the expression of the *fur* gene (24), to identify additional sites for proteins such as Fis, SoxS, OxyR and PXR/RXR (23,25–27), to characterize splice junctions (28), and to discover T7 islands, a unique class of mobile genetic elements (29,30). In this study, information theory has allowed us to identify new Fur binding sites, 13 of which we confirmed experimentally.

MATERIALS AND METHODS

Programs

Programs used in this study are available at `http://www.ccrnp.ncifcrf.gov/~toms/`. A web-based tool for searching for Fur sites is available at `http://www.ccrnp.ncifcrf.gov/~toms/delilaserver.html`.

Creating the Fur model

Eleven footprinted *E. coli* Fur binding sequences were extracted from the *E. coli* K-12 genome (NC_000913) by the **delila** program (31). These sites are from the promoters of the genes *cir*, *fecA*, *fecI*, *fur*, *sodA*, *tonB* and *iucA*, along with two bidirectional promoter regions for the genes *fepA-fes* and *fepB-entC* (8,32–39). The promoter region at *fepB-entC* has two distinctly protected regions; each region was included in the data set as individual sequences (*fepB* and *entC*). The promoter of *iucA* has an exceptionally long secondary footprint and so two regions were used from the *E. coli* plasmid ColV-K30 (M10930, from 347 to 370, and from 365 to 393) (7). The complement of each footprinted sequence was also included, since Fur binds as a dimer (6,7,9,40). The program **malign** was used to obtain an alignment of the sequences that maximizes the information content within defined windows of the alignment (17).

As the initial model described above contains only 11 sites, we then refined the model by including 13 more sites that were identified in the genome by a search and confirmed by gel shift experiments in this study (see subsequently). The validity of the refined model was tested by scanning it across all available *E. coli* Fur footprinted regions, using the programs **scan** and **lister** to create sequence walkers (19,20). The model is verified if the walkers correspond to DNase I footprint regions. For further verification, the model was also scanned across synthetic oligonucleotides that contain GATAAT repeats and which had been previously footprinted (5).

To compare Fur models of different bacteria, we also built *Pseudomonas aeruginosa* and *Bacillus subtilis* models by using footprinted Fur binding sequences from these two species (41,42).

Scanning for Fur binding sites

Fur in high concentration can bind weak sites, while in low concentration it binds only strong sites. The affinity of Fur to its binding site also varies significantly with the availability of metal. Furthermore, as demonstrated by footprint data, Fur binding can extend to flanking regions that bear little resemblance to the strong Fur binding sites. This suggests that it may not be practical to use a fixed cutoff for a Fur binding site model under different

conditions. In our study, we used two cutoff values for different scans.

For the whole *E. coli* genome scan, we used the lowest information in the set of sites used to build the model as the cutoff. This should minimize false positives in our predictions, as the model was built from experimentally proven sites. Groups of sites that were within 100 bases of each other were identified using the **localbest** program, and the strongest one was selected to represent the region. This ensured that each region in the data set was unique. All sites with R_i values greater than 17.0 bits were extracted (Table 4). This value was chosen simply to allow for a manageable set of regions for further analysis.

As Fur can extend binding to much weaker regions in the footprints, we should use a lower cutoff for scanning footprinted sequences. The second law of thermodynamics sets a theoretical lower bound for the information content of a binding site (R_i) at zero bits (19). Therefore we used a zero-bit cutoff when scanning footprinted sequences.

Fur has been suggested to repress genes by binding to their promoter elements (−35 and −10 regions), thereby blocking the access of polymerase to the promoters (33,43–46). To determine how many of the predicted Fur binding clusters overlap with promoters, we used an information theory-based flexible σ^{70} binding model to scan the Fur clusters. The σ^{70} model was built from 401 experimentally proven *E. coli* promoters by uniformly taking into account the information present in the −10, the −35, and the uncertainty of the spacing between them (47); the same method has been successfully used to model prokaryotic ribosome binding sites (48). The information content (average conservation) of the promoter model is low ($\overline{x} = 6.5$ bits, SD = 2.8 bits for the R_i distribution), and the individual information for a site in the model ranges from 0.3 to 12.6 bits. The probability that a site is less than 4 bits is only 18.6%, so we used 4 bits as a cutoff to predict reasonably strong non-activated promoters (47).

Gel mobility shift assays

Two sets of oligonucleotides containing predicted Fur binding sites in *E. coli* were designed and synthesized (Oligos Etc). All oligonucleotides were self-complementary, had a 5′-GCTA-3′ overhang on the 5′ end, and contained a hairpin loop. Oligos *exbB*, *exbD* and *fhuF* contained a hairpin of the sequence 5′-C GCGAAGC G-3′, while the other 12 oligos [*yoeA*, *fepD-entS* (formerly *ybdA*), *gpmA*, *ryhB*, *fhuA*, *nohA*, *oppA*, *gspC*, *garP* (formerly *yhaU*), *yahA*, *fadD* and *ygaC*] contained a hairpin of the sequence 5′-ACGATCGC GCGAAGC GCGATCGT-3′ in the center. Such loops form a structure that is stabilized by base pairing between G_3 and A_5 of the central seven bases of each loop (49), and they are convenient for use in DNA mobility shift assays because of the exact equimolar concentrations of the complementary strands and their high stability (25).

Three oligos containing the promoter regions for *exbB*, *ygaC* and the upstream region of *exbD* were created to test previously published consensus sequence predictions (50,51). Eight oligos contained potential Fur-controlled promoters identified using both an 11-site Fur model and a σ^{70} model as described above [*yoeA*, *fepD-entS* (formerly *ybdA*), *gpmA*, *ryhB*, *fhuA*, *nohA*, *oppA* and *fhuF*]. We were also interested in whether Fur binds intragenically, so four oligos were synthesized that contained strong predicted sites found within gene coding regions (*gspC*, *garP*, *yahA* and *fadD*). These sequences can be found in Supplementary Table S1 and Figure S1.

Gel mobility shift assays were performed using the 15 oligos (Figure 3). The oligonucleotides were labeled by a fill-in reaction with *Taq* DNA polymerase. The 20 µl of labeling mixture [20 pmol oligo, 2 mM $MgCl_2$, 1 × PCR buffer, 5 units *Taq* polymerase (GibcoBRL®), and 10 µM tetramethylrhodamine-6-dUTP (NEN)] was incubated at 72°C for 1 h. 80 µl of ddH_2O was added to the mixture, followed by two phenol extractions, a phenol/chloroform extraction and a chloroform extraction. The labeled oligonucleotides were diluted 1:5 in TE (10 mM Tris–Cl, pH 7.5, 1 mM EDTA), boiled for 10 min and then placed on ice to prevent dimerization and trimerization.

The labeled oligonucleotides (5 µl each, 0.2 pmol) were then incubated in Fur binding buffer (10 mM Bis–Tris–HCl, pH 7.5, 5 µg/ml sonicated salmon sperm DNA, 5% glycerol, 100 µM $MnCl_2$, 100 µg/ml BSA, 1 mM $MgCl_2$, 40 mM KCl) with 150 nM Fur protein at 37°C for 13 min (7). Fur was a gift from T. O'Halloran and C. Outten. Samples were electrophoresed on a 5% polyacrylamide gel in Fur electrophoresis buffer (0.1 M Bis–Tris–HCl, pH 7.5, 10 mM $MnCl_2$) at 120 V for ∼1 h.

Bands were visualized with an FMBIO II fluorescent scanner (Hitachi), with an excitation wavelength of 532 nm and a 585 nm filter for detection of tetramethylrhodamine.

Footprinting

To generate the *fhuF* promoter construct (pGSO129) used to test for Fur binding, a 240-bp fragment amplified by PCR from genomic DNA (using the primers 5′-GCG GCT GGA GAT GAA TTC GCC AGA TG and 5′-GCC CTG CAA TCA GGG ATC CCG GCA GC) was cloned into the BamHI and EcoRI sites of pUC18 (27). Purified Fur protein, generously provided by C. Outten and T. O'Halloran, was incubated with a Mn^{2+}-containing buffer according to de Lorenzo *et al.* (7). DNase I footprinting then was carried out as described previously (52). The 240-bp BamHI–EcoRI fragment of pGSO129 was labeled with [γ-^{32}P]ATP at either the BamHI site (top strand) or the EcoRI site (bottom strand). The labeled fragments were incubated with 0, 50, 100, 200 and 400 nM purified Fur protein at room temperature for 5 min. The samples then were subjected to limited DNase I digestion, purified and separated on 8% polyacrylamide sequencing gels.

RESULTS

E. coli Fur binding model

Eleven footprinted Fur binding sites (see Table 3 and Supplementary Figure S1) (8,32–39) from *E. coli* K-12 (53) were used to create an initial information theory model.

