Toxin–antitoxin regulation: bimodal interaction of YefM–YoeB with paired DNA palindromes exerts transcriptional autorepression

Barbara Kędzierska, Lu-Yun Lian[1] and Finbarr Hayes*

Faculty of Life Sciences and Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK and [1]School of Biological Sciences, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK

ABSTRACT

Toxin–antitoxin (TA) complexes function in programmed cell death or stress response mechanisms in bacteria. The YefM–YoeB TA complex of *Escherichia coli* consists of YoeB toxin that is counteracted by YefM antitoxin. When liberated from the complex, YoeB acts as an endoribonuclease, preferentially cleaving 3′ of purine nucleotides. Here we demonstrate that *yefM-yoeB* is transcriptionally autoregulated. YefM, a dimeric protein with extensive secondary structure revealed by circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy, is the primary repressor, whereas YoeB is a repression enhancer. The operator site 5′ of *yefM-yoeB* comprises adjacent long and short palindromes with core 5′-TGTACA-3′ motifs. YefM binds the long palindrome, followed sequentially by short palindrome recognition. In contrast, the repressor–corepressor complex recognizes both motifs more avidly, impying that YefM within the complex has an enhanced DNA-binding affinity compared to free YefM. Operator interaction by YefM and YefM–YoeB is accompanied by structural transitions in the proteins. Paired 5′-TGTACA-3′ motifs are common in *yefM-yoeB* regulatory regions in diverse genomes suggesting that interaction of YefM–YoeB with these motifs is a conserved mechanism of operon autoregulation. Artificial perturbation of transcriptional autorepression could elicit inappropriate YoeB toxin production and induction of bacterial cell suicide, a potentially novel antibacterial strategy.

INTRODUCTION

Bacteria populate and thrive in remarkably diverse environmental conditions and ecological niches. However, bacterial species may need to adapt rapidly to dramatic changes in nutritional or physiological circumstances. For example, intestinal bacteria must adjust to cycles of nutriment excess and starvation, and soil and other microorganisms may need to cope with oscillating periods of dessication and hydration (1,2). Moreover, when passaging between exponential growth and stationary phase, the production of diverse macromolecules and cell components decelerates at different rates (3). Furthermore, bacteria generally exist as multicellular colonies (4) or as biofilms (5). Within these microenvironments, intercellular signalling and coordinated multicellular processes are mediated by quorum-sensing mechanisms that regulate a diverse array of physiological activities (6). Even at the colonial level bacteria maintain discrete, ordered spatial structures (4).

The concept that bacteria possess programmed cell death or cell cycle arrest mechanisms that interconnect with complex physiological networks and multicellular organization has emerged relatively recently, but has been validated by observations that the genome of *Escherichia coli* K12, for example, harbours a number of toxin–antitoxin (TA) modules. These modules are involved in the response to nutrient deprivation or other stresses (7). Other bacterial genomes also contain multiple putative TA cassettes (8). Remarkably, some bacteria harbour 10s of TA modules, including some species that possess >40 distinct TA loci (9). These modules are organized similarly and are homologous to TA cassettes that promote plasmid maintenance by post-segregational killing of plasmid-free cells, indicating extensive module transfer between plasmid and chromosome genomes (7). In plasmid-specified complexes, the antitoxin is more susceptible to host proteases than is the toxin. When a plasmid-free cell arises

the toxin is liberated from its tight interaction with depleted antitoxin and targets an essential intracellular host factor to cause cell death or severe growth impairment. Chromosomal TA activity in bacteria might be triggered in response to starvation conditions, bacteriophage infection exposure to antimicrobial agents, DNA damage and other physiological stresses. Under these circumstances it may be beneficial to the community to sacrifice a portion of the cells within it so that a sub-population can persist, perhaps even cannabilizing nutrients that dead cells have released (10,11). Alternatively, by acting as reversible cell cycle arrest agents, TA factors might allow cells to enter a dormant or semidormant state as a protection against temporary nutrient limitation and to revive when physiological conditions become more conducive (12).

The two most well characterized chromosomal TA factors are RelBE and MazEF of *E.coli*. RelE toxin is a global inhibitor of translation that is activated during amino acid starvation but which is otherwise counteracted by RelB antitoxin (13). RelE is an endoribonuclease that, although it does not degrade free RNA, cleaves mRNA in the ribosomal A site with high-codon specificity (12). Production of the MazEF proteins is regulated by the alarmone guanosine-3′5′-bispyrophosphate (ppGpp) that is synthesized by RelA protein under conditions of amino acid starvation. Moreover, overproduction of ppGpp induces MazEF-mediated cell death (14). MazF, like RelE, is an endoribonuclease that cleaves mRNA site-specifically, although without a requirement that the mRNA be associated with the ribosome (15,16). MazEF, like RelBE, might be responsible for death in starving cultures of *E.coli* (14). Cell death mediated by MazEF can also be triggered by several antibiotics that are general inhibitors of transcription and/or translation. In contrast, it has been proposed that the MazF toxin does not induce cell killing, but instead is a bacteriostatic agent from which cells can recover when more conducive conditions prevail (17).

Pomerantsev *et al.* first proposed that *yefM-yoeB* in *E.coli* specified a TA complex, based on homology between YefM and the Phd antitoxin encoded by bacteriophage P1 (18). YefM–YoeB was subsequently shown to be a functional TA related to the Axe–Txe complex encoded by enterococcal, multidrug resistance plasmid pRUM, and that YefM–YoeB homologues are widely-distributed on bacterial genomes (19). The YefM antitoxin forms a heterotrimeric complex with the YoeB toxin (20,21). YoeB is an endoribonuclease, like RelE and MazF, cleaving preferentially at the 3′ side of adenine and guanine nucleotides (21,22). Overproduction of Lon ATP-dependent protease specifically activates this cleavage. This probably occurs due to the liberation of YoeB from its tight association with YefM, the latter failing to be replenished due to Lon-mediated translation inhibition of a YoeB-independent pathway (22). Interestingly, the *yefM* gene is upregulated during growth of *E.coli* in biofilms (23).

The tertiary structures of YoeB and the YefM$_2$–YoeB complex have recently been described (21). One of the C-termini in the YefM homodimer is unstructured, whereas the other C-terminus adopts an α-helical conformation within the heterotrimeric complex and conceals the atypical ribonuclease fold of YoeB. The N-terminal regions of YefM form a

symmetrical dimer within the YefM$_2$–YoeB complex and do not contact YoeB directly in the crystal structure (21), although the YefM recognition determinant that interacts most strongly with YoeB was previously mapped to the N-terminal segment of the protein (24). The three residues at the C-terminal tip of YoeB that form part of the atypical endoribonucleolytic fold are rearranged into a less favourable conformation when in complex with the YefM dimer, partly explaining the mechanism by which the antitoxin blocks the toxic activity of YoeB (21).

Controlled activation of the toxin factor is paramount to the function of TA complexes. Transcriptional autoregulation of TA cassettes is one level at which this control is exerted (25–33). Here we dissect the role of the YefM–YoeB complex in modulating its own synthesis: YefM is the primary transcriptional repressor of the *yefM-yoeB* cassette, with YoeB acting as a repression enhancer. DNA binding is achieved by the association of the proteins with a pair of palindromes that comprise the *yefM-yoeB* operator site. Understanding the molecular basis of *yefM-yoeB* regulation may suggest strategies for perturbing this control, leading to the production of excess intracellular toxin and therefore controlled bacterial cell suicide.

MATERIALS AND METHODS

Strains

E.coli DH5α was used for plasmid construction and RNA isolation, BL21(DE3) for recombinant YefM and YefM–YoeB overproduction, and SC301467 (22) for β-galactosidase assays. Bacteria were grown in Luria–Bertani (LB) medium at 37°C. Antibiotics were added at final concentrations of 100 μg/ml (ampicillin) and 34 μg/ml (chloramphenicol).

Plasmids and oligonucleotides

Oligonucleotides and plasmids used in this study are listed in Tables 1 and 2, respectively.

Protein production and purification

The *yefM* and the *yefM-yoeB* genes were amplified by PCR using oligonucleotides 8/10 and 8/13, respectively, and cloned separately between NdeI and XhoI restriction ezyme sites in the pET-22b(+) overexpression vector (Novagen) to produce proteins C-terminally-tagged with a hexahistidine motif. The *yefM* gene was also amplified using oligonucleotides 8/9 and cloned in pET-16b for production of an N-terminally-tagged derivative and, using oligonucleotides 8/11, as an NdeI–SapI fragment in pTYB1 (New England Biolabs) to generate an intein fusion protein. The *yefM-yoeB* cassette was also amplified with oligonucleotides 8/12 for insertion in pET-16b. His-tagged proteins were overproduced in *E.coli* BL21(DE3) and purified by Ni^{2+} affinity chromatography essentially according to the Novagen technical manual. Protein concentrations in samples containing purified YefM–YoeB complex were estimated using a 2:1 ratio of YefM:YoeB (21).

The YefM–intein fusion protein was overproduced and the intein tag cleaved as follows. 300 ml of strain BL21 harbouring the expression plasmid was grown at 37°C until

$OD_{600} \approx 0.8$, expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and shifted to 25°C, and growth continued overnight. Cells were harvested at 1250 *g* at 4°C for 10 min. The pellet was resuspended in 10 ml of

Table 1. Oligonucleotides used in this study

Oligonucleotide	Sequence (5′–3′) (length [nt])[a]
1	GTCGAGAATTCTCTACAAACTAATTAATAAATA-GTTAATTAACGCTCATCATTGTACAATGAACT-GTACAAAAGAGGAGATTGAC**ATG**GGATCCAG-TGC (99)
2	GCACTGGATCCCATGTCAATCTCCTCTTTTGTAC-AGTTCATTGTACAATGATGAGCGTTAATTAAC-TATTTATTAATTAGTTTGTAGAGAATTCTC-GAC (99)
3	GTCGAGAATTCTCTACAAACTAATTAATAAATA-GTTAATTAACGCTCATCATTGTACAATGAAAT-CGCTCAAAGAGGAGATTGAC**ATG**GGATCCAG-TGC (99)
4	GCACTGGATCCCATGTCAATCTCCTCTTTGAGC-GATTTCATTGTACAATGATGAGCGTTAATTAA-CTATTTATTAATTAGTTTGTAGAGAATTCTCG-AC (99)
5	CCTCGAGCTCATCATTGTACAATGAACTG (29)
6	TCCCAAGCTTCCTCAGACCAGATTAGTTTC (30)
7	TCCCAAGCTTAGATCTATAAGGCTACGCTA-GC (32)
8	GCGATACAT**ATG**CGTACAATTAGCTACAGCG-AA (33)
9	GCCTCGAGATTAGTT**TCA**CTCAATGATGTCC (31)
10	ATACTCGAGCTCAATGATGTCCTTTTCCGTT-CC (33)
11	GGTTGCTCTTCCGCACTCAATGATGTCCTTTTCC-GTTTC (38)
12	CCACTCGAGT**TCA**ATAATGATAACGACATGCT-GC (34)
13	CCACTCGAGATAATGATAACGACATGCTGC (30)
14	CGCGGGAATTCGGGGAAAGGAGGGGG (26)
15	CGGTCGGATCCTGACGCGCTTCGCTG (26)

[a]Restriction enzyme recognition sites are underlined, and translation start and stop codons are highlighted in bold.

