

Research article

Open Access

## DNA-binding by *Haemophilus influenzae* and *Escherichia coli* YbaB, members of a widely-distributed bacterial protein family

Anne E Cooley<sup>1,6</sup>, Sean P Riley<sup>1,7</sup>, Keith Kral<sup>2</sup>, M Clarke Miller<sup>3,8</sup>, Edward DeMoll<sup>3,4</sup>, Michael G Fried<sup>5</sup> and Brian Stevenson\*<sup>1</sup>

Address: <sup>1</sup>Department of Microbiology, Immunology, and Molecular Genetics, University of Kentucky College of Medicine, Lexington, Kentucky, USA, <sup>2</sup>Paul Laurence Dunbar High School Math, Science and Technology Center, Lexington, Kentucky, USA, <sup>3</sup>Department of Chemistry, University of Kentucky, Lexington, KY, USA, <sup>4</sup>Department of Biology, University of Kentucky, Lexington, KY, USA, <sup>5</sup>Department of Molecular and Cellular Biochemistry, Center for Structural Biology, University of Kentucky College of Medicine, Lexington, Kentucky, USA, <sup>6</sup>Current address: Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA, <sup>7</sup>Current address: Dept of Microbiology, University of Chicago, Chicago, Illinois, USA and <sup>8</sup>Current address: Brown Cancer Center, University of Louisville, Louisville, Kentucky, USA

Email: Anne E Cooley - a-cooley@northwestern.edu; Sean P Riley - sriley@bsd.uchicago.edu; Keith Kral - kakral@gmail.com; M Clarke Miller - clarke.miller@louisville.edu; Edward DeMoll - eldemol@uky.edu; Michael G Fried - mgfrie2@uky.edu; Brian Stevenson\* - brian.stevenson@uky.edu

\* Corresponding author

Published: 13 July 2009

Received: 21 November 2008

BMC Microbiology 2009, 9:137 doi:10.1186/1471-2180-9-137

Accepted: 13 July 2009

This article is available from: <http://www.biomedcentral.com/1471-2180/9/137>

© 2009 Cooley et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Genes orthologous to the *ybaB* loci of *Escherichia coli* and *Haemophilus influenzae* are widely distributed among eubacteria. Several years ago, the three-dimensional structures of the YbaB orthologs of both *E. coli* and *H. influenzae* were determined, revealing a novel "tweezer"-like structure. However, a function for YbaB had remained elusive, with an early study of the *H. influenzae* ortholog failing to detect DNA-binding activity. Our group recently determined that the *Borrelia burgdorferi* YbaB ortholog, EbfC, is a DNA-binding protein. To reconcile those results, we assessed the abilities of both the *H. influenzae* and *E. coli* YbaB proteins to bind DNA to which *B. burgdorferi* EbfC can bind.

**Results:** Both the *H. influenzae* and the *E. coli* YbaB proteins bound to tested DNAs. DNA-binding was not well competed with poly-dI-dC, indicating some sequence preferences for those two proteins. Analyses of binding characteristics determined that both YbaB orthologs bind as homodimers. Different DNA sequence preferences were observed between *H. influenzae* YbaB, *E. coli* YbaB and *B. burgdorferi* EbfC, consistent with amino acid differences in the putative DNA-binding domains of these proteins.

**Conclusion:** Three distinct members of the YbaB/EbfC bacterial protein family have now been demonstrated to bind DNA. Members of this protein family are encoded by a broad range of bacteria, including many pathogenic species, and results of our studies suggest that all such proteins have DNA-binding activities. The functions of YbaB/EbfC family members in each bacterial species are as-yet unknown, but given the ubiquity of these DNA-binding proteins among Eubacteria, further investigations are warranted.