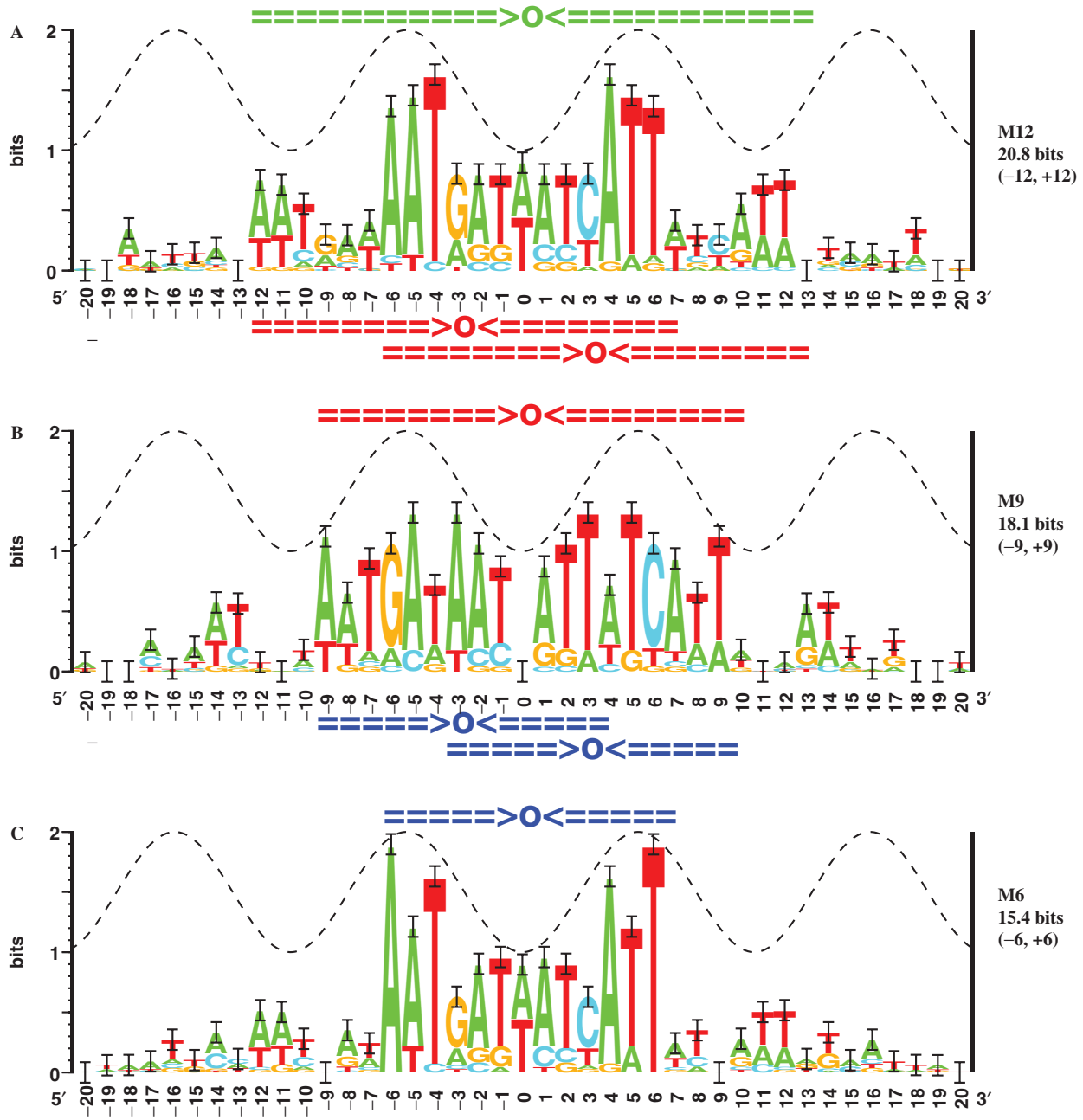

Figure 1. Different alignments of footprinted *E. coli* Fur binding sites. Sequence alignments were done using the program **malign** with different window sizes [from (−15, +15) to (−5, +5)] (17) on 11 footprinted *E. coli* Fur binding sequences and their complements (see Supplementary Figure S1 for sequences). Three classes of alignments, M12 (**A**), M9 (**B**) and M6 (**C**) were obtained by using window sizes of (−12, +12), (−9, +9) and (−6, +6), respectively. In these logos, the height of each letter is proportional to the frequency of that base at each position, and the height of the letter stack is the conservation in bits (18). The dashed sine wave on each logo represents the 10.6 base helical twist of B-form DNA (31,81). The double dashed arrows on the top of each logo mark the inverted repeats in each model. Note that the M12 logo corresponds to two overlapping M9 sites, as indicated by the two red double dashed arrows below the M12 logo. A similar relationship occurs between the M9 and M6 logos.

Because Fur binds as a dimer (9,54), the model included both the footprinted Fur sequences and their complements. The sequences were aligned using the program **malign**, which maximizes the information content within a region of the alignment (a 'window') by shuffling the sequences back and forth (17). Multiple alignment with a window size of (−12, +12) gave a sequence logo that shows base conservation in the range of (−12, +12) (Figure 1A). For convenience we named this logo M12.

Table 1. Information theory test of 11-site *E. coli* Fur models M12, M9 and M6

Oligo[1]	Number of observed[1] Fur dimers (nM)	Number of M12[2] walkers (bits)	Number of M9[2] walkers (bits)	Number of M6[2] walkers (bits)
F-F	1 (250.00)	• 1 (5.5)	○ 2 (9.0, 9.4)	• 1 (11.5)
F-F-F	2 (36.10)	○ 1 (14.8)	• 2 (11.8, 15.3)	• 2 (11.9, 9.9)
F-F-F-F	4 (6.30)	○ 2 (16.6, 19.2)	○ 3 (11.8, 18.9, 15.3)	○ 3 (11.9, 11.9, 9.9)
F-F-x-R	2 (0.95)	• 2 (22.1, 4.3)	• 2 (17.7, 17.6)	○ 1 (20.4)
F-F-x-R-R	3 (0.78)	○ 2 (29.6, 22.7)	• 3 (17.7, 27.5, 10.8)	○ 2 (20.4, 16.4)
fepB	2 (8.60)	○ 1 (15.7)	• 2 (10.9, 8.2)	○ 1 (15.2)
entS	3 (3.01)	○ 2 (15.6, 17.2)	• 3 (7.5, 17.2, 9.5)	○ 2 (8.3, 12.2)

[1]Oligos and observed Fur dimers are from (55). The apparent K_d is also given (nM, in parentheses) for each oligo.
[2]Number of sequence walkers of each model, followed by the information (bits) of each walker, are shown for each sequence. The predictions that match the number of observed Fur dimers are marked with a solid circle, while those that do not are marked with an open circle. Only major walkers (strong walkers with 6-bp separation) are counted (see the text). A full depiction of all walkers (>0 bits) can be found in Supplementary Figures S3 and S5.

The M12 logo appears to consist of two overlapping direct repeats, one is from −12 to +6 and the other from −6 to +12. Because the multiple alignment process can incorporate one or the other of these repeats, this suggests that different alignments may be obtained when other windows are used to align the sequences. Because two Fur dimers can bind overlapping sites with 6-base separation (40,55), we then multiply aligned the sequences by using window sizes ranging from (−15, +15) to (−5, +5). These sizes should cover the sites recognized by one or two Fur dimers (9). A total of seven different alignments were obtained; the corresponding sequence logos are shown in Supplementary Figure S2. According to the patterns shown in the logos, the seven alignments that were obtained can be further consolidated into and assigned to three basic, alignment classes; these alignment classes are represented by logos M12, M9 and M6 (Figure 1). Each of the three alignments itself is an inverted repeat; the M12 logo also appears to consist of two overlapping M9 sites (from −12 to +6 and from −6 to +12), and the M9 logo looks like two overlapping M6 sites (from −9 to +3 and from −3 to +9) (Figure 1).

To determine which of the three alignment classes best represents a single-dimer binding model, some published experimental data was analyzed. Lavrrar and McIntosh (55) used the Ferguson method (56) to determine how many *E. coli* Fur dimers can bind on five synthetic oligos and two natural sites, *fepB* and *entS*. We applied the three models to their sequences and used sequence walkers to show the predicted sites of each model. A full list of the predictions (with walkers >0 bits) is given in Supplementary Figure S3. The results show that the models frequently found overlapping sites that are separated by 3 or 6 bases. The sites that are 6 bases apart can both be strong, while 3-base separated sites have one strong site and one weak site (Supplementary Figure S3). Furthermore, it has been shown that Fur dimers can simultaneously bind two overlapping sites with 6-bp spacing (40), but there is no experimental data showing that 3-base separated overlapping sites can be bound at the same time. Therefore to analyze the results, we only counted major walkers (strong walkers that are separated by 6 bases) for each model in each oligo. A summary of the results (with major walkers) is available in Table 1; two examples of the prediction (with major walkers) are shown in Figure 2A and B. The results show that the M9 model made correct predictions for five of the seven oligos, while the M12 and M6 models both made only two correct predictions. Furthermore, the M12 model always predicted one less site than the M9 model. These results suggest that the M9 model may be a single-dimer binding model, the M12 model may represent a two-dimer binding model, while the M6 alignment may be just a result of compression of the binding sequences into a small alignment window.

Confirmation of 13 predicted sites and model refinement

Since our *E. coli* Fur model had only 11 sites, we wanted to refine the model by including more sites. A genome scan with the 11-site M9 model revealed 389 sites above 11 bits (the lowest information of a site in the model is 11.1 bits). Within these, 13 sites (from 12 to 26 bits) were selected for experimental confirmation (Figure 3, Supplementary Table S1 and Figure S1). These include eight sites located in promoter regions [*yoeA*, *fepD-entS* (43), *gpmA* (= *pgm*) (51), *ryhB* (= *yhhX-yhhY*) (51,57), *fhuA* (13), *nohA* (51), *oppA* (13) and *fhuF* (51)], and four sites inside genes (*gspC*, *garP*, *yahA* and *fadD*). We also included two sites that were predicted by consensus models to be in *ygaC* and *exbD* (50,51), since our model only detected a weak site (1.4 bits) in *ygaC*, and a negative site (−5.0 bits) in *exbD* (Figure 3). Gel mobility shift assays showed that all 13 sites predicted by the M9 model shifted when incubated with Fur protein, while the *exbD* site did not shift, and the *ygaC* site only showed extremely weak binding (Figure 3).

McHugh *et al.* (13) found that both *exbB* and *exbD* were regulated by Fur, but their model did not detect a binding site in the *exbB* promoter while our information theory model successfully predicted that Fur binding site, and our gel shift results confirmed this (Figure 3). The *exbD* gene is located only seven bases downstream of the *exbB* gene. Both genes have the same orientation, so they may belong to a single operon cotranscribed from the *exbB* promoter and coregulated by Fur.

To strengthen our model, we added the 13 confirmed sites to the 11 previously footprinted sites (Figure 4A).

