ORIGINAL ARTICLE PEER REVIEWED | OPEN ACCESS

Differences in the energy of DNA strands interaction in the near start-codon sequences of three bacterial genomes

Papirny Maxim A., Pasiuga Vladimir N., Shckorbatov Yuriy G.

ABSTRACT

Aims: The energy of DNA strands interaction (DSI) can regulate binding of enzymes and regulatory proteins to DNA in gene promoter and upstream region. The purpose of this work was to investigate the mean value of the energy DSI in the near promoter zone (± 200 nucleotides) of three bacterial genomes. Methods: We used the computer analysis for assessment of the energy of DSI in 90 in direct orientation and 90 genes in reverse orientation for every genome in proximity of start codon for *E. coli*, *B. subtilis*, and *S. typhimurium* genomes. Results: The DSI energy in genes of three tested bacterial species decreases in the prepromoter zone and comes to the elevated level in the coding zone of gene. The assessed mean energy of DSI differs in all tested species. It is higher (in absolute value) in genes of *E. coli* and *S. typhimurium* than in genes of *B. subtilis* in the prepromoter and promoter zone (area from -200 to -1). The DSI energy in the promoter and prepromoter zones (from -150 to -1 nucleotides) and in coding zone (from + 51 to + 150 nucleotides) is lower in genes of *S. typhimurium* and *B. subtilis* with reverse orientation than in directly oriented genes. In genes of E. coli such regularity was not observed. Conclusion: The proposed approach— the assessment of the DSI energy in the genomes is a perspective tool for determination differences in genotype between the organisms of different systematic position. For instance, the observed similarity in DSI in the near promoter zone between *E. coli* and *S. typhimurium* and difference in DSI to *B. subtilis* may be connected with the relative systematic proximity between *E. coli* and *S. typhimurium* and relative systematic distance of *B. subtilis* from these species.

Keywords: *Bacillus subtilis*, DNA strands interaction energy, *Escherichia coli*, Gene activity regulation, Gene orientation, *Salmonella typhimurium*

How to cite this article

Papirny MA, Pasiuga VN, Shckorbatov YG. Differences in the energy of DNA strands interaction in the near start-codon sequences of three bacterial genomes. Edorium J Bioinforma 2015;1:1–8.

Article ID: 100001B02PA2015

****

doi:10.5348/B02-2015-1-OA-1

Papirny Maxim A.[1], Pasiuga Vladimir N.[2], Shckorbatov Yuriy G.[3]

Affiliations: [1]Post-graduate student, National Scientific Centre "ISSAR" NAAS, Kharkiv, Ukraine; [2]Researcher, Research Institute of Biology, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine; [3]Head of Department of Genetics, Research Institute of Biology, V. N. Karazin Kharkiv National University, Kharkiv, Ukraine.

Corresponding Author: Yuriy G. Shckorbatov, Office - V. N Karazin Kharkiv National University, Institute of Biology, pl. Svobody 4, Kharkiv 61022, Ukraine; Ph:+380(57)7075262 (office); E-mail: yuriy. shckorbatov@gmail. com

Received: 24 May 2015
Accepted: 24 July 2015
Published: 08 September 2015

INTRODUCTION

The physical properties of DNA promoter sequences namely the energy of interaction of DNA strands in double helix is very important for regulation of gene activity [1]. The definite sites in bacterial promoter have very

conservative composition connected with their specific function. Stress-induced DNA duplex destabilization is closely associated with specific promoter regions [2]. The -10 motifs binds the transcription initiation complex, while the -35 motif participate in regulation of transcription activity of promoters. In the region around position -50 is located A+T rich UP element [3]. Most of the promoter prediction programs search sequences for conserved -10 motifs, and in some cases also include -35 motifs [4, 5]. The ability to DNA melting in the promoter zone is regarded as very important characteristics that influence on gene activity [6]. The theoretical model proposed by Peyrard, Bishop, and Dauxois [7], accurately describes experiments with DNA denaturation, and the formation and role of bubble states in the premelting regime [8].