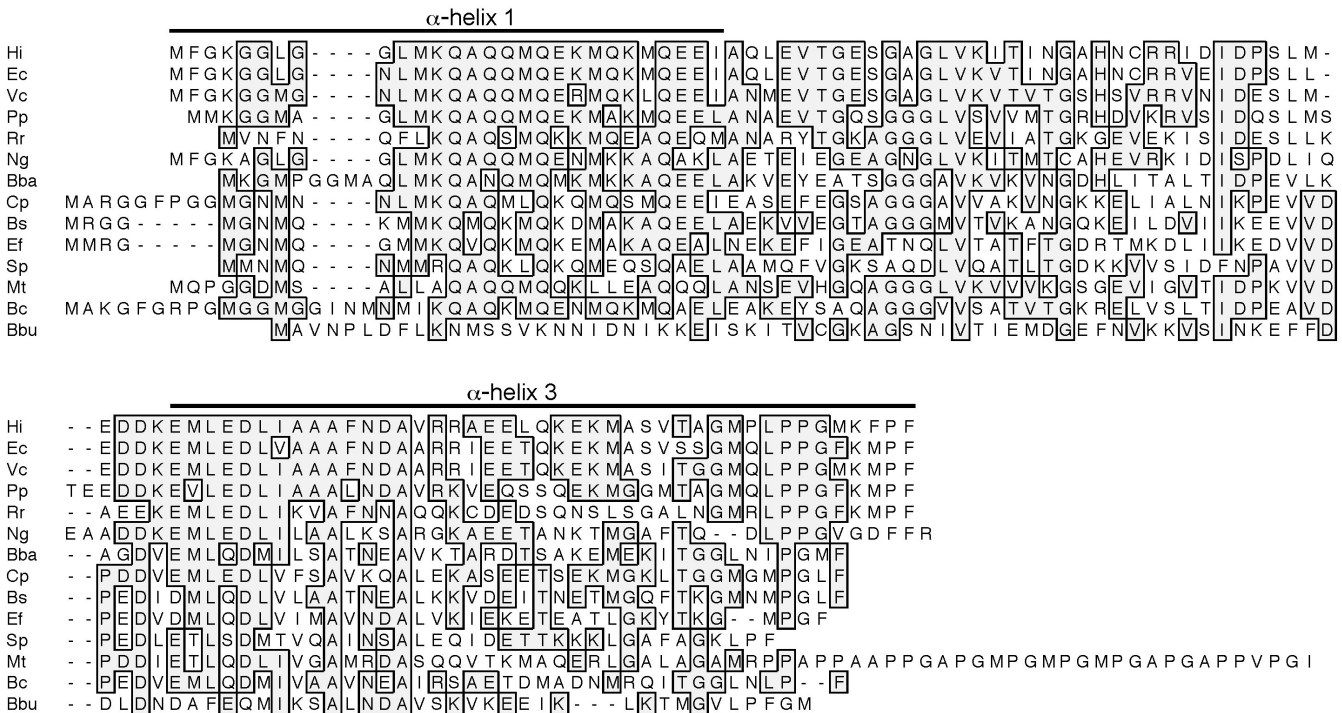
**Background**

Genome sequencing of diverse bacterial species has revealed widespread distribution of conserved gene products with as-yet unknown functions. Among these are a family of small proteins with approximate molecular masses of 12 kDa, which have been variously classed as "domain of unknown function" (DUF) 149, Pfam 2575 and COG-0718 [1]. Such genes have been identified in a wide variety of bacterial phyla, a list that includes many significant pathogens of humans, domestic animals and plants (Fig. 1).

After the genome sequence of *H. influenzae* strain KW20 rd (also known as *H. influenzae* Rd) was determined in 1995 [2], the "Structure 2 Function Project" was established to crystallize recombinant proteins from *H. influenzae* genes of unknown function <http://s2f.umbi.umd.edu/>. Among these orphan gene products was the *H. influenzae* DUF 149 group member annotated as open reading frame (ORF) HI0442, and tentatively named "YbaB" [3]. *H.*

*influenzae* YbaB (YbaB<sub>Hi</sub>) crystallized as a homodimer, with the central portion forming 3 antiparallel  $\beta$ -strands, long  $\alpha$ -helices at the amino- and carboxy-termini ( $\alpha$ -helices 1 and 3, respectively), and a short  $\alpha$ -helix bridging the  $\beta$ -folded region and  $\alpha$ -helix 3 ( $\alpha$ -helix 2). The two subunits of the homodimer interface at the  $\beta$ -strand region,  $\alpha$ -helix 2 and the initial residues of  $\alpha$ -helix 3, while  $\alpha$ -helix 1 and the terminal portion of  $\alpha$ -helix 3 project away from the dimerization region. This distinctive structure that has been described as resembling a set of tweezers [3]. Although the researchers who initially characterized YbaB<sub>Hi</sub> speculated that it may be a DNA-binding protein, studies conducted at that time failed to detect binding to any of their analyzed DNA probes [3].

The *Escherichia coli* chromosome carries an orthologous gene that has been referred to as "ORF 12" (Fig. 1) [4-6]. Recombinant *E. coli* YbaB (YbaB<sub>Ec</sub>) has also been crystallized and information about its unpublished three-dimensional structure is available on-line <http://>



**Figure 1**

**Alignment of the predicted amino acid sequences of YbaB/Ebfc orthologs of *H. influenzae* (Hi), *E. coli* (Ec), *Vibrio cholerae* (Vc), *Pseudomonas putida* (Pp), *Rickettsia rickettsiae* (Rr), *Neisseria gonorrhoeae* (Ng), *Bdellovibrio bacteriovorus* (Bba), *Clostridium perfringens* (Cp), *Bacillus subtilis* (Bs), *Enterococcus faecalis* (Ef), *Streptococcus pneumoniae* (Sp), *Mycobacterium tuberculosis* (Mt), *Bacteroides capillosus* (Bc), and *B. burgdorferi* (Bbu). Identical amino acids are boxed and shaded. Amino acid residues of YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> that comprise  $\alpha$ alpha-helices 1 and 3 of their determined protein structures are identified.**

[www.rcsb.org/pdb/explore.do?structureId=1PUG](http://www.rcsb.org/pdb/explore.do?structureId=1PUG). The determined structures of YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> are nearly identical. A function for YbaB<sub>Ec</sub> appears not to have been investigated prior to the current work.

The spirochete *Borrelia burgdorferi* produces a protein named EbfC that shares 29% identical and 56% similar amino acids with YbaB<sub>Hi</sub> (Fig. 1). Our laboratories recently discovered that EbfC binds a specific DNA sequence 5' of the spirochete's *erp* loci [7-10]. Those results suggested that orthologous proteins may also be DNA-binding proteins. We therefore re-examined the properties of YbaB<sub>Hi</sub> and found that it does bind to certain DNAs. YbaB<sub>Ec</sub> was also demonstrated to be a DNA-binding protein.

## Results and discussion

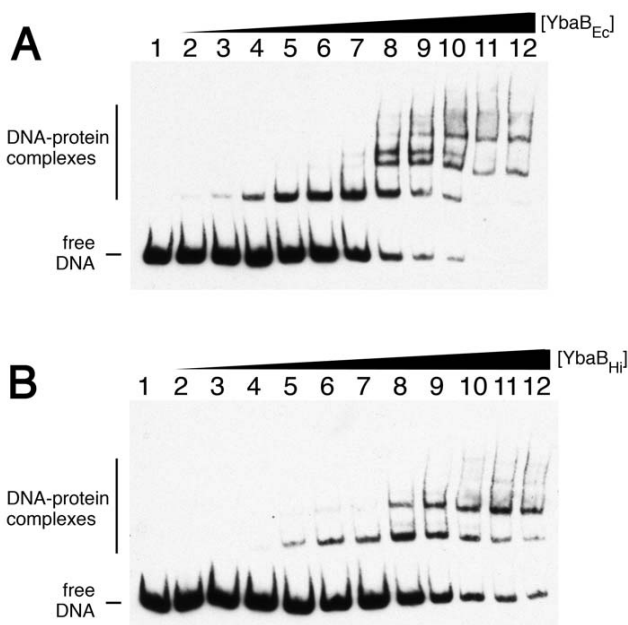
The abilities of YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> to bind DNA were first tested using a labeled DNA probe corresponding to sequences surrounding *B. burgdorferi* *erpAB* Operator 2 (Fig. 2). This DNA was chosen because the *B. burgdorferi* YbaB ortholog, EbfC, binds specifically to sequences within that region of DNA [7,8]. Both the *E. coli* and *H. influenzae* orthologs bound this DNA probe, each forming multiple DNA-protein complexes (Fig. 3). The simplest interpretation of these data is that each ladder of gel bands represents a stoichiometric series with higher stoichiometry (lower mobility) products formed from lower stoichiometry (higher mobility) precursors as protein concentration is increased. Similar patterns have been reported for other molecular systems (e.g., lac repressor-DNA complexes and CAP-DNA complexes) for which this interpretation has been found to be correct [11,12]. The EMSA assay does not provide information about the nature of the macromolecular interactions that stabilize each protein-DNA complex. Thus while the formation of the first complex must involve protein-DNA contacts, the interactions that stabilize higher-order complexes may include protein-protein contacts or protein-DNA contacts or both. The simplest model, and the one we favor, is one

