FOR THE RECORD

Solution structure of a BolA-like protein from *Mus musculus*

TAKUMA KASAI,[1,2] MAKOTO INOUE,[1] SEIZO KOSHIBA,[1] TAKASHI YABUKI,[1] MASAAKI AOKI,[1] EMI NUNOKAWA,[1] EIKO SEKI,[1] TAKAYOSHI MATSUDA,[1] NATSUKO MATSUDA,[1] YASUKO TOMO,[1] MIKAKO SHIROUZU,[1,3] TAKAHO TERADA,[1,3] NAOMI OBAYASHI,[1] HIROAKI HAMANA,[1] NAOKO SHINYA,[1] AYAKO TATSUGUCHI,[1] SATOKO YASUDA,[1] MAYUMI YOSHIDA,[1] HIROSHI HIROTA,[1] YO MATSUO,[1] KAZUTOSHI TANI,[1] HARUKAZU SUZUKI,[1] TAKAHIRO ARAKAWA,[1] PIERO CARNINCI,[1] JUN KAWAI,[1] YOSHIHIDE HAYASHIZAKI,[1] TAKANORI KIGAWA,[1] AND SHIGEYUKI YOKOYAMA[1,2,3]

[1]RIKEN Genomic Sciences Center, Yokohama, Japan
[2]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo, Japan
[3]RIKEN Harima Institute at SPring-8, Hyogo, Japan

(RECEIVED September 1, 2003; FINAL REVISION September 1, 2003; ACCEPTED October 7, 2003)

Abstract

The BolA-like proteins are widely conserved from prokaryotes to eukaryotes. The BolA-like proteins seem to be involved in cell proliferation or cell-cycle regulation, but the molecular function is still unknown. Here we determined the structure of a mouse BolA-like protein. The overall topology is $\alpha\beta\beta\alpha\alpha\beta\alpha$, in which β_1 and β_2 are antiparallel, and β_3 is parallel to β_2. This fold is similar to the class II KH fold, except for the absence of the GXXG loop, which is well conserved in the KH fold. The conserved residues in the BolA-like proteins are assembled on the one side of the protein.

Keywords: NMR structure; conserved protein; BolA; cell proliferation; cell-cycle regulation; KH fold; structural genomics

Escherichia coli BolA and its homologs constitute a widely conserved protein family, called the BolA-like protein family. BolA-like proteins are found in most bacteria and eukaryotes, and in an archaeon. The *E. coli bolA* gene was isolated as the morphogene induced in stationary phase (Aldea et al. 1988, 1989, 1990; Lange and Hengge-Aronis 1991), and was also found to be induced by various stresses (Santos et al. 1999). The mutation or the overexpression of the *bolA* gene affects the transcription level of the penicillin binding protein (PBP) genes, which are involved in cell wall biosynthesis (Aldea et al. 1989; Santos et al. 2002). In *Schizosaccharomyces pombe*, a *bolA* homolog, *uvi31*$^+$, which was isolated as a UV-inducible gene, shows cell-cycle and growth phase-dependent expression (Lee et al. 1994; Kim et al. 1997). In addition, the deletion mutant of *uvi31* proliferates faster than the wild-type cell, and shows abnormal septation after UV-induced cell cycle arrest (Kim et al. 2002). Although these results imply that BolA-like proteins play some roles in cell proliferation by controlling the transcription of other genes, their molecular functions remain unknown.

Most of the BolA-like proteins consist of about 100 amino acid residues, as do the three BolA-like proteins found within the mouse full-length cDNA libraries, FANTOM and FANTOM2 (Kawai et al. 2001; Okazaki et al. 2002). We designated the mouse BolA-like proteins as BolA1 (gi|12841442), BolA2 (gi|26325949), and BolA3 (gi|26389531). Among the three proteins, BolA1 shows the highest homology to *E. coli* BolA. Close homologs of the mouse BolA2 were found in most eukaryotes, including *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis*

Reprint requests to: Shigeyuki Yokoyama, Protein Research Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan; e-mail: yokoyama@biochem.s.u-tokyo.ac.jp; fax: 81-45-503-9195.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.03401004.

elegans, *Arabidopsis thaliana*, and *Saccharomyces cerevisiae*. Therefore, the BolA2 proteins constitute a eukaryotic subfamily of the BolA-like proteins (Fig. 1A). Mouse BolA3 and its homologs compose another eukaryotic subfamily, with lower similarity to *E. coli* BolA than the BolA2 subfamily.