There are differences in nucleotide composition between genes in direct and reverse orientation. Leading and lagging strands of DNA are asymmetric in nucleotide contents. In the broad study of bacterial genomes of different species done by Lobry and Sueoka was demonstrated that most of the 43 chromosomes taken under consideration were relatively enriched in G over C and in T over A and slightly depleted in G+C in the leading strand compared with the lagging strand [9]. The problems of strand asymmetry are analyzed in many reviews [9–14]. In bacterial genomes some of genes are located in the leading strand, the direction of their transcription is the same as direction of replication (directly oriented genes) and some of genes are located in the lagging strand (genes with reverse orientation) [13]. The frequency of leading strand genes is ~75 % in *B. subtilis* [15], but only ~55 % in *E. coli* genome [16]. Such differences in gene location may be connected with RNA polymerase properties of these species. In *B. subtilis* there are two different RNA polymerases for leading and lagging strands [17]. In *E. coli* the two core polymerases are not pre-dedicated to one strand and can be interchanged [18]. As known the promoter region enriched with AT-nucleotide which enables linkage RNA-synthesizing machinery. In our previous investigation, it was demonstrated that promoters with minimal and maximal energy have low activity (strengths), and the strongest promoters are characterized by intermediate mean energy of double strand interaction (DSI) values [19]. The minimum of DSI is applied as a parameter for search of promoters in DNA sequences [19]. In connection with the mentioned above findings we studied the energy of DSI in the promoter and near-promoter area in genes with direct and reverse orientation of *E. coli*, *S. typhimurium*, and *B. subtilis* [20].

Two of bacterial species – *E. coli* and *S. typhimurium* are evolutionary closer, and the other species – *B. subtilis* located more distantly from these in evolutionary sense. The systematic position and evolutionary connections between the species analysed in this work are presented in Figure 1 [21]. The investigation of DSI in these species is a convenient model for application of the proposed experimental approach.

MATERIALS AND METHODS

The DNA sequences of *E. coli* K12, *S. typhimurium*, and *B. subtilis*, obtained from KEGG database (http:// www. genome. jp/kegg/kegg2. html) were analyzed. The DNA sequences of 90 genes with direct orientation and 90 genes with reverse orientation were investigated for every species. The number 90 was selected because it is the maximal number of genes with long (200 nucleotides) intergenic sequences in *S. typhimurium*, and *B. subtilis*. In *E. coli* genes with 200 nucleotides intergenic sequences were taken from the beginning of chromosome without any selection. The DNA sequences located in a gene promoter zone from -200 to +200 nucleotides were investigated. The nucleotide composition and the energy of DSI of promoter and coding DNA sequences in direct and reverse orientation were analyzed. For assessment of the energy of DSI in the DNA double helix we assessed two types of interaction between nucleotides: hydrogen bonds and base staking interaction. The energy of interaction due to hydrogen bonds was assumed: T-A -68, 8 kJ/ mol and G-C -115, 1 kJ/mol [22]. Values of free energy of coaxial stacking range from -0.85 kcal/mol for TA dinucleotide stack to -2.76 kcal/mol for GC, free energy of base-stacking interactions is presented in Table 1 [23]. The analogous approach was applied for DNA strands interaction energy oligonucleotides [24]. The impact of stacking energy in the total DSI is relatively small compared to hydrogen-bond energy, but it also take into consideration properties of the neighbouring nucleotides. We calculated the mean energy of DSI per one nucleotide in sequences of 50 nucleotides (spans: -200 to -151; -150 to -101; 100 to -51; 50 to -1 and 1 to 50; 51 to 100; 101 to 150; 151 to 200). We used sliding window with breadth 10 nucleotides and mean DSI energy was summarized for every 50-nucleotide span. The mean data of DSI energy are presented in Figures 2–4.

To show the distribution of energy of DSI along DNA the nucleotide sequence was applied the method of sliding window with 10 nucleotides width. The mean energy of DSI was calculated for 10 nucleotides and was ascribed to the fifth nucleotide in the window. After this we calculated the mean data for 90 sequences (Figures 5–8). As a result we have obtained the energy distribution along nucleotide sequence for directly and reversely oriented genes separately. To determine the probability of the difference between spans of DNA we applied a statistical processing by *t*-student and ANOVA tests. The difference between variances with probability level $p<0.05$ are marked with asterisks (*).