in which similar mechanisms direct the binding of each protein unit to DNA or pre-existing protein-DNA complex. Affinity data for the first two binding steps (described below) are consistent with this picture, but do not rule out more heterogeneous binding mechanisms.

Binding distributions were graphed (Fig. 4A) and analyzed according to Eqs. 3–5 (see the Methods section). These data are consistent with models in which 2 molecules of YbaB<sub>Hi</sub> bind free DNA to form the first complex, and in which the second binding step involves the concerted binding of 2 additional YbaB<sub>Hi</sub> molecules. For these binding models, the association constants for the first and second binding steps are  $K_{a,1} = 1.7 \pm 0.7 \times 10^{13} \text{ M}^{-2}$  and  $K_{a,2} = 3.0 \pm 1.4 \times 10^{12} \text{ M}^{-2}$ . Assuming equipartition of binding free energies, these values correspond to apparent, monomer-equivalent dissociation constants  $K_{d,1} = 2.4 \pm 0.4 \times 10^{-7} \text{ M}$  and  $K_{d,2} = 5.8 \pm 1.0 \times 10^{-7} \text{ M}$ . These values indicate that the two best YbaB<sub>Hi</sub> binding sites on this DNA are of nearly equal affinity; the ~2-fold difference in affinity between first and second binding steps is just what would be expected on a statistical basis for independent binding to identical sites [13]. Parallel measurements were made for the binding of YbaB<sub>Ec</sub> to the b-WT DNA fragment (Fig. 4B). These data also indicate that 2 molecules of YbaB<sub>Ec</sub> bound free DNA to form the first complex and two more bound to form the second complex. The association constants for the first and second binding steps are  $K_{a,1} = 1.7 \pm 0.8 \times 10^{14} \text{ M}^{-2}$  and  $K_{a,2} = 2.9 \pm 0.5 \times 10^{13} \text{ M}^{-2}$ . Assuming equipartition of binding free energies as before, these correspond to monomer-equivalent dissociation constants  $K_{d,1} = 7.7 \pm 0.4 \times 10^{-8} \text{ M}$  and  $K_{d,2} = 1.9 \pm 0.3 \times 10^{-7} \text{ M}$ . As with the *H. influenzae* protein, the ~2-fold difference in affinity is what would be expected for independent binding to two identical sites. We note that these binding constants reflect binding under our standard *in vitro* conditions and should not be interpreted to represent the corresponding affinities for binding *in vivo*. None of our binding data suggests that either protein can bind DNA as a monomer. YbaB<sub>Hi</sub> and YbaB<sub>Ec</sub> proteins crystallized as

b-WT	TTGTAATGAGTAGTGCAATTTGCAATGGAGAGATTTGGGGAGTTGTTTAAAAATTACATTTGCGTTTTGTTAAAAATGTAACAGCTGAATGTAAACAAAATATATATATTTAAATCTTTGAAATATTGC
b-C2	TTGTAATGAGTAGTGCAATTTGCAATGGAGAGATTTGGGGAGTTGTTTAAAAATTACATTTGCGTTTTGTTAAAAacAAACAGCTGAATGTAAACAAAATATATATATTTAAATCTTTGAAATATTGC
b-C20	TTGTAATGAGTAGTGCAATTTGCAATGGAGAGATTTGGGGAGTTGTTTAAAAATTACATTTGCGTTTTGTTAAAAacAAACAGCTGAATGTAAACAAAATATATATATTTAAATCTTTGAAATATTGC
b-C30	TTGTAATGAGTAGTGCAATTTGCAATGGAGAGATTTGGGGAGTTGTTTAAAAATTACATTTGCGTTTTGTTAAAAacAAACAGCTGAATGTAAcTAAATATATATATTTAAATCTTTGAAATATTGC
b-C40	TTGTAATGAGTAGTGCAATTTGCAATGGAGAGATTTGGGGAGTTGTTTAAAAATTACATTTGCGTTTTGTTAAAAacAAACAGCTGAATGTgCAAAAATATATATATTTAAATCTTTGAAATATTGC
b-C50	TTGTAATGAGTAGTGCAATTTGCAATGGAGAGATTTGGGGAGTTGTTTAAAAATTACATTTGCGTTTTGTTAAAAacAAACAGCTGAATGTATCAAAAATATATATATTTAAATCTTTGAAATATTGC

**Figure 2**  
**Nucleotide sequences (5' to 3') of DNA probes used for EMSA in these studies, based on the operator 2 sequences of *B. burgdorferi* *erpAB* [7,8,10]. Underlined nucleotides identify the wild-type (GTnAC) and mutated sequences to which *B. burgdorferi* EbfC will either bind or not bind, respectively (see Fig. 5). Mutated nucleotides are indicated by lower case letters. All probes used in EMSAs were labeled with a biotin moiety at the one 5' end.**

