The Journal of Microbiology, September 2004, p.200-204 Vol. 42, No. 3
Copyright © 2004, The Microbiological Society of Korea

A Proteomic Approach to Study msDNA Function in *Escherichia coli*

Mi-Ae Jeong and Dongbin Lim*

Department of Bioinformatics and Life Science Soongsil University Sangdodong, Dongjak, Seoul 156-743, KOREA

(Received May 31, 2004 / Accepted August 13, 2004)

Retron is a prokaryotic genetic element that produces multicopy single-stranded DNA covalently linked to RNA (msDNA) by a reverse transcriptase. It was found that cells producing a large amount of msDNA, rather than those that did not, showed a higher rate of mutation. In order to understand the molecular mechanism connecting msDNA production to the high mutation rate the protein patterns were compared by two dimensional gel electrophoresis. Ten proteins were found to be differentially expressed at levels more than three fold greater in cells with than without msDNA, nine of which were identified by MALDI TOF MS. Eight of the nine identified proteins were repressed in msDNA-producing cells and, surprisingly, most were proteins functioning in the dissimilation of various carbon sources. One protein was induced four fold greater in the msDNA producing cells and was identified as a 30S ribosomal protein S2 involved in the regulation of translation. The molecular mechanism underlying the elevated mutation in msDNA-producing cell still remains elusive.

Key words: msDNA, retron, reverse transcriptase, *Escherichia coli,* mutation rate

Retron is a generic name for a prokaryotic genetic element having reverse transcriptase (RT) (Temin, 1989). It produces a small single-stranded DNA (msDNA) covalently linked to RNA (msdRNA) (Furuichi *et al.*, 1987; Lim and Maas, 1989). Retrons are composed of genes for msDNA *(msd)*, msdRNA *(msr)* and RT (*ret*). These are located on the chromosome in the order of *msr, msd* and *ret*, and form a single operon (Lim and Maas, 1989). Some retrons have extra open reading frames, but their functions are largely unknown (Lim, 1991).

msDNA is synthesized thrbygh reverse transcription. A primary transcript covering *msr*, *msd* and *ret* serves as both a primer and template for reverse transcription (Lim and Maas, 1989). Since the 2'-OH group of a special guanine residue located at the end of a pair of inverted repeats of the template RNA functions as a primer for DNA synthesis, msDNA is covalently linked to the template via a 2'-5' phosphodiester bond. DNA synthesis is terminated by an unknown mechanism approximately at the middle of the template, leaving a part of the template RNA as an msdRNA (Lim and Maas, 1989). Recent whole genome sequencings of various bacteria have revealed that retron-type RT is present in a variety of prokaryotes including an Archea *Methanococcoides burtonii*. It is unknown if these RTs produce msDNA.

The biological function of msDNA that is relevant to the retron itself, or to the host bacterium, is unknown. The sole phenotypic difference observed so far is the elevated mutation occurring in msDNA producing cells (Maas *et al.*, 1994). When plasmid pT-14 producing a large amount of msDNA was introduced into *E. coli* CC107 with a frameshift mutation in *lacZ* (Cupples *et al.*, 1990), far more revertants were produced than in *E. coli* CC107 containing a control plasmid. The molecular mechanism connecting msDNA production to the high frequency of mutation is totally unknown.

In order to investigate the molecular bases of this msDNA-induced mutation, the protein pattern of *E. coli* CC107 (pT-14), the strain producing a large amount of msDNA with a high observed frequency of mutation, was compared to that of a control strain containing vector using two dimensional gel electrophoresis. Herein, the identification of differentially expressed proteins and their possible functions are presented and discussed.