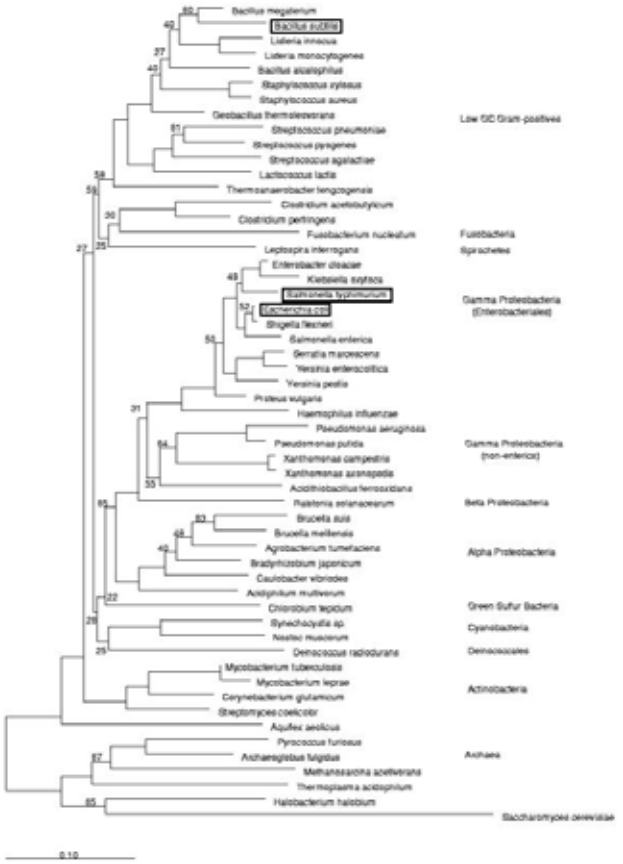


Figure 1: Phylogenetic tree based on 16S rRNA gene sequences [21].

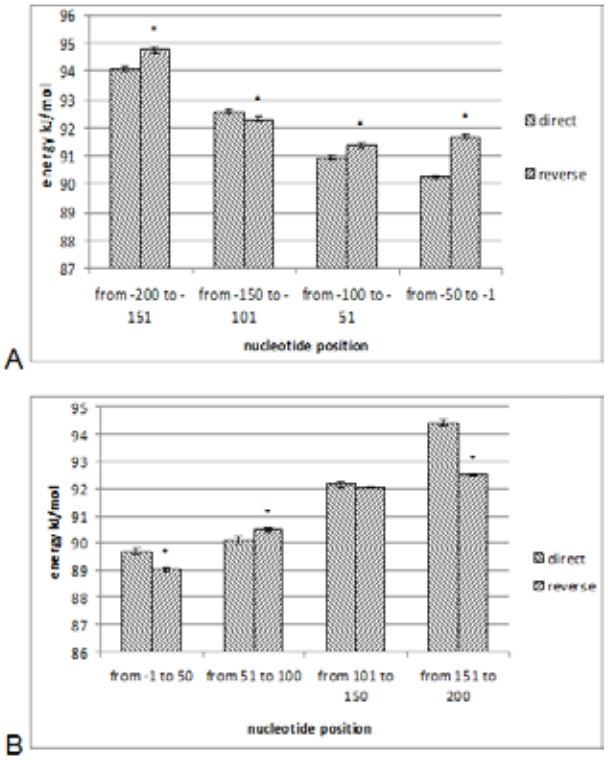


Figure 2: The DSI energy (absolute value) in the non-coding region (A) and the coding region, (B) in genes of *Escherichia coli* with direct and reverse orientation, mean data for DNA sequences of 50 nucleotides long.

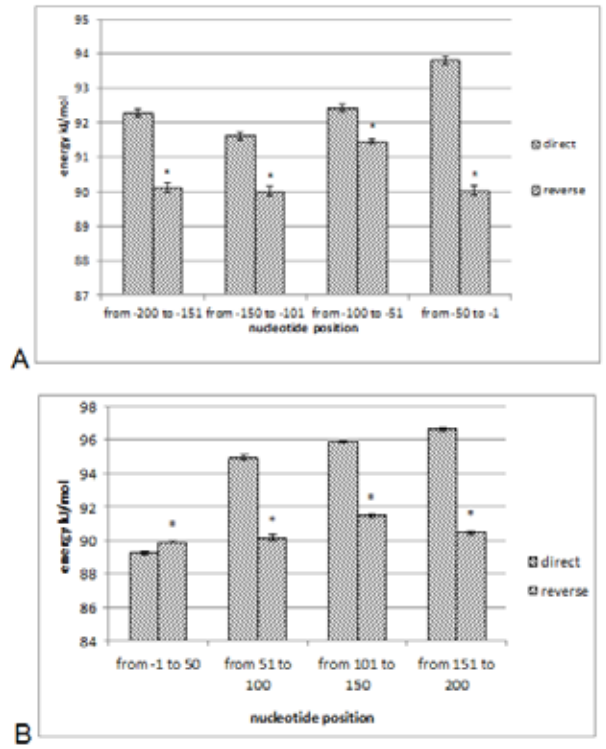


Figure 3: The energy of DSI (absolute value) in the non-coding region (A) and the coding region, (B) of genes of *Salmonella typhimurium* with direct and reverse orientation, mean data for segments of DNA sequences of 50 nucleotides long.