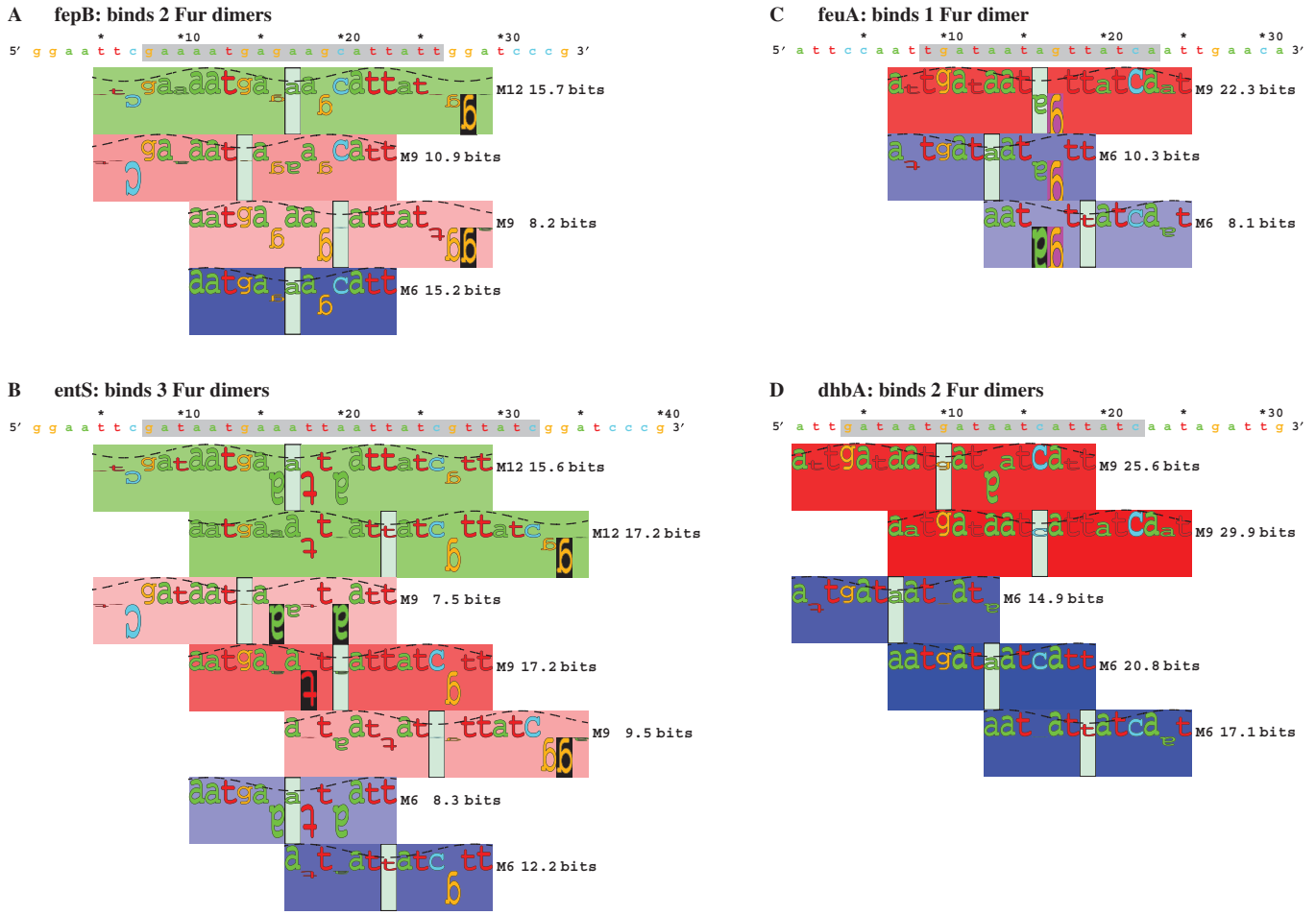

Figure 2. Sequence walkers to test *E. coli* (**A** and **B**) and *B. subtilis* (**C** and **D**) Fur models. (A and B) The three *E. coli* Fur models (M12, M9 and M6) were tested with two *E. coli* Fur binding sites, *fepB* and *entS* (55). Sequence walkers of each model (green, M12; red, M9; blue, M6) are shown for each site. (C and D) Two B. subtilis Fur models (M9 and M6) were tested with two *B. subtilis* Fur binding sites, *feuA* and *dhbA* (40). Gray bars indicate natural sequence, surrounded by synthetic DNA. Different colored rectangles indicate different Fur models (by hue: green, M12; red, M9; blue, M6) and their strengths (by saturation) (30). Only major walkers are shown in these figures, for a full list of walkers (Ri>0 bits) please see Supplementary Figures S3, S5 and S7.

As with the 11 sites, multiple alignment of the 24 sites using different window sizes also gave three classes of alignments, M12, M9 and M6 (Supplementary Figure S4). When these models were tested against the same set of sequences as that used to test the 11-site models (Table 1, Supplementary Figure S3), similar, but cleaner results were obtained, i.e. the weak walkers that were not counted previously become weaker (some are even below 0 bits), but the strong walkers still have similar R_i values (Supplementary Figure S5). This confirms that our counting of major walkers predicted by the initial 11-site models was reasonable.

P. aeruginosa and *B. subtilis* Fur models

To compare *E. coli* Fur models with other bacterial Fur models, we also built *P. aeruginosa* and *B. subtilis* Fur binding site models. As with *E. coli* Fur binding sites, we applied the same method to align footprinted Fur binding sites from *P. aeruginosa* (41) and *B. subtilis* (42). For both bacteria we obtained two classes of alignments (Supplementary Figure S6), which correspond to the *E. coli* M9 and M6 logos (Figure 1B and C), respectively. No M12 alignments were obtained for these two bacteria, probably because some sites used to make the models are single-dimer binding sites (see Discussion Section). Both *P. aeruginosa* and *B. subtilis* M9 models are highly similar to the *E. coli* M9 model (Figure 4). The *P. aeruginosa* M6 logo looks almost identical to the *E. coli* M6 logo, while the *B. subtilis* M6 logo is less similar to the *E. coli* M6 logo (Figure 1C, Supplementary Figure S6B and D).

As with Lavrrar and McIntosh's work on *E. coli* Fur binding sites (55), Baichoo and Helmann (40) performed similar experiments on eight synthetic oligos and two natural sites (*feuA* and *dhbA*) with *B. subtilis* Fur, and determined how many Fur dimers can bind each of these oligos. We tested the *B. subtilis* M9 and M6 models with these 10 sequences (Table 2, Figure 2C and D, and Supplementary Figure S7). The results show that the M9 model made correct predictions for all 10 sequences, while the M6 model was correct for only four oligos.

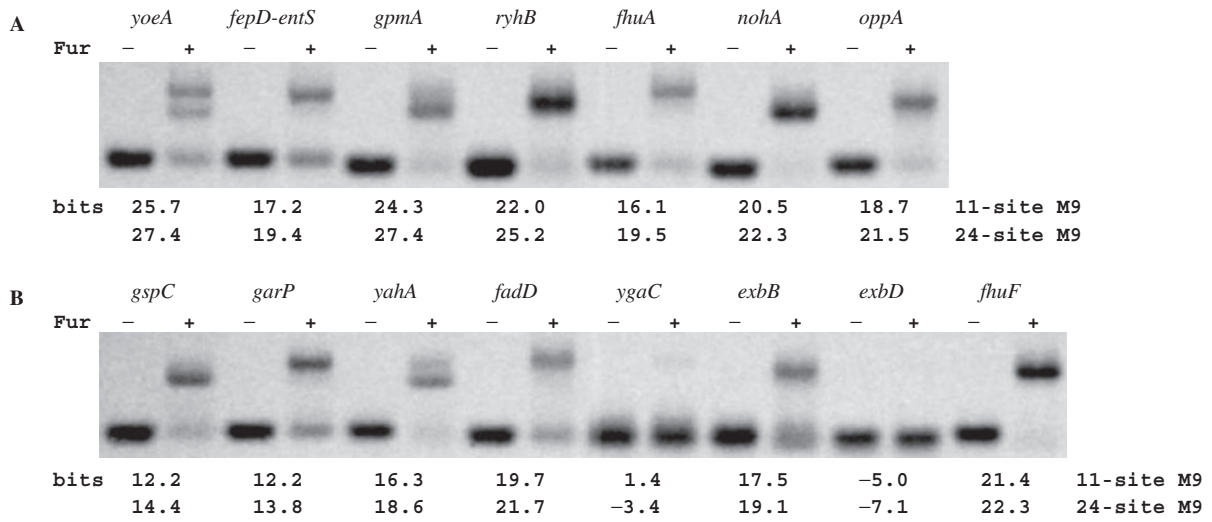

Figure 3. Gel mobility shift assays to test predicted Fur binding sites. Oligonucleotides containing predicted Fur binding sites were incubated without (−) or with (+) 150 nM E. coli Fur protein and gel electrophoresed to test the 11-site M9 model (Figure 1B). Below each set of lanes is the strength in bits of the strongest sequence walker found on each oligo by the 11-site M9 model. Predictions by the 24-site M9 model, which includes 13 of the sites tested here (not *ygaC* or *exbD*), are also given for comparison. (**A**) The first set of oligos contain seven predicted Fur sites located in promoter regions (*yoeA, fepD-entS, gpmA, ryhB, fhuA, nohA* and *oppA*). (**B**) The second set of oligos contain four predicted Fur sites located within genes (*gspC*, *garP*, *yahA* and *fadD*. See Supplementary Figures S10–S13 for sequence walkers.), two sites located in promoter regions (*exbB* and *fhuF*) and two consensus-predicted sites (*ygaC* and *exbD*) (50,51). As expected, all oligos that contain predicted sites (by the 11-site M9 model) exhibit one or more mobility shifts following incubation with 150 nM Fur; the oligo *ygaC* (1.4 bits) shows extremely weak binding, and *exbD* (−5.0 bits) does not shift.

For most of the oligos, the M6 model predicted one more site than the M9 model did. Similar to our observations in *E. coli*, these results strongly suggest that the M9 model is a single-dimer binding model, while the M6 alignment is due to compression of the binding sequences.

Relative to *E. coli* Fur (NP_415209), the *P. aeruginosa* Fur (NP_253452) is moderately diverged (58% identical in protein sequence), and the *B. subtilis* Fur (NP_390233) is distantly diverged (33% identity). However, a comparison of the three M9 models from these three species shows that these models are highly similar to each other (Figure 4), suggesting that the Fur–DNA recognition mechanism is conserved in even distantly related species.

Scans of footprinted regions

Gel shift experiments cannot give precise Fur binding regions. Originally using the 11-site M9 model and later using the 24-site M9 *E. coli* Fur model, we predicted Fur sites in the *fhuF* promoter region and then, to validate our model, we performed DNase I footprinting on that region. To further confirm the 24-site M9 model, we also scanned it across all other available footprinted *E. coli* Fur binding regions, and correlated our predictions with the published footprints. Fur has been documented to exhibit 'secondary footprinting', protecting extended regions of DNA under higher protein concentrations (Supplementary Figure S8) (33,36,46). To reveal the extended regions we used a cutoff at zero bits, since that is the theoretical lowest information for binding (19).

The *E. coli* 24-site M9 model predicts two separated Fur binding clusters in the *fhuF* promoter region; one cluster contains three strong sites of 16.1, 20.8 and 22.3 bits, and the other contains two weaker sites of 14.6 and 9.3 bits (Figure 5A). The *fhuF* oligo used in the gel shift experiments in Figure 3 only contains the strong Fur cluster. To determine if both predicted clusters are protected by Fur, we performed DNase I footprinting on a larger *fhuF* promoter region (see Materials and Methods Section) that contains both predicted Fur binding clusters. The results show that there are indeed two Fur-protected regions in the *fhuF* promoter, one has high affinity with Fur (protected by 50 nM of Fur), and the other has a lower affinity (weakly protected by 50 nM of Fur) (Figure 5B). The predicted strong Fur binding cluster covers 91% of the high-affinity footprinted region, and the weak Fur cluster covers 66% of the low-affinity region (Table 3, Figure 5A).

