

УДК 577.21

ПОИСК МИНИМУМОВ ЭНЕРГИИ ВЗАИМОДЕЙСТВИЯ КОМПЛЕМЕНТАРНЫХ ЦЕПЕЙ ДНК В ПРОМОТОРНОЙ И КОДИРУЮЩЕЙ ЧАСТИ ГЕНОВ *E. COLI*

Ю.Г. Шкорбатов¹, М.А. Папирный², М.В. Кулик²

¹ Харьковский национальный университет, институт биологии

г. Харьков, пл. Свободы 4, 61077, Украина

yury.g.shckorbatov@univer.kharkov.ua

² Харьковский национальный университет, кафедра молекулярной биологии и биотехнологии

г. Харьков, пл. Свободы 4, 61077, Украина

max_papirny@mail.ru

Поступила в редакцию 30 ноября 2009 г.

Принята 8 апреля 2010 г.

Проведен компьютерный анализ нуклеотидного состава и энергии взаимодействия цепей ДНК в промоторных и кодирующих участках генов *E. coli*. Было показано увеличение содержания тимина и аденина в зоне промотора. Значительных отличий между генами в прямой и обратной ориентации мы не обнаружили. Также обнаружены минимумы энергии взаимодействия комплементарных цепей ДНК в прямой и обратной ориентациях в промоторной и кодирующей областях генов *E. coli*; вблизи следующих позиций: -100, -70, -40, +5, +20, +30. Участки -40, +5 близки к известным консенсусным последовательностям нуклеотидов в промоторе, влияющим на активность гена. Участки: -100, -70, +20, +30, по нашим данным, также имеют сниженную энергию взаимодействия комплементарных цепей ДНК, что позволяет предположить, что они имеют регуляторное значение. Энергия взаимодействия комплементарных цепей ДНК снижена не только в области промотора, но и в начале кодирующей части гена. Рост в энергии взаимодействия ДНК в кодирующей части гена наблюдается после 100-ой позиции. Энергия взаимодействия цепей ДНК ниже в обратных, чем в прямых последовательностях. Отличия между прямыми и обратными последовательностями статистически достоверны в области от -150 до -101 и от 1 до 51. Гены в обратной ориентации имеют дополнительные минимумы энергий в позициях: -140, -120, -25 и +40. Кроме того, в генах с обратной ориентацией имеется минимум энергии взаимодействия цепей ДНК в области -85, в то время как в генах с прямой ориентацией минимум энергий имеется в области -80.

КЛЮЧЕВЫЕ СЛОВА: энергия взаимодействия цепей ДНК, состав ДНК, промотор, ориентация генов.

ПОШУК МІНІМУМІВ ЕНЕРГІЇ ВЗАЄМОДІЇ КОМПЛЕКСІВ КОМПЛЕМЕНТАРНИХ ЛАНЦЮГІВ ДНК У ПРОМОТОРНИХ І КОДУЮЧИХ ЧАСТИНАХ ГЕНІВ *E. COLI*

Ю.Г. Шкорбатов, М.А. Папирний, М.В. Кулік

Проведено комп'ютерний аналіз нуклеотидного складу та енергії взаємодії ланцюгів ДНК в промоторних і кодуючих ділянках генів *E. coli*. Було показано збільшення вмісту тиміну і аденіну в зоні промотора. Значних відмінностей між генами в прямій і зворотній орієнтації не виявлено. Виявлені мінімуми енергії взаємодії комплементарних ланцюгів ДНК в прямій і зворотній орієнтаціях в промоторних і кодуючих областях генів *E. coli* поблизу наступних позицій: -100, -70, -40, +5, +20, +30. Сайти -40, +5 близьки до відомих консенсусних послідовностей нуклеотидів в промоторі, що впливають за активність гена. Ділянки: -100, -70, +20, +30, за нашими даними, також мають знижену енергію взаємодії комплементарних ланцюгів ДНК, що дозволяє припустити, що вони мають регуляторний значення. Енергія взаємодії комплементарних ланцюгів ДНК знижена не тільки в області промотора, а й на початку кодуючої частини гена. Зростання в енергії взаємодії ДНК в кодуючої частини гена спостерігається після 100-ї позиції. Енергія взаємодії ланцюгів ДНК нижче в зворотних, ніж у прямих послідовностях. Відмінності між прямими та зворотними послідовностями статистично достовірні в області від -150 до -101 та від 1 до 51. Гени у зворотній орієнтації мають додаткові мінімуми енергій в позиціях: -140, -120, -25 і +40. Крім того, в генах зі зворотним орієнтацією є мінімум енергії взаємодії ланцюгів ДНК в області -85, в той час, як у генах з прямою орієнтацією мінімум енергій є в області -80.

КЛЮЧОВІ СЛОВА: енергія взаємодії ланцюгів ДНК, склад ДНК, промотор, орієнтація генів.

SEARCH OF THE MINIMUMS OF ENERGY OF INTERACTION BETWEEN COMPLEMENTARY CHAINS OF DNA IN PROMOTER AND CODING PARTS OF GENES OF *E. COLI*

Y.G. Shckorbatov, M.A. Papirny, M.V. Kulik

By means of computer analysis of nucleotide composition and the energy of interaction between DNA chains in promoter and coding regions of genes of *E. coli* was studied. It was shown an increase in the content of adenine and thymine in the promoter region. There were no significant differences between the genes in the direct and reverse orientation. It was found the minimums in the energy of interaction of complementary DNA strands in genes of direct and reverse orientations of the promoter and coding regions are near the positions: -100, -70, -40, +5, +20, +30. Sites -40, +5 are close to the known consensus sequences of nucleotides in the promoter affecting the activity of the gene. Sites -100, -70, +20, +30, according to our data, also have the energy decrease of the interaction of complementary strands of DNA, suggesting that they have regulatory significance. The interaction energy of the complementary strands of DNA is reduced not only in the promoter, but also at the beginning of the coding gene. Growth in the energy of interaction of DNA in the coding region of gene occurs after the 100th position. The interaction energy of DNA chains interaction is lower in genes of reverse orientation. Differences between direct and reverse sequences are statistically significant in the region from -150 to -101 and from 1 to 51. Genes in the reverse orientation have the additional energy minimums in positions: -140, -120, -25 and +40. In addition, genes with a reverse orientation have a minimum of energy of interaction between DNA strands in the site -85, while in the genes with a direct orientation the minimum of energy is in position about -80.