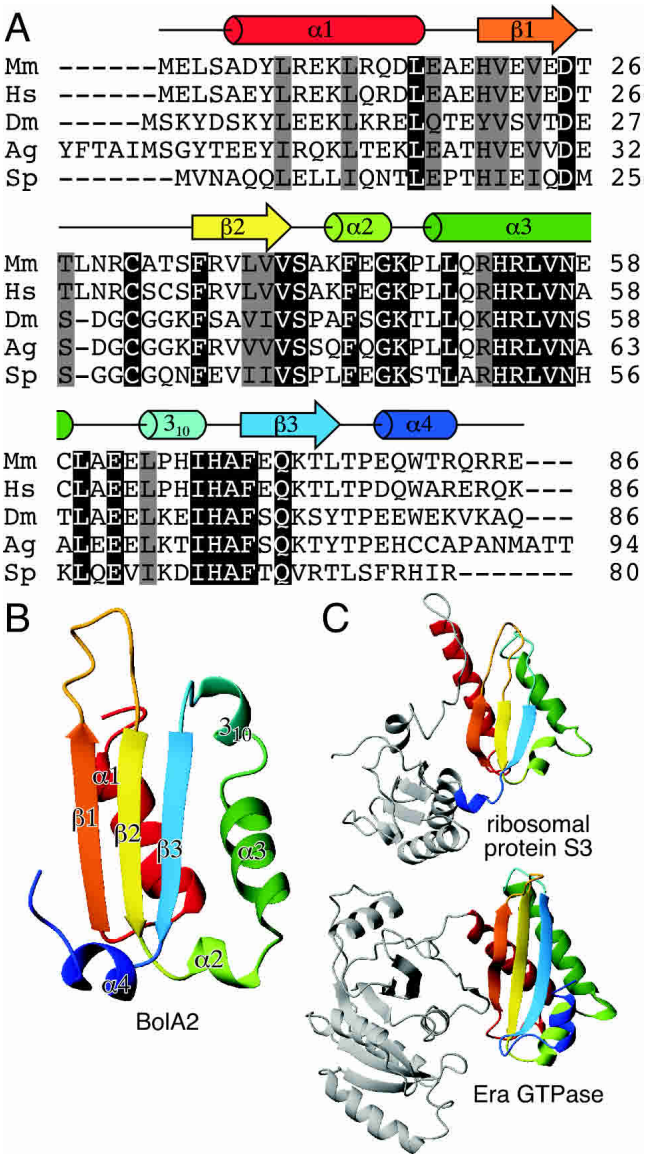

Figure 1. (*A*) Multiple sequence alignment of the BolA2 subfamily. Highlighted and shaded residues indicate identical and similar residues, respectively. Mm: BolA2 from *Mus musculus.* Hs: My016 from *Homo sapiens.* Dm: CG16804-PB from *Drosophila melanogaster.* Ag: agCP4532 from *Anopheles gambiae.* Sp: SPAC8C9.11 from *Schizosaccharomyces pombe*. The multiple alignment was achieved with CLUSTAL_X (Thompson et al. 1997). The secondary structures of the mouse BolA2 are indicated at the *top* of the alignment. (*B*) Ribbon diagram of BolA2. (*C*) Class II KH domains of ribosomal protein S3 (*top*, PDB ID: 1J5E) and Era GTPase (*bottom*, PDB ID: 1EGA). Each class II KH domain is colored in the rainbow order, from red (N terminus) to blue (C terminus).

In the present study, we determined the solution structure of the mouse BolA2, which is the first tertiary structure of a BolA-like protein family member. The BolA2 protein has the αββααβα topology, which is the same as that of the class II KH fold. However, the BolA-like proteins lack the GXXG sequence, which is involved in binding to nucleic acids in the authentic KH-fold proteins. Instead of the GXXG loop, BolA2 has a turn, which results in a helix-turn-helix (HTH) motif. The surface residues around the HTH motif are well conserved in the BolA2 subfamily.

Results and Discussion

The overall topology of BolA2 is αββααβα, in which β_1 and β_2 are antiparallel, and β_3 is parallel to β_2 (Fig.1B). A nine-residue loop, inserted between β_1 and β_2, has several amide protons with very low intensities in the $^1H/^{15}N$-HSQC spectrum. The α_4 helix is anchored on one side of the β-sheet, while the other helices are on the other side. A one-turn 3_{10}-like helix exists between α_3 and β_3. The α_2 and α_3 helices form an HTH motif, and are arranged at an angle of about 60° to each other. In *E. coli* BolA, another HTH motif had been predicted by the Chou-Fasman algorithm (Aldea et al. 1989). However, the corresponding region of BolA2 comprises the β_1/β_2 loop and the β_2 strand.