There are 11 other footprinted *E. coli* Fur binding regions, including the 10 regions used to build the initial 11-site models, and the *fepD-entS* promoter region (58). When these footprints were scanned, similar coverage was obtained (Table 3, Supplementary Figure S8). All 11 footprinted sequences displayed clusters of multiple overlapping Fur walkers; in these clusters, the majority of the walkers were separated by six bases; some were separated by three bases (Table 3, Supplementary Figure S8). For 7 of the 11 regions (*fepA-fes*, *fepB*, *entC*, *fur*, *tonB*, *cir* and *iucA*), sequence walkers cover ~83% (72–93%) of the footprints. No sequence walkers cover the secondary footprints in the *fecA*, *fepD-entS* and *sodA* regions, resulting in low coverage of these three footprints (66%, 54% and 56%, respectively). The secondary *sodA* footprint was only protected with a high concentration of purified Fur (200 and 500 nM); it was not protected by crude extracts containing overproduced Fur (36). In the other region, *fecI*, several low-information content

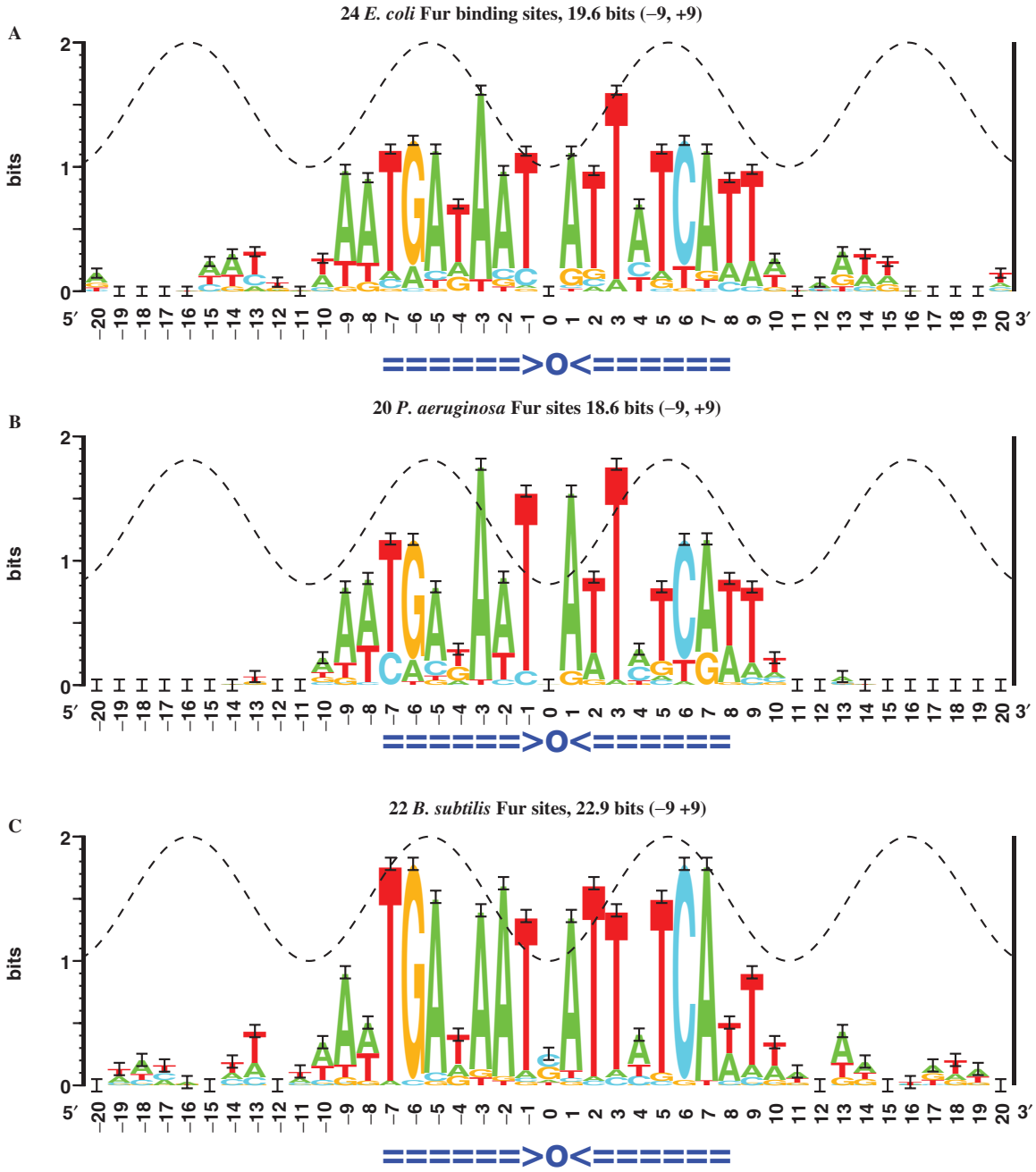

Figure 4. Comparison of M9 Fur models of three bacterial species. Experimentally proven Fur binding sites from *E. coli* (**A**), *P. aeruginosa* (**B**) and *B. subtilis* (**C**) were used to build the models for these three bacteria. The marked region from −7 to +7 in each logo indicates the 7-1-7 model that was proposed by Baichoo and Helmann (40).

walkers (0.2 and 0.8 bits) extend past the range of the footprint, resulting in 144% coverage of the footprint.

In a previously published study, Escolar *et al.* synthesized oligonucleotides that contained repeats of the sequence GATAAT and determined Fur binding by footprint experiments (5). No footprint was seen with one insert; correspondingly, no sequence walkers were observed in that sequence (Supplementary Figure S9). For the oligos with three, four and five inserts, sequence walkers cover 83–90% of the footprints. The 2-insert sequence had a weak interaction with Fur at high protein concentrations and two overlapping walkers of 3.0 and 4.2 bits appeared, which cover the repeats (5). These results demonstrate that the Fur model accurately predicts Fur binding on both natural and synthetic sequences.

Whole genome scan

The lowest information of any site in the 24-site M9 model is 10.1 bits (Supplementary Figure S1). Since all sites in the model are proven binding sites, we used 10 bits as the cutoff for a whole genome scan. This presumably

Table 2. Information theory test of *B. subtilis* Fur models M9 and M6

Oligo[1]	Number of observed[1] Fur dimers (nM)	Number of M9[2] walkers (bits)	Number of M6[2] walkers (bits)
6-mer	0	• 0	• 0
6-6	1 (500)	• 1 (2.1)	○ 2 (6.7, 5.4)
6-1-6	1 (200)	• 1 (9.7)	○ 2 (5.4, 7.4)
7-1-7	1 (100)	• 1 (20.7)	○ 2 (10.7, 12.8)
8-1-7	2 (100, 1000)	• 2 (7.6, 20.2)	• 2 (17.4, 10.8)
8-1-8	2 (100, 1000)	• 2 (7.6, 23.0)	• 2 (17.4, 16.0)
9-1-9	2 (50, 100)	• 2 (11.9, 30.1)	• 2 (20.8, 19.4)
2 (7-1-7)	2 (10, 100)	• 2 (16.6, 25.7)	○ 3 (6.1, 20.8, 14.3)
feuA	1 (10)	• 1 (22.3)	○ 2 (10.3, 8.1)
dhbA	2 (10, 100)	• 2 (25.6, 29.9)	○ 3 (14.9, 20.8, 17.1)

[1]Oligos and observed Fur dimers are from (40). The Fur concentration (nM) at which one or two Fur dimer-DNA complexes appeared is given in parentheses.
[2]Number of sequence walkers of each model, followed by the information (bits) of each walker, are shown for each sequence. The predictions that match the number of observed Fur dimers are marked with a solid circle, while those that do not are marked with an open circle. Only major walkers are counted. A full depiction of all walkers can be found in Supplementary Figure S7.

minimizes false positives in our predictions. When the model was scanned across the *E. coli* K-12 genome (NC_000913), a total of 412 sites were found above 10 bits. These sites belong to 363 unique clusters. A list of the clusters is available at `http://www.ccrnp.ncifcrf.gov/~toms/papers/fur/`.

The 363 regions were extracted from −50 to +50 around the strongest site in each region, and scanned with the 24-site M9 model using a cutoff of 0 bits. We then determined the exact region of each cluster that is covered by walkers greater than 0 bits. The results show that these clusters contain 1–14 walkers, covering a region from 19–142 bases (35.0 bases on average). This is comparable to the known footprints (30–103 bases) (Table 3, Supplementary Figure S8). Of these clusters, 155 are in intergenic regions, 143 are entirely inside gene coding regions and 65 overlap the boundary between coding and intergenic regions.

To determine how many of the Fur clusters overlap with reasonably strong promoters, we scanned these 363 Fur clusters with a σ^{70} promoter model (47) using a 4-bit cutoff. The results showed that 303 clusters overlap with one or more strong promoters ($R_i > 4$ bits). The other 60 clusters do not overlap with a strong promoter. When we scanned the flanking regions of these 60 clusters (−50 to 0 of the left edge, and 0 to +50 of the right edge of each cluster), only 23 clusters were found to have strong promoters ($R_i > 4$ bits) in the flanking regions. We examined the locations of Fur sites relative to known promoter elements (transcriptional start site, −10 and −35) and found that they are more tightly distributed around the promoter elements than observed with the DNA binding proteins Fis, H-NS and IHF (47) (data not shown). These results suggest that Fur may mainly act as a direct repressor of genes by binding to and blocking the promoters, and that direct activation by Fur through binding to the region upstream of a promoter, as suggested in *Neisseria meningitidis* by Delany *et al.* (59), may not be a common mechanism of Fur regulation in *E. coli*.

Out of the 363 clusters, 42 were found containing at least one predicted Fur binding site above 17.0 bits (our arbitrary cutoff used to locate significant regions) (Table 4). Of the 42 clusters, 39 were found to overlap with one or more strong σ^{70} promoters (>4 bits). The other three clusters, *fepD-entS*, *yddA* and *fecA*, do not overlap. Scanning the flanking regions of these three clusters with the σ^{70} promoter model revealed that these Fur clusters are located between the respective promoters and translational starts. There may be only weak repression of these genes by Fur, as none of them were significantly repressed by Fur according to microarray analyses (13,60).

Scans of other proposed Fur-regulated genes

Many genes have been proposed to be Fur-regulated by comparing conventional consensus sequences to promoter regions and also by homology to systems in other organisms. Kammler *et al.* annotated two Fur binding sites in the promoter region of *feoA* in *E. coli* by comparison to a consensus sequence (10). The 24-site M9 model predicts two clusters of sites in the same region with each cluster covering one of the marked Fur boxes. One cluster has one major site of 10.5 bits (at 3538006), the other cluster has two major overlapping sites of 10.9 (at 3538059) and 13.4 bits (at 3538065), respectively. The same authors have confirmed that Fur does bind to this region *in vivo*, but the exact Fur binding regions have not been determined by footprint experiments.