Table 2. Plasmids used in this study

Plasmid	Description	Reference
pRS415	Vector for generating transcriptional fusion to *lacZ*	(34)
pBAD33	Arabinose-inducible expression vector	(35)
pET16b	IPTG-inducible expression vector allowing fusion of a N-terminal His$_{10}$ tag to a target protein	Novagen
pET22b(+)	IPTG-inducible expression vector allowing fusion of a C-terminal His$_6$ tag to a target protein	Novagen
pTYB1	IPTG-inducible expression vector allowing fusion of a bifunctional tag, consisting of an intein and a chitin binding domain, to the C-terminus of a target protein	New England Biolabs
pRSyy_wt	*yefM-yoeB* promoter-operator region cloned as annealed oligonucleotides 1/2 between BamHI and EcoRI restriction sites of pRS415	This study
pRSyy_Smut	*yefM-yoeB* promoter-operator region with mutagenized S repeat cloned as annealed oligonucleotides 3/4 between BamHI and EcoRI restriction sites of pRS415	This study
pBAD*yefM*	*yefM* amplified with oligonucleotides 5/6, digested with XhoI–HindII and cloned between the equivalent sites in pBAD33	This study
pBAD*yefMyoeB*	*yefM-yoeB* amplified with oligonucleotides 5/7, digested with XhoI–HindII and cloned between the equivalent sites in pBAD33	This study
pET16*yefM*	*yefM* amplified with oligonucleotides 8/9, digested with NdeI–XhoI and cloned between the equivalent sites in pET16b	This study
pET22*yefM*	*yefM* amplified with oligonucleotides 8/10, digested with NdeI–XhoI and cloned between the equivalent sites in pET22b(+)	This study
pTYB*yefM*	*yefM* amplified with oligonucleotides 8/11, digested with NdeI–SapI and cloned between the equivalent sites in pTYB1	This study
pET16*yefMyoeB*	*yefM-yoeB* amplified with oligonucleotides 8/12, digested with NdeI–XhoI and cloned between the equivalent sites in pET16b	This study
pET22*yefMyoeB*	*yefM-yoeB* amplified with oligonucleotides 8/13, digested with NdeI–XhoI and cloned between the equivalent sites in pET22b(+)	This study

binding buffer A [20 mM Tris–HCl (pH 8.5), 500 mM NaCl and 1 mM EDTA], the cells were sonicated, and then centrifuged for 1 h at 25 000 *g* at 4°C. The supernatant was applied to a column containing 5 ml of chitin resin (New England Biolabs) and equilibrated with buffer A. Binding of the fusion protein to the chitin resin was allowed to continue for 3 h at 4°C after which the column was washed with 80 ml of buffer A, and then quickly flushed with 9 ml of buffer A with 50 mM DTT. The column was closed to allow cleavage of the intein tag for 21 h at 4°C. Elution was performed with 7 ml of buffer A and 1 ml fractions were collected. Fractions containing native YefM protein were combined and dialysed against 1 l of storage buffer [50 mM Tris–HCl (pH 8.5), 150 mM NaCl and 10% glycerol], and then aliquoted and stored at −80°C.

For purification of untagged YoeB, 300 ml of *E.coli* BL21 harbouring a pET16b plasmid producing His$_{10}$-YefM–YoeB were grown at 37°C until $OD_{600} \approx 0.8$. Expression of the complex was induced with 1 mM IPTG and incubation continued for 3 h. Cells were harvested at 1250 *g* at 4°C for 10 min. The pellet was resuspended in 10 ml of buffer A with lysozyme (0.1 mg/ml) and phenylmethylsulfonyl fluoride (PMSF) (1 mM), the cells were sonicated, and then centrifuged for 1 h at 25 000 *g* at 4°C. The supernatant was applied to a column consisting of 3 ml of His-tag resin (Novagen) equilibrated with buffer B [20 mM Tris–HCl (pH 8.0), 10 mM imidazole and 500 mM NaCl]. Binding of His$_{10}$-YefM–YoeB to the resin was continued for 1–2 h at 4°C. The column was washed with 60 ml of buffer B, and then with 50 ml of wash buffer C [20 mM Tris–HCl (pH 8.0), 100 mM imidazole and 500 mM NaCl]. Denaturation and elution of YoeB from the column was performed with 10 ml of buffer D [20 mM Tris–HCl (pH 8.0), 6 M guanidine-hydrochloride, 10 mM imidazole and 500 mM NaCl]. 1 ml fractions were collected, and fractions containing the highest concentrations of denatured YoeB were pooled and dialysed successively for 2 h against 500 ml volumes of 20 mM

sodium acetate (pH 5.5), 200 mM NaCl, 1 mM DTT, 10% glycerol containing 3, 2 or 1 M urea, followed by buffer without urea. The renatured YoeB sample was then dialysed again for 16 h against buffer without urea, aliquoted and stored at $-80°C$. For purification of YoeB-His$_6$, 300 ml of *E.coli* BL21 harbouring a pET22b(+) plasmid producing YefM–YoeB–His$_6$ were grown at $37°C$ until $OD_{600} \approx 0.8$ and subsequent steps followed the procedure described for purification of untagged YoeB with the following differences: first, buffer C contained 50 mM, instead of 100 mM, imidazole. Second, untagged YefM was denatured and eluted with buffer D, followed by YoeB-His$_6$ with buffer D that contained 200 mM imidazole. YoeB-His$_6$ was renatured by dialysis into successively more dilute concentrations of urea as described for untagged YoeB. Refolding of denatured proteins was monitored by circular dichroism (CD).

Electrophoretic mobility shift assays (EMSA)

DNA substrates consisted of 5′-biotinylated, double-stranded oligonucleotides 1/2 (Table 1) that included the *yefM* start codon and 74 bp of the *yefM-yoeB* regulatory region. Oligonucleotides 3/4 consist of the same sequence, but with mutations in the S repeat. Reactions containing 0.1 nM of biotin-labelled DNA and the protein concentrations indicated in the legend to Figure 2 were assembled in binding buffer [10 mM Tris–HCl (pH 7.5), 50 mM NaCl, 1 mM DTT, 5 mM MgCl$_2$, 1 μg of poly(dI·dC), 2.5% glycerol] in final volumes of 20 μl and incubated for 20 min at $22°C$. For YefM–YoeB reconstitution experiments, the two untagged proteins were first coincubated for 20 min prior to adding DNA. Samples were electrophoresed on 6% native polyacrylamide gels in 0.5× TBE buffer for 90 min at 80 V at $22°C$. DNA was transferred by capillary action or electroblotting to positively-charged nylon membranes (Roche), and the transferred DNA fragments were immobilized onto the membrane by ultraviolet (UV) cross-linking. Detection of the biotin end-labelled DNA was performed using the LightShift™ chemiluminescent EMSA kit (Pierce).

DNase I footprinting

220 bp PCR fragments in which either the top or bottom strand was 5′ biotinylated were generated with oligonucleotides 14/15, and were electrophoresed on 7.5% polyacrylamide gels with 7.5% glycerol in 1× TBE buffer. Fragments were excised from gels and electroeluted in 0.1× TBE for 30 min at 100 V. The DNA was extracted with phenol:chloroform (1:1) and precipitated with ethanol. DNA was harvested by centrifugation, and the pellets dried and resuspended in sterile water. Reactions containing 2 nM of biotin-labelled DNA and YefM or YefM–YoeB–His$_6$ protein at concentrations indicated in the legend to Figure 3 were mixed in binding buffer [20 mM Tris–HCl (pH 7.5), 50 mM NaCl, 5 mM MgCl$_2$, 1 mM DTT, 5% glycerol, 1 μg of poly(dI·dC) and 10 μg of BSA] in a final volume of 20 μl and incubated for 20 min at $22°C$. Each reaction was treated with 0.0075 U of DNase I (Roche, RNase free, 10 U/μl diluted in buffer [20 mM Tris–HCl (pH 7.5), 50 mM NaCl, 7.5 mM MgCl$_2$ and 5 mM CaCl$_2$]) for 45 s at $22°C$. Reactions were stopped by addition of 200 μl of stop solution (10 mM EDTA and 300 mM sodium acetate) followed by extraction with an

equal volume of phenol:chloroform (1:1). The upper phase was collected and 1 μl of glycogen (Roche) and 500 μl of ethanol were added. Samples were precipitated at $-80°C$ for 30 min, harvested by centrifugation, and the pellets washed with 70% ethanol. Pellets were dried and resuspended in 10 μl of loading buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol). Samples were heated at $99°C$ for 10 min and loaded on 6% sequencing gels (SequaGel; GeneFlow), which were pre-run for at least 100 min. Samples were electrophoresed at 60 W in 1× TBE buffer for 2 h for fragments in which the bottom strand was biotin-labelled, or for 3 h for fragments in which the top-strand was biotin-labelled. DNA was transferred by capillary action to positively-charged nylon membranes (Roche), and the transferred DNA fragments were immobilized onto the membrane by UV cross-linking. Detection of the biotin end-labelled DNA was performed using the LightShift™ chemiluminescent EMSA kit (Pierce).

Maxam–Gilbert sequencing

A total of 30 nM of DNA was diluted in water to a final volume of 12 μl, and 50 μl of formic acid was added. The sample was incubated for 2.5 min at $22°C$; the reaction stopped by adding 200 μl of 300 mM sodium acetate (pH 7.0) and 700 μl of ice-cold ethanol and precipitated for 15 min at $-80°C$. The DNA was harvested by centrifugation for 15 min at $4°C$. The pellet was washed three times with 70% ethanol, each time centrifuging for 10 min at $4°C$ and discarding the supernatant. The pellet was dried, resuspended in 100 μl of a 1 M solution of piperidine and incubated for 30 min at $90°C$. A total of 10 μl of 3 M sodium acetate (pH 7.0) and 300 μl of ice-cold ethanol were added and incubated for 20 min at $-80°C$. The pellet was washed twice with 1 ml of 70% ethanol, dried and resuspended in 20 μl of loading buffer. Sequencing reactions were analysed on 6% sequencing gels alongside the corresponding DNase I footprinting reactions.

Primer extension analysis

Total cellular RNA from strain DH5α harbouring a pRS415-based plasmid possessing a transcriptional fusion of the *yefM-yoeB* promoter-operator region to the *lac* operon (pRSyy_wt), and primer extension mapping were performed essentially as described previously (36) using 5′-biotinylated oligonucleotide 15 (Table 1).

Assays of β-galactosidase activity

Strain SC301467 harbouring the pRS415 plasmid with a *lacZ* gene under transcriptional control of the *yefM-yoeB* promoter (pRSyy_wt), or with mutations in the S repeat (pRSyy_Smut), was cotransformed with pBAD33 plasmids encoding *yefM* or *yefM-yoeB* genes under control of an arabinose-inducible promoter (pBAD*yefM* and pBAD*yefMyoeB*). At $OD_{600} \approx 0.2$, synthesis of antitoxin or TA was induced by addition of 0.2% arabinose for 1 h. β-Galactosidase assays were performed with cells permeabilized with chloroform and SDS as described by Miller (37). Plasmids pBAD*yefM* and pBAD*yefMyoeB* were generated using oligonucleotides 5/6 and 5/7 in amplification of the *yefM* and *yefM-yoeB*

genes, respectively, for cloning as HindIII–XhoI fragments in the equivalent sites in pBAD33. Oligonucleotides 1/2 and 3/4 were annealed, digested with EcoRI–BamHI and inserted in pRS415 to produce pRSyy_wt and pRSyy_Smut, respectively.

CD spectroscopy

YefM or YefM–YoeB–His$_6$ were buffer-exchanged using Microcon 3 kDa cut off filters (Millipore) into 20 mM Tris–HCl (pH 8.5), 50 mM NaCl. For CD, YefM and YefM–YoeB–His$_6$ were used at concentrations of 10 and 5 μM, respectively. 99 bp double-stranded oligonucleotides 1/2 and 3/4 were used at concentrations of 2 μM. CD scans were performed in a Jasco J-810 spectropolarimeter at 20 nm/min with a 0.2 nm data pitch and 1 s response using a 1 mm band width for eight accumulations at 20°C. Temperature scans were performed at 222 nm, with a temperature change of 1°C/min from 5 to 80°C. Following a 1 s rest interval, reverse scans were performed at the same speed.

