doi:10.1016/j.jmb.2007.07.037 J. Mol. Biol. (2007) 373, 11–26

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structure and Function of the *Escherichia coli* Protein YmgB: A Protein Critical for Biofilm Formation and Acid-resistance

Jintae Lee[1]†, Rebecca Page[2]†, Rodolfo García-Contreras[1] Jeanne-Marie Palermino[3], Xue-Song Zhang[1], Ojus Doshi[2] Thomas K. Wood[1,4] and Wolfgang Peti[3]*

[1]*Artie McFerrin Department of Chemical Engineering Texas A & M University College Station TX 77843-3122, USA*

[2]*Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence RI 02912, USA*

[3]*Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence RI 02912, USA*

[4]*Department of Biology Texas A & M University College Station TX 77843-3258, USA*

Received 23 April 2007; received in revised form 12 July 2007; accepted 16 July 2007
Available online 2 August 2007

Edited by J. Karn

The *Escherichia coli* gene cluster *ymgABC* was identified in transcriptome studies to have a role in biofilm development and stability. In this study, we showed that YmgB represses biofilm formation in rich medium containing glucose, decreases cellular motility, and protects the cell from acid indicating that YmgB has a major role in acid-resistance in *E. coli*. Our data show that these phenotypes are potentially mediated through interactions with the important cell signal indole. In addition, gel mobility-shift assays suggest that YmgB may be a non-specific DNA-binding protein. Using nickel-enrichment DNA microarrays, we showed that YmgB binds, either directly or indirectly, *via* a probable ligand, genes important for biofilm formation. To advance our understanding of the function of YmgB, we used X-ray crystallography to solve the structure of the protein to 1.8 Å resolution. YmgB is a biological dimer that is structurally homologous to the *E. coli* gene regulatory protein Hha, despite having only 5% sequence identity. This supports our DNA microarray data showing that YmgB is a gene regulatory protein. Therefore, this protein, which clearly has a critical role in acid-resistance in *E. coli*, has been renamed as AriR for regulator of acid resistance influenced by indole.

Keywords: biofilm; protein structure; gene regulatory protein; acid resistance; *E. coli*

Introduction

Bacteria seldom live individually,[1,2] and swarming and biofilm formation are two important examples of bacterial multicellular behavior, which is significant because it enhances the chance for survival in competitive environments.[3,4] However, it can cause serious problems, such as infection and biofouling. In fact, 80% of human bacterial infections involve biofilms.[5] In recent years, a number of studies have been initiated with the aim of unraveling the

*Corresponding author. E-mail address: wolfgang_peti@brown.edu.

† J.L. and R.P. contributed equally to this work.

Abbreviations used: β-OG, β-octyl-β-D-glucopyranoside; EMSA, electrophoretic mobility-shift assay; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

development and regulation of these multicellular structures in more detail. Specifically, genes important for biofilm formation and propagation have been identified using gene chips (for *Pseudomonas aeruginosa*[6] and *Escherichia coli*[7–10]), proteome analysis,[11] and classical knockout studies, and include genes that code for proteins involved in bacterial motility, quorum sensing, and the induction of polysaccharide synthesis. Notably, a significant subset of these identified genes code for hypothetical proteins of unknown function.[12] One of these hypothetical proteins is the 88 amino acid residue *E. coli* protein YmgB, which we have now functionally and structurally characterized using a powerful cross-disciplinary approach.

Suggestions of the function of the YmgB protein were first obtained from transcriptome studies of the *ymgABC* gene cluster. Specifically, we found that the hypothetical *E. coli* gene cluster *ymgABC* was differentially expressed in biofilm cells. Additionally, we found that this expression was influenced by cell signaling.[10,13–15] For example, the furanosyl borate diester or derivative quorum signal autoinducer 2 (AI-2)[16] repressed *ymgAB* threefold.[14] In contrast, the biofilm inhibitor furanone from the alga *Delisea pulchra*, which masks autoinducer 2 (AI-2) signaling,[17] induced *ymgA* twofold.[14] Furthermore, deleting the AI-2 transporter gene *tqsA* also affected *ymgABC* expression. Specifically, it repressed *ymgBC* fourfold.[13] Also, *ymgABC* were induced 14-fold at 15 h relative to 7 h biofilms,[10] and the stationary-phase biofilm signal indole repressed *ymgABC* two- to fivefold.[15] Therefore, these results all strongly suggested that the *ymgABC* gene cluster, and thus likely the YmgB protein itself, has an important role in *E. coli* biofilm formation as a result of AI-2 or indole signaling.