* To whom correspondence should be addressed.
(Tel) 82-820-0452; (Fax) 82-828-7190
(E-mail) dblim@ssu.ac.kr

Materials and Methods

Sample preparation for 2-DE

Escherichia coli CC107 (pKK223) and CC107 (pT-14) were grown aerobically at 37°C in Luria Bertain (LB) broth containing ampicillin (50 μg/ml) (Lim and Maas, 1989). The growth was stopped in the late exponential phase at an OD_{600} of 0.8. Two hundred ml of the culture medium was centrifuged at 6,000 rpm for 15 min at 4°C, and the pellet was washed twice with a solution containing 10 mM Tris-HCl, pH 8.0, 1.5 mM KH_2PO_4, 68 mM NaCl and 9 mM NaH_2PO_4. The sample was resuspended

in 2 ml of buffer containing 10 mM Tris-HCl, pH 8.0, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 5 mM Pefabloc (Roche, USA) and 0.1% SDS (Lee *et al.*, 2002). The suspension was sonicated 4 times (5 s burst), 150 U of endonuclease (Sigma, USA) was added and left for 15 min at room temperature. After centrifugation at 12,000 rpm for 1 h at 4°C, the supernatant was recovered and subjected to 2-DE.

Two-dimensional gel electrophoresis

Isoelectric focusing was performed using a Pharmacia IPG-Phor Isoelectric focusing system. The sample was mixed with an appropriate amount of rehydration buffer, containing 8 M urea, 2% 3-[(3-chloramidopropyl) dimethylamino]-1-propanesulfonate (CHAPS), 0.5% IPG buffer (Amersham Pharmacia Biotech, USA), 20 mM dithiothreitol (DTT) and a few grains of bromophenol blue, and applied to 13 cm Immobiline DryStrips (Amersham Pharmacia Biotech, USA) by overnight rehydration in a strip holder. The isoelectric focusing was carried out for 44,700 Vh (Lee *et al.*, 2002). The gels were equilibrated for 10 min with equilibration buffer I containing 50 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol, 2% SDS, a grain of bromophenol blue and 65 mM DTT; then with equilibration buffer II that contained 50 mM Tris-Hc*l*, pH 8.8, 6 M urea, 30% glycerol, 2% SDS and 260 mM iodoacetamide, for 15 min. The proteins were separated onto 7.5-17.5% polyacrylamide gradient gels and then stained with silver (Blum *et al.*, 1987).

Image analysis

Gels were digitized using a GS710 Densitometer (Bio-Rad, USA) and analyzed by the PDQuest software package (version 6.2, BioRad, USA). The optical density of a given spot was normalized by dividing the absolute intensity of each spot by the sum of the intensities of all the spots on a particular gel. The normalization was carried out in such a way that the sum of the total intensity of a gel would be equal to 1,000,000. The intensity of an individual spot was expressed in ppm.

Protein identification by mass spectrometry

Protein spots were excised from a gel, washed with 50% acetonitrile in 0.1 M NH_4HCO_3 and dried under vacuum. The gel particles were reduced for 45 min at 56°C in 10 mM DTT/0.1 M NH_4HCO_3. After cooling, the DTT solution was immediately replaced with 55 mM iodoacetamide/0.1 M NH_4HCO_3. After 30 min incubation at room temperature in the dark, the gel particles were washed with 50% acetonitrile in 0.1 M NH_4HCO_3 and dried in a SpeedVac evaporator. The dried gel pieces were swollen in a minimum volume of 10 μl digestion buffer, containing 50 mM NH_4HCO_3, 5 mM $CaCl_2$ and 12.5 ng/μl of trypsin (Boehringer Mannheim, sequencing grade, USA), in an ice-cold bath. After 45 min, the supernatant was removed and replaced with the same buffer, without trypsin, to keep the gel pieces wet during enzymatic cleavage (37°C, overnight). One volume of a solution, composed of distilled water:acetonitrile:trifluoroacetic acid (93 : 5 : 2), was mixed with the digestion mixture. Then the sample was sonicated for 5 min and centrifuged for 2 min. Samples were analyzed using a PerSeptive Biosystems Voyager-DE STR MALDI-TOF mass spectrometer (Applied Biosystems, USA), equipped with a 337 nm N2 UV laser (Lee *et al.*, 2002). A database search was performed using ProteinProspector (http://prospector.ucsf.edu) and/or PROWL.