Nuclear magnetic resonance (NMR) spectroscopy

Proton NMR spectra of YefM were recorded at 30°C on a Bruker Avance DRX 600 MHz spectrometer equipped with a CryoProbe, in 50 mM Tris (pH 8.5), 150 mM NaCl in 95/5% $^1H_2O/^2H_2O$. Spectral data were processed using TopSpin (Bruker). Standard pulse sequences, with WATER-GATE water suppression, were used.

Chemical cross-linking

Dimethyl pimelimidate (DMP) (Sigma) was added to reactions at a final concentration of 10 mM. Proteins were diluted to 20 μM (YefM) or 14 μM (YefM–YoeB) in buffer [20 mM HEPES-NaOH (pH 8.5), 50 mM NaCl and 5 mM MgCl$_2$]. The final reaction volume was 20 μl. Reactions were incubated at 22°C as indicated in Figure 7, and stopped by the addition of 1 μl of 0.5 M Tris–HCl (pH 6.8) followed by 2× SDS loading buffer. The samples were heated at 95°C for 5 min and analysed by SDS–PAGE (15% polyacrylamide).

Hydrodynamic properties

Molecular mass and hydrodynamic radius of YefM were determined using a combination of size exclusion chromatography, multi-angle light scattering (MALS) and quasi-elastic light scattering (QELS). YefM [in 10 mM Tris and 100 mM NaCl (pH 8.5)] was applied to a Superdex 75 10/30 column that had been pre-equilibrated in the same buffer, and eluted at room temperature at a flow rate of 0.710 ml/min. The column was attached downstream to a multiangle laser light (690.0 nm) scattering DAWN EOS photometer (Wyatt). QELS data were collected using a Wyatt-QELS instrument. The concentration of the eluted protein was estimated using values of 0.180 for the refractive index increment (dn/dc) and 1.330 for the solvent for the solvent refractive index. Molecular weights were determined using a Zimm plot. Data were analysed using Astra 4.90.08 software (Wyatt) as recommended by the manufacturer.

Bioinformatics

YefM–YoeB homologs were identified in the National Center for Biotechnology Information (NCBI) database using PSI- and PHI-BLAST searches (38). The nucleotide sequences of the regions upstream of the corresponding genes were aligned using ClustalW (39). GlobPlot (40) was used to predict intrinsic disorder in YefM.

RESULTS

YefM and YoeB are the transcriptional repressor and corepressor, respectively, of the *yefM-yoeB* genes

During *in vitro* transcriptional analysis of the histidine biosynthesis (*his*) genes in *E.coli* K-12, a cryptic transcript was detected from a promoter orientated divergently to the *his* operon promoter (41,42). In hindsight, this transcript is likely to be derived from the recently identified *yefM-yoeB* cassette, which is located upstream of the *his* genes, but which apparently is transcribed in the opposite direction (19). Sequencing of the *in vitro* transcript mapped its 5′ end to position -3 of the *yefM-yoeB* promoter region illustrated in Figure 1A (41). However, primer extension analysis performed here unambiguously delineated the *in vivo* transcription start point to a position three nucleotides 3′ of that previously assessed (Figure 1B). Sequences with close matches to consensus −10 and −35 hexamer promoter boxes, and which are separated by an optimal 17 bp, are located 5′ of this transcription start point (Figure 1A). The discrepancy between the primer extension analysis here and previous transcript sequencing experiments might reflect differences in transcription start site selectivity *in vivo* and *in vitro*, or 5′ end processing of the *in vivo* transcript which is absent *in vitro*. Nevertheless, both datasets strongly indicate that the first position in the *yefM* translational start codon is at position +21, and not at another potential start codon at −7 as suggested previously (24) (Figure 1A). The region 5′ of position +21 also possesses a sequence that resembles a ribosome binding site separated by a window of 6 bp from the putative *yefM* ATG initiation codon. These observations agree with findings that overexpressed *yefM* which includes the potential start codon at −7 to −5 produces a translation product that commences at position +21 (21).

A 99 bp fragment encompassing the *yefM-yoeB* promoter and *yefM* start codon was inserted upstream of a promoterless *lac* operon in the transcription fusion vector pRS415. This fusion produced 4099 ± 547 U of β-galactosidase activity in strain SC301467, which is deleted of chromosomal *yefM-yoeB* genes (22), whereas pRS415 alone produced <100 units. Thus, the region 5′ of *yefM-yoeB* possesses substantial promoter activity. Subsequently, YefM protein was provided in trans from an arabinose-inducible promoter, and its effect on β-galactosidase production by the *yefM–lacZ* fusion was examined: β-galactosidase levels were reduced ∼5.5-fold to 739 ± 164 U in the presence of YefM (Figure 1C). Moreover, coexpression of *yefM-yoeB* in trans further reduced β-galacosidase levels to background levels (126 ± 33 U). Thus, YefM is a transcriptional autorepressor of the *yefM-yoeB* promoter, with YoeB acting as a corepressor.

YefM and YefM–YoeB recognize DNA palindromes with common core sequences

Three versions of YefM were purified to >95% homogeneity and tested in EMSA with a 99 bp double-stranded
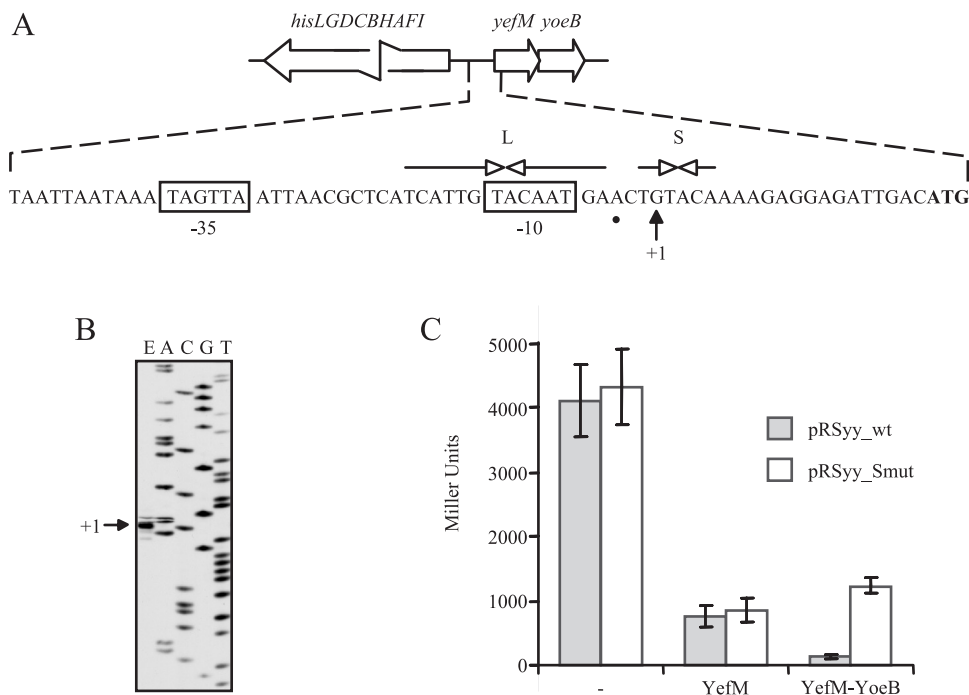
Figure 1. Autoregulation of the *yefM-yoeB* module. (**A**) Nucleotide sequence of the region 5′ of *yefM-yoeB*. The transcription start-site mapped by primer extension is marked as +1. Hexameric promoter motifs are boxed and the *yefM* start codon is in bold. The 5′ end of the *yefM-yoeB* transcript previously determined by RNA sequencing (41,42) is indicated by the filled circle. Long (L) and short (S) palindromes recognized by YefM-YoeB are denoted by inverted arrows. (**B**) Primer extension analysis of the *yefM-yoeB* module. Total RNA from *E.coli* DH5α harbouring a plasmid possessing the *yefM-yoeB* operon was subjected to primer extension analysis (E) using a 5′-biotinylated primer that anneals within flanking vector sequences. Reactions were performed and analysed as outlined in Materials and Methods, and electrophoresed on a denaturing 6% polyacrylamide gel in parallel with nucleotide sequencing reactions (A, C, G, T) carried out with the same biotinylated primer. The major product from the primer extension is marked as +1. (**C**) Autoregulation of *yefM-yoeB* expression by YefM and YefM-YoeB. A transcriptional fusion of the *yefM-yoeB* regulatory region to the *lacZYA* operon in plasmid pRS415 (pRSyy_wt) was transformed into *E.coli* SC301467, which is deleted of five chromosomal TA cassettes including *yefM-yoeB*, and β-galactosidase levels determined without YefM, and with YefM (pBAD*yefM*) or YefM-YoeB (pBAD*yefMyoeB*) supplied *in trans* from the pBAD33 arabinose-inducible vector (filled columns). Similar experiments were performed with a transcriptional fusion in which the S palindrome was mutated (pRSyy_Smut) (open columns).

oligonucleotide substrate encompassing the *yefM-yoeB* promoter region: native YefM isolated following cleavage from a fusion of the C-terminus of the protein to the N-terminus of an intein tag; YefM with a hexahistidine tag at its C-terminus (YefM-His$_6$); and YefM with a 21 amino acid extension, including a decahistidine tag, at its N-terminus (His$_{10}$-YefM). Native YefM weakly bound the 99 bp oligonucleotide, producing a single-retarded complex in EMSA (Figure 2A). Although complex formation apparently was inefficient, the interaction of YefM with the *yefM-yoeB* promoter region was specific as no retarded species were observed in EMSA with an unrelated substrate. Similar DNA binding patterns were observed with YefM-His$_6$ (data not shown) indicating that the C-terminal tag does not detectably impede interaction of YefM with DNA. In contrast, His$_{10}$-YefM produced a smear of higher molecular weight complexes with the 99 bp oligonucleotide suggesting that the N-terminal tag significantly altered the association of YefM with DNA (data not shown). This observation correlates with the suggestion that a conserved basic patch in the N-terminal domain of YefM might be involved in DNA recognition (21).

The 3′ end of *yefM* overlaps the 5′ of *yoeB* by 1 bp. These overlapping genes were cloned in an expression vector to permit purification of a YefM–YoeB–His$_6$ complex in which the two proteins are present at a physiological ratio. The

YefM-YoeB-His$_6$ complex produced a major retarded species at >200-fold lower protein concentrations than that generated by YefM alone (Figure 2B). Furthermore, at YefM–YoeB concentrations >12 nM, >90% of the substrate was bound into nucleoprotein complex(es), in contrast with YefM alone which retarded <1% of the substrate at this concentration. The nucleoprotein complex formed by YefM–YoeB–His$_6$ migrated slightly more slowly in EMSA than that produced by YefM suggesting that the former incorporates a greater number of protein protomers than the latter and/or that the conformation of the DNA in the complexes differ.

