JOURNAL OF BACTERIOLOGY, June 1999, p. 3842–3844 Vol. 181, No. 12
0021-9193/99/$04.00+0

Identification of Two New Proteins in Spermidine Nucleoids Isolated from *Escherichia coli*

LIZABETH D. MURPHY, JUDAH L. ROSNER, STEVEN B. ZIMMERMAN,* AND DOMINIC ESPOSITO

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0560

Received 9 February 1999/Accepted 14 April 1999

The *Escherichia coli* nucleoid contains DNA in a condensed but functional form. Analysis of proteins released from isolated spermidine nucleoids after treatment with DNase I reveals significant amounts of two proteins not previously detected in wild-type *E. coli*. Partial amino-terminal sequencing has identified them as the products of *rdgC* and *yejK*. These proteins are strongly conserved in gram-negative bacteria, suggesting that they have important cellular roles.

The *Escherichia coli* nucleoid maintains its DNA in a compact form yet permits the DNA to carry out vital functions including replication, recombination, gene regulation, and expression (1, 7, 8, 11, 12). We have been characterizing spermidine nucleoids (5, 6) for their structure and stability under various conditions and for their protein content. The proteins released from the spermidine nucleoids by DNase I treatment are separated by polyacrylamide gel electrophoresis and identified either by comigration with known standards or by partial N-terminal amino acid sequencing. Over half of the protein released is found in only five proteins, namely, RNA polymerase, HU, H-NS, Fis, and residual lysozyme introduced during the cell lysis procedure (5). Of these, Fis, RNA polymerase, and H-NS are bound particularly strongly to the nucleoids (6); the role of these proteins in nucleoid structure is currently being examined.

In the present experiments, isolated spermidine nucleoids were incubated with 0 or 10 μg of DNase I per ml at 0°C for 60 min, and the nonsedimentable fraction was isolated (6). After precipitation with trichloroacetic acid (5), samples were redissolved, electrophoresed on a 10% NuPage Bis-Tris gel with a MOPS (morpholinepropanesulfonic acid)-sodium dodecyl sulfate running buffer, and transferred to a polyvinylidene difluoride membrane according to the manufacturer's specifications (Novex, San Diego, Calif.). The membrane was rinsed with water and absolute methanol and then stained with Coomassie blue R-250 (Bio-Rad, Hercules, Calif.) for 45 s and destained in 50% methanol. Amino-terminal sequences were determined with a Hewlett-Packard 241 protein sequencer by using the polyvinylidene difluoride 4.0 method.

Two previously uncharacterized protein bands with apparent molecular masses of approximately 35 and 37 kDa can be seen on a membrane from a typical experiment (Fig. 1). Partial N-terminal amino acid sequencing identified the two polypeptides as the protein products of the genes *rdgC* (also *yaiD*/*orf303*; GenBank accession no. AE000145) and *yejK* (GenBank accession no. AE000308) (Table 1). The abundance of these two polypeptides in the DNase I-released fraction from several strains of *E. coli* was determined and expressed relative to the amount of the α subunit of RNA polymerase (RpoA) in the same fraction (Table 1). RpoA was used as a reference point because it is a well-defined major component of the proteins released by DNase I, it has been identified by its N-terminal amino acid sequence, it is present in a relatively constant amount in this fraction in all strains tested, and it is similar in electrophoretic mobility (apparent molecular mass of 39 kDa) to the polypeptides in question. The quantitation of the two polypeptides shown in Table 1 is based upon the intensity of Coomassie blue staining of bands on the membrane and so is

