

## **Analysis of nucleotide sequences in promoter regions of supercoiling-sensitive genes: How to relate microarray expression data with genomic sequences**

Galina I. Kravatskaya

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilov str., 32, Moscow 119991, Russia, gk@eimb.ru*

Vladimir R. Chechetkin,

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilov str., 32, Moscow 119991, Russia, chechet@eimb.ru*

Yury V. Kravatsky,

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilov str., 32, Moscow 119991, Russia, jiri@eimb.ru*

Vladimir G. Tumanyan

*Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Vavilov str., 32, Moscow 119991, Russia, tuman@eimb.ru*

The level of supercoiling in the chromosome can affect gene expression. Peter et al. [1] used microarrays representing nearly the entire genome of *E. coli* MG1655 and identified statistically significant changes in expression of 306 genes (about 7% of the genome) under relaxation of negative supercoiling. The expression of 106 genes increased upon chromosome relaxation (relaxation-induced genes), whereas the expression of 200 genes decreased (relaxation-repressed genes). To clarify the basis of supercoiling sensitivity, we analyzed the structural features of nucleotide sequences in the vicinity of promoters for the genes with expression enhanced and decreased in response to loss of chromosomal supercoiling in *Escherichia coli*. Fourier analysis [2,3,4] of promoter sequences for supercoiling-sensitive genes reveals the tendency in selection of sequences with helical periodicities close to 10 nt for relaxation-induced genes and to 11 nt for relaxation-repressed genes. The experimental data by Peter et al. [1] provide an opportunity to relate the expression level with the features observed in nucleotide sequences in the vicinity of the promoters. The relevant quantitative analysis is non-trivial: (i) the expression data are noisy; (ii) the level of expression varies (generally, non-monotonously) with time, temperature, or concentration of agents; and (iii)

the genes may be transcribed from several different promoters, and it is unknown which promoter is active in particular conditions. A special procedure [5] was developed for the study of correlations between the intensities of periodicities in the promoter sequences and the expression levels of corresponding genes. Significant correlations of expression with the AT content and with AT periodicities about 10, 11, and 50 nt indicate their role in regulation of supercoiling-sensitive genes. The developed scheme of correlational analysis is universal and may be applied to data mining in gene expression analysis with expression microarrays. Taking into account the huge data on DNA sequences and expression levels stored in databanks, we hope that our study may initiate the further regular investigations on the relationship between structural features in DNA sequences and gene expression levels.

This work was partially supported by the Molecular and Cellular Biology Program of the Presidium of the Russian Academy of Sciences and by RFBR, research project №12-04-01776 a.

1. B.J. Peter, J. Arsuaga, A. Breier, A.B. Khodursky, P.O. Brown, N.R. Cozzarelli (2004) Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*, *Genome Biol.*, **5**: R87.
2. V.R. Chechetkin, A.Y. Turygin, (1995) Search of hidden periodicities in DNA sequences, *J. Theor. Biol.*, **175**: 477–494.
3. V.V. Lobzin, V.R. Chechetkin (2000) Order and correlations in genomic DNA sequences. The spectral approach, *Physics-Uspokhi*, **43**: 55–78.
4. G.I. Kravatskaya, Y.V. Kravatsky, V.R. Chechetkin, V.G. Tumanyan (2011) Coexistence of different base periodicities in prokaryotic genomes as related to DNA curvature, supercoiling, and transcription, *Genomics*, **98**: 223–231.
5. G.I. Kravatskaya, V.R. Chechetkin, Y.V. Kravatsky, V.G. Tumanyan (2013) Structural attributes of nucleotide sequences in promoter regions of supercoiling-sensitive genes: How to relate microarray expression data with genomic sequences, *Genomics*, **101**: 1–11.