YoeB protein was purified as a hexahistidine-tagged (YoeB–His$_6$) version from the YefM–YoeB–His$_6$ complex, as well as in native form from the His$_{10}$–YefM–YoeB complex. Neither version of YoeB generated a shifted complex with the *yefM-yoeB* promoter DNA in EMSA revealing that the protein apparently does not bind directly to DNA (Figure 2C). Mixing experiments in which native, untagged YefM and YoeB were preincubated in different ratios and subsequently used in binding experiments with the 99 bp double-stranded oligonucleotide showed that the reconstituted YefM–YoeB complex produced a major retarded species with migration analogous to that observed with the YefM–YoeB–His$_6$ complex. Thus, co-purified and reconstructed YefM–YoeB complexes exhibit similar DNA binding properties. Finally, the optimal YefM:YoeB ratio at
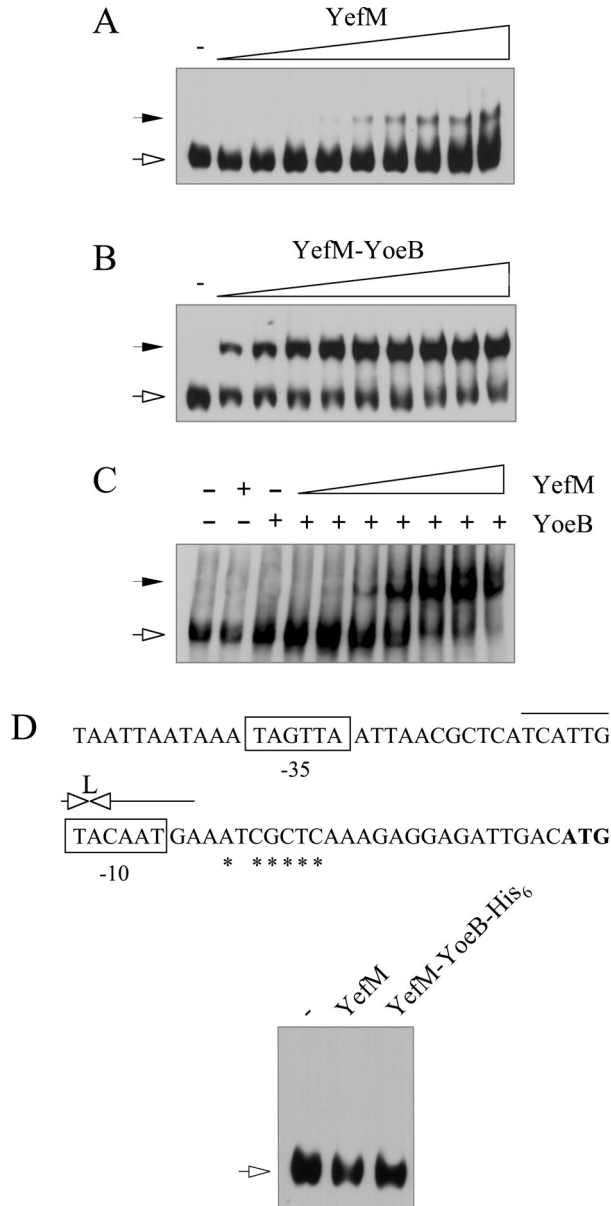
Figure 2. YefM and YefM-YoeB binding to the *yefM-yoeB* promoter-operator region. (**A**) A 99 bp double-stranded oligonucleotide (0.1 nM) that was 5′ biotinylated on one strand and that included the *yefM* translation start codon and 74 bp upstream was subjected to EMSA using increasing amounts of native YefM. YefM concentrations used (left to right) (μM): 0, 0.032, 0.064, 0.125, 0.25, 0.5, 1.0, 2.0, 4.0 and 8.0. Reactions were processed as outlined in Materials and Methods. Open and filled arrows denote positions of unbound oligonucleotide and YefM–DNA complexes, respectively. (**B**) EMSA of the same biotinylated oligonucleotide as in (A) using increasing amounts of YefM–YoeB–His₆. Protein concentrations used (left to right) (μM): 0, 0.003, 0.006, 0.012, 0.024, 0.048, 0.1, 0.2, 0.4 and 0.8. Open and filled arrows denote positions of unbound oligonucleotide and YefM–YoeB–His₆–DNA complexes, respectively. (**C**) EMSA of the same biotinylated oligonucleotide as in (A) with YefM–YoeB reconstituted from individual native proteins. The first three lanes contained no protein, YefM only (1 μM) or YoeB only (1 μM). The following lanes contained YoeB (1 μM) and increasing concentrations of YefM (μM): 0.125, 0.250, 0.5, 1.0, 2.0, 4.0 and 8.0. Open and filled arrows denote positions of unbound oligonucleotide and YefM–YoeB–DNA complexes, respectively. (**D**) Top, nucleotide sequence of the *yefM-yoeB* promoter–operator region with substitution mutations (stars) that disrupt the S palindrome. Bottom, EMSA of a 99 bp oligonucleotide substrate harbouring the S palindrome mutations without added protein, with native YefM (8 μM) and with YefM–YoeB–His₆ (6 μM).

which complete DNA binding was observed was between 1:2 and 2:1 (Figure 2C). This correlates with biophysical and structural studies which demonstrated that purified YefM–YoeB complex consisted of either a 1:2 or 2:1 mix of the two proteins (21,24).

The binding sites for YefM and YefM–YoeB–His₆ in the *yefM-yoeB* promoter region were localized by DNase I footprinting (Figure 3). YefM first specifically protected the region between −17 and −3 on the upper strand from DNase I digestion, with an extended footprint between −2 and +12 at higher protein concentrations. Protection of the bottom strand was staggered by 3 or 4 nt in the 3′ direction with protection initially occurring from −21 to −6 and then between −5 and +8. The observed 3′-stagger potentially reflects minor groove coverage at these ends (43). Position −14 on the lower strand is hypersensitive to DNase I cleavage in the presence of YefM indicating that the DNA structure is perturbed at this point. The extent of protection on both strands by the YefM–YoeB–His₆ complex was indistinguisable from that provided by YefM alone at higher protein concentration. However, protection against DNase I digestion by YefM–YoeB-His₆ occurred at ∼100-fold lower protein concentration than by YefM, assuming a 2:1 ratio of YefM:YoeB. In addition, the enhancement at position −14 on the bottom strand was more pronounced in the presence of YefM–YoeB–His₆ (Figure 3). The DNase I footprinting results agree with EMSA data demonstrating that the *yefM-yoeB* promoter region is bound more avidly by the two-protein complex than by YefM alone.

The YefM and YefM–YoeB–His₆ DNase I footprints cover the −10 hexamer promoter box as well as 3′ and 5′ flanking regions, suggesting that the proteins inhibit transcription by blocking the access of RNA polymerase to the *yefM-yoeB* promoter (44). The primary region protected by YefM against DNase I attack includes a 5′-TCATTGTACAATGA-3′ palindrome (L [long] repeat). The 5′-TGTACA-3′ core of this inverted repeat is also present in the secondary region of YefM protection (S [short] repeat) (Figure 3). Preferential binding of YefM to the L repeat might be facilitated by the additional palindromic nucleotides that flank the hexameric core sequence within this repeat, but that are absent from the S repeat. In summary, the operator site for *yefM-yoeB* autoregulation encompasses a pair of palindromic sequences that possess a common 5′-TGTACA-3′ core with a centre-to-centre distance of 12 bp.

The S repeat plays a crucial role in transcriptional repression and DNA binding by YefM–YoeB

Multiple substitution mutations were introduced into the S repeat in the operator site (Figure 2D) to assess the contribution of this motif to regulation of *yefM-yoeB* expression. Disruption of the S repeat did not detectably perturb either unregulated expression from the resulting *yefM(S′)-lacZ* fusion or repression of this fusion by YefM *in vivo* (Figure 1C). However, the YefM–YoeB complex failed to exert additional downregulation of the *yefM(S′)-lacZ* fusion, which was observed with the wild-type *yefM–lacZ* fusion. Indeed, YefM–YoeB reproducibly repressed the *yefM(S′)–lacZ* fusion slightly less efficiently than did YefM alone. Interactions between YefM–YoeB complexes assembled at
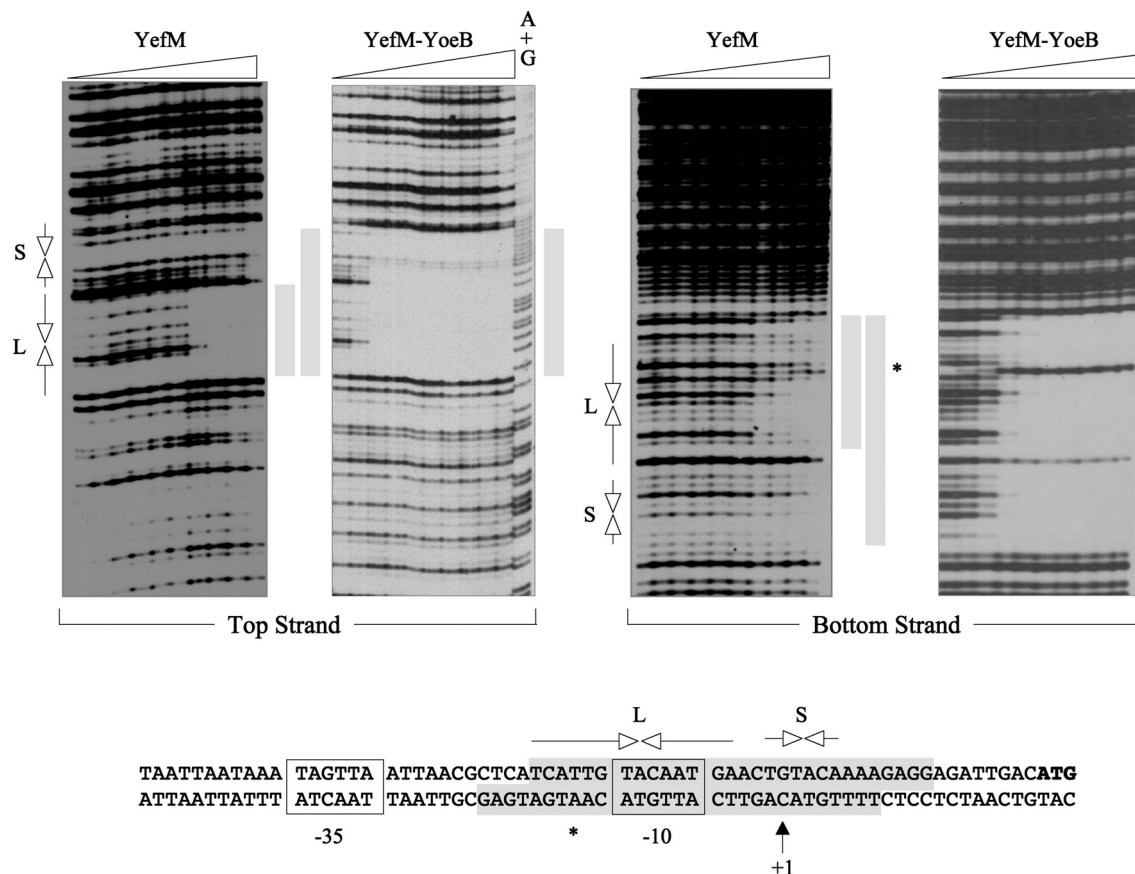
Figure 3. DNase I footprinting of the *yefM-yoeB* promoter-operator region. Footprinting reactions were performed as outlined in Materials and Methods using PCR fragments biotinylated at the 5′ ends of either upper or lower strands. YefM concentrations (µM, left to right): 0, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1.0, 2.5 and 5.0. YefM–YoeB–His$_6$ concentrations (µM, left to right): 0, 0.007, 0.018, 0.036, 0.072, 0.18, 0.36, 0.72, 1.8 and 3.6. The locations of the L and S repeats are marked by inverted arrows. Shaded boxes denote the regions protected from DNase I digestion by YefM and YefM–YoeB–His$_6$. A + G, Maxam–Gilbert sequencing reactions. A position on the lower strand that is hypersenstive to DNase I cleavage in the presence of YefM and YefM–YoeB–His$_6$ is highlighted by the star. The relative dispositions of regions on the upper and lower strands that are protected from DNase I digestion and other features of the *yefM-yoeB* promoter–operator region are illustrated in the lower panel.

the L and S repeats might be sufficiently perturbed by mutation of the S palindrome that cooperativity is weakened resulting in reduced binding of the complex to the palindromes. In accord with the *in vivo* results, both the weak binding by YefM and the more proficient interaction of YefM–YoeB–His$_6$ with a substrate encompassing the *yefM-yoeB* promoter region *in vitro* were abolished with an analogous 99 bp double-stranded oligonucleotide harbouring the S repeat mutations (Figure 2D). In combination, these results emphasize that the S repeat is critical for association of the YefM–YoeB complex with the *yefM-yoeB* operator site, and that its disruption dramatically impairs the interaction of the protein complex with the site.