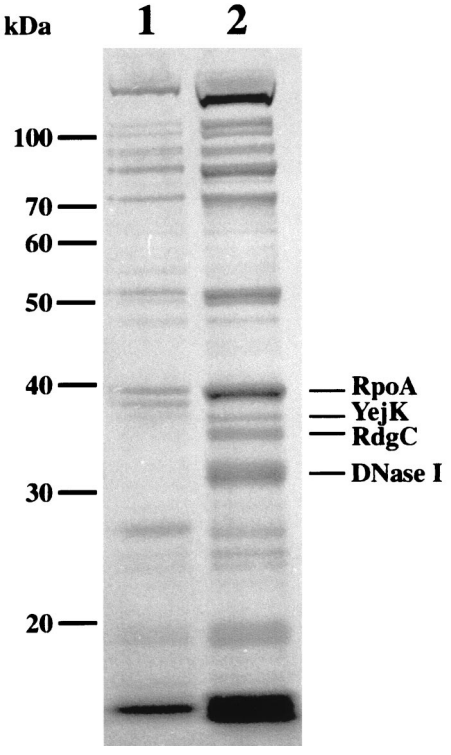

FIG. 1. Polypeptides from *E. coli* M182 spermidine nucleoids. A typical membrane is shown, with lanes 1 and 2 representing polypeptides isolated from nucleoids treated with 0 and 10 μg of DNase I per ml, respectively. Locations of protein products described in the text and sizes of molecular mass standards (BenchMark protein ladder; Life Technologies) are indicated in the margins. Fis, H-NS, HU, and lysozyme migrate into a single band at the bottom of the membrane. Unlabelled protein bands are currently under investigation.

* Corresponding author. Mailing address: Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bldg. 5, Room 328W, 5 Center Dr. MSC0560, Bethesda, MD 20892-0560. Phone: (301) 496-2208. Fax: (301) 496-0201. E-mail: stevenz@bdg5.niddk.nih.gov.

TABLE 1. Identification of RdgC and YejK in the DNase I-released fraction of isolated spermidine nucleoids

Gene (predicted N-terminal amino acid sequence)	*E. coli* strain	Observed N-terminal amino acid sequence[a]	Relative amount of protein[b]
rdgC (MLWFKNLMVYRLSRE)	C600	MLXFK(N/V)LMVYRLS	0.59
	C600 *fis*[c]	MLXFKXLMVYRL	0.73
	M182[d]	MLXFKXLMVYRL	0.39
	3284[e]	M(L/V)XFKXLMVYRLSRE	0.33
yejK (MSLDINQIALHQLIKRD)	C600	SLDI(G/N)QIAL(H/G)QLI[f]	0.31
	C600 *fis*[c]	SLDINQIALHXLXKRD[f]	0.32
	M182[d]	S(L/Q)DIXQIALHQLI[f]	0.26
	3284[e]	SLDINQIALHQLIKR[f]	0.42

[a] N-terminal sequences were determined as described in the text.
[b] The amount of each protein relative to that of RpoA from the same lane of the membrane as determined by Coomassie staining. RpoA represented 6.3 to 8.8% of the protein released from the nucleoid preparations by DNase I.
[c] C600 *fis-767*::Kan was made by P1 transduction of *fis-767*::Kan (3) into C600.
[d] The genotype of strain M182 is Δ(*lac*)*74 galU galK rpsL* (13).
[e] Strain 3284 is a Δ*stpA*::Tet Δ*hns*::Kan derivative of strain M182.
[f] The observed N-terminal sequences of YejK indicate that a cleavage of the initial methionine has taken place during posttranslational processing. This is expected due to the presence of serine at the penultimate amino acid position (2).

subject to some variability due to intrinsic differences in staining and transfer. It is, nonetheless, clear that the two polypeptides are relatively abundant components of the DNase I-released fraction in all parental and mutant strains tested and represent 2 to 3% of the total protein in this fraction. In the absence of DNase I treatment, there was a greater than 20-fold decrease in the levels of these polypeptides. This strongly suggests that the two polypeptides are associated with nucleoid