According to the DALI (Holm and Sander 1993) search, BolA2 shows significant structural similarity to the ribosomal protein S3 N-terminal domain (Wimberly et al. 2000) and the Era GTPase C-terminal domain (Chen et al. 1999), with Z-values of 5.4 and 4.7, respectively, which belong to class II KH fold (Fig. 1C). The KH fold is commonly found among nucleic acid binding proteins, and consists of two classes with distinct topologies, βααββα for class I and αββααβ for class II (Grishin 2001). Most of the KH-fold proteins have a well-conserved GXXG sequence on the loop between two adjacent helices. Some of the class I KH-fold proteins bind to nucleic acids using their GXXG loop (Lewis et al. 2000; Liu et al. 2001; Braddock et al. 2002a,b). The corresponding region of BolA2 forms an HTH motif, and the BolA-like proteins lack the conserved GXXG sequence in this region.

Figure 2A shows the conserved surface residues of BolA2. Most of the conserved residues are assembled on one side of the protein, especially around the HTH motif. The "conserved" side consists of an electrically neutral region surrounded by several basic residues (Fig. 2B). There are significantly conserved solvent-exposed hydrophobic residues (L50, L55) near the HTH motif. On the other hand, the other side of BolA2 is highly acidic. Some residues are still conserved on the "variable" side of BolA2.

The nucleic-acid binding property of the BolA-like proteins is implicated by the structural similarity to the KH fold and the assembly of conserved surface residues around the HTH motif. However, the existence of some conserved sur-

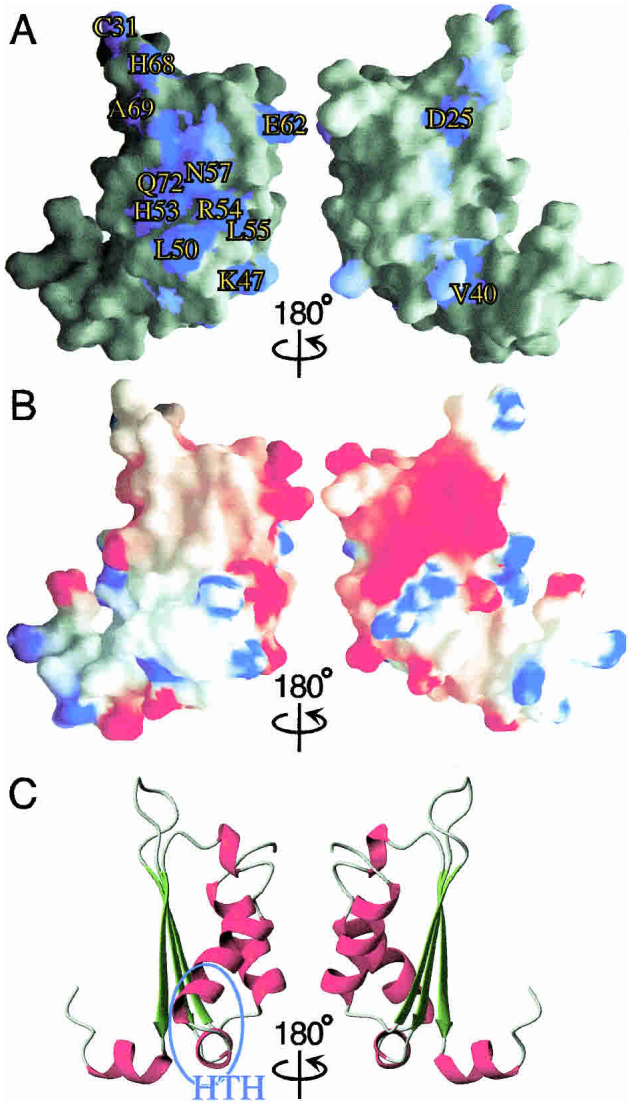

Figure 2. (*A*) The conserved residues on the surface of the mouse BolA2. The identical and the similar residues defined in Figure 1A are colored dark blue and light blue, respectively. The *right* panel is viewed from the opposite side of the *left* panel. (*B*) The surface electrostatic potential of BolA2. (*C*) Ribbon diagram in the same orientation as *A* and *B*. The HTH motif is indicated by the blue circle.

face residues on the "variable" side and the absence of the GXXG sequence suggest the possibility that the BolA-like proteins have other functions than those of the KH-fold proteins.