YefM possesses extensive secondary structure

CD analysis suggested that YefM is an unfolded protein that entirely lacks secondary structure (24). However, native YefM purified here exhibited distinct CD minima at ∼208 and ∼222 nm that are characteristic of α-helix content and helix–helix interactions, respectively, and which indicate that the protein was at least partly folded (Figure 4A). The relatively shallow peak at 222 nm in comparison with

the deeper trough at 208 nm implies that the protein lacks extensive coiled-coil formation (45). Deconvolution of the data with CDSSTR software (46) suggested that the α-helical helical content of YefM was ∼40%, with ∼30% β-strands, 5–10% turns, and the remainder unordered. Analogous results were obtained with His$_{10}$-YefM (data not shown). The recently-described tertiary structure of the YefM–YoeB heterotrimeric complex consists of one YoeB monomer associated with a YefM homodimer. The YefM homodimer is asymmetric with one of the YefM monomers possessing a disordered C-terminal region, whereas the equivalent region of the second monomer is folded (21). The α-helical content of the asymmetric YefM homodimer within the complex was 45%, with 20% β-strands, 18% turns and 17% disordered: these values are not dissimilar to those obtained from CD studies of YefM here. In the absence of high-resolution information about the tertiary structure of free YefM, it remains to be clarified whether YefM undergoes structural transitions between its free state and when associated with YoeB, although this seems likely.

In view of the inconsistency between the CD spectra for YefM observed here (Figure 4A) and those described recently (24), NMR spectra were acquired to provide another
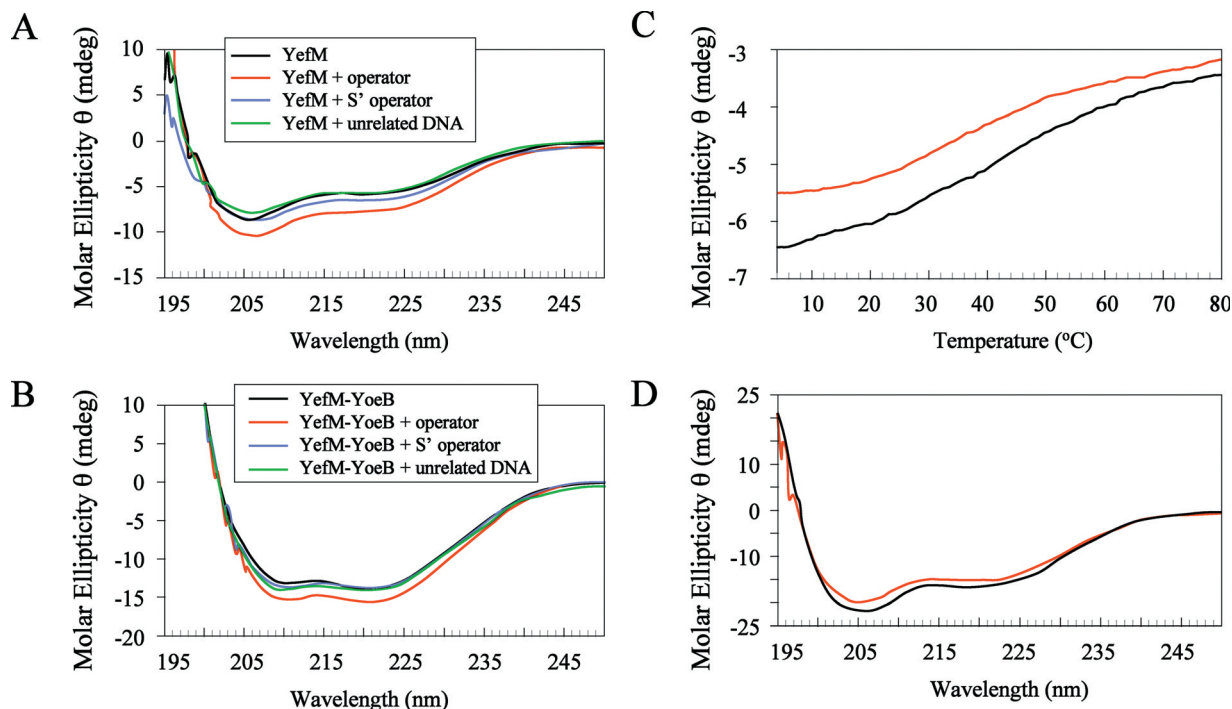
Figure 4. CD analysis of the YefM and YefM–YoeB–His$_6$ proteins in the absence and presence of DNA. (**A**) Far UV CD spectrum of YefM alone (10 µM), and in the presence of 2 µM 99 bp oligonucleotides with the *yefM-yoeB* promoter–operator region, the same region but with mutations in the S palindrome (Figure 2D), and without the promoter–operator sequences. Spectra of complexes are difference spectra as any contribution of the oligonucleotides to the spectra is subtracted from the spectrum of the complex. (**B**) Far UV CD spectrum of YefM–YoeB–His$_6$ alone (5 µM), and in the presence of 99 bp oligonucleotides with the *yefM-yoeB* promoter–operator region, the same region but with mutations in the S palindrome (Figure 2D), and without the promoter–operator sequences. (**C**) Thermal denaturation of native YefM (10 µM) monitored by CD at 222 nm. Denaturation of YefM was followed from 5–80°C (black line). YefM renaturation was subsequently analysed (red line). (**D**) Far UV spectra of native YefM (20 µM) before (black line) and after (red line) thermal denaturation.
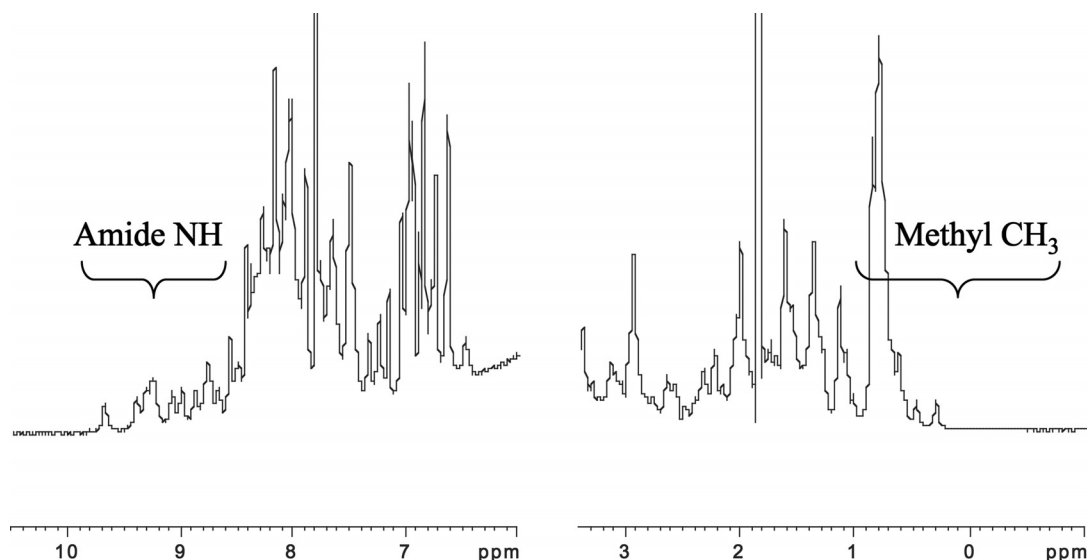

Figure 5. The 1D proton spectrum of YefM (150 µM) in 50 mM Tris (pH 8.5), 150 mM NaCl at 30°C. Regions of resolved methyl and amide resonances are highlighted.

assessment of the protein's conformational state. The 1D spectrum of YefM-His$_6$ is shown in Figure 5 and the 2D NOESY in Figure 6. Close examination of the spectra reveals a mixture of resonances in which well-dispersed signals are superimposed upon resonances that are clustered together; these characteristics are clearly seen in the amide and methyl

proton regions (Figure 5). In addition, resonances are also observed in the region just downfield of the water resonance, these being typically from residues in β-strands. The 2D NOESY spectra showed cross-peaks which are indicative of secondary structures: cross-peaks between amide protons suggest the presence of helices, cross peaks between amide
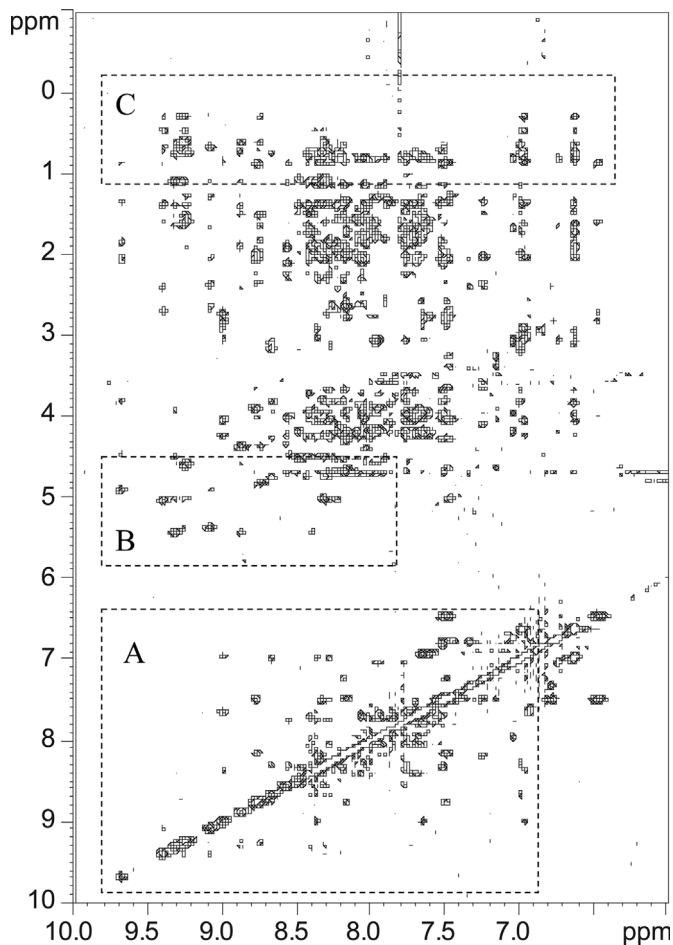
Figure 6. A section of the 600 MHz 2D ^{1}H-^{1}H NOESY spectrum (τ_m 100 ms) of YefM (150 µM) in 50 mM Tris (pH 8.5), 150 mM NaCl at 30°C. Regions containing cross peaks from regular secondary and tertiary structures are shown: (**A**) NH-NH, (**B**) NH-αH and (**C**) aromatic/NH-methyl protons.

protons and alpha protons downfield of 5 p.p.m. (Figure 6) indicate the presence of β-strands. The NOESY peaks between the upfield methyl protons and downfield amide and aromatic protons additionally suggest the presence of tertiary structures. Therefore, the NMR data show that the recombinant YefM-His$_6$ used in the present studies has secondary and tertiary structures. A series of spectra were collected at increasing temperatures from 15 to 30°C. Apart from systematic shifts due to temperature variation and a general decreased in linewidths resulting from the faster tumbling of the protein in solution at higher temparatures, the spectra of YefM-His$_6$ remained essentially the same, with little signs of thermal unfolding; this is further evidence that YefM-His$_6$ adopts a somewhat stable conformation within this temperature range. Overall, the NMR data suggests that the YefM protein exists either as a mixture of folded and unfolded states, or that the protein has folded and unfolded regions in its tertiary structure.