Vassinova and Kozyrev used an *in vivo* selection to locate Fur sites on *Sau3A* fragments from the *E. coli* genome (51). The five regions from Figure 2 of their paper were analyzed using sequence walkers. Using an unidentified consensus sequence, *yhhX* (from 3578828 to 3577791) was predicted by Vassinova and Kozyrev to have two Fur binding sites. In the same region we found one strong site of 25.2 bits (at 3579054) (Table 4), overlapping two weaker sites (8.2 and 8.3 bits). This cluster of Fur sites is actually located right in front of an sRNA gene, *ryhB* (from 3579039 to 3578946) (Table 4) (61,62), which has been shown to be repressed by Fur (57). The promoter region of *gpmA* (from 786818 to 786066, named *pgm* in reference 51), was predicted by consensus to contain two sites. One of the highest sites (27.4 bits) was found in this region (at 786853), overlapping two weaker sites of 7.5 and 5.6 bits. The consensus sequence-predicted Fur site by Vassinova and Kozyrev (51) is located about 600 bases upstream of the *ygaC* gene. In the same region, our 11-site M9 model only found a 1.4-bit site (Figure 3), and the 24-site M9 model found a negative site of −3.4 bits (Supplementary Figure S1). Instead, in the nearby regions the 24-site M9 model found a 7.4-bit site (11 bases upstream of *ygaC*) and a 10.1-bit site (728 bases upstream of *ygaC*). For *nohA* (from 1634391 to 1633822), two overlapping sites of 22.3 and 12.6 bits were found at 1634624 and 1634630 (Table 4). The fifth region was *fhuF* (from 4603686 to 4602898). Figure 5 shows the DNase I footprints and predicted walkers in this region. Only the high-affinity region (*fhuF1*) was identified by the

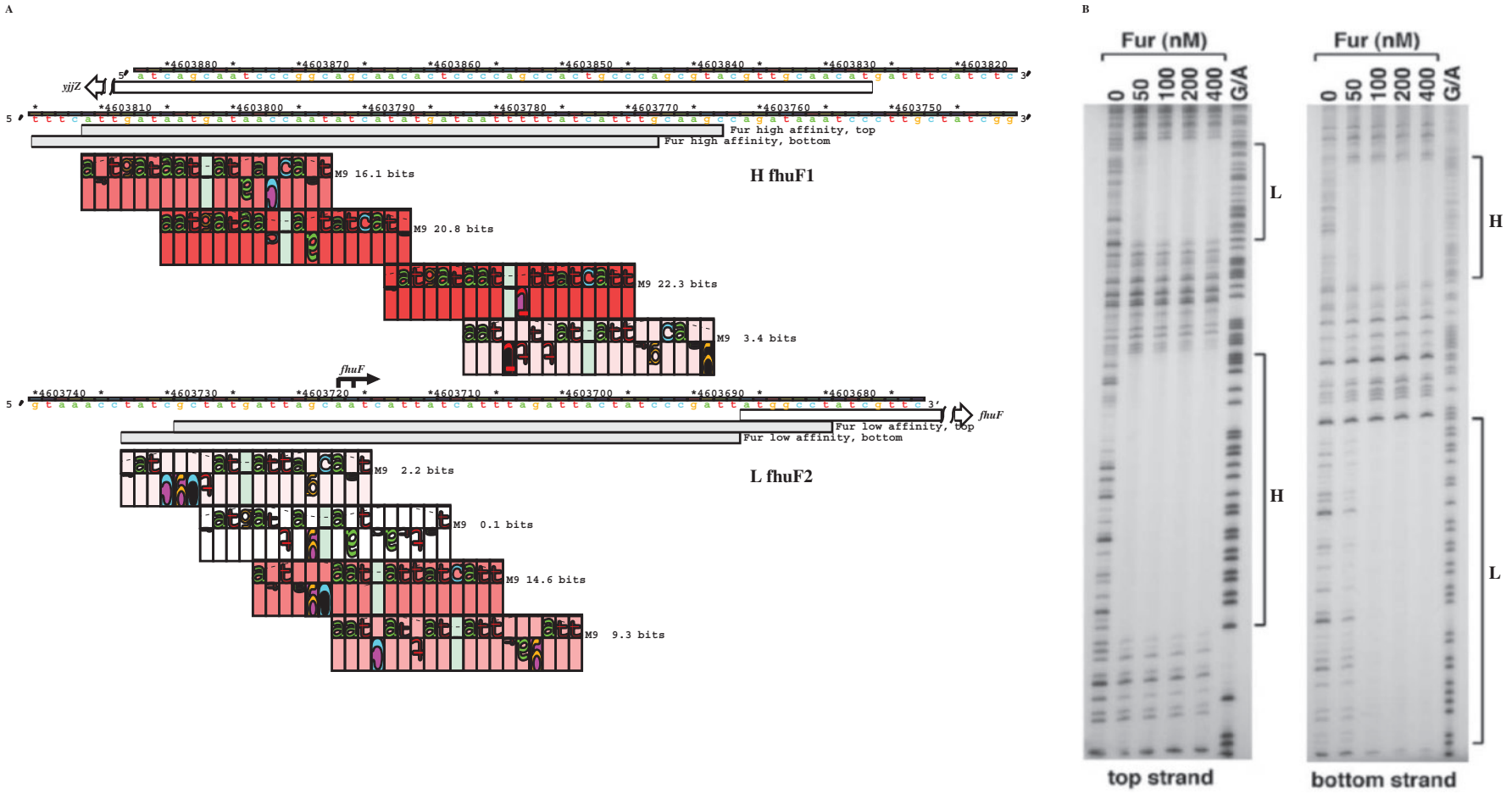

Figure 5. Fur model scan and DNase I footprints of the *fhuF* promoter region. (**A**) The promoter region was scanned with the 24-site M9 model, and two clusters of sequence walkers ($R_i > 0$ bits) were detected, with each closely corresponding to one of the two Fur protected regions (gray bars, *fhuF1* and *fhuF2*). The solid black arrow with two tails above the DNA at coordinates 4603717 and 4603716 indicates the *fhuF* transcription starts. The open arrows starting at coordinates 4603827 and 4603686 indicate the *yjjZ* and *fhuF* translation starts, respectively. The saturation of colored rectangles behind each sequence walker is proportional to the information of that site. For reference, colored lines cycle with 6-base periodicity. (**B**) DNase I footprinting on the *fhuF* promoter by Fur showed two regions protected by the protein, marked in the figure by brackets. The high-affinity and low-affinity regions are marked with H and L, respectively. The footprinting samples were run in parallel with Maxam–Gilbert sequencing ladders (marked by G/A).

Table 3. Scan of footprinted Fur binding sites in *E. coli*

Accession	Gene	Footprinted region	References	Number	R_i	Percentage
NC_000913	*fhuF1*	4603763 – 4603815	This work	4	22.3	91
NC_000913	*fhuF2*	4603680 – 4603733	This work	4	14.6	66
NC_000913	*fepA-fes**	611868 – 611915	(35)	4	18.7	77
NC_000913	*fepB**	623939 – 623969	(39)	2	16.7	81
NC_000913	*entC**	624054 – 624102	(39)	4	23.1	76
NC_000913	*fur**	709931 – 709960	(34)	2	12.3	83
NC_000913	*tonB**	1309040 – 1309072	(8,38)	3	11.5	88
NC_000913	*cir**	2244957 – 2244999	(32)	3	20.7	72
NC_000913	*sodA**	4098726 – 4098780	(36)	3	23.3	56
NC_000913	*fecA**	4514752 – 4514789	(33)	2	17.6	66
NC_000913	*fecI**	4516300 – 4516340	(33)	7	23.1	144
M10930	*iucA**	291 – 393	(37)	13	20.3	93
NC_000913	*fepD-entS*	621394 – 621462	(58)	4	20.1	54

The footprinted regions were scanned with the 24-site M9 model (R_i>0) to test the model's validity. This table lists the GenBank accession number, the gene promoter that was footprinted, the coordinates of the footprinted region, the references for the footprints, the number of walkers in the region, the strongest R_i value in the same region and the percent of the footprint covered by sequence walkers. This table is a summary of Figure 5A and Supplementary Figure S8. Sites in the promoters that are marked by an asterisk were used to build the 11-site models (Figure 1, Supplementary Figure S1); in the case of *iucA*, two sites were used in the model (see Materials and Methods Section).

consensus sequence. Thus four of the five *Sau3A* fragments (*yhhX*, *gpmA*, *nohA* and *fhuF*) selected to have Fur binding sites *in vivo* were identified in our genome scan (Table 4), and our gel shift results confirmed this (Figure 3); the remaining site, *ygaC*, had a weak walker of 1.4 bits (by the 11-site M9 model) and showed extremely weak binding (Figure 3).

Eick-Helmerich and Braun matched a Fur consensus sequence to the promoters of *exbB* and *exbD* (50). One strong walker of 19.1 bits was found in the *exbB* promoter region, overlapping the 'Fur box'. The region upstream of *exbD* showed no walkers with R_i greater than 0 bits, even though it had also been predicted to contain a Fur binding site using the same consensus sequence as used in the *exbB* promoter. The highest walker that we could detect in the region upstream of the *exbD* gene is −7.1 bits (by the 24-site M9 model, Figure 3 and Supplementary Figure S1), significantly lower than the theoretical lower bound of a binding site (0 bits) (19). This indicates that Fur should not bind this region, and our gel shift results confirmed this (Figure 3).

A microarray analysis by McHugh *et al*. found 101 genes to be regulated by Fur, 53 of which were repressed and 48 activated (13). The 53 Fur-repressed genes belong to 32 transcription units (individual genes or operons), 18 of which are involved in iron metabolism, 14 in energy metabolism and other functions. Using the 24-site M9 model, we predicted strong Fur binding clusters (higher than 13 bits) for all 18 iron metabolism-related transcription units. Of the other 14 transcription units, only two have a Fur binding site higher than 10 bits (*nrdHIEF*, 10.1 bits; *fimE*, 10.8 bits). The 48 Fur-activated genes belong to 34 transcription units, only three of these (*garPLRK*, *ynaE* and *ydfK*) have a predicted Fur binding cluster. The *garPLRK* operon has a weak Fur cluster that contains one major site of 10.4 bits. The *ynaE* and *ydfK* genes are almost identical in both promoter and coding regions, and the two Fur clusters (17.1 bits, centered at +8 from the translational start) are the same (Table 4). These results strongly suggest that in *E. coli* Fur is a direct repressor of iron metabolism genes and an indirect repressor or activator of other genes.

McHugh *et al*. (13) used an altered information theory approach for modeling Fur binding (63) to identify which genes are under direct Fur control. Our model identified sites in all genes that their model did, as well as four others, *fhuE* (12.4 bits, at 1160880), *exbB* (19.1 bits, at 3150076), *fimE* (10.8 bits, at 4539697) and *ydfK* (17.1 bits, at 1631071).

The indirect activation by Fur requires an intermediate regulator, which could directly target Fur-activated genes. A small regulatory RNA gene, *ryhB* is one such important player. Earlier study revealed several genes that are directly repressed by *ryhB* (57). Recently, a microarray analysis of *ryhB* control by Masse *et al*. (60) expanded this direct target set to 18 operons. They also identified an additional 10 indirectly down-regulated operons and 10 up-regulated operons. The 10 *ryhB* indirectly repressed operons are directly repressed by Fur, and for all of these operons we found strong Fur binding clusters (higher than 13 bits) in the corresponding promoter regions. In contrast, only a few of the other 28 *ryhB*-controlled operons (18 directly repressed operons and 10 up-regulated operons) have a predicted Fur binding cluster above 10 bits. These include *acnA* (17.6 bits, at 1333484), *ydhD* (10.9 bits, at 1732367) and *oppA* (21.5 bits, at 1298973). The *acnA* and *oppA* Fur clusters are both strong; the former cluster has three overlapping sites of 17.6, 9.8 and 14.2 bits, and the latter has two overlapping sites of 21.5 and 9.6 bits (Table 4). However, microarray analysis did not find these two genes to be repressed by Fur (13,60), suggesting that the Fur binding clusters may not be the primary control elements of these genes.