Using similar transcriptome studies, we present in this manuscript that the gene cluster *ymgABC* has a significant role in acid-resistance. Interestingly and not currently understood, the transcription of a number of acid-resistance genes is down-regulated upon biofilm formation. One example is the *gadABC* operon, which is regulated by GadE. GadABC protects *E. coli* under extremely acidic conditions (pH 2 and below) and allows the bacterium to colonize the gastrointestinal tract.[18] Using biofilm array studies, we showed that biochemical signals correlated with enhanced biofilm formation result in the repression of *gadABC,* and *vice versa*. Specifically, AI-2 signaling, which increases biofilm formation 30-fold,[19] represses *gadABC* sevenfold to 12-fold,[14] while furanone does the opposite, repressing biofilm formation and inducing *gadB* transcription threefold.[14] Also, the deletion of the *tqsA* gene, causes biofilm formation to increase 7000-fold, while *gadABCX* expression is repressed 11-fold to 13-fold.[13] In addition, when the biofilm inhibitor/stress regulator *bhsA* is deleted, which leads to a fivefold increase in biofilm formation, *gadABCE* are also repressed three- to sixfold.[20] Similar observations were made upon the deletion of *bssR*.[21] Finally, indole repressed *gadABCEX* two- to fourfold.[15]

The transcription levels of other identified acid-resistance genes, such as *hdeABD* (which function as chaperones to prevent aggregation of periplasmic proteins under extremely acidic conditions[22]) vary in a manner similar to that of *gadABCEX*. *hdeABD* are also repressed three- to fivefold by indole,[15] sixfold to 18-fold[14] by AI-2 and in all mutants (*tqsA, bssR*) that lead to elevated biofilm formation.[13,21] In addition, the deletion of *bhsA,* which increases biofilm formation fivefold, induces *hdeABD* expression three- to fivefold.[20] Since the uncharacterized gene cluster *ymgABC* is regulated in these data sets in a manner almost identical with those of the known acid-resistance genes *gadABCEX* and *hdeABD*, we deduced that this locus most likely encodes an acid-resistance locus.[10,13–15]

Here, we show that YmgB, one protein of this gene cluster, represses biofilm formation in rich medium containing glucose, protects the cell from acid, and decreases cell motility. In addition, our data show that all of these phenotypes may be mediated through interactions with the cellular signaling molecule indole.[15,23] However, the biological function of YmgB is unknown and, in fact, is annotated as a hypothetical protein with no assigned function. Because a 3-D structure of a protein is more highly conserved than its sequence, it can frequently lead to identification of its biological function. Therefore, we used X-ray crystallography to determine the high-resolution structure of YmgB. On the basis of our structure, we were able to determine that YmgB is structurally and functionally similar to the well-characterized *E. coli* protein Hha, despite having only 5% sequence identity. This prediction, based on structural analysis, was then confirmed using *in vitro* functional gel mobility-shift experiments and *in vivo* nickel-enrichment DNA microarrays. Thus, using these powerful combined genetic and structural studies, we show that YmgB functions as a regulatory protein that binds DNA promoter sequences that regulate genes important in acid-resistance and biofilm formation.

Results

YmgB and its function in biofilm formation and acid-resistance

YmgB regulation in biofilms and motility

Microarray experiments have shown that the *ymg* locus is highly regulated in response to biofilm formation.[10,13–15] Therefore, we tested the effect of deleting the *ymg* locus on biofilm formation, and compared these results with those observed for other known acid-resistance gene mutants. Two independent cultures of ten isogenic *E. coli* K-12 mutants (mutants *ycgZ, ymgA, ymgB*, and *ymgC*, along with the known acid-eresistance mutants *gadA, gadB, gadE, hdeA, hdeB*, and *hdeD* as controls) were tested for biofilm formation using the 96-well