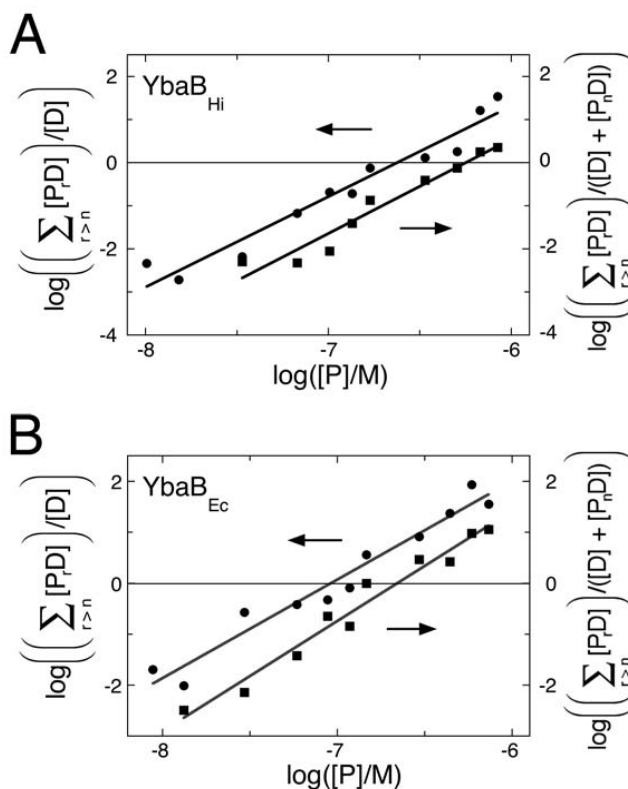


**Figure 3**  
**YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> are DNA-binding proteins.** (A) Representative EMSA using labeled probe b-WT and increasing concentrations of recombinant YbaB<sub>Ec</sub>. Lane 1 lacked YbaB<sub>Ec</sub>, and lanes 2 through 12 contained 0.14, 0.21, 0.47, 0.93, 1.4, 1.8, 2.3, 4.7, 7.0, 9.4 or 12 μg/ml YbaB<sub>Ec</sub>, respectively. (B) Representative EMSA using labeled probe b-WT and increasing concentrations of recombinant YbaB<sub>Hi</sub>. Lane 1 lacked YbaB<sub>Hi</sub>, and lanes 2 through 12 contained 0.18, 0.26, 0.59, 1.2, 1.8, 2.3, 2.9, 5.9, 8.8, 12 or 15 μg/ml YbaB<sub>Hi</sub>, respectively.

dimers, and both previous sedimentation analyses and our gel filtration analyses indicated that YbaB<sub>Hi</sub> exists primarily as a homodimer in solution [data not shown and [3]]. Taken together, these data indicate that the homodimer is the basic unit of DNA-binding activity for this family of proteins.

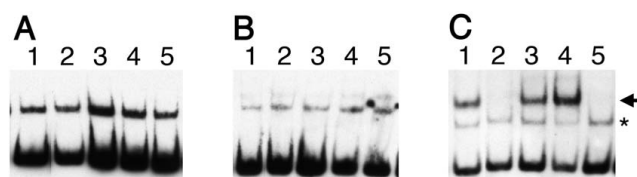
In control experiments, purified YbaB proteins were treated either by incubation with 1 mg/ml proteinase K for 30 min or by heating in a boiling water bath for 10 min. EMSA of either protease-treated or boiled YbaB preparations did not yield reduced-mobility complexes or reduce the levels of free DNA probe (data not shown), demonstrating that the DNA-binding activity in the purified YbaB preparations was due to the native forms of the proteins.

*B. burgdorferi* EbfC binds specifically to the tetrad GTnAc, and mutation of any of those 4 bases eliminates specific DNA binding (Fig. 5, [8,10]). To assess the requirements for those nucleotides on YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> binding, EMSAs were performed using as probes either a derivative of *B. burgdorferi* *erpAB* operator 2 that contains only 1 con-



**Figure 4**  
**Analysis of stoichiometries and affinities of YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> binding to b-WT DNA.** Data from the experiments shown in Fig. 3. (A) Binding of YbaB<sub>Ec</sub>. Symbols: (black circle), first binding step; (black square), second binding step. The lines are least-squares fits to Eqs 4 and 5, returning stoichiometry values of  $1.93 \pm 0.14$  for the first binding step and  $2.16 \pm 0.14$  for the second. From the logarithm of the free protein concentration at the midpoint of each binding transition we estimate that  $K_{a,1} = 1.7 \pm 0.8 \times 10^{14} \text{ M}^{-2}$  and  $K_{a,2} = 2.9 \pm 0.5 \times 10^{13} \text{ M}^{-2}$ . The ranges given for these parameters are 95% confidence limits calculated for the least squares fits. (B) Binding of YbaB<sub>Hi</sub>. Symbols: (black circle), first binding step; (black square), second binding step. The lines are least-squares fits to Eqs 4 and 5, returning stoichiometry values of  $2.09 \pm 0.16$  for the first binding step and  $2.18 \pm 0.19$  for the second. From the logarithm of the free protein concentration at the midpoint of each binding transition we estimate  $K_{a,1} = 1.7 \pm 0.7 \times 10^{13} \text{ M}^{-2}$  and  $K_{a,2} = 3.0 \pm 1.4 \times 10^{12} \text{ M}^{-2}$ . The ranges given for these parameters are 95% confidence limits calculated for the least squares fits.

sensus EbfC-binding site (probe b-C2) or that DNA containing single bp mutations (probes b-C20, 30, 40 and 50, Fig. 2). For each protein, a concentration of one half its  $K_d$  was utilized in order to show either increases or decreases in binding. Note that both YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> produced one protein-DNA complex at these protein concentrations, whereas EbfC yielded two mobility complexes. Other studies from our laboratories demonstrated that the upper (more slowly migrating) EbfC-DNA complex repre-