DISCUSSION

Fur binding models

In this study, we used experimentally proven sites to create information theory models of Fur binding. Our models,

Table 4. Strong Fur clusters in the *E. coli* genome predicted by the 24-site M9 model

R_i (bits)	Number	Size (bp)	Coordinate	Genes	Experimental data	Position	References
27.4	12	95	2066611	*yoeA*	G	−48	NA
27.4	5	31	786853	*gpmA*	G	−35	(13,51)
25.3	4	28	579824	*nohB*		−133	(1)
25.2	4	31	3579054	*ryhB*	G	−15	(57)
23.3	3	31	4098746	*sodA*	F	−87	(60,69)
23.1	7	59	4516309	*fecI*	F	−51	(13,60)
23.1	4	37	624072	*entC*	F	−36	(39)
22.8	2	25	1903412	*yebN*		−300	NA
22.3	4	48	4603779	*fhuF*[a]	F,G	−93	(13,51,60)
22.3	3	25	1634624	*nohA*	G	−233	(51)
21.7	2	25	1886651	*fadD*	G	+1119[b]	NA
21.5	2	25	1298973	*oppA*	G	−233	(70,71)
21.3	3	31	1762463	*sufA*		−53	(13,60)
20.7	3	31	2244980	*cir*	F	−189	(13,72)
20.6	5	40	1118358	*yceJ*		−89	NA
20.2	3	31	3465053	*bfd*		−40	(13,60)
20.1	10	53	1787602	*ydiE*		−35	(13)
20.1	2	35	1080513	*ycdN*		−57	(13,60)
19.9	8	106	3214572	*yqjH*		−59	(13)
19.9	7	103	1752756	*ydhY*		−255	NA
19.9	2	25	2510784	*mntH*		−56	(73,74)
19.6	3	25	2231892	*yeiT*		−163	NA
19.6	2	25	1577358	*yddA*		+8[c]	(13)
19.5	3	31	167439	*fhuA*	G	−45	(13,60)
19.4	4	37	621437	*fepD – entS*	F,G	−25, −86	(13,75)
19.3	3	52	675768	*ybeQ*		+2[c]	NA
19.1	3	31	3150076	*exbB*	G	−70	(13,60,72)
18.7	8	96	2784036	*ygaQ*		−383	NA
18.7	4	37	611889	*fepA – fes*	F	−172, −149	(13)
18.6	7	38	3266404	*yhaC*		−33	NA
18.6	4	28	331901	*yahA*	G	+306[e]	NA
18.3	4	61	331085	*yahA*		−510[e]	NA
17.6	6	41	1333484	*acnA*		−371	(57,76–78)
17.6	2	25	4514779	*fecA*	F	−79	(13,60)
17.6	1	19	2254041	*yeiL*		+5[d]	NA
17.5	3	25	490057	*priC – ybaN*		−21, −49	(79,80)
17.3	2	29	2454142	*yfcV*		−474	NA
17.2	8	41	3582772	*yrhB*		−10	NA
17.2	4	37	840877	*fiu*		−123	(13)
17.1	3	40	1631071	*ydfK*		+8[c]	(13)
17.1	3	40	1432273	*ynaE*		+8[c]	(13)
17.1	2	25	4570225	*yjiT*		−212	NA

The strongest individual information value (R_i, bits) in each cluster is shown, followed by the number of walkers ($R_i > 0$ bits), the size of the cluster, the coordinate of the strongest site, the genes that are presumably controlled, experimental data: footprinted (F) and/or gel shifted (G) sites, the location of the strongest Fur site relative to the gene start or stop codon (as noted) and references about Fur regulation of the genes (NA means no reference is available).

[a]This cluster corresponds to the high-affinity region (*fhuF1*) in the *fhuF* promoter (Figure 5).

[b]The coordinate is high because Fur cluster is inside the *fadD* coding region.

[c]These Fur clusters overlap with translational starts.

[d]This Fur cluster overlaps with the stop codon.

[e]The *yahA* gene has two Fur clusters, one is inside the coding region, another is 510 bases upstream of the translational start. The Fur cluster inside the gene was tested by gel shift experiments (Figure 3).

especially the 24-site M9 model, approximate the binding characteristics of Fur more fully than previous models that depended on conventional consensus sequences, data from multiple species and sequences that were not footprinted (1,32,51). The rigorous approach used here revealed new binding sites, disproved two sites predicted by a consensus sequence, and clarified the manner in which the protein binds.

To produce these models, binding sequences were multiply aligned to maximize the information content in different window sizes. Because multiple Fur sites overlap, our analysis of proven Fur binding sequences from three bacterial species gave two or three different alignments, depending on the species (Figure 1, Supplementary Figure S6). By testing these models against published data, we showed clearly that the M9 model represents single-dimer binding sites, the M12 model is for two-dimer binding, while the compact M6 alignment was caused by compression of the binding sequences into a small window (Tables 1 and 2, Figures 1 and 2, Supplementary Figures S3, S5 and S7).

As evaluated by their respective M9 models, our initial set of 11 footprinted *E. coli* Fur binding sites all contain overlapping sites of 6-base spacing

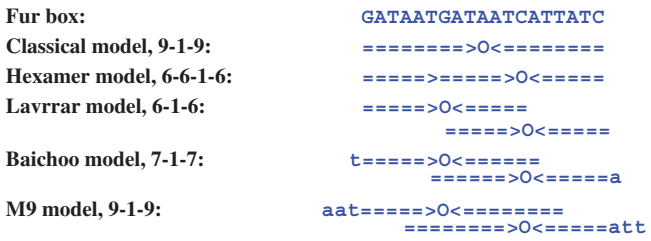

Figure 6. Different Fur binding models to interpret the Fur box.

(Supplementary Figure S8), while only 7 of the 20 *P. aeruginosa* Fur binding sites, and 6 of the 22 *B. subtilis* Fur binding sites are single-dimer binding sites (data not shown). This scarcity of multiple sites explains why an M12 alignment can be obtained for *E. coli*, but not for *P. aeruginosa* or *B. subtilis*. When we performed multiple alignment by only using overlapping sites of 6-bp spacing, we obtained M12 alignments for both *P. aeruginosa* and *B. subtilis* (data not shown). These observations also support the use of M9 as a single-dimer model.

Our multiple alignment method should apply to any DNA binding sites that contain internal direct repeats, as repetition within binding sites may cause alternative alignments of the sites. In the RegulonDB (`http://regulondb.ccg.unam.mx`) (64), a Fur binding model was created by multiply aligning 47 Fur sites, although it is not clear what these 47 sites are and how the alignment was made. Based on their alignment (`http://regulondb.ccg.unam.mx/html/matrix_Alignment.jsp`), we made a sequence logo (data not shown), which we found to be similar to our M6 logos. Although the M6 models appear to be able to predict some sites (e.g. Tables 1 and 2, Supplementary Figures S3, S5 and S7), the models only represent the internal part of overlapping Fur binding sites, i.e. from −6 to +6 of the M12 model (Figure 1A).

A 1.8 Å crystal structure for *P. aeruginosa* Fur was obtained by Pohl *et al.* and based on this, the authors proposed a single-dimer binding model (9). The *P. aeruginosa* crystal structure contains a putative DNA binding α-helix H4 and a loop between helices H1 and H2, both of which were proposed to be involved in binding DNA through contacting bases in the major groove. Thus, a single dimer would protect two consecutive major grooves on the same face of DNA, and the minimal recognition unit of Fur should be close to 20 bp (9). Our M9 binding site models (19 bp, a 9-1-9 inverted repeat) fit with this single-dimer binding model closely. Pohl *et al.* also proposed a two-dimer binding model, in which two Fur dimers bind two overlapping sites that are 5-bp apart (9). However, in our whole genome scans, and footprint scans as well, we did not observe any case of 5-bp separated overlapping sites above 5 bits. Instead, we mostly found 3- or 6-bp separated overlapping sites. Above five bits, we found 95 cases of 3-bp spacing and 92 cases of 6-bp spacing.

Several other consensus-based Fur binding site models have been proposed to interpret a 19-bp consensus 'Fur box' (Figure 6) (65). Within these, two earlier models, the classical model and hexamer model (5,7,37,54), interpreted the 'Fur box' as a single recognition unit of a Fur dimer. Later, Lavrrar and McIntosh suggested that a 13-bp inverted repeat (6-1-6) is the minimal unit recognized by a single Fur dimer, and that two overlapping '6-1-6' motifs correspond to the Fur box, which is required for high-affinity binding of Fur (55,58). Baichoo and Helmann reinterpreted the Fur box as two overlapping 7-1-7 inverted repeats with a 6-bp spacing, and suggested that a 7-1-7 site, but not 6-1-6, represents the minimal recognition unit of Fur (40). Baichoo and Helmann also demonstrated that high-affinity binding by Fur can happen on a single-dimer binding site, i.e. an inverted 7-1-7 site.

The 7-1-7 model basically agrees with our M9 models (Figure 6). The main difference is that the 7-1-7 model did not count the weakly conserved bases at positions −9, −8, +8 and +9 in the *B. subtilis* Fur binding site alignment (Figure 4C). As the *E. coli* and *P. aeruginosa* M9 models have highly conserved bases at these four positions (Figure 4A and B), and also the Fur binding mechanisms for these three species are similar (9), we suggest that a minimal Fur binding unit of 19 bp (9-1-9), as represented by our M9 models, is more reasonable.

Relative to our M9 models, the Fur box only represents the internal part of two overlapping sites that are separated by 6 bases, i.e. from −9 to +9 of the M12 logo (Figures 1A and 6), thus any predictions based on the Fur box may not be accurate. Two sites (*exbD* and *ygaC*) that were predicted by matching to the Fur box were shown to be non-binding (Figure 3).

Many difficulties in understanding Fur binding sites can be attributed to the choice of consensus sequences as a model (66). In contrast to the individual information weight matrix model, the consensus sequence method ignores the varying importance of bases by treating mismatches equivalently. In addition, the consensus method does not have a criterion for an acceptable number of mismatches.

Sequence walkers predicted by the 24-site M9 model cover ~83% of the footprints (Table 3, Figure 5, Supplementary Figure S8). Similar results were obtained for both the *P. aeruginosa* and *B. subtilis* M9 models (data not shown). These results strongly suggest that our M9 models can accurately predict Fur binding.

Fur Binding Site Clusters

The whole genome scan with the 24-site M9 model revealed 363 Fur binding clusters in *E. coli*, most of which contain multiple walkers (up to 14), covering regions up to about 140 bases. This is comparable to the size of known footprints, which ranges from 30 to 103 bases (Table 3, Supplementary Figure S8). Most of the clusters (303 out of 363, 83%) were found to overlap with one or more strong σ^{70} promoters. Indeed, the four Fur clusters in the middle of coding regions *gspC*, *garP*, *yahA* and *fadD*, that we tested by gel mobility shift assays (Figure 3), all have one or more potential σ^{70} sites ($R_i > 4$ bits) exactly overlapping them (see Figures S10–S13). The function of these control elements is unknown.