Further investigations into these options were carried out using CD by examination of the temperature dependence of ellipticity at 222 nm in 0.2°C steps over the range 5–80°C at a heating rate of 1°C/min (Figure 4C). The protein showed a gradual melting transition and dimunition of the CD signal

indicating a non-cooperative unfolding process, again reflective either of a mobile, partially unfolded protein and/or that YefM exists in a variety of folded states. The midpoint of the denaturation curve revealed a melting temperature (T_m) of ~45°C. No visible precipitation of the protein sample was apparent at 80°C. When the sample was cooled gradually to 5°C and monitored by CD ellipticity at 222 nm, a pattern very similar in reverse to that of the denaturation profile was observed (Figure 4C). Moreover, the CD spectra of native and renatured YefM were virtually indistinguishable indicating that heat denaturation of native YefM is fully reversible (Figure 4D). The CD spectrum of YefM–YoeB–His$_6$ observed here (Figure 4B) is similar to that described previously (20), as is the complex's thermal denaturation curve over the range 5–80°C which confirms a pronounced T_m at ~60°C, and lack of renaturation following heating to 80°C (data not shown).

CD spectra in the near- and far-UV regions can be used to discriminate between DNA and protein within a nucleoprotein complex as the far UV region of the spectrum is dominated by contributions of amide moieties from the peptide backbone, whereas secondary structure alterations in nucleic acids upon formation of the nucleoprotein complex are evident in the non-overlapping region from 240 to 320 nm. This distinction was used to assess whether YefM or YefM-YoeB-His$_6$ underwent detectable structural changes when bound to operator DNA, and vice versa. To allow comparison of protein spectra free and in the DNA bound state, the contribution of the DNA was subtracted from the curves. Conversely, any contribution of the protein in the 240–320 nm region was subtracted separately to allow assessment of spectral alterations indicative of changes in DNA when complexed with protein. Using a 5:1 molar concentration of YefM and a 99 bp DNA fragment encompassing the operator site, the CD minima at 208 and 222 nm became more pronounced than in the absence of DNA (Figure 4A). An equivalent DNA fragment harbouring mutations in the S repeat (Figure 2D) induced less profound alterations in the far UV region, whereas YefM spectra in the presence and absence of a DNA fragment without a YefM recognition site were indistinguishable. The results suggest that YefM undergoes structural transitions when bound to DNA, and that the alterations are induced specifically by its cognate binding site. Analogous CD results were noted for YefM-YoeB-His$_6$ in the absence and presence of the three different DNA fragments, except that the fragment containing S repeat mutations elicited a weaker response with YefM-YoeB-His$_6$ than with YefM (Figure 4B).

B-form DNA presents a typical positive CD maximum centred at 275 nm, a minimum near 245 nm, with a zero point transition around 258 nm (47). The *yefM-yoeB* operator DNA showed no significant alterations in these characteristics in the presence YefM or YefM-YoeB-His$_6$ (data not shown), suggesting that the operator site does not undergo major structural transitions when bound by its cognate proteins.

YefM is dimeric in solution

Cross-linking experiments with DMP were performed to characterize the oligomeric state of native YefM (9.3 kDa).
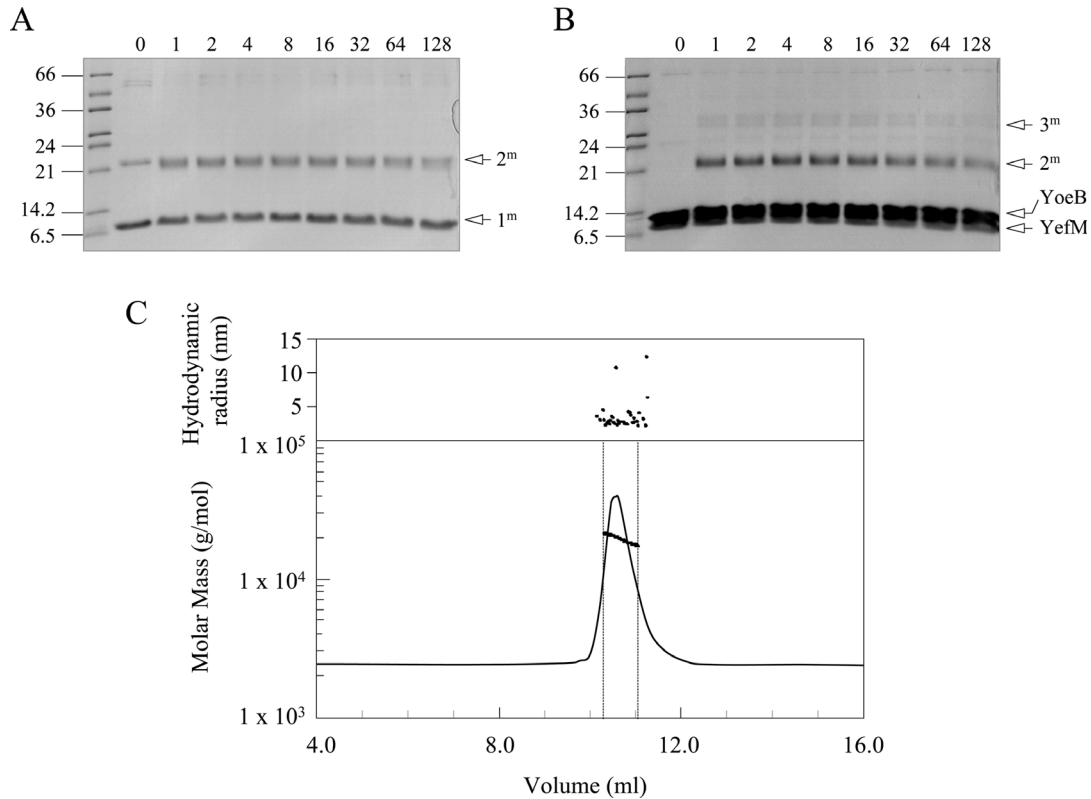
Figure 7. Analysis of the solution oligomeric state of YefM. (**A**) Timecourse (minutes) of YefM cross-linking performed at 22°C in the presence of DMP (10 mM). Samples were electrophoresed on a 15% SDS–polyacrylamide gel. Different species formed are indicated by arrows (right) and molecular marker weights are expressed in kDa (left). (**B**) Timecourse (minutes) of YefM–YoeB–His$_6$ cross-linking performed at 22°C in the presence of DMP (10 mM). Samples were electrophoresed on a 15% SDS–polyacrylamide gel. Different species formed are indicated by arrows (right) and molecular marker weights are expressed in kDa (left). (**C**) Molar mass distribution of YefM. Bottom, the solid line is the trace from the refractive index indicator. Peak area selected for analysis is between the two dashed lines. Dots within the peak area are the weight average molecular weights for each slice, i.e. measured every second. Top, hydrodynamic radius (nm) versus volume (ml) across the peak.

Although predominantly monomeric, a species with the molecular mass of a dimer was frequently evident in untreated samples analysed by SDS–PAGE. Moreover, a significant fraction of YefM was rapidly fixed into covalently bound dimers with 10 mM DMP at 22°C (Figure 7A). Similar results were noted with DMP concentrations as low as 0.1 mM (data not shown). These results demonstrate that YefM forms dimers in solution. In the case of YefM-YoeB-His$_6$ (29.9 kDa), DMP cross-linking produced a ~22 kDa species that is likely to be dimeric YefM, and a doublet at ~30 kDa that correlates with a trimeric YefM–YoeB–His$_6$ complex (Figure 7B), as noted previously (20,21).

YefM eluted predominantly (>98%) as a single peak in size exclusion chromatography on a Superdex 75 10/30 column. MALS of this peak material was consistent with the presence of a major dimeric species: the molecular weight distribution across the peak area was 19.70 ± 0.79 kDa (Figure 7C). Thus, both cross-linking and MALS studies conclusively demonstrated that YefM is a dimer in solution. Moreover, the hydrodynamic radius of native YefM determined by QELS performed in parallel with MALS was 2.6 nm (Figure 7C). This is relatively elongated for a molecule of this size suggesting that unstructured regions might contribute to the protein's extended conformation, in agreement with CD (Figure 4) and NMR (Figures 5 and 6) results.