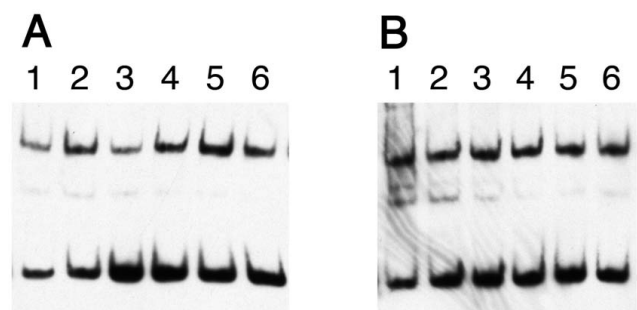


**Figure 5**  
**Neither YbaB<sub>Ec</sub> nor YbaB<sub>Hi</sub> specifically binds the same nucleotide sequence as does *B. burgdorferi* EbfC.** For all panels, lanes 1 contain probe b-C2, lanes 2 contain probe b-C20, lanes 3 contain b-C30, lanes 4 contain b-C40, and lanes 5 contain b-C50. **(A)** YbaB<sub>Ec</sub>. **(B)** YbaB<sub>Hi</sub>. **(C)** EbfC, with the arrowhead indicating the specific EbfC-DNA complex and the asterisk indicating a non-specific EbfC-DNA complex [8,10].

sents specific binding to the GTnAC sequence, while the lower (more rapidly-migrating) complex reflects a sequence-nonspecific interaction [10]. None of the single mutations had any detectable effect on binding by either YbaB<sub>Ec</sub> or YbaB<sub>Hi</sub> (Fig. 5A & 5B). Point mutations that disrupted the GTnAC sequence eliminated specific binding of EbfC, but did not affect non-specific binding by that protein (Fig. 5C).

The specificity of YbaB binding was further addressed by EMSA using progressively greater concentrations of poly(dI-dC), which acts as a competitor for non-specific DNA binding activities [14]. Addition of even 500-fold excesses of poly(dI-dC) had no measurable effect on either YbaB<sub>Ec</sub> or YbaB<sub>Hi</sub> binding to the *B. burgdorferi* *erpAB* operator 2 probe (Fig. 6).

A previous study did not detect binding of YbaB<sub>Hi</sub> to any tested DNA, leading to the conclusion that this protein does not bind DNA in a completely sequence-independent



**Figure 6**  
**Addition of increasing concentrations of poly(dI-dC) did not detectably alter DNA-binding by either YbaB ortholog.** **(A)** YbaB<sub>Ec</sub>. **(B)** YbaB<sub>Hi</sub>. For both panels, lanes 1 did not contain any poly(dI-dC), and lanes 2 through 6 contained 0.1, 0.5, 1, 2 or 4 ng per reaction, respectively.

ent manner [3]. The present work demonstrated that YbaB<sub>Hi</sub> and the homologous protein of *E. coli*, do bind to certain DNAs. EbfC, the orthologous protein of the spirochete *B. burgdorferi*, binds specifically to the DNA sequence GTnAC and, with a lower affinity, to DNA lacking that sequence [8,10]. The *E. coli* and *H. influenzae* YbaB proteins both exhibited preferences for certain tested DNA sequences, but neither showed the same high affinity for GTnAC as did the spirochetal ortholog. Both YbaB proteins also showed a marked preference for DNA derived from the *B. burgdorferi* *erpAB* promoter over poly(dI-dC). Such large differences in affinities for target and non-target sequences may account for the previous failure to detect DNA-binding by YbaB<sub>Hi</sub> [3]. These results suggest that YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> have higher affinities for some DNA sequences than for others, but whether those preferences depend upon a specific nucleotide sequence(s), A+T content, and/or DNA topology remain to be determined. The three-dimensional structure of dimeric YbaB resembles "tweezers", with  $\alpha$ -helices 1 and 3 of each monomeric subunit protruding from the dimerization domains [3]. The spacing between the  $\alpha$ -helical protrusions is approximately 15 Å at the base of the dimerization domain and approximately 22 Å at the distal ends of the  $\alpha$ -helices [3], similar to the diameter of B-form duplex DNA (~20 Å [3]). Site-directed mutagenesis studies of the orthologous *B. burgdorferi* EbfC demonstrated that certain amino acid substitutions in either  $\alpha$ -helix 1 or 3 of EbfC eliminate DNA-binding, without affecting dimerization [10]. It is noteworthy that many of the  $\alpha$ -helix 1 and 3 residues of EbfC are distinct from residues in both YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> (Fig. 1), consistent with the differences in DNA preferences between the *E. coli* and *H. influenzae* YbaB proteins and their spirochetal ortholog. YbaB/EbfC orthologs of other bacterial species likewise exhibit sequence variations in their  $\alpha$ -helices 1 and 3, suggesting that they may also possess unique DNA-binding properties.

The function(s) of YbaB/EbfC proteins remains to be determined. Many bacterial *ybaB/ebfC* orthologs are located between *dnaX* and *recR*, a synteny that has led to suggestions of roles in DNA replication or recombination [3,5,6,15-18]. While the abilities of the examined orthologs to bind DNA may support those hypotheses, several lines of evidence suggest that YbaB/EbfC proteins perform functions that are independent of DNA recombination or replication. Proteomic analyses of cultured *H. influenzae* detected production of YbaB without accompanying production of DNA repair proteins [19]. A *ybaB recR* double mutant of *Streptomyces coelicolor* exhibited recombination defects that could be complemented with *recR* alone [18]. The *ybaB/ebfC* orthologs of some bacterial species are not linked to *recR* or any other recombination-related gene and some, such as the *B. burgdorferi*, do not

even encode RecR [8,20]. Several bacteria, such as *H. influenzae*, have *ybaB* genes located distantly from their *dnaX* [2]. Moreover, some *ybaB* family genes can be transcribed independently of their upstream genes, using promoter elements within the 5' gene [4,6,21-23].