We correlated our whole genome scan with published microarray data about Fur in *E. coli* (13,60). The results showed that all operons that are involved in iron metabolism have a predicted strong Fur binding cluster (higher than 13 bits), suggesting direct repression of these genes by Fur. In contrast, only a few of other Fur-activated or repressed operons have a predicted Fur cluster higher than 10 bits, suggesting indirect control of these genes by Fur. It has been shown that RNA gene *ryhB* directly targets and represses a set of other genes and *ryhB* itself is directly repressed by Fur, thus *ryhB* plays an important role in indirect activation of genes by Fur in *E. coli* (13,57,60).

It has been suggested that in *N. meningitidis* Fur can also directly activate genes by binding to the region upstream of the promoters (59). However, our whole genome analysis did not support this activation mechanism in *E. coli*. The *E. coli ryhB* gene is located between genes *yhhX* and *yhhY*. No corresponding homologs of the genes *ryhB*, *yhhX* or *yhhY* can be found in *N. meningitidis*. This suggests that Fur activation mechanisms may be different in different species.

Disentangling Fur models

Each binding site that we have studied in detail with information theory has had unique properties and challenges for analysis. For example, ribosome binding sites were initially modeled as rigid objects (16) but when enough sites became available, it was clear that the Shine–Dalgarno sequence was being lost because it has a variable distance to the initiation codon (67) and a flexible model was required (48). The flexible model was later successfully applied to σ^{70} promoters (47) and that was used to study Fur sites. Likewise, information theory analysis revealed that Fis binding sites are commonly found to overlap each other (68), leading to some uncertainty as to how much nearby sites influence the sites being studied in the alignments. Overlapping Fur sites that are bound by Fur protein dimers create an even more complex situation.

Although the 24-site M9 model was quite successful, the model is comprised of information from multiple Fur sites because the Fur dimers overlap on the DNA. Since the clusters frequently display 6-base separation of sites, positions −3 to +9 of the model contain data from positions −9 to +3 of the adjacent Fur sites, e.g. see the logo of Figure 4 and the M9 walkers of Figure 2A. In addition, information from sites separated by 3 bases and the complements of all of these, also contribute to the overall information content of the model. It is not obvious how to disentangle these effects. We suggest that the M9 model can be scanned across the genome to locate strong single-dimer binding sites, which could then be experimentally confirmed and used to construct a disentangled model.

Unfortunately the distribution of the individual information in these sites would be determined by the model used to locate them and this could introduce biases. For example, the mean of the distribution can be altered just by selecting different subsets of sites from the distribution. A similar bias has occurred in binding site databases where the consensus sequence was used to locate new sites (23). Thus, although the sequence logo would no longer contain self-overlapping data, it would not necessarily represent the natural distribution of site strengths. To approach such an ultimate model may require further scanning of footprinted regions. Incorporation of the weaker sites in the footprints would, of course, re-entangle the model so such scans might be best used only to determine whether the model can detect the weaker sites. A viable solution may be to obtain all single-dimer binding sites in the genome that have more than zero bits and can be confirmed experimentally. If the resulting model matches the observed footprints, it may be considered complete. On the other hand, if the resulting model does not match the footprints, the mode of binding by Fur could be different between single dimers and clusters.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Tom O'Halloran and Caryn Outten for Fur protein; Pete Rogan for helping to develop the local best concept; Elaine Bucheimer for writing the **localbest** program and Lakshmanan Iyer, Brent Jewett, Danielle Needle, Xiao Ma, Shu Ouyang, Denise Rubens and Bruce Shapiro for their helpful comments and discussions. K.A.L. also thanks the National Cancer Institute at Frederick for sponsoring the Werner H. Kirsten Student Intern Program. This publication has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract #NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Funding to pay the Open Access publication charges for this article was provided by National Cancer Institute.

Conflict of interest statement. None declared.

REFERENCES

1. Stojiljkovic,I., Baumler,A.J. and Hantke,K. (1994) Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated *Escherichia coli* genes by a *fur* titration assay [published erratum appears in *J. Mol. Biol.* 1994 Jul 15;240(3):271]. *J. Mol Biol.*, **236**, 531–545.
2. Le Cam,E., Frechon,D., Barray,M., Fourcade,A. and Delain,E. (1994) Observation of binding and polymerization of Fur repressor onto operator-containing DNA with electron and atomic force microscopes. *Proc. Natl Acad. Sci. USA*, **91**, 11816–11820.
3. Calderwood,S.B. and Mekalanos,J.J. (1987) Iron regulation of Shiga-like toxin expression in *Escherichia coli* is mediated by the *fur* locus. *J. Bacteriol.*, **169**, 4759–4764.

4. Niederhoffer,E.C., Naranjo,C.M., Bradley,K.L. and Fee,J.A. (1990) Control of *Escherichia coli* superoxide dismutase (*sodA* and *sodB*) genes by the ferric uptake regulation (*fur*) locus. *J. Bacteriol.*, **172**, 1930–1938.
5. Escolar,L., Perez-Martin,J. and de Lorenzo,V. (1998) Binding of the Fur (Ferric Uptake Regulator) repressor of *Escherichia coli* to arrays of the GATAAT sequence. *J. Mol. Biol.*, **283**, 537–547.
6. Bagg,A. and Neilands,J.B. (1987) Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry*, **26**, 5471–5477.
7. de Lorenzo,V., Wee,S., Herrero,M. and Neilands,J.B. (1987) Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (*fur*) repressor. *J. Bacteriol.*, **169**, 2624–2630.
8. Althaus,E.W., Outten,C.E., Olson,K.E., Cao,H. and O'Halloran,T.V. (1999) The ferric uptake regulation (Fur) repressor is a zinc metalloprotein. *Biochemistry*, **38**, 6559–6569.
9. Pohl,E., Haller,J.C., Mijovilovich,A., Meyer-Klaucke,W., Garman,E. and Vasil,M.L. (2003) Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator. *Mol. Microbiol.*, **47**, 903–915.
10. Kammler,M., Schon,C. and Hantke,K. (1993) Characterization of the ferrous iron uptake system of *Escherichia coli*. *J. Bacteriol.*, **175**, 6212–6219.
11. Desai,P.J., Angerer,A. and Genco,C.A. (1996) Analysis of Fur binding to operator sequences within the *Neisseria gonorrhoeae fbpA* promoter. *J. Bacteriol.*, **178**, 5020–5023.
12. Heidrich,C., Hantke,K., Bierbaum,G. and Sahl,H.G. (1996) Identification and analysis of a gene encoding a Fur-like protein of *Staphylococcus epidermidis*. *FEMS Microbiol. Lett.*, **140**, 253–259.
13. McHugh,J.P., Rodriguez-Quinones,F., Abdul-Tehrani,H., Svistunenko,D.A., Poole,R.K., Cooper,C.E. and Andrews,S.C. (2003) Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis. *J. Biol Chem.*, **278**, 29478–29486.
14. Shannon,C.E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 379–423, 623–656. http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html
15. Pierce,J.R. (1980) *An Introduction to Information Theory: Symbols, Signals and Noise*, 2nd edn Dover Publications, Inc., New York.
16. Schneider,T.D., Stormo,G.D., Gold,L. and Ehrenfeucht,A. (1986) Information content of binding sites on nucleotide sequences. *J. Mol Biol.*, **188**, 415–431. http://www.ccrnp.ncifcrf.gov/~toms/paper/schneider1986/
17. Schneider,T.D. and Mastronarde,D. (1996) Fast multiple alignment of ungapped DNA sequences using information theory and a relaxation method. *Discrete Appli. Math.*, **71**, 259–268. http://www.ccrnp.ncifcrf.gov/~toms/paper/malign
18. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100. http://www.ccrnp.ncifcrf.gov/~toms/paper/logopaper/
19. Schneider,T.D. (1997) Information content of individual genetic sequences. *J. Theor Biol.*, **189**, 427–441. http://www.ccrnp.ncifcrf.gov/~toms/paper/ri/
20. Schneider,T.D. (1997) Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. *Nucleic Acids Res.*, **25**, 4408–4415. http://www.ccrnp.ncifcrf.gov/~toms/paper/walker/, erratum: NAR 26(4): 1135, 1998.
21. Panina,E.M., Mironov,A.A. and Gelfand,M.S. (2001) Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res.*, **29**, 5195–5206.
22. Schneider,T.D. (2000) Evolution of biological information. *Nucleic Acids Res.*, **28**, 2794–2799. http://www.ccrnp.ncifcrf.gov/~toms/paper/ev/
23. Vyhlidal,C.A., Rogan,P.K. and Leeder,J.S. (2004) Development and refinement of pregnane X receptor (PXR) DNA binding site model using information theory: insights into PXR-mediated gene regulation. *J. Biol Chem.*, **279**, 46779–46786.
24. Zheng,M., Doan,B., Schneider,T.D. and Storz,G. (1999) OxyR and SoxRS regulation of *fur*. *J. Bacteriol.*, **181**, 4639–4643. http://www.ccrnp.ncifcrf.gov/~toms/paper/oxyrfur/
25. Hengen,P.N., Bartram,S.L., Stewart,L.E. and Schneider,T.D. (1997) Information analysis of Fis binding sites. *Nucleic Acids Res.*, **25**, 4994–5002. http://www.ccrnp.ncifcrf.gov/~toms/paper/fisinfo/
26. Wood,T.I., Griffith,K.L., Fawcett,W.P., Jair, K.-W., Schneider,T.D. and Wolf,R.E. (1999) Interdependence of the position and orientation of SoxS binding sites in the transcriptional activation of the class I subset of *Escherichia coli* superoxide-inducible promoters. *Mol. Microbiol.*, **34**, 414–430.
27. Zheng,M., Wang,X., Doan,B., Lewis,K.A., Schneider,T.D. and Storz,G. (2001) Computation-directed identification of OxyR-DNA binding sites in *Escherichia coli*. *J. Bacteriol.*, **183**, 4571–4579.
28. Rogan,P.K., Faux,B.M. and Schneider,T.D. (1998) Information analysis of human splice site mutations. *Hum. Mutat.*, **12**, 153–171. Erratum in: Hum Mutat 1999;13(1):82. http://www.ccrnp.ncifcrf.gov/~toms/paper/rfs/
29. Chen,Z. and Schneider,T.D. (2005) Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases. *Nucleic Acids Res.*, **33**, 6172–6187. http://www.ccrnp.ncifcrf.gov/~toms/papers/t7like/
30. Chen,Z. and Schneider,T.D. (2006) Comparative analysis of tandem T7-like promoter containing regions in enterobacterial genomes reveals a novel group of genetic islands. *Nucleic Acids Res.*, **34**, 1133–1147. http://www.ccrnp.ncifcrf.gov/~toms/papers/t7island/
31. Schneider,T.D. (1996) Reading of DNA sequence logos: prediction of major groove binding by information theory. *Meth. Enzymol*, **274**, 445–455. http://www.ccrnp.ncifcrf.gov/~toms/paper/oxyr/
32. Griggs,D.W. and Konisky,J. (1989) Mechanism for iron-regulated transcription of the *Escherichia coli cir* gene: metal-dependent binding of Fur protein to the promoters. *J. Bacteriol.*, **171**, 1048–1052.
33. Angerer,A. and Braun,V. (1998) Iron regulates transcription of the *Escherichia coli* ferric citrate transport genes directly and through the transcription initiation proteins. *Arch. Microbiol.*, **169**, 483–490.
34. de Lorenzo,V., Herrero,M., Giovannini,F. and Neilands,J.B. (1988) Fur (ferric uptake regulation) protein and CAP (catabolite-activator protein) modulate transcription of *fur* gene in *Escherichia coli*. *Eur. J. Biochem.*, **173**, 537–546.
35. Hunt,M.D., Pettis,G.S. and McIntosh,M.A. (1994) Promoter and operator determinants for *fur*-mediated iron regulation in the bidirectional *fepA-fes* control region of the *Escherichia coli* enterobactin gene system. *J. Bacteriol.*, **176**, 3944–3955.
36. Tardat,B. and Touati,D. (1993) Iron and oxygen regulation of *Escherichia coli* MnSOD expression: competition between the global regulators Fur and ArcA for binding to DNA. *Mol. Microbiol.*, **9**, 53–63.
37. Escolar,L., Perez-Martin,J. and de Lorenzo,V. (2000) Evidence of an unusually long operator for the fur repressor in the aerobactin promoter of *Escherichia coli*. *J. Biol. Chem.*, **275**, 24709–24714.
38. Young,G.M. and Postle,K. (1994) Repression of *tonB* transcription during anaerobic growth requires Fur binding at the promoter and a second factor binding upstream. *Mol. Microbiol.*, **11**, 943–954.
39. Brickman,T.J., Ozenberger,B.A. and McIntosh,M.A. (1990) Regulation of divergent transcription from the iron-responsive *fepB-entC* promoter-operator regions in *Escherichia coli*. *J. Mol Biol.*, **212**, 669–682.
40. Baichoo,N. and Helmann,J.D. (2002) Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence. *J. Bacteriol.*, **184**, 5826–5832.
41. Ochsner,U.A. and Vasil,M.L. (1996) Gene repression by the ferric uptake regulator in *Pseudomonas aeruginosa*: cycle selection of iron-regulated genes. *Proc. Natl Acad. Sci. USA*, **93**, 4409–4414.
42. Baichoo,N., Wang,T., Ye,R. and Helmann,J.D. (2002) Global analysis of the *Bacillus subtilis* Fur regulon and the iron starvation stimulon. *Mol. Microbiol.*, **45**, 1613–1629.
43. Christoffersen,C.A., Brickman,T.J., Hook-Barnard,I. and McIntosh,M.A. (2001) Regulatory architecture of the iron-regulated *fepD-ybdA* bidirectional promoter region in *Escherichia coli*. *J. Bacteriol.*, **183**, 2059–2070.
44. Chai,S., Welch,T.J. and Crosa,J.H. (1998) Characterization of the interaction between Fur and the iron transport promoter of the virulence plasmid in *Vibrio anguillarum*. *J. Biol. Chem.*, **273**, 33841–33847.