Paired L and S palindromes in *yefM-yoeB* regulatory regions in diverse genomes

Homologues of *yefM-yoeB* are widely disseminated in bacteria (19). To assess whether L and S palindromes might also be implicated in transcriptional autoregulation of these homologues, the regions upstream of the cassettes in diverse genomes were scrutinized for the presence of 5′-TGTACA-3′ motifs with a centre-to-centre distance of 12 bp, as in *E.coli* K-12. Many *yefM-yoeB* genes were accompanied by paired 5′-TGTACA-3′ boxes within 80 bp of the *yefM* translational start codon (Figure 8). Palindromicity in the more 5′ motif often extended beyond the core hexamer sequence, analogous to the L repeat in strain K-12. Although the paired hexamer boxes were consistently separated by 6 bp, the sequences of the spacers were relatively diverse, as were both the sequences between the S repeat and the translational start, and 5′ of the L repeat. Although paired L and S palindromes are embedded within the regulatory regions of numerous *yefM-yoeB* operons, many of the genomes that were analysed did not harbor 5′-TGTACA-3′ motifs upstream of their *yefM-yoeB* genes suggesting that YefM–YoeB complexes in these bacteria recognize different regulatory motifs, or that regulation occurs by mechanisms that do not involve YefM–YoeB.
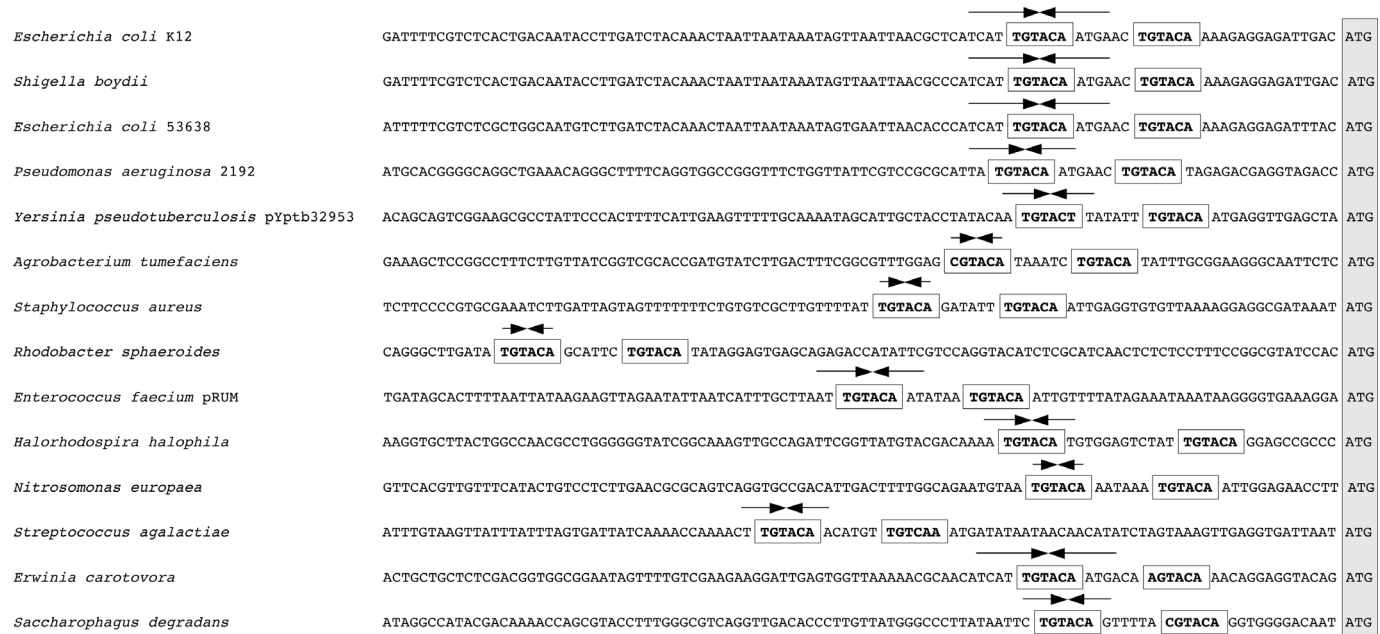
Escherichia coli K12	GATTTTCGTCTCACTGACAATACCTTGATCTACAAACTAATTAATAAATAGTTAATTAACGCTCATCAT TGTACA ATGAAC TGTACA AAAGAGGAGATTGAC ATG
Shigella boydii	GATTTTCGTCTCACTGACAATACCTTGATCTACAAACTAATTAATAAATAGTTAATTAACGCCCATCAT TGTACA ATGAAC TGTACA AAAGAGGAGATTGAC ATG
Escherichia coli 53638	ATTTTTCGTCTCGCTGGCAATGTCTTGATCTACAAACTAATTAATAAATAGTGAATTAACACCCATCAT TGTACA ATGAAC TGTACA AAAGAGGAGATTTAC ATG
Pseudomonas aeruginosa 2192	ATGCACGGGGCAGGCTGAAACAGGGCTTTTCAGGTGGCCGGGTTTCTGGTTATTCGTCCGCGCATTA TGTACA ATGAAC TGTACA TAGAGACGAGGTAGACC ATG
Yersinia pseudotuberculosis pYptb32953	ACAGCAGTCGGAAGCGCCTATTCCCACTTTTCATTGAAGTTTTTGCAAAATAGCATTGCTACCTATACAA TGTACT TATATT TGTACA ATGAGGTTGAGCTA ATG
Agrobacterium tumefaciens	GAAAGCTCCGGCCTTTCTTGTTATCGGTCGCACCGATGTATCTTGACTTTCGGCGTTTGGAG CGTACA TAAATC TGTACA TATTTGCGGAAGGGCAATTCTC ATG
Staphylococcus aureus	TCTTCCCCGTGCGAAATCTTGATTAGTAGTTTTTTTTCTGTGTCGCTTGTTTTAT TGTACA GATATT TGTACA ATTGAGGTGTGTTAAAAGGAGGCGATAAAT ATG
Rhodobacter sphaeroides	CAGGGCTTGATA TGTACA GCATTC TGTACA TATAGGAGTGAGCAGAGACCATATTCGTCCAGGTACATCTCGCATCAACTCTCTCCTTTCCGGCGTATCCAC ATG
Enterococcus faecium pRUM	TGATAGCACTTTTAATTATAAGAAGTTAGAATATTAATCATTTGCTTAAT TGTACA ATATAA TGTACA ATTGTTTTATAGAAATAAATAAGGGGTGAAAGGA ATG
Halorhodospira halophila	AAGGTGCTTACTGGCCAACGCCTGGGGGGTATCGGCAAAGTTGCCAGATTCGGTTATGTACGACAAAA TGTACA TGTGGAGTCTAT TGTACA GGAGCCGCCC ATG
Nitrosomonas europaea	GTTCACGTTGTTTCATACTGTCCTCTTGAACGCGCAGTCAGGTGCCGACATTGACTTTTGGCAGAATGTAA TGTACA AATAAA TGTACA ATTGGAGAACCTT ATG
Streptococcus agalactiae	ATTTGTAAGTTATTTATTTAGTGATTATCAAAACCAAAACT TGTACA ACATGT TGTCAA ATGATATAATAACAACATATCTAGTAAAGTTGAGGTGATTAAT ATG
Erwinia carotovora	ACTGCTGCTCTCGACGGTGGCGGAATAGTTTTGTCGAAGAAGGATTGAGTGGTTAAAAACGCAACATCAT TGTACA ATGACA AGTACA AACAGGAGGTACAG ATG
Saccharophagus degradans	ATAGGCCATACGACAAAACCAGCGTACCTTTGGGCGTCAGGTTGACACCCTTGTTATGGGCCCTTATAATTC TGTACA GTTTTA CGTACA GGTGGGGACAAT ATG

Figure 8. Paired 5′-TGTACA-3′ motifs in the promoter regions of *yefM-yoeB* operons from diverse bacteria. Sequences are aligned at the ATG start codons of the *yefM* homologues. Paired 5′-TGTACA-3′ motifs, or motifs that possess no more than one mismatch, and that are separated by centre-to-centre distances of 12 bp are boxed. In one case, the spacing is 18 bp.

DISCUSSION

The toxin components of TA systems are intracellular molecular time bombs whose release from complexes with their cognate antitoxins can trigger bacterial programmed cell death or cell cycle arrest (7). Understanding the mechanisms by which expression and activation of these toxins are controlled could allow the development of artificial means for toxin detonation, and therefore novel antibacterial strategies. For example, chemical genetics approaches may reveal innovative antibiosis strategies based on small molecule perturbation of TA module expression. Among chromosomal TA operons that have been analysed in *E.coli*, transcriptional autoregulation has been demonstrated for the *relBE* and *mazEF* modules. In both cases, the antitoxin acts as the primary repressor and the toxin as a co-repressor (27,28), which are also features of plasmid TA complexes. The chromosomal *chpBI-chpBK* TA operon is also autoregulated (48). Among chromosomal systems, regulation of the *mazEF* operon has been examined most closely: repression involves binding of the MazE antitoxin–MazF toxin complex to two alternating palindromes that overlap a pair of promoters that drive *mazEF* expression. Factor for inversion stimulation (FIS) weakly activates *mazEF* by interacting with sequences 5′ of the palindromes (30). In comparison, the *yefM-yoeB* operator site consists of L and S palindromes that possess a common hexameric core motif. The YefM antitoxin is the major transcriptional autoregulator of the operon, preferentially recognizing the L palindrome followed by the S repeat at elevated protein concentrations *in vitro*. Dimerization of YefM (Figure 7) probably reflects its interaction with these palindromic DNA sites, and suggests that the YefM$_2$–YoeB heterotrimer observed in crystallographic studies (21) is more likely to represent the repressor–corepressor complex

for transcriptional regulation than the YefM–YoeB$_2$ complex that has also been described (20).

YoeB is a corepressor that permits improved, probably cooperative, DNA binding by YefM most likely by either enhancing the stability of YefM, by altering YefM conformation to one that is more favourable for DNA binding, and/or by stabilizing the nucleoprotein complex at the operator site. Enhanced operator site binding by antitoxin when complexed with cognate toxin is a general characteristic of TA complexes. Furthermore, the binding of YefM alone or YefM–YoeB to DNA induces structural alterations in the proteins: examination of the far UV region in CD spectra revealed alterations in molar ellipticity of the proteins within the nucleoprotein complex. Although CD analysis suggests that the operator site does not undergo major structural transitions when bound to either protein, the presence of a DNase I hypersensitive cleavage site in the YefM- and YefM–YoeB-operator complexes nevertheless suggests that the operator site within the nucleoprotein complex undergoes deformations. DNA conformational changes such as bending, major groove opening and kinking are not uncommon in repressor–operator interactions (49–52), and reflect the formation of DNA structures that interfere with assembly or progression of the transcriptional machinery.

Bacteriophage P1 specifies a TA complex comprised of the Doc toxin and Phd antitoxin. Phd is the primary transcriptional repressor of the *phd-doc* operon, with Doc acting as a corepressor. The *phd-doc* operator site consists of two 8 bp palindromes that are 13 bp apart, centre to centre. The palindromes are bound sequentially, each by one Phd dimer, with Doc acting to promote cooperative binding of Phd to the sites by an undefined mechanism (26,28,53). YefM and Phd are homologues, although Doc and YoeB
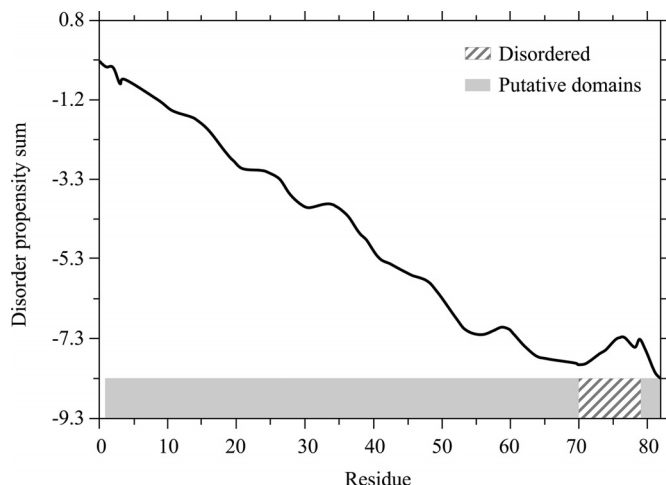
Figure 9. GlobPlot prediction (40) of intrinsic disorder in YefM. Descending regions correspond to putative domains, whereas the ascending region between residues ~70–80 is predicted to be disordered.

are not (18,19). Intriguingly, the inverted repeats bound by Phd include core 5′-GTAC-3′ motifs (33), identical to the central tetranucleotides of the L and S repeats recognized by YefM (Figure 3). Furthermore, the N-terminal end of Phd is required for DNA binding (33), which has also been proposed to be the case for YefM (21). The interesting parallels in operator organization, and apparently in the operator-binding properties of the N-terminal regions of YefM and Phd, attest to their common evolutionary origin. In contrast, the C-terminal halves of the two proteins are implicated in interactions with their cognate toxins and share very little similarity (21,33,54), which has been proposed to reflect modular exchange of DNA binding and toxin recognition domains among proteins related to Phd (54), including YefM homologues.

Equidistantly-spaced L and S repeats are common features in the *yefM-yoeB* regulatory regions of numerous genomes suggesting that YefM–YoeB homologues in diverse backgrounds exert transcriptional regulation by a common mechanism. Axe–Txe are YefM–YoeB homologues encoded by the pRUM multiresistance plasmid of *Enterococcus faecium* (19). L and S palindromes are located 5′ of *axe-txe* (Figure 8). However, the YefM homologue Axe does not repress the *yefM-yoeB* promoter *in vivo*. Conversely, YefM fails to down-regulate expression from the *axe* promoter (B. Kędzierska and F. Hayes, unpublished data). Despite the specificity observed *in vivo*, YefM and Axe recognize the non-cognate operator sites *in vitro* with approximately equal affinities (B. Kędzierska and F. Hayes, unpublished data). Although sequences that flank the palindromes likely influence the binding of YefM–YoeB to the repeats (Figure 7), additional factors, e.g. architectural proteins might also be significant for repression of certain *yefM-yoeB* promoters *in vivo*, albeit not for other promoters. Many genomes that do not possess identifiable L and S repeats upstream of *yefM-yoeB* nevertheless harbour the 5′-GTAC-3′ tetranucleotides, which are located at the centres of the L and S palindromes, and are separated by 11–13 bp (data not shown). These abbreviated repeats might represent minimal binding sites for the cognate YefM–YoeB proteins. In contrast, other *yefM-yoeB*

homologues possess no upstream motifs that are obviously related to L and S palindromes: the mechanism of regulation of these *yefM-yoeB* operons remains to be investigated.