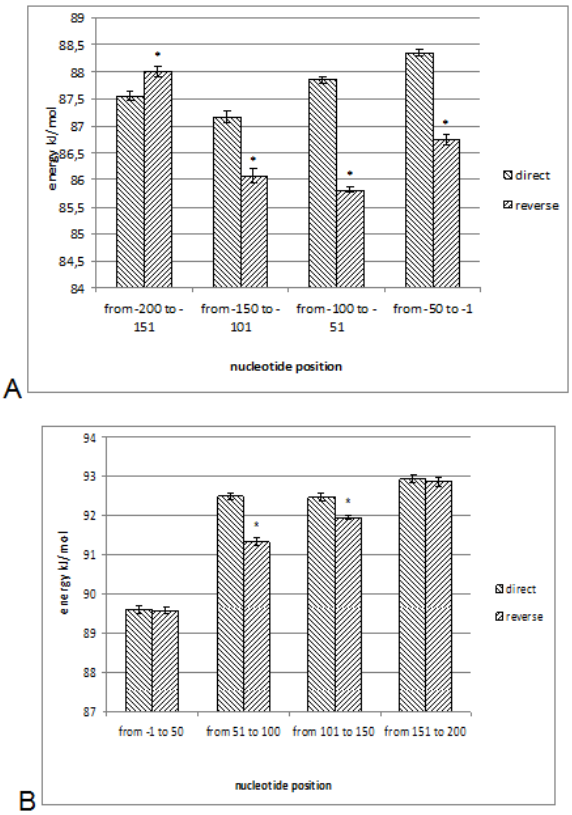


Figure 4: The energy of DSI (absolute value) in the non-coding region (A) and the coding region (b) of genes of *Bacillus subtilis* with direct and reverse orientation, mean data for segments of DNA sequences of 50 nucleotides long.

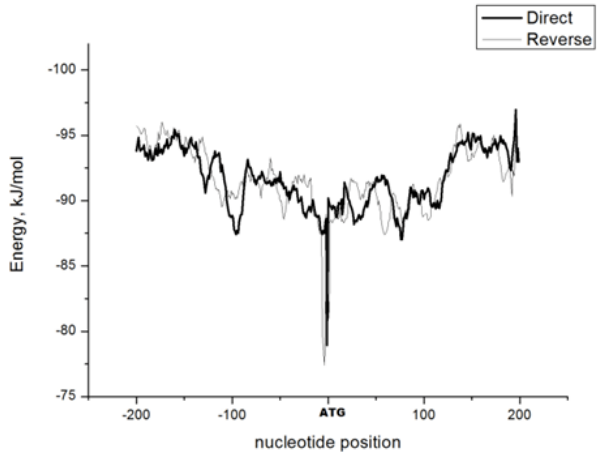


Figure 5: The DSI energy in DNA strands in genes with direct and reverse orientation (*Escherichia coli*).

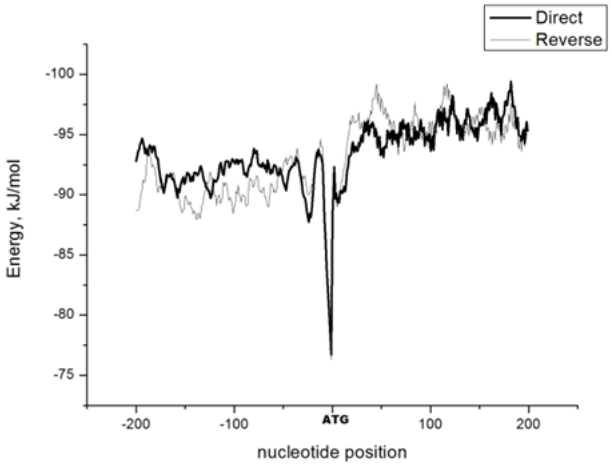


Figure 6: The DSI energy in genes with direct and reverse orientation in *Salmonella typhimurium*.