Materials and methods

Sample preparation and NMR spectroscopy

The mouse BolA2 protein was produced as the 113-amino-acid recombinant protein with a 27-amino-acid purification tag sequence (MKGSSHHHHH HSSGASLVPR GSEGAAT) at the N terminus of the 86-amino-acid coding sequence. The uniformly ^{13}C- and ^{15}N-labeled protein was produced by the *E. coli* cell-free protein synthesis system (Kigawa et al. 1999), and then 5.0 mg of the purified protein were obtained by polyhistidine affinity chromatography followed by anion exchange chromatography. All NMR spectra were acquired on a Bruker AVANCE600 spectrometer at 298K, with a 1.2 mM $^{13}C/^{15}N$ BolA2 protein solution in 50 mM Na-phosphate buffer (pH 6.0) containing 100 mM NaCl, 0.02% NaN_3, 1 mM DTT-d_{10}, and 10% ^2H_2O. Spectra were processed and analyzed with NMRPipe (Delaglio et al. 1995) and NMRview (Johnson and Blevins 1994).

Structure determinaiton

The chemical shifts were assigned using $^1H/^{15}N$-HSQC, $^1H/^{13}C$-HSQC, HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CO)NH, CC(CO)NH, H(CCCO)NH, HCCH-COSY, HCCH-TOCSY, NOESY-$^1H/^{15}N$-HSQC (τ_m = 100 msec), and NOESY-$^1H/^{13}C$-HSQC (τ_m = 100 msec) spectra. In the BolA2 coding region, 97% of the observable proton signals were assigned. The NOESY cross-peaks were assigned with DYANA/CANDID (Herrmann et al. 2002), following manual correction. The total number of assigned NOEs was 1546, but 931 meaningful distance constraints were used for the final structure

Table 1. *Calculation statistics for the ensemble of 20 structures*

NOE distance constraints	
All	931
Intraresidual	467
Sequential	214
Medium range ($2 \leq \|i-j\| \leq 4$)	97
Long range ($\|i-j\| > 4$)	157
Hydrogen bond constraints	29
Dihedral angle constraints	
ϕ	54
ψ	54
χ	25
Number of violations	
Distance violations (>0.2 Å)	0
Dihedral angle violations (>2°)	0
Coordinate RMSD (Å)[a]	
Backbone atoms	0.38 ± 0.09
All nonhydrogen atoms	0.87 ± 0.08
Ramachandran plot[b]	
Most favored regions	92.4%
Additional allowed regions	7.5%
Generously allowed regions	0.1%
Disallowed regions	0.0%
CNS energy term (kcal · mole^{-1})	
$E_{overall}$	46.9 ± 0.6
E_{bond}	0.735 ± 0.03
E_{angle}	39.1 ± 0.2
E_{improp}	1.05 ± 0.03
E_{vdW}	4.53 ± 0.29
E_{NOE}	1.39 ± 0.28
E_{dihed}	0.0260 ± 0.0079

[a] For residues 5–14, 19–26, 35–41, 51–62, 69–75. RMSD: root mean square deviation.
[b] From PROCHECK-NMR (Laskowski et al. 1996). For residues 4–25, 35–82.

calculation. The dihedral-angle constraints of the main chain were used in the secondary structure regions predicted by TALOS (Cornilescu et al. 1999; $\phi = -60° \pm 15°$, $\psi = -45° \pm 15°$ for the helical regions, $\phi = -120° \pm 30°$, $\psi = 150° \pm 30°$ for the β-strand regions). The side-chain χ angle constraints (180°, 60°, or −60° with a tolerance of ±30°) were used for the residues for which the χ angles could be obviously determined from the NOESY spectra. To obtain hydrogen bond information, $^{1}H/^{15}N$-HSQC spectra were measured in $^{2}H_2O$. Twenty-nine peaks in the α-helix or the β-strand regions were used for the structure calculation, while a total of 37 peaks were observed. Hydrogen bond constraints were used as two distance constraints (3.5 Å for N and O, 2.5 Å for HN and O) for each hydrogen bond. The final structure was calculated with CNS (Brünger et al. 1998). The calculation statistics are summarized in Table 1. The coordinates of the best 20 structures have been deposited in the Protein Data Bank, under the accession ID 1IW5. The ribbon diagrams were drawn with MOLMOL (Koradi et al. 1996). The molecular surface presentations were generated by GRASP (Nicholls et al. 1991).