Results and Discussion

Comparison of bacterial growth

The only known difference so far produced by the presence of retron in *E. coli* is the elevated frequency of mutation observed in strains overproducing msDNA (Maas *et al.*, 1994). The reason for this high mutation rate caused by msDNA is unknown. In order to see if msDNA affects the growth of *E. coli*, the growth curves of *E. coli* CC107 harboring a vector plasmid pKK223 or msDNA-producing plasmid pT-14 (Lim and Maas, 1989) were compared. As shown in Fig. 1, the growth rate of *E. coli* CC107 harboring pT-14 [*E. coli* CC107(pT-14)] was somewhat lower that that of *E. coli* CC107 or CC107(pKK223). Moreover, the maximal cell density of *E. coli* CC107(pT-14) was only about 64% of that of *E. coli* CC107(pKK223) (OD_{600}=3.65 vs. 5.67). This growth defect may be due to the continuous cell death generated by mutation of an essential gene because of the high rate of mutation. The presence of an excess amount of single-stranded DNA in cells may cause unknown stress, which interferes with cell

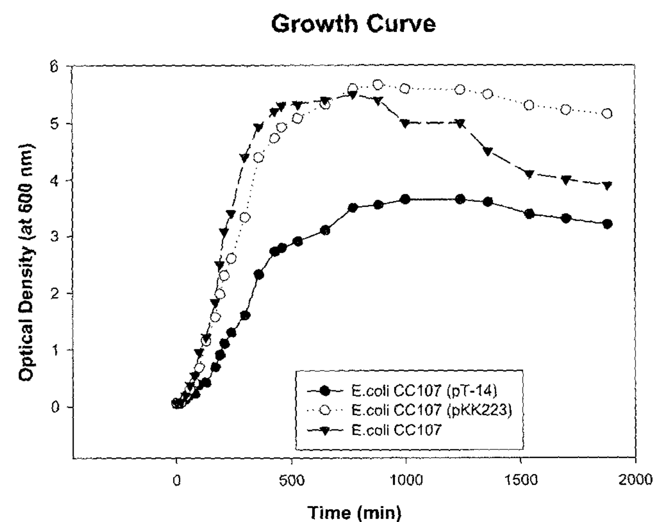

Fig. 1. Growth of *E. coli* CC107(pT-14) producing a large amount of msDNA. Growth of *E. coli* CC107(pT-14) was compared with those of *E. coli* CC107 and CC107(pKK223) to see the effect of msDNA on the bacterial growth.

growth. Microscopic observations showed no difference in the cell morphologies between the three strains.

Identification of differentially expressed proteins

To investigate why the mutation frequency was much higher in msDNA-producing cell than in the control, the protein pattern of msDNA producing *E. coli* CC107(pT-14) was compared with that of the control strain *E. coli* CC107 (pKK223) using two dimensional gel electrophoresis (2-DE). The above two strains were grown aerobically until the late logarithmic phase and the total proteins were separated by 2-DE (Fig. 2). The gels were scanned and analyzed by PDQuest, as described in Material and Methods.

Nine of the identified spots were decreased more than four-fold and a single spot increased more than four fold in the msDNA producing strain. Areas containing differentially expressed proteins were boxed and zoomed, as shown in Figs. 2 and 3, respectively. Among the 10 differentially expressed proteins, 9 were identified by peptide finger printing, using MALDI-TOF MS, and the results are summarized in Table 1.