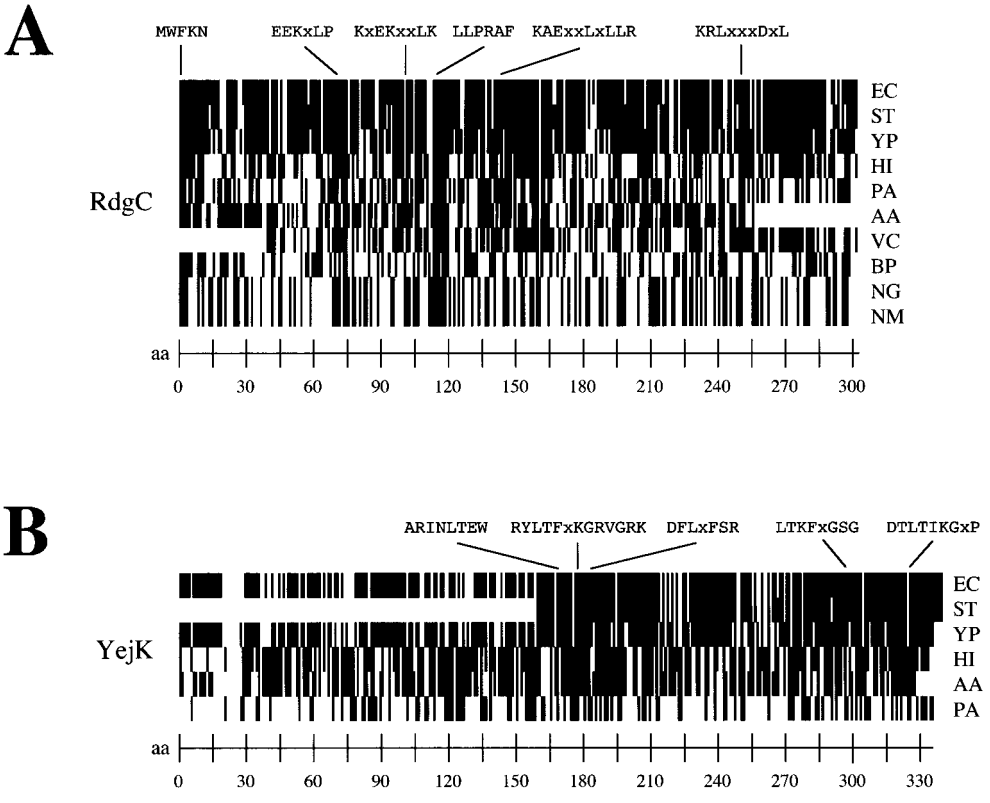

FIG. 2. Groupwise comparison of the bacterial homologs of RdgC (A) and YejK (B). The groups of protein homologs from different bacteria are shown as sets of stacked horizontal bars. The scale bars along the horizontal axes in each panel identify amino acid (aa) positions. A blackened position indicates identity of the amino acid at the given position with the amino acid which occurs most frequently at that position among the homologs. If no single amino acid occurs in at least 33% of the sequences at a given position in the group of homologs, no identity is indicated at that position. Sequences along the top of each panel are strongly conserved motifs which are present in at least 80% of the sequences; none of these regions correspond to known protein motifs. Bacterial identifiers and sequence accession numbers are as follows: *E. coli* (EC) (RdgC, AE000145; YejK, AE000308), *Salmonella typhimurium* (ST) (genomic sequence data), *Yersinia pestis* (YP) (genomic sequence data), *Haemophilus influenzae* (HI) (RdgC, U32716; YejK, U32765), *Pseudomonas aeruginosa* (PA) (genomic sequence data), *Actinobacillus actinomycetemcomitans* (AA) (genomic sequence data), *Vibrio cholerae* (VC) (RdgC, AF043352), *Bordetella pertussis* (BP) (genomic sequence data), *N. gonorrhoeae* (NG) (RdgC, AF058711), and *Neisseria meningitidis* (NM) (genomic sequence data). Genomic sequence data was obtained by searching the NCBI unfinished microbial genomes databases (6a). Significant regions at the N terminus of *V. cholerae* RdgC and *S. typhimurium* YejK and the C terminus of *A. actinomycetemcomitans* RdgC were not present in the currently available genomic sequence data. Sequence alignments were carried out using the MacDNASIS program (Hitachi Software).

DNA; however, factors other than DNA that are affected by DNA degradation could incorrectly indicate an association with DNA in a structure as complex as the spermidine nucleoid.

While the cellular functions of the two polypeptides are unclear, their conservation in other bacteria indicates their importance. The *rdgC* gene encodes a slightly acidic polypeptide (predicted pI of 5.1) of 303 amino acids (predicted mass of 33.9 kDa). This protein had not been observed in wild-type *E. coli*; however, a product of the correct size was expressed from a plasmid containing the *rdgC* gene and studied previously (9). That study identified *rdgC* as a gene required for the proper replication of DNA in cells deficient in the recombination enzymes RecABC and SbcCD. The authors proposed that RdgC may be an exonuclease that is involved in the removal of stalled replication forks; such a role would be consistent with our isolation of the polypeptide in a DNA-associated fraction. However, no biochemical evidence currently exists to confirm this proposed activity. We have identified no detectable protein motifs in the predicted RdgC sequence which might provide insight into its function; most notably, it lacks any of the known nucleotide binding sites that one might expect to find in a nuclease. The RdgC protein has significant homologs in nine diverse gram-negative bacteria, with levels of amino acid identity ranging from 35 to 92% and levels of overall similarity ranging from 48 to 95% (Fig. 2A). However, no homologous proteins were found in the complete genomes of four archaebacteria (*Archaeoglobus fulgidus*, *Methanococcus jannaschii*, *Pyrococcus horikoshii*, and *Methanobacterium thermoautotrophicum*) or in the complete genome of the gram-positive bacterium *Bacillus subtilis*. The absence of an RdgC homolog in *B. subtilis*, which contains all the components of the RecABC-SbcCD system (10), suggests that the RdgC protein family may carry out a different function than that previously proposed for the *E. coli* protein. Recent unpublished data states that the *Neisseria gonorrhoeae* homolog of RdgC is involved in pilin antigenic variation (4), which implies a possible role for RdgC in recombination.