Based on CD analysis, YefM has been described as an intrinsically unfolded protein that typifies a novel family of proteins that entirely lack any secondary structure (24). In contrast, native YefM examined here is dimeric and exhibits a CD spectrum that is consistent with a protein that is relatively well-folded and which possesses extensive α-helix and β-sheet features. This contention is supported by NMR spectra of YefM, which display signatures characteristic of a well-ordered protein containing both α-helical and β-strand secondary structure. Prediction software tools for disordered regions within proteins also suggest that YefM is well-structured, with only a patch of amino acids close to the C-terminus expected to be unfolded (Figure 9). CD analysis of the Axe homologue of YefM shows that it is also highly-structured (B. Kędzierska and F. Hayes, unpublished data). It is also worth emphasizing that the YefM dimer within the YefM$_2$–YoeB complex is largely structured, possessing α-helix, β-sheet and random coil elements with only the C-terminal region of one YefM monomer (corresponding to ~18% of YefM$_2$) being disordered (21). Moreover, the bulk of the antitoxin dimer within the YefM$_2$–YoeB complex projects tens of angströms from YoeB and is highly organized. It is difficult to envisage how YefM might achieve dimerization and become extensively structured principally by the interaction of a single α-helix at the C-terminal end of one YefM monomer with YoeB. Structure determination of free YefM$_2$, and of YefM and YefM-YoeB structures bound to DNA, will provide further crucial insights into the mechanism by which YefM–YoeB exerts transcriptional regulation.

ACKNOWLEDGEMENTS

REFERENCES

1. Potts,M. (2001) Desiccation tolerance: a simple process? *Trends Microbiol.*, **9**, 553–559.
2. Chang,D.E., Smalley,D.J., Tucker,D.L., Leatham,M.P., Norris,W.E., Stevenson,S.J., Anderson,A.B., Grissom,J.E., Laux,D.C., Cohen,P.S. *et al.* (2004) Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc. Natl Acad. Sci. USA*, **101**, 7427–7432.
3. Nystrom,T. (2004) Stationary-phase physiology. *Annu. Rev. Microbiol.*, **58**, 161–181.
4. Shapiro,J.A. (1998) Thinking about bacterial populations as multicellular organisms. *Annu. Rev. Microbiol.*, **52**, 81–104.
5. Hall-Stoodley,L., Costerton,J.W. and Stoodley,P. (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nature Rev. Microbiol.*, **2**, 95–108.
6. Camilli,A. and Bassler,B.L. (2006) Bacterial small-molecule signaling pathways. *Science*, **311**, 1113–1116.

7. Hayes,F. (2003) Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science*, **301**, 1496–1499.

8. Anantharaman,V. and Aravind,L. (2003) New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.*, **4**, R81.

9. Pandey,D.P. and Gerdes,K. (2005) Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.*, **33**, 966–976.

10. Engelberg-Kulka,H. and Glaser,G. (1999) Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol.*, **53**, 43–70.

11. Hazan,R., Sat,B. and Engelberg-Kulka,H. (2004) *Escherichia coli mazEF*-mediated cell death is triggered by various stressful conditions. *J. Bacteriol.*, **186**, 3663–3669.

12. Pedersen,K., Zavialov,A.V., Pavlov,M.Y., Elf,J., Gerdes,K. and Ehrenberg,M. (2003) The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell*, **112**, 131–140.

13. Christensen,S.K., Mikkelsen,M., Pedersen,K. and Gerdes,K. (2001) RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc. Natl Acad. Sci. USA*, **98**, 14328–14333.

14. Engelberg-Kulka,H., Sat,B., Reches,M., Amitai,S. and Hazan,R. (2004) Bacterial programmed cell death systems as targets for antibiotics. *Trends Microbiol.*, **12**, 66–71.

15. Zhang,Y., Zhang,J., Hara,H., Kato,I. and Inouye,M. (2005) Insights into the mRNA cleavage mechanism by MazF, an mRNA interferase. *J. Biol. Chem.*, **280**, 3143–3150.

16. Munoz-Gomez,A.J., Santos-Sierra,S., Berzal-Herranz,A., Lemonnier,M. and Diaz-Orejas,R. (2004) Insights into the specificity of RNA cleavage by the *Escherichia coli* MazF toxin. *FEBS Lett.*, **567**, 316–320.

17. Christensen,S.K., Pedersen,K., Hansen,F.G. and Gerdes,K. (2003) Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *J. Mol. Biol.*, **332**, 809–819.

18. Pomerantsev,A.P., Golovliov,I.R., Ohara,Y., Mokrievich,A.N., Obuchi,M., Norqvist,A., Kuoppa,K. and Pavlov,V.M. (2001) Genetic organization of the *Francisella* plasmid pFNL10. *Plasmid*, **46**, 210–222.

19. Grady,R. and Hayes,F. (2003) Axe-Txe, a broad-spectrum proteic toxin-antitoxin system specified by a multidrug-resistant, clinical isolate of *Enterococcus faecium*. *Mol. Microbiol.*, **47**, 1419–1432.

20. Cherny,I., Rockah,L. and Gazit,E. (2005) The YoeB toxin is a folded protein that forms a physical complex with the unfolded YefM antitoxin. Implications for a structural-based differential stability of toxin-antitoxin systems. *J. Biol. Chem.*, **280**, 30063–30072.

21. Kamada,K. and Hanaoka,F. (2005) Conformational change in the catalytic site of the ribonuclease YoeB toxin by YefM antitoxin. *Mol. Cell*, **19**, 497–509.

22. Christensen,S.K., Maenhaut-Michel,G., Mine,N., Gottesman,S., Gerdes,K. and Van Melderen,L. (2004) Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the *yefM-yoeB* toxin-antitoxin system. *Mol. Microbiol.*, **51**, 1705–1717.

23. Ren,D., Bedzyk,L.A., Thomas,S.M., Ye,R.W. and Wood,T.K. (2004) Gene expression in *Escherichia coli* biofilms. *Appl. Microbiol. Biotechnol.*, **64**, 515–524.

24. Cherny,I. and Gazit,E. (2004) The YefM antitoxin defines a family of natively unfolded proteins: implications as a novel antibacterial target. *J. Biol. Chem.*, **279**, 8252–8261.

25. Ruiz-Echevarria,M.J., Berzal-Herranz,A., Gerdes,K. and Diaz-Orejas,R. (1991) The *kis* and *kid* genes of the *parD* maintenance system of plasmid R1 form an operon that is autoregulated at the level of transcription by the co-ordinated action of the Kis and Kid proteins. *Mol. Microbiol.*, **5**, 2685–2693.

26. Magnuson,R. and Yarmolinsky,M.B. (1998) Corepression of the P1 addiction operon by Phd and Doc. *J. Bacteriol.*, **180**, 6342–6351.

27. Gotfredsen,M. and Gerdes,K. (1998) The *Escherichia coli relBE* genes belong to a new toxin-antitoxin gene family. *Mol. Microbiol.*, **29**, 1065–1076.

28. Gazit,E. and Sauer,R.T. (1999) Stability and DNA binding of the Phd protein of the phage P1 plasmid addiction system. *J. Biol. Chem.*, **274**, 2652–2657.

29. Afif,H., Allali,N., Couturier,M. and Van Melderen,L. (2001) The ratio between CcdA and CcdB modulates the transcriptional

30. Marianovsky,I., Aizenman,E., Engelberg-Kulka,H. and Glaser,G. (2001) The regulation of the *Escherichia coli mazEF* promoter involves an unusual alternating palindrome. *J. Biol. Chem.*, **276**, 5975–5984.

31. Dao-Thi,M.H., Charlier,D., Loris,R., Maes,D., Messens,J., Wyns,L. and Backmann,J. (2002) Intricate interactions within the *ccd* plasmid addiction system. *J. Biol. Chem.*, **277**, 3733–3742.

32. Lemonnier,M., Santos-Sierra,S., Pardo-Abarrio,C. and Diaz-Orejas,R. (2004) Identification of residues of the Kid toxin involved in autoregulation of the *parD* system. *J. Bacteriol.*, **186**, 240–243.

33. Zhao,X. and Magnuson,R.D. (2005) Percolation of the Phd repressor-operator interface. *J. Bacteriol.*, **187**, 1901–1912.

34. Simons,R.W., Houman,F. and Kleckner,N. (1987) Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.

35. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.*, **177**, 4121–4130.

36. Hayes,F., Radnedge,L., Davis,M.A. and Austin,S.J. (1994) The homologous operons for P1 and P7 plasmid partition are autoregulated from dissimilar operator sites. *Mol. Microbiol.*, **11**, 249–260.

37. Miller,J.H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual for Escherichia coli and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 72–74.

38. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

39. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

40. Linding,R., Russell,R.B., Neduva,V. and Gibson,T.J. (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.*, **31**, 3701–3708.

41. Frunzio,R., Bruni,C.B. and Blasi,F. (1981) *In vivo* and *in vitro* detection of the leader RNA of the histidine operon of *Escherichia coli* K-12. *Proc. Natl Acad. Sci. USA*, **78**, 2767–2771.

42. Verde,P., Frunzio,R., di Nocera,P.P., Blasi,F. and Bruni,C.B. (1981) Identification, nucleotide sequence and expression of the regulatory region of the histidine operon of *Escherichia coli* K-12. *Nucleic Acids Res.*, **9**, 2075–2086.

43. Patzer,S.I. and Hantke,K. (2000) The zinc-responsive regulator Zur and its control of the *znu* gene cluster encoding the ZnuABC zinc uptake system in *Escherichia coli*. *J. Biol. Chem.*, **275**, 24321–24332.

44. Rojo,F. (2001) Mechanisms of transcriptional repression. *Curr. Opin. Microbiol.*, **4**, 145–151.

45. Zhou,N.E., Kay,C.M. and Hodges,R.S. (1992) Synthetic model proteins. Positional effects of interchain hydrophobic interactions on stability of two-stranded α-helical coiled-coils. *J. Biol. Chem.*, **267**, 2664–2670.

46. Sreerama,N. and Woody,R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Anal. Biochem.*, **287**, 252–260.

47. Sprecher,C.A., Baase,W.A. and Johnson,W.C.,Jr (1979) Conformation and circular dichroism of DNA. *Biopolymers*, **18**, 1009–1019.

48. Santos-Sierra,S., Giraldo,R. and Díaz-Orejas,R. (1998) Functional interactions between *chpB* and *parD*, two homologous conditional killer systems found in the *Escherichia coli* chromosome and in plasmid R1. *FEMS Microbiol. Lett.*, **168**, 51–58.

49. Erie,D.A., Yang,G., Schultz,H.C. and Bustamante,C. (1994) DNA bending by Cro protein in specific and nonspecific complexes: implications for protein site recognition and specificity. *Science*, **266**, 1562–1566.

50. Spronk,C.A., Folkers,G.E., Noordman,A.M., Wechselberger,R., van den Brink,N., Boelens,R. and Kaptein,R. (1999) Hinge-helix formation and DNA bending in various *lac* repressor-operator complexes. *EMBO J.*, **18**, 6472–6480.

51. Akakura,R. and Winans,S.C. (2002) Mutations in the *occQ* operator that decrease OccR-induced DNA bending do not

cause constitutive promoter activity. *J. Biol. Chem.*, **277**, 15773–15780.

52. Schumacher,M.A., Miller,M.C., Grkovic,S., Brown,M.H., Skurray,R.A. and Brennan,R.G. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J.*, **21**, 1210–1218.

53. Magnuson,R., Lehnherr,H., Mukhopadhyay,G. and Yarmolinsky,M.B. (1996) Autoregulation of the plasmid addiction operon of bacteriophage P1. *J. Biol. Chem.*, **271**, 18705–18710.

54. Smith,J.A. and Magnuson,R.D. (2004) Modular organization of the Phd repressor/antitoxin protein. *J. Bacteriol.*, **186**, 2692–2698.