Identified proteins

Spot 1 in *E. coli* CC107(pT-14) was repressed about four times than in the strain harboring the control vector (Fig. 3). It was identified by peptide fingerprinting as the glycerol dehydrogenase, the product of *gldA*. The expression of this enzyme was higher during the stationary phase than logarithmic growth, with full expression of *gldA* only achieved by induction with hydroxyacetone and under stationary-phase growth conditions (Truniger and Boos, 1994). No factors other than those are known to regulate *gldA* expression. The four fold repression observed in the msDNA producer implies that the msDNA system may

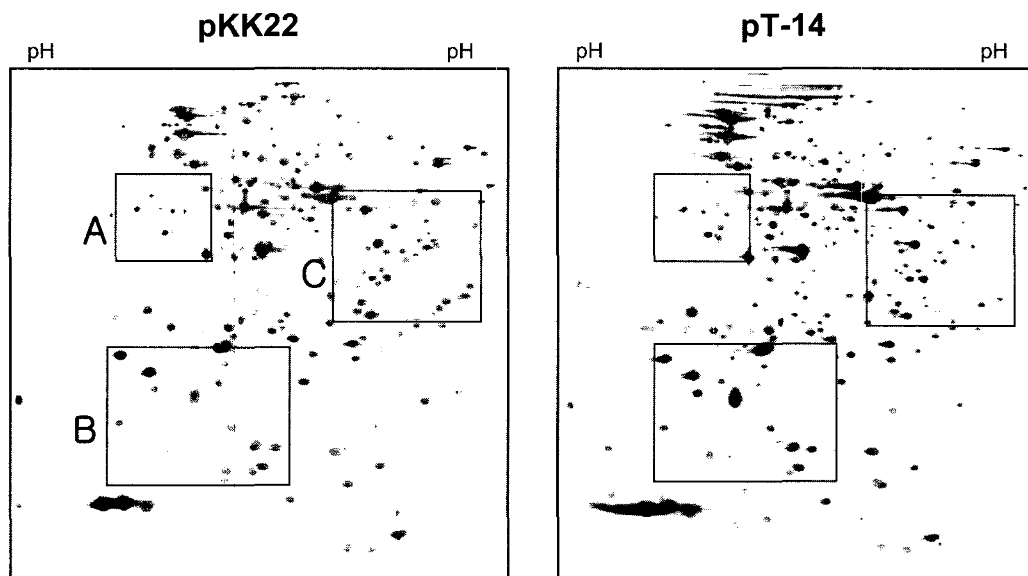

Fig. 2. Comparison of *E. coli* proteins between cells with and without msDNA production. Protein profiles of *E. coli* CC107(pKK223) and CC107(pT-14) were compared by 2-DE. About 60 μg of proteins were loaded on pH 4-7 IPG gel strips (130×3×0.5 mm) and separated onto 10-17.5% SDS-PAGE gel (140×160×0.75 mm), as described in Materials and Methods. Proteins were detected by silver staining.

Table 1. Identification of proteins differentially expressed in *E. coli* CC107(pKK223) and CC107(pT-14)

Spot No	Protein name	Accession No	Mr(Da)	pI	S.C (%)*
1	Glycerol dehydrogenase	NP_418380	38712.5	4.81	21
2	30 ribosomal protein S2	P02351	2673.9	6.61	26
3	TDCF protein	P42631	14007.9	5.06	29
4	Threonine dehydratase	NP_417587	35232.7	5.75	57
5	Mannose-permease enzyme IIAB component	P08186	35047.9	5.74	34
6	Maltose operon periplasmic protein precursor	P03841	29897.91	5.93	22
7	UDP-glucose-4-epimerase	BAA35421	37265.4	5.89	47
8	N-acetylneuraminate lyase subunit	P06995	32593.8	5.61	35
9	Unidentified				
10	Tagatose-bisphosphate aldolase GATY	NP_416599	30812.2	5.87	28