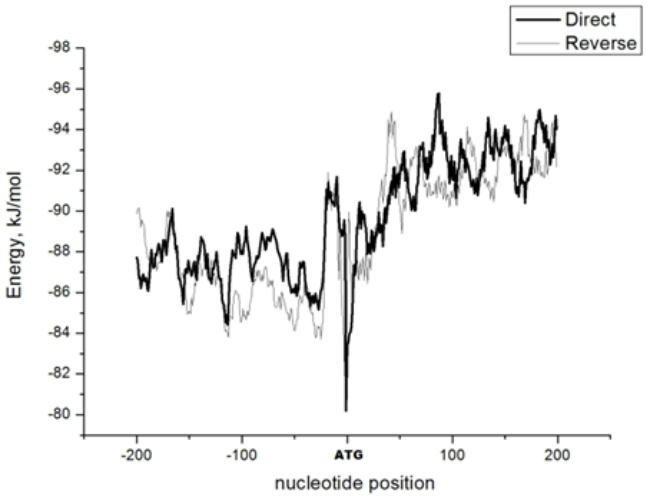
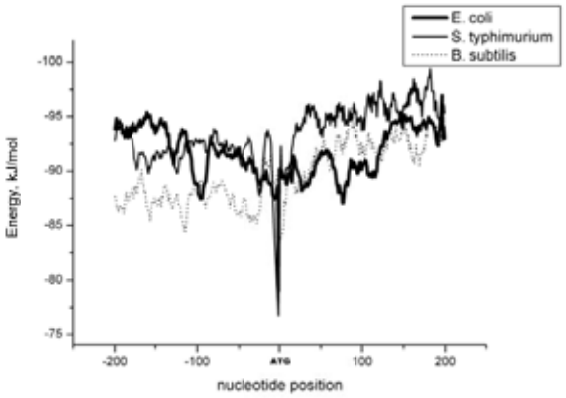
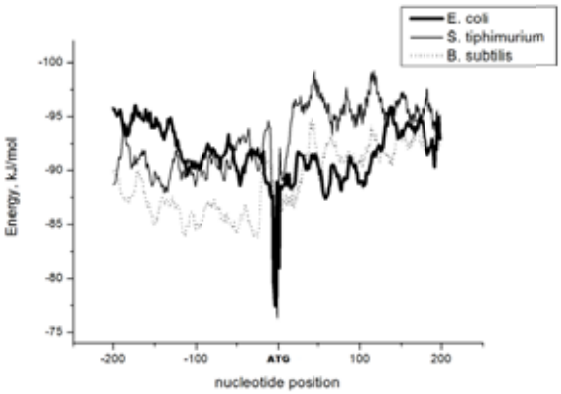


Figure 7: The DSI energy in DNA of genes with direct and reverse orientation in *Bacillus subtilis*.



(a)



(b)

Figure 8: The DSI energy of three organisms (*E. coli, S. typhimurium, B. subtilis*) in genes with direct (a) and reverse (b) orientation.

RESULTS AND DISCUSSION

The Energy of DNA Strands Interaction in *Escherichia coli* Genes

The data of DSI energy in *E. coli* are presented in Figure 5. As one can see, it gradually decreases in the near promoter and promoter area from -150 to +1 nucleotide and it is minimal in promoter area. The data of analysis of DNA strands energy interaction in 50 nucleotides spans are presented in Figure 3. These data indicate a statistically significant decrease of absolute value the energy of interaction of DNA strands in the region from -100 to -1 (Figure 2A) that supports conclusion based on the data of Figure 5. For convenience, we present the absolute value of DSI energy. As one can see, in the coding region the energy of DNA strands is decreased in the spans from 1 to 50 and from 50 to 100 nucleotides for both types of genes. The energy of DSI is lower in genes with direct orientation in spansfrom-200 to – 151, from-100 to -1, and from +50 to +101 regions, but in the region from-150 to -101, from + 1 to +50, and from + 151 to +200 the DSI energy in directly oriented genes is higher.

The energy of DNA strands interaction in *Salmonella typhimurium* genes

Figure 6 shows data of analysis of DNA DSI energy of *Salmonella typhimurium* LT2 genes in the direct and reverse orientation. From the data of Figure 6 it may be seen that the energy of DSI has two distinct minima (in absolute value) at the -35 nucleotide positions, and around -10, in sites of the binding of regulatory proteins and RNA polymerase.

As one can see from Figure 3, in genes of reverse orientation the DSI energy is less in genes with reverse orientation than in genes with direct orientation from -200 to-1 and from +51 to +200, in absolute value. On the contrary, the energy of DNA interaction in genes with direct orientation is less in the zone from -1 to +50, in absolute value.

The energy of DNA strands interaction in *Bacillus subtilis* genes

The data of analysis of the energy of DSI in Bacillus subtilis are presented in Figures 4 and 7.