The protein product of the *yejK* gene had not been previously detected. It encodes an acidic polypeptide (predicted pI of 4.4) of 335 amino acids (predicted mass of 37.8 kDa) which also contains no obvious protein motifs. There are strong homologies (28 to 92% identity and 45 to 95% similarity) between *E. coli* YejK and open reading frames from five bacteria (Fig. 2B). No homology was detected with any sequences from the genomes of the four archaebacteria previously mentioned or with sequences from the genome of *B. subtilis*. The level of sequence conservation of YejK argues that it plays a significant role in a shared cellular process; its isolation from spermidine nucleoids suggests that this function may involve interactions with nucleic acids.

We thank M. Belfort and R. Johnson for kindly supplying strains used in these studies.

REFERENCES

1. **Drlica, K.** 1987. The nucleoid, p. 91–103. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
2. **Hirel, P.-H., J.-M. Schmitter, P. Dessen, G. Fayat, and S. Blanquet.** 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA **86:**8247–8251.
3. **Johnson, R. C., C. A. Ball, D. Pfeffer, and M. I. Simon.** 1988. Isolation of the gene encoding the Hin recombinational enhancer binding protein. Proc. Natl. Acad. Sci. USA **85:**3484–3488.
4. **Mehr, I. J., and H. S. Seifert.** 1998. Unpublished data.
5. **Murphy, L. D., and S. B. Zimmerman.** 1997. Isolation and characterization of spermidine nucleoids from *Escherichia coli*. J. Struct. Biol. **119:**321–335.
6. **Murphy, L. D., and S. B. Zimmerman.** 1997. Stabilization of compact spermidine nucleoids from *Escherichia coli* under crowded conditions: implications for *in vivo* nucleoid structure. J. Struct. Biol. **119:**336–346.

6a. **National Center for Biotechnology Information.** January 1999, revision date. Microbial genomes databases. [Online.] http://www.ncbi.nlm.nih.gov/BLAST/unfin_databases.html. [20 January 1999, last date accessed.]

7. **Pettijohn, D. E.** 1990. Bacterial chromosome structure. Nucleic Acids Mol. Biol. **4:**152–162.
8. **Pettijohn, D. E., and R. R. Sinden.** 1985. Structure of the isolated nucleoid, p. 199–227. *In* N. Nanninga (ed.), Molecular cytology of *Escherichia coli*. Academic Press, London, England.
9. **Ryder, L., G. J. Sharples, and R. G. Lloyd.** 1996. Recombination-dependent growth in exonuclease-depleted *recBC sbcBC* strains of *Escherichia coli* K-12. Genetics **143:**1101–1114.
10. **Sharples, G. J., and R. G. Lloyd.** 1993. Location of the *Bacillus subtilis sbcD* gene downstream of *addAB*, the analogues of *E. coli recBC*. Nucleic Acids Res. **21:**2010.
11. **Trun, N. J., and J. F. Marko.** 1998. Architecture of a bacterial chromosome. ASM News **64:**276–283.
12. **Woldringh, C. L., and N. Nanninga.** 1985. Structure of nucleoid and cytoplasm in the intact cell, p. 161–197. *In* N. Nanninga (ed.), Molecular cytology of *Escherichia coli*. Academic Press, London, England.
13. **Zhang, A., S. Rimsky, M. E. Reaban, H. Buc, and M. Belfort.** 1996. *Escherichia coli* protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics. EMBO J. **15:**1340–1349.