*Sequence coverage

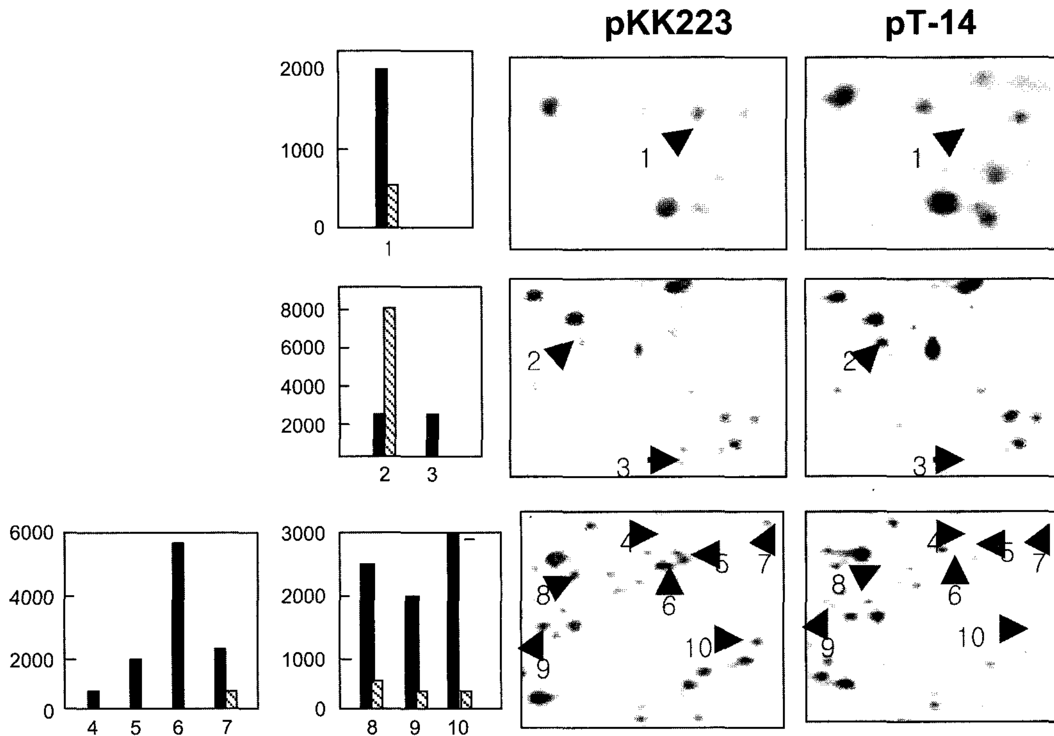

Fig. 3. Identification of proteins differentially expressed in *E. coli* CC107(pKK223) and CC107(pT-14). Parts of the two dimensional gels showing protein spots differentially expressed between *E. coli* CC107(pKK223) and in CC107(pT-14) were shown. Top, middle and bottom panels correspond to the region indicated by A, B and C in Fig. 2, respectively. The relative amount of each spot was calculated as a fraction of the optical density of the total spots on a gel, and the spot intensities, expressed as ppm, are given on the Y-axis of histograms (see Material and Methods). Filled and hatched bars represent the relative intensities of the matched spots from *E. coli* CC107(pKK223) and CC107(pT-14), respectively. The numbers on the X-axis indicate the spot numbers.

have some connection with the stationary control system.

Spot 2 was the only protein increased in the strain with msDNA. The spot density analysis showed that this spot was increased about four times more in *E. coli* CC107(pT-14) than in the strain with the vector only (Fig. 3). Mass spectrometric analysis revealed that it was a 30S ribosomal protein S2 encoded by *rpsB* (Tooze *et al.*, 1991) and depletion of this protein was shown to stimulate the translation of the leaderless mRNA of the phage λ *cI* gene (Shean and Gottesman, 1992, Moll *et al.*, 2002). Proteomic analysis revealed that, together with heat shock proteins, the level of RpsB was increased in *Bacillus subtilis* accumulating an outer membrane protein of *Neisseria meningitides* in the form of inclusion bodies (Jurgen *et al.*, 2001). An increase of RpsB protein in *E. coli* accumulating msDNA suggested that the large amount of short single-stranded DNA in the cytoplasm may cause a stress similar to that from the inclusion bodies formed of the overexpressed protein.