The DSI energy is less in prepromoter region than in coding region of gene and minimum of DSI energy is observed in promoter region (Figure 7).

It the spans from -150 to +1 and from +51 to +150 the genes with direct orientation have higher DSI energy (in absolute value), as it is seen in Figure 4A–B. In the region from -150 to + 1 the DSI energy is lower (in absolute value) in genes with reverse orientation.

The DSI energy of DNA 50 nucleotide stretches was analysed by means of ANOVA statistics. It was shown the significant difference (p>0, 001) in DSI energy of subsequent DNA 50 nucleotide stretches (factor P) in three bacterial species (Table 2). The significant differences in DSI energy were also revealed between three bacterial species (dependent factor B in Table 2). The direction of gene (independent factor D in Table 2) also significantly (p=0, 044) influences on the DSI energy.

If we compare the DSI energy of 50 nucleotide stretches in bacterial species by pairs: 1 – *S. typhimurium* – *B. subtilis*; 2 – *S. typhimurium* – *E. coli*; 3 – *B. subtilis* – *E. coli* by ANOVA statistical method we can see the significant differences in mean DSI energy between bacterial species. It is true for genes as in direct and also in reverse orientation (Tables 3–5).

The comparative analysis of energy of DNA strands interaction in three bacterial species

The DSI energy in genes with direct and reverse orientation for three organisms: *Escherichia coli*, *Salmonella typhimurium*, and *Bacillus subtilis* are

Table 1: The values of stacking free energy (ΔG, kJ/mol)

Nucleotides	A	T	G	C
A	-4. 65	-5. 61	-4. 44	-7. 58
T	-0. 80	-4. 64	-2. 30	-5. 99
G	-5. 98	-7. 57	-6. 02	-9. 08
C	-2. 30	-4. 43	-3. 80	-6. 02

Table 2: ANOVA statistics for DNA strands interaction (DSI) energy in 50-nucleotide stretches

Vari ants	F	p
D	4	0, 044347
P	935	0, 000000
D*P	11	0, 000000
B	5354	0, 000000
B*D	29	0, 000000
B*P	66	0, 000000
B*D*P	18	0, 000000

Independent variable D – direction of gene: 1 – direct, 2 – reverse; independent variable P – stretches of nucleotides: 1 – from -200 to -151; 2 – from -150 to -101; 3 – from -100 to -51; 4 – from -50 to -1; 5 – from 1 to 50; 6 - from 51 to 100; 7 – from 101 to 150; 8 – 151 to 200); dependent variable B – species of bacteria: 1 – *E. coli*; 2 – *B. subtilis*; 3 – *S. typhimurium*

Table 3: ANOVA statistics for DNA strands interaction (DSI) energy of S. typhimurium and B. subtilis genes

Variants	F	p
Direct		
P	362	0, 000000
B	5743	0, 000000
B*P	17	0, 000000
Reverse		
P	485	0, 000000
B	5861	0, 000000
B*P	20	0, 000000

Independent variables P – groups of nucleotide positions: 1 – from -200 to -151; 2 – from -150 to -101; 3 – from -100 to -51; 4 – from -50 to -1; 5 – from 1 to 50; 6 - from 51 to 100; 7 – from 101 to 150; 8 – 151 to 200; dependent variables B: 1 – *S. typhimurium*; 2– *B. subtilis*

Table 4: ANOVA statistics for DNA strands interaction (DSI) energy of *S. typhimurium* and *E. coli* genes

Variants	F	p
Direct		
P	342	0, 000000
B	296	0, 000000
B*P	55	0, 000000
Reverse		
P	387	0, 000000
B	129	0, 000000
B*P	33	0, 000000

Independent variable P – groups of nucleotide stretches: 1 – from -200 to -151; 2 – from -150 to -101; 3 – from -100 to -51; 4 – from -50 to -1; 5 – from 1 to 50; 6 - from 51 to 100; 7 – from 101 to 150; 8 – 151 to 200; dependent variable B: 1 – S. typhimurium 2- E. coli

Table 5: ANOVA statistics for DNA strands interaction (DSI) energy in *B. subtilis* and *E. coli* genes

Variants	F	p
Direct		
P	372	0, 000000
B	2176	0, 000000
B*P	43	0, 000000
Reverse		
P	456	0, 000000
B	3734	0, 000000
B*P	70	0, 000000

Independent variable P – groups of nucleotide stretches: 1 – from -200 to -151; 2 – from -150 to -101; 3 – from -100 to -51; 4 – from -50 to -1; 5 – from 1 to 50; 6 - from 51 to 100; 7 – from 101 to 150; 8 – 151 to 200. Dependent variable B: 1– *B. subtilis*; 2– *E. coli*

presented for genes in direct (Figure 8a) and reverse orientation (Figure 8b).