## Conclusion

We demonstrated that YbaB<sub>Hi</sub> is in fact a DNA-binding protein. It exhibits an element of specificity, in that the protein preferentially bound to *B. burgdorferi* *erp* Operator 2 DNA over poly-dI-dC and, apparently, the DNA sequences examined by an earlier research group [3]. Consistent with those data, the *E. coli* YbaB ortholog was also determined to be a DNA-binding protein. For both orthologs, the basic unit of DNA-binding is a homodimer, consistent with results from analyses of soluble proteins and crystallization data. The solved structures of YbaB<sub>Ec</sub> and YbaB<sub>Hi</sub> are distinct from any other known DNA-binding protein. Genes encoding orthologs of YbaB/EbfC proteins are found throughout the Eubacteria, including many important human pathogens, suggesting that these proteins perform important function(s). Thus, continued study of these unique proteins may provide insight regarding critical bacterial processes that might be exploited for infection control.

## Methods

### Bacterial gene sequences

Bacterial protein sequences orthologous to YbaB<sub>Hi</sub>, YbaB<sub>Ec</sub> and *B. burgdorferi* EbfC were identified by BlastP, using the predicted sequences of those three proteins as queries <http://blast.ncbi.nlm.nih.gov/Blast.cgi>. Amino acid sequences were aligned using Clustal X, with default parameters [24]. Orthologs from the following bacteria were chosen as representative of different bacterial classifications: the  $\alpha$  proteobacterium *Rickettsia rickettsiae* (accession number NC\_009882), the  $\beta$  proteobacterium *Neisseria gonorrhoeae* (NC\_002946.2), the gamma proteobacteria *Vibrio cholerae* (NC\_002505.1) and *Pseudomonas putida* (NC\_010501.1), the delta proteobacterium *Bdellovibrio bacteriovorus* (NC\_005363.1), the firmicutes *Clostridium perfringens* (NC\_003366.1), *Bacillus subtilis* (NC\_000964.2), *Enterococcus faecalis* (NC\_004668.1), and *Streptococcus pneumoniae* (NC\_003098.1), the actinomycete *Mycobacterium tuberculosis* (NC\_000962.2), and the bacteroidete *Bacteroides capillosus* (NZ\_AAXG02000011.1).

### Recombinant proteins

Recombinant YbaB<sub>Hi</sub> protein was produced from pET15b-HI0442 (a gift of Osnat Herzberg, University of Maryland) [3]. Recombinant YbaB<sub>Ec</sub> was produced by first PCR amplifying the *ybaB*<sub>Ec</sub> gene from total genomic DNA using the oligonucleotide primers 5'-CACCCGTGATTGAG-GAGGAAACCTATG-3' and 5'-CAGCGGGCTGGTTT-GCATCAG-3'. The resulting amplicon was cloned into

pET200-TOPO (Invitrogen, Carlsbad, CA), and the insert completely sequenced on both strands. Recombinant *B. burgdorferi* EbfC was produced using the previously-described plasmid construct p462-M5 [8].

Each plasmid was individually used to transform *E. coli* Rosetta pLysS (Novagen, San Diego, CA), and production of recombinant proteins induced by addition of isopropylthiogalactopyranoside. Bacteria were lysed by sonication in 30 mM imidazole, 0.5 M NaCl, 20 mM NaPO<sub>4</sub>, pH = 7.4, and cleared by centrifugation. The recombinant proteins were purified using His-Trap HP columns and an AKTA-FPLC equipped with a UPC-900 UV absorbance monitor and a Frac920 fraction collector (GE Healthcare, Piscataway, NJ). Proteins were eluted with a constantly increasing gradient between the lysis buffer and 0.75 M imidazole, 20 mM NaPO<sub>4</sub>, 0.5 M NaCl, pH = 7.4. Proteins were then dialyzed against 1 × e0 buffer (50 mM Tris [pH = 7.5], 1 mM dithiothreitol, 1 mM phenylmethanesulfonyl fluoride, and 100  $\mu$ l/l Tween-20). Glycerol was added to a final concentration of 10% (vol/vol), and aliquots were snap frozen in liquid nitrogen and stored at -80°C. Purity of protein preparations was assessed by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE), followed by staining with Coomassie brilliant blue. BCA (bicinchoninic acid) protein assays (Pierce, Rockford, IL), calibrated with bovine serum albumin (Pierce), were used to determine protein concentrations.

### Electrophoretic mobility shift assays (EMSA)

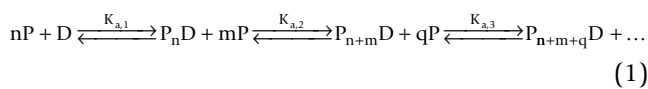
All EMSAs were performed at least three times. Biotin-labeled DNA probes were produced based upon the sequence of the *B. burgdorferi* strain B31 *erpAB* 5'-noncoding DNA, to which the orthologous EbfC protein is known to bind [7,8,10]. Probe b-WT corresponds with bp -160 through -36 (relative to the start of translation) of the *erpAB* operon, and contains two consensus EbfC-binding sites [8,10] (Fig. 2). Probe b-WT was produced by PCR using oligonucleotide primers bio-A14A (5'-biotin-TTG-TAATGAGTAGTGCATTG-3') and R8 (5'-GCAATATTCAAAGATTTAAA-3') from DNA template pBLS591 [7]. That same oligonucleotide primer pair was used to produce probe b-C2 from mutant template pSRJ-2, a derivative of pBLS591 in which EbfC-binding site II was changed to CACAACA (Fig. 2) [10]. Probes b-C20, b-C30, b-C40 and b-C50 were also produced using primers bio-A14A and R8, from mutant templates pSRJ-20, pSRJ30, pSRJ40 and pSRJ50, respectively, derivatives of pSRJ-2 in which single bp mutations were introduced to site I (Fig. 2) [10]. Each PCR reaction product was separated by agarose gel electrophoresis and DNA visualized by ethidium bromide staining. Amplicons were extracted from gels into nuclease-free water using Wizard SV (Promega, Madison, WI), and quantified by spectrophotometric determination of absorbance at 260 nm.