Acknowledgments

We are grateful to Yukiko Fujikura, Yoko Motoda, Miyuki Saito, Yukako Miyata, Atsuo Kobayashi, Noriko Hirakawa, Noriko Sakagami, Masaomi Ikari, Fumiko Hiroyasu, and Megumi Watanabe for their technical assistance. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aldea, M., Hernández-Chico, C., de la Campa, A.G., Kushner, S.R., and Vicente, M. 1988. Identification, cloning, and expression of *bolA*, an *ftsZ*-dependent morphogene of *Escherichia coli*. *J. Bacteriol.* **170:** 5169–5176.

Aldea, M., Garrido, T., Hernández-Chico, C., Vicente, M., and Kushner, S.R. 1989. Induction of a growth-phase-dependent promoter triggers transcription of *bolA*, an *Escherichia coli* morphogene. *EMBO J.* **8:** 3923–3931.

Aldea, M., Garrido, T., Pla, J., and Vicente, M. 1990. Division genes in *Escherichia coli* are expressed coordinately to cell septum requirements by gearbox promoters. *EMBO J.* **9:** 3787–3794.

Braddock, D.T., Louis, J.M., Baber, J.L., Levens, D., and Clore, G.M. 2002a. Structure and dynamics of KH domains from FBP bound to single-stranded DNA. *Nature* **415:** 1051–1056.

Braddock, D.T., Baber, J.L., Levens, D., and Clore G.M. 2002b. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: Solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* **21:** 3476–3485.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Chen, X., Court, D.L., and Ji, X. 1999. Crystal structure of ERA: A GTPase-dependent cell cycle regulator containing an RNA binding motif. *Proc. Natl. Acad. Sci.* **96:** 8396–8401.

Cornilescu, G., Delaglio, F., and Bax A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13:** 289–302.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277–293.

Grishin, N.V. 2001. KH domain: One motif, two folds. *Nucleic Acids Res.* **29:** 638–643.

Herrmann, T., Güntert, P., and Wüthrich, K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol. Biol.* **319:** 209–227.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233:** 123–138.

Johnson, B.A. and Blevins, R.A. 1994. NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4:** 603–614.

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* **409:** 685–690.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* **442:** 15–19.

Kim, S.H., Kim, M., Lee, J.K., Kim, M.J., Jin, Y.H., Seong, R.H., Hong, S.H., Joe, C.O., and Park, S.D. 1997. Identification and expression of *uvi31*$^{+}$, a UV-inducible gene from *Schizosaccharomyces pombe*. *Environ. Mol. Mutagen.* **30:** 72–81.

Kim, M.J., Kim, H.S., Lee, J.K., Lee, C.B., and Park, S.D. 2002. Regulation of septation and cytokinesis during resumption of cell division requires *uvi31*$^{+}$, a UV-inducible gene of fission yeast. *Mol. Cells* **14:** 425–430.

Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14:** 51–55.

Lange, R. and Hengge-Aronis, R. 1991. Growth phase-regulated expression of *bolA* and morphology of stationary-phase *Escherichia coli* cells are controlled by the novel sigma factor σ^{S}. *J. Bacteriol.* **173:** 4474–4481.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8:** 477–486.

Lee, J.K., Park, E.J., Chung, H.K., Hong, S.H., Joe, C.O., and Park, S.D. 1994. Isolation of UV-inducible transcripts from *Schizosaccharomyces pombe*. *Biochem. Biophys. Res. Commun.* **202:** 1113–1119.

Lewis, H.A., Musunuru, K., Jensen, K.B., Edo, C., Chen, H., Darnell, R.B., and Burley, S.K. 2000. Sequence-specific RNA binding by a Nova KH domain: Implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100:** 323–332.

Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A., and Sattler, M. 2001. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294:** 1098–1102.

Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11:** 281–296.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420:** 563–573.

Santos, J.M., Freire, P., Vicente, M., and Arraiano, C.M. 1999. The stationary-phase morphogene *bolA* from *Escherichia coli* is induced by stress during early stages of growth. *Mol. Microbiol.* **32:** 789–798.

Santos, J.M., Lobo, M., Matos, A.P., De Pedro, M.A., and Arraiano, C.M. 2002. The gene *bolA* regulates *dacA* (PBP5), *dacC* (PBP6) and *ampC* (AmpC), promoting normal morphology in *Escherichia coli*. *Mol. Microbiol.* **45:** 1729–1740.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25:** 4876–4882.

Wimberly, B.T., Brodersen, D.E., Clemons Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407:** 327–339.