As it can be seen from Figure 8, showing the mean DSI values in prepromoter region (from -200 to -1) it may be seen that the pattern of energy distribution along DNA is closer in *E. coli* and *S. typhimurium*, than in *B. subtilis*. The mean DSI energy from -200 to -1 in genes with direct and reverse orientation is lower (in absolute value) in genes of *B. subtilis*, than in genes of *E. coli* and *S. typhimurium* (Figure 9). This presumably is connected with relatively distant systematic position of *B. subtilis* from *E. coli* and S. typhimurium.

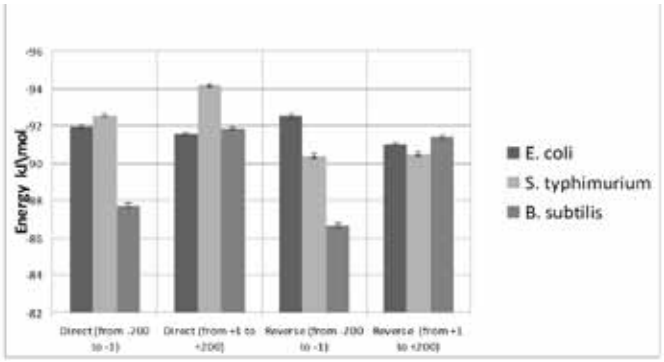


Figure 9: The mean DSI energy of three organisms (*E. coli*, *S. typhimurium*, *B. subtilis*) in genes with direct and reverse orientation. The mean values for 200 nucleotide sequences (n = 90) ± SE are presented. Differences between all variants are statistically safe ($p < 0.05$).

The differences in the mean DSI energy between bacterial species may be connected with functional activity ofgenes, because the analysis of DSI revealed DSI energy decrease (in absolute value) near around -55, -35, -10 and +6 nucleotide positions, as it is known, these positions are very important in connection to transcriptional activity regulation [25].

The further investigation of the patterns of DSI energy may be done in connection with the ecologically conditions in which the species exist. Also such investigation may be done in connection with the evolutionary history of species.

CONCLUSIONS

1. The energy of DNA strands interaction (DSI) in genes of three tested bacterial species decreases in the prepromoter zone and comes to the elevated level in the coding zone of gene.
2. The analyzed groups of genes of three bacterial species: *E. coli*, *S. typhimurium* and *B. subtilis* significantly differ in the mean values of DSI energy in areas before and after beginning of the gene (from -200 to +200 nucleotides).
3. The DSI energy in the promoter and prepromoter zones (from -150 to -1 nucleotides) and also in coding zone (from + 51 to + 150 nucleotides) is lower in genes of *S. typhimurium* and *B. subtilis* with reverse orientation than in directly oriented genes. In genes of *E. coli* such regularity was not observed.
4. The mean DSI energy in the prepromoter and promoter zones (from -200 to -1 nucleotides) is lower (in absolute value) in *B. subtilis* genes than in genes of *E. coli* and *S. typhimurium*. The observed difference may be connected with relative systematic proximity of *E. coli* and *S. typhimurium* and relative systematic distance of *B. subtilis*.
5. The proposed approach: the assessment of the

DSI energy in the genomes is a perspective tool for determination differences in genotype between the organisms of different systematic position. For instance, the observed similarity in DSI in the near promoter zone between *E. coli* and *S. typhimurium* and difference in DSI to *B. subtilis* correlates with their systematic position.

Source of Support

Kharkiv National University, Kharkiv, Ukraine

****

Author Contributions

Papirny Maxim A. – Substantial contributions to conception and design, Acquisition of data, Analysis and interpretation of data, Drafting the article, Critical revision of the article, Final approval of the version to be published
Pasiuga Vladimir N. – Substantial contributions to conception and design, Acquisition of data, Analysis and interpretation of data, Drafting the article, Critical revision of the article, Final approval of the version to be published
Shckorbatov Yuriy G. – Substantial contributions to conception and design, Acquisition of data, Analysis and interpretation of data, Drafting the article, Critical revision of the article, Final approval of the version to be published

Guarantor

The corresponding author is the guarantor of submission.

Conflict of Interest

Authors declare no conflict of interest.