Spots 3 and 4 were low abundant proteins in the control strain and were not detected in *E. coli* CC107(pT-14) (Fig. 3). Mass spectrometric analysis identified spot 3 as a hypothetical protein TdcF. This protein is a member of the yjgF family, a group of proteins conserved throughout prokaryotes. In the protein database TdcF was annotated as a putative translation initiation inhibitor, but its function is unknown. Spot 4 was identified as anaerobic threonine dehydratase which catalyzes the dehydration of both L-threonine and L-serine to ammonia and the corresponding α-keto acids. Genes for threonine dehydratase, *tdcB* and *tdcF*, are components of the *tdc* operon for threonine and serine utilization (Hagewood *et al.*, 1994). Expression of the tdcABCDEFG operon was regulated in a complex way involving many factors including TdcA, TdcR, cAMP receptor protein (CRP), integration host factor (IHF) and histone-like protein HU (Hagewood *et al.*, 1994; Hesslinger, *et al.*, 1998). Repression of both *tdcB* and *tdcF* in *E. coli* CC107 (pT-14) clearly showed that overexpression of msDNA repressed the *tdc* operon, but the mechanism of regulation exerted by msDNA system is unclear.

Spot 5 was a low abundant protein in the control strain, but was not detected in the strain producing a high amount of msDNA (Fig. 3). This protein is a component of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) specific to mannose.

Interestingly, there was another protein that was repressed by msDNA and that also functions in the sugar transport system. Spot 6 was highly repressed by msDNA (Fig. 3), which was identified as the maltose operon periplasmic protein, MalM encoded by the last gene of malK

operon consisting of malK-lamB-malM, but the function is also unknown.

Spot 7, which was repressed about five fold by msDNA (Fig. 3), was identified as UDP-galactose 4-epimerase, a product of the *galE* gene. *GalE* is the first gene of the *galETKM* operon that codes for proteins required in galactose metabolism. This operon is activated by the cAMP-CRP complex and repressed by GalR and HU (Weickert and Adhya 1993; Choy *et al.*, 1995; Lia *et al.*, 2003), but how msDNA production affects its regulation is unclear.

Spot 8, identified as N-acetylneuraminate lyase or sialic acid lyase, was repressed about 5 fold in the msDNA producing strain. This enzyme catalyzes the cleavage of N-acetylneuraminic acid (sialic acid) to form pyruvate and N-acetylmannosamine. It is not clear how and why msDNA production represses this catabolic enzyme, whose only known function is the utilization of N-acetylneuraminate. The spot 9 was repressed about five fold in the msDNA producing strain than the control, but the protein could not be identified by mass spectrometry.

Spot 10, which was repressed seven-fold in *E. coli* CC107 (pT-14), was identified as tagatose-1,6-bisphospahte aldolase, a product of the gene *gatY. GatY* is the first gene of the *gatYZABCDR* operon for galactitol utilization (Nobelmann and Lengeler, 1995, 1996), and is regulated by *gatR* and catabolite repression mediated by cAMP-CRP (Ozbudak *et al.*, 2004).

Conclusion

Ten proteins with different expression levels between the msDNA-producing cell and control cell were identified. Interestingly, eight of the nine identified proteins were related to the utilization of various carbon sources, and repressed more than four fold in the msDNA-producing cells. How the msDNA production affects the expression of these diverse proteins remains to be elucidated. One possibility is that msDNA may affect the DNA-protein interaction of a global regulator, such as the HU protein. A single spot was induced in msDNA-producing cells, and identified as a 30 ribosomal protein S2 involved in the control of protein translation. The phenotypic differences observed in msDNA-producing cells were slow growth, low cell density in the stationary phase and a higher mutation rate. The mechanism underlying these physiological changes caused by msDNA production still remains elusive.