EMSAs were performed using 100 pM biotin-labeled DNA fragment and varying concentrations of purified recombinant YbaB<sub>Ec</sub> or YbaB<sub>Hi</sub>. Binding conditions consisted of 50 mM Tris-HCl (pH = 7.5), 1 mM dithiothreitol, 8 µl/ml protease inhibitor (Sigma-Aldrich, St. Louis, MO), 2 µl/ml phosphatase inhibitor cocktail II (Sigma-Aldrich), and 10% glycerol. Protein and DNA were mixed together, in final volumes of 10 ml, and allowed to proceed toward equilibrium for 20 minutes at room temperature, then subjected to electrophoresis through 6% DNA retardation gels (Invitrogen) for 9000 V-min. DNA was electrotransferred to Biodyne B nylon membranes (Pierce), cross-linked by ultraviolet light, and biotinylated DNA detected using Chemiluminescent Nucleic Acid Detection Modules (Pierce).

Competition for DNA binding by poly(dI-dC) was assessed using the above binding conditions, 2 fmol (0.082 ng) labeled probe b-WT and either 1.2 µg/ml YbaB<sub>Ec</sub> or 2.1 µg/ml YbaB<sub>Hi</sub>. After 20 min incubation at room temperature, either no or 0.1, 0.5, 1, 2 or 4 ng poly(dI-dC) was added to each tube, followed by an additional 20 min incubation at room temperature. DNA-protein mixtures were subjected to electrophoresis and detection as described above.

**Binding analyses**

Exposed films were scanned in 8 bit depth at 1200 dpi resolution using Image J 1.37 v <http://rsbweb.nih.gov/ij/>. Band intensities were converted into mole fractions as previously described [11]. Binding was analyzed according to a model in which several molecules of protein can bind the target DNA according to the general mechanism



here n, m and q are n numbers of protein monomers that associate at the first, second and third binding steps, characterized by association constants  $K_{a,1}$ ,  $K_{a,2}$  and  $K_{a,3}$ , respectively. As indicated by the ellipsis, this model can include > 3 binding steps, as necessary. For the first binding step

$$K_{a,1} = \frac{[P_nD]}{[P]^n[D]} \tag{2}$$

When not complicated by subsequent binding events, the evaluation  $K_{a,1}$  can be done according to standard procedures [12,25]. However, when higher-stoichiometry complexes accumulate before the first step reaches saturation, as is the case for the binding reactions shown in Fig. 3, it is necessary to account for all of the species in the equilibrium mixture that are formed from  $P_nD$ . When this is done, the equilibrium constant for the first binding step becomes

$$K_{a,1} = \frac{\sum [P_rD]}{[P]^n[D]} \tag{3}$$

Here the subscript r denotes the protein stoichiometry of the corresponding complex. Rearranging Eq. 3 and taking logs gives

$$\log \frac{\sum [P_rD]}{[D]} = n \log [P] + \log K_{a,1} \tag{4}$$

Thus, a graph of  $\log((\sum_r P_rD)/[D])$  as a function of  $\log [P]$  will have a slope equal to the stoichiometry n and an x-intercept at which  $-n \log [P] = \log K_a$ . For the binding of m protein molecules to a  $P_nD$  complex, the corresponding expression is

$$\log \frac{\sum_{r>n} [P_rD]}{([D] + [P_nD])} = m \log [P] + \log K_{a,2} \tag{5}$$

It is important to note that in this approach, values of stoichiometry and equilibrium constant are not fully independent (fitted values of  $K_a$  and n are related by  $-n \log [P] = \log K_a$ ). As a result, the parameters returned are the most likely values (in the least squares sense) that are internally-consistent. A similar analysis strategy has been described previously [12].

In studies of this kind, accurate measurement of  $K_a$  values require good estimates of the free protein concentration, [P]. In the present experiments, the protein concentrations (range  $\sim 10^{-8}$  M to  $\sim 10^{-6}$  M) exceeded by far the total DNA concentration ( $10^{-10}$  M). Thus, even in the presence of additional DNA binding (up to  $\sim 10$  protein molecules/DNA), free protein concentration [P] is well-approximated by the total protein concentration,  $[P]_{total}$ .

**Size-exclusion chromatography**

A Superdex 75 10/300 GL column (GE Healthcare) was prepared with a mobile phase consisting of 200 mM NaCl, 50 mM Tris-HCl (pH 7.5), and 1% (vol/vol) glycerol. The column was run with a flow rate of 0.20 ml per min using a Waters 600 pump and controller equipped with a Waters 996 photodiode array UV/Vis detector (Waters, Milford, MA). A calibration curve was created using an MW-GF-70 low-molecular-weight calibration kit (Sigma-Aldrich, St. Louis, MO), and the void volume,  $V_0$ , was determined by injection of 200 µl of 1 mg/ml blue dextran in elution buffer with 5% glycerol. The remaining protein standards, bovine lung aprotinin (6.5 kDa), horse heart cytochrome c (12.4 kDa), bovine carbonic anhy-

drase (29 kDa), and bovine serum albumin (66 kDa), were individually prepared in elution buffer with 5% glycerol to total concentrations of 0.3 mg/ml each, and the volume with which the protein eluted,  $V_e$ , was determined. The molecular-mass calibration curve was generated by plotting the log (molecular mass) versus  $V_e/V_o$  (5). A 200- $\mu$ l sample of recombinant YbaB<sub>HI</sub> (approximately 0.2 mg/ml) was then injected and its elution profile compared to the established curve to determine molecular masses of each elution peak.

### Authors' contributions

AEC, ED, MGF and BS designed the experiments. AEC, SPR and KK performed EMSA analyses. MCM and ED conducted size exclusion chromatography. AEC, SPR, ED, MGF and BS interpreted the results. All authors read and approved the manuscript.