Copyright

REFERENCES

1. Alexandrov BS, Gelev V, Yoo SW, et al. DNA dynamics play a role as a basal transcription factor in the positioning and regulation of gene transcription initiation. Nucleic Acids Res 2010 Apr;38(6):1790–5.
2. Wang H, Benham CJ. Promoter prediction and annotation of microbial genomes based on DNA sequence and structural responses to superhelical stress. BMC Bioinformatics 2006 May 5;7:248.
3. Harley CB, Reynolds RP. Analysis of E. coli promoter sequences. Nucleic Acids Res 1987 Mar 11;15(5):2343–61.
4. Helmann JD, deHaseth PL. Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. Biochemistry 1999 May 11;38(19):5959–67.
5. Huerta AM, Collado-Vides J. Sigma70 promoters in Escherichia coli: specific transcription in dense regions of overlapping promoter-like signals. J Mol Biol 2003 Oct 17;333(2):261–78.
6. Hawley DK, McClure WR. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res 1983 Apr 25;11(8):2237–55.
7. Dauxois T, Peyrard M, Bishop AR. Entropy-driven DNA denaturation. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 1993 Jan;47(1):R44–R47.
8. Zeng Y, Montrichok A, Zocchi G. Bubble nucleation and cooperativity in DNA melting. J Mol Biol 2004 May 21;339(1):67–75.
9. Lobry JR, Sueoka N. Asymmetric directional mutation pressures in bacteria. Genome Biol 2002 Sep 26;3(10):RESEARCH0058.
10. Tillier ER, Collins RA. Replication orientation affects the rate and direction of bacterial gene evolution. J Mol Evol 2000 Nov;51(5):459–63.
11. Morton RA, Morton BR. Separating the effects of mutation and selection in producing DNA skew in bacterial chromosomes. BMC Genomics 2007 Oct 12;8:369.
12. Rocha EP, Danchin A, Viari A. Universal replication biases in bacteria. Mol Microbiol 1999 Apr;32(1):11–6.
13. Tillier ER, Collins RA. The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. J Mol Evol 2000 Mar;50(3):249–57.
14. Mrázek J, Karlin S. Strand compositional asymmetry in bacterial and large viral genomes. Proc Natl Acad Sci U S A 1998 Mar 31;95(7):3720–5.
15. Kunst F, Ogasawara N, Moszer I, The complete genome sequence of the gram-positive bacterium Bacillus subtilis. Nature 1997 Nov 20;390(6657):249–56.
16. Blattner FR, Plunkett G 3rd, Bloch CA, et al. The complete genome sequence of Escherichia coli K-12. Science 1997 Sep 5;277(5331):1453–62.
17. Dervyn E, Suski C, Daniel R, et al. Two essential DNA polymerases at the bacterial replication fork. Science 2001 Nov 23;294(5547):1716–9.
18. Yuzhakov A, Turner J, O'Donnell M. Replisome assembly reveals the basis for asymmetric function in leading and lagging strand replication. Cell 1996 Sep 20;86(6):877–86.
19. Berezhnoy AY, Shckorbatov YG. Dependence of the E. coli promoter strength and physical parameters upon the nucleotide sequence. J Zhejiang Univ Sci B 2005 Nov;6(11):1063–8.
20. Morey C, Mookherjee S, Rajasekaran G, Bansal M. DNA free energy-based promoter prediction and comparative analysis of Arabidopsis and rice genomes. Plant Physiol 2011 Jul;156(3):1300–15.

21. Jackson CR, Dugas SL. Phylogenetic analysis of bacterial and archaeal arsC gene sequences suggests an ancient, common origin for arsenate reductase. BMC Evol Biol 2003 Jul 23;3:18.
22. Sponer J, Jurecka P, Hobza P. Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. J Am Chem Soc 2004 Aug 18;126(32):10142–51.
23. Protozanova E, Yakovchuk P, Frank-Kamenetskii MD. Stacked-unstacked equilibrium at the nick site of DNA. J Mol Biol 2004 Sep 17;342(3):775–85.
24. Pyshnyi DV, Pyshnaya I, Levina A, et al. Thermodynamic analysis of stacking hybridization of oligonucleotides with DNA template. J Biomol Struct Dyn 2001 Dec;19(3):555–70.
25. Berezhnoy AY, Shckorbatov YG, Hisanori K. Regularities in the E. coli promoters composition in connection with the DNA strands interaction and promoter activity. J Zhejiang Univ Sci B 2006 Dec;7(12):969–73.