Acknowledgement

This work was supported by the Soongsil University Research Fund.

References

Blum, H., H. Beier, and H.J. Gross. 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8, 93-99.

Choy, H.E., S.W Park, P. Parrack, and S. Adhya. 1995. Transcription regulation by inflexibility of promoter DNA in a looped complex. *Proc. Natl. Acad. Sci. USA* 92, 7327-7331.

Cupples, C.G., M. Cabrera, C. Cruz, and J.H. Miller. 1990. A set of *lacZ* mutations in *Escherichia coli* that allow rapid detection of specific frameshit mutations. *Genetics* 125, 275-280.

Furuichi, T., S. Inouye, and M. Inouye. 1987. Biosynthesis and structure of stable branched RNA covalently linked to the 5' end of multicopy single-stranded DNA of *Stigmatella aurantiaca. Cell* 48, 55-62.

Hagewood, B.T., Y.L. Ganduri, and P. Datta. 1994. Functional analysis of the tdcABC promoter of *Escherichia coli*: roles of TdcA and TdcR. *J. Bacteriol.* 176, 6214-6220.

Hesslinger, C., S.A. Fairhurst, and G. Sawers. 1998. Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L-threonine to propionate. *Mol. Microbiol.* 27, 477-492.

Jurgen, B., R. Hanschke, M. Sarvas, M. Hecker, and T. Schweder. 2001. Proteome and transcriptome based analysis of *Bacillus subtilis* cells overproducing an insoluble heterologous protein. *Appl. Microbiol. Biotechnol.* 55, 326-332.

Lee, K., D. Bae, and D. Lim. 2002. Evaluation of parameters in peptide mass fingerprinting for protein identification by MALDI-TOF mass spectrometry. *Mol. Cells.* 13, 175-184.

Lia, G., D. Bensimon, V. Croquette, J.F. Allemand, D. Dunlap, D.E. Lewis, S. Adhya, and L. Finzi. 2003. Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc. Natl. Acad. Sci. USA* 100, 11373-11377.

Lim, D. and W.K. Maas. 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell* 56, 891-904.

Lim, D. 1991. Structure of two retrons of *Escherichia coli* and their common chromosomal insertion site. *Mol. Microbiol.* 5, 1863-1872.

Maas, W. K., C. Wang, T. Lima, G. Zubay, and D. Lim. 1994. Multicopy single-stranded DNAs with mismatched base pairs are mutagenic in *Escherichia coli. Mol. Microbiol.* 14, 437-441.

Moll, I., S. Grill, A. Grundling, and U. Blasi. 2002. Effects of ribosomal proteins S1, S2 and the DeaD/CsdA DEAD-box helicase on translation of leaderless and canonical mRNAs in *Escherichia coli. Mol. Microbiol.* 44, 1387-1396.

Nobelmann, B. and J.W. Lengeler. 1995. Sequence of the gat operon for galactitol utilization from a wild-type strain EC3132 of *Escherichia coli. Biochim. Biophys. Acta* 1262, 69-72.

Nobelmann, B. and J.W. Lengeler. 1996. Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. *J. Bacteriol.* 178, 6790-6795.

Ozbudak, E.M., M. Thattai H.N. Lim, B.I. Shraiman, and A. van Oudenaarden. 2004. Multistability in the lactose utilization network of *Escherichia coli. Nature* 427, 737-740.

Shean, C.S. and M.E. Gottesman 1992. Translation of the prophage lambda cl transcript. *Cell* 70, 513-522.

Temin, H. M. 1989. Retrons in bacteria. *Nature* 339, 254-255.

Truniger, V. and W. Boos. 1994. Mapping and cloning of *gldA*, the structural gene of the *Escherichia coli* glycerol dehydrogenase. *J. Bacteriol.* 176, 1796-1800.

Weickert, M.J. and S. Adhya. 1993. The galactose regulon of *Escherichia coli. Mol. Microbiol.* 10, 245-251.