### Acknowledgements

The work was funded by NIH grant R01-AI044254 to Brian Stevenson and R01-GM070662 to Michael Fried. Sean Riley was supported in part by NIH Training Grant in Microbial Pathogenesis T32-AI49795 and a University of Kentucky Graduate School Dissertation Year Fellowship. We thank Osnat Herzberg for the generous gift of the YbaB-producing plasmid, and Amy Bowman, Catherine Brissette, Logan Burns, Tomasz Bykowski, Ashutosh Verma, Erin Welsh, and Michael Woodman for assistance during these studies and comments on the manuscript.

### References

1. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwartz M, He S, Hurwitz DI, Jackson JD, Ke Z, et al.: **CDD: a conserved domain database for protein classification.** *Nucleic Acids Res* 2005, **33**:D192-196.
2. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al.: **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science* 1995, **269**:496-512.
3. Lim K, Tempczyk A, Parsons JF, Bonander N, Toedt J, Kelman Z, Howard A, Eisenstein E, Herzberg O: **Crystal structure of YbaB from *Haemophilus influenzae* (HI0442), a protein of unknown function coexpressed with the recombinational DNA repair protein RecR.** *Proteins* 2003, **50**:375-379.
4. Flower AM, McHenry CS: **Transcriptional organization of the *Escherichia coli* dnaX gene.** *J Mol Biol* 1991, **220**:649-658.
5. Mahdi AA, Lloyd RG: **The recR locus of *Escherichia coli* K-12: molecular cloning, DNA sequencing and identification of the gene product.** *Nucleic Acids Res* 1989, **17**:6781-6794.
6. Yeung T, Mullin DA, Chen K, Craig EA, Bardwell JCA, Walker JR: **Sequence and expression of the *Escherichia coli* recR locus.** *J Bacteriol* 1990, **172**:6042-6047.
7. Babb K, McAlister JD, Miller JC, Stevenson B: **Molecular characterization of *Borrelia burgdorferi* erp promoter/operator elements.** *J Bacteriol* 2004, **186**:2745-2756.
8. Babb K, Bykowski T, Riley SP, Miller MC, DeMoll E, Stevenson B: ***Borrelia burgdorferi* EbfC, a novel, chromosomally-encoded protein, binds specific DNA sequences adjacent to erp loci on the spirochete's resident cp32 prophages.** *J Bacteriol* 2006, **188**:4331-4339.
9. Stevenson B, Bykowski T, Cooley AE, Babb K, Miller JC, Woodman ME, von Lackum K, Riley SP: **The Lyme disease spirochete Erp lipoprotein family: structure, function and regulation of expression.** In *Molecular Biology of Spirochetes* Edited by: Cabello FC, Godfrey HP, Hulinska D. Amsterdam: IOS Press; 2006:354-372.
10. Riley SP, Bykowski T, Cooley AE, Burns LH, Babb K, Brissette CA, Bowman A, Rotondi M, Miller MC, DeMoll E, et al.: ***Borrelia burgdorferi* EbfC defines a newly-identified, widespread family of**

**bacterial DNA-binding proteins.** *Nucleic Acids Res* 2009, **37**:1973-1983.

11. Fried MG, Crothers DM: **Equilibria and kinetics of Lac repressor-operator interactions by polyacrylamide gel electrophoresis.** *Nucl Acids Res* 1981, **9**:6505-6525.
12. Fried MG, Crothers DM: **Equilibrium studies of the cyclic AMP receptor protein-DNA interaction.** *J Mol Biol* 1984, **172**:241-262.
13. Klotz IM: **Ligand-Receptor Interactions.** New York: Wiley; 1997.
14. Varshavsky A: **Electrophoretic assay for DNA-binding proteins.** *Methods Enzymol* 1987, **151**:551-565.
15. Bork JM, Cox MM, Inman RB: **The RecOR proteins modulate RecA protein function at 5' ends of single-stranded DNA.** *EMBO J* 2001, **20**:7313-7322.
16. Morimatsu K, Kowalczykowski SC: **RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair.** *Mol Cell* 2003, **11**:1337-1347.
17. Flower AM, McHenry CS: **The  $\gamma$  subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting.** *Proc Natl Acad Sci USA* 1990, **87**:3713-3717.
18. Peláez AI, Ribas-Aparicio RM, Gómez A, Rodicio MR: **Structural and functional characterization of the recR gene of *Streptomyces*.** *Mol Genet Genomics* 2001, **265**:663-672.
19. Kolker E, Purvine S, Galperin MY, Stolyar S, Goodlett DR, Nesvizhskii AI, Keller A, Xie T, Eng JK, Yi E, et al.: **Initial proteome analysis of model microorganism *Haemophilus influenzae* strain Rd KW20.** *J Bacteriol* 2003, **185**:4593-4602.
20. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, et al.: **Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.** *Nature* 1997, **390**:580-586.
21. Chen K, Saxena P, Walker JR: **Expression of the *Escherichia coli* dnaX gene.** *J Bacteriol* 1993, **175**:6663-6670.
22. Rezuchova B, Miticka H, Homerova D, Roberts M, Kormanec J: **New members of the *Escherichia coli*  $\sigma^E$  regulon identified by a two-plasmid system.** *FEMS Microbiol Lett* 2003, **225**:1-7.
23. Engels S, Ludwig C, Schweitzer J, Mack C, Bott M, Schaffer S: **The transcriptional activator CigR controls transcription of genes involved in proteolysis and DNA repair in *Corynebacterium glutamicum*.** *Mol Microbiol* 2005, **57**:576-591.
24. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The Clustal X window interface: flexible strategies for multiple sequence alignment aided by quality analyses tools.** *Nucleic Acids Res* 1997, **24**:4876-4882.
25. Adams CA, Fried MG: **Analysis of protein-DNA equilibria by native gel electrophoresis.** In *Protein interactions: Biophysical approaches for the study of complex reversible systems* Edited by: Schuck P. New York: Academic Press; 2007:417-446.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
http://www.biomedcentral.com/info/publishing\_adv.asp