45. Escolar,L., Perez-Martin,J. and de Lorenzo,V. (1998) Coordinated repression in vitro of the divergent *fepA-fes* promoters of *Escherichia coli* by the iron uptake regulation (Fur) protein. *J. Bacteriol.*, **180**, 2579–2582.
46. Escolar,L., de Lorenzo,V. and Perez-Martin,J. (1997) Metalloregulation in vitro of the aerobactin promoter of *Escherichia coli* by the Fur (ferric uptake regulation) protein. *Mol. Microbiol.*, **26**, 799–808.
47. Shultzaberger,R.K., Chen,Z., Lewis,K.A. and Schneider,T.D. (2007) Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res.*, **35**, 771–788. http://www.ccrnp.ncifcrf.gov/~toms/paper/flexprom/
48. Shultzaberger,R.K., Bucheimer,R.E., Rudd,K.E. and Schneider,T.D. (2001) Anatomy of *Escherichia coli* ribosome binding sites. *J. Mol Biol.*, **313**, 215–228. http://www.ccrnp.ncifcrf.gov/~toms/paper/flexrbs/
49. Hirao,I., Kawai,G., Yoshizawa,S., Nishimura,Y., Ishido,Y., Watanabe,K. and Miura,K. (1994) Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat. *Nucleic Acids Res.*, **22**, 576–582.
50. Eick-Helmerich,K. and Braun,V. (1989) Import of biopolymers into *Escherichia coli*: nucleotide sequences of the *exbB* and *exbD* genes are homologous to those of the *tolQ* and *tolR* genes, respectively. *J. Bacteriol.*, **171**, 5117–5126.
51. Vassinova,N. and Kozyrev,D. (2000) A method for direct cloning of Fur-regulated genes: identification of seven new Fur-regulated loci in *Escherichia coli*. *Microbiology*, **146**, 3171–3182.
52. Toledano,M.B., Kullik,I., Trinh,F., Baird,P.T., Schneider,T.D. and Storz,G. (1994) Redox-dependent shift of OxyR-DNA contacts along an extended DNA binding site: a mechanism for differential promoter selection. *Cell*, **78**, 897–909.
53. Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K. *et al.* (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
54. Bagg,A. and Neilands,J.B. (1987) Molecular mechanism of regulation of siderophore-mediated iron assimilation. *Microbiol. Rev.*, **51**, 509–518.
55. Lavrrar,J.L. and McIntosh,M.A. (2003) Architecture of a Fur binding site: a comparative analysis. *J. Bacteriol.*, **185**, 2194–2202.
56. Orchard,K. and May,G.E. (1993) An EMSA-based method for determining the molecular weight of a protein – DNA complex. *Nucleic Acids Res.*, **21**, 3335–3336.
57. Masse,E. and Gottesman,S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl Acad. Sci USA*, **99**, 4620–4625.
58. Lavrrar,J.L., Christoffersen,C.A. and McIntosh,M.A. (2002) Fur–DNA interactions at the bidirectional *fepDGC-entS* promoter region in *Escherichia coli*. *J. Mol. Biol.*, **322**, 983–995.
59. Delany,I., Rappuoli,R. and Scarlato,V. (2004) Fur functions as an activator and as a repressor of putative virulence genes in *Neisseria meningitidis*. *Mol. Microbiol.*, **52**, 1081–1090.
60. Masse,E., Vanderpool,C.K. and Gottesman,S. (2005) Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J. Bacteriol.*, **187**, 6962–6971.
61. Argaman,L., Hershberg,R., Vogel,J., Bejerano,G., Wagner,E.G., Margalit,H. and Altuvia,S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.*, **11**, 941–950.
62. Wassarman,K.M., Repoila,F., Rosenow,C., Storz,G. and Gottesman,S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.*, **15**, 1637–1651.
63. Robison,K., McGuire,A.M. and Church,G.M. (1998) A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.*, **284**, 241–254.
64. Salgado,H., Santos-Zavaleta,A., Gama-Castro,S., Millan-Zarate,D., Diaz-Peredo,E., Sanchez-Solano,F., Perez-Rueda,E., Bonavides-Martinez,C. and Collado-Vides,J. (2001) RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res.*, **29**, 72–74.
65. Andrews,S.C., Robinson,A.K. and Rodriguez-Quinones,F. (2003) Bacterial iron homeostasis. *FEMS Microbiol. Rev.*, **27**, 215–237.
66. Schneider,T.D. (2002) Consensus Sequence Zen. *Appl. Bioinformatics*, **1**, 111–119. http://www.ccrnp.ncifcrf.gov/~toms/papers/zen/
67. Rudd,K.E. and Schneider,T.D. (1992) Compilation of *E. coli* ribosome binding sites. In Miller,J.H. (ed.), *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, New York, pp. 17.19–17.45.
68. Hengen,P.N., Lyakhov,I.G., Stewart,L.E. and Schneider,T.D. (2003) Molecular flip-flops formed by overlapping Fis sites. *Nucleic Acids Res.*, **31**, 6663–6673.
69. Fee,J.A. (1991) Regulation of *sod* genes in *Escherichia coli*: relevance to superoxide dismutase function. *Mol. Microbiol.*, **5**, 2599–2610.
70. Igarashi,K., Saisho,T., Yuguchi,M. and Kashiwagi,K. (1997) Molecular mechanism of polyamine stimulation of the synthesis of oligopeptide-binding protein. *J. Biol. Chem.*, **272**, 4058–4064.
71. Yoshida,M., Meksuriyen,D., Kashiwagi,K., Kawai,G. and Igarashi,K. (1999) Polyamine stimulation of the synthesis of oligopeptide-binding protein (OppA). Involvement of a structural change of the Shine-Dalgarno sequence and the initiation codon aug in oppa mRNA. *J. Biol. Chem.*, **274**, 22723–22728.
72. Litwin,C.M. and Calderwood,S.B. (1993) Role of iron in regulation of virulence genes. *Clin. Microbiol. Rev.*, **6**, 137–149.
73. Makui,H., Roig,E., Cole,S.T., Helmann,J.D., Gros,P. and Cellier,M.F. (2000) Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter. *Mol. Microbiol.*, **35**, 1065–1078.
74. Patzer,S.I. and Hantke,K. (2001) Dual repression by Fe(2+)-Fur and Mn(2+)-MntR of the *mntH* gene, encoding an NRAMP-like Mn(2+) transporter in *Escherichia coli*. *J. Bacteriol.*, **183**, 4806–4813.
75. Furrer,J.L., Sanders,D.N., Hook-Barnard,I.G. and McIntosh,M.A. (2002) Export of the siderophore enterobactin in *Escherichia coli*: involvement of a 43 kDa membrane exporter. *Mol. Microbiol.*, **44**, 1225–1234.
76. Cunningham,L., Gruer,M.J. and Guest,J.R. (1997) Transcriptional regulation of the aconitase genes (*acnA* and *acnB*) of *Escherichia coli*. *Microbiology*, **143 (Pt 12)**, 3795–3805.
77. Fuentes,A.M., Diaz-Mejia,J.J., Maldonado-Rodriguez,R. and Amabile-Cuevas,C.F. (2001) Differential activities of the SoxR protein of *Escherichia coli*: SoxS is not required for gene activation under iron deprivation. *FEMS Microbiol. Lett.*, **201**, 271–275.
78. Varghese,S., Tang,Y. and Imlay,J.A. (2003) Contrasting sensitivities of *Escherichia coli* aconitases A and B to oxidation and iron depletion. *J. Bacteriol.*, **185**, 221–230.
79. Zavitz,K.H., DiGate,R.J. and Marians,K.J. (1991) The PriB and PriC replication proteins of *Escherichia coli*. Genes, DNA sequence, overexpression, and purification. *J. Biol. Chem.*, **266**, 13988–13995.
80. Sandler,S.J. (2005) Requirements for replication restart proteins during constitutive stable DNA replication in *Escherichia coli* K-12. *Genetics*, **169**, 1799–1806.
81. Schneider,T.D. (2001) Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. *Nucleic Acids Res.*, **29**, 4881–4891. http://www.ccrnp.ncifcrf.gov/~toms/paper/baseflip/