KEY WORDS: energy of interaction of DNA chains, composition of DNA, promoter, orientation of genes.

Проблема связи между структурой управляющей части гена – его промоторной и предпромоторной областей и активностью транскрипции исследуется давно. Было показано, что в особых частях промотора гена – так называемых консенсусных последовательностях, расположенных вблизи нуклеотидных позиций – 10, -35, -45 – повышено содержание адениловых и тимидиловых нуклеотидов (А и Т) [1]. Это, естественно, приводит к снижению энергии взаимодействия комплементарных цепей ДНК. Такое снижение (по абсолютной величине) энергии взаимодействия имеет большое биологическое значение – снижение энергии взаимодействия приводит к облегчению расплетания цепей ДНК, что необходимо для специфического узнавания последовательностей ДНК регуляторными белками [2]. Сайты таких взаимодействий известны, однако возможно существование еще не определенных сайтов ДНК-белкового взаимодействия характерных для ряда бактериальных генов.

Мы изучали последовательности ДНК в промоторах генов *E. coli* с прямой и обратной ориентацией. Прямая ориентация – это такая, при которой направление транскрипции и трансляции совпадает. Было проанализировано содержание нуклеотидов и энергия взаимодействия цепей в кодирующей части ДНК.

Известно, что ведущая и отстающая цепь ДНК асимметричны по составу нуклеотидов. При исследовании геномов бактерий разных видов продемонстрировано, что в большинстве из 43 видов микроорганизмов, взятых к рассмотрению, в ведущей цепи относительно больше G сравнительно с C и T [3]. Проблемы асимметрии состава нуклеотидов в ведущей и отстающей цепях проанализированы в обзорах [2 - 6]. Число генов в обеих цепях неравное. Большинство из генов расположено в лидирующей цепи [7,8]. Частота генов в лидирующей цепи составляет 75 % у *B. subtilis* [9], и только 55 % у *E. coli* [10]. Такие отличия в распределении генов, возможно, связаны со свойствами ДНК-полимеразы этих видов. У *B. subtilis* есть две ДНК-полимеразы - для прямых и обратных цепей [11], у *E. coli* - одна [12].

МАТЕРИАЛЫ И МЕТОДЫ

Мы проанализировали 90 прямых и обратных последовательностей ДНК *E. coli* K12 длиной 150 нуклеотидных пар до и 200 после старта транскрипции, полученных из сайта KEGG (<http://www.genome.jp/kegg/kegg2.html>). Нуклеотидные

последовательности анализировали с помощью компьютерной программы, разработанной на визуальном языке программирования Delphi 7, используя работу с массивами и строками. Алгоритм вычисления частот нуклеотидных оснований консенсусной последовательности:

Algorithm : Calculate frequency

Input String, len_str, k

Output freq, cons_str

1. For i=1 to k

2. Str:=string[i]

3. For j=1 to len_str

4. str:=number of nucleotide(a=1,t=2,c=3,g=4)

5. array[i,j]:=str

6. if array[i,j]= number of nucleotide, calculate frequency:

7. kol [number of nucleotide]:=(kol[number of nucleotide]+1)/k

8. freq:=kol

//For output consensus sequences

9. array[i]:=kol

10. cons_str:=max{array[i]}

11. compare_sq={kons_direct==kons_revers}

Мы проанализировали энергию взаимодействия комплементарных нуклеотидов в ДНК, взяв энергию взаимодействия Т-А $\sim -68,8$ кДж/моль, и энергию взаимодействия G-C $\sim -115,137$ кДж/моль [13]. Данные значения энергии взаимодействия нуклеотидов рассчитаны с учетом энергии водородных связей между нуклеотидами, при их расчете не учитывались взаимодействия внутри цепочки нуклеотидов и между нуклеотидами в комплементарных цепях, в частности, стэкинг-взаимодействия, поэтому они имеют ориентировочный характер, что, однако, не влияет на результаты данной работы, поскольку нас интересовали не абсолютные значения энергии взаимодействия нуклеотидов, а различия в энергии взаимодействия А-Т и G-C пар. Алгоритм вычисления энергии взаимодействия комплементарных цепей ДНК:

Algorithm : Calculate energy

Input String, len_str, k

Const e(at), e(gc)

Output usredn_e

1. For i=1 to k

2. Str:=string[i]

3. n:=1

4. For j=n to len_str

5. s:=cut for fragment

6. n:=n+50

7. For l=n to len(s)

8. if s=nucleotide[a,c,g,t], kol[a,c,g,t]:=k[a,c,g,t]+1

9. E:=(kol[a,c,g,t]*e[a,c,g,t])/len[fragment]

10. E=0

11. For i=1 to k

12. For j=1 to len_str

13. Array[i,j]:=E

14. usredn_e=(usredn_e+mas[1,j])/k

Статистические расчеты проводили с помощью программ Excel 2003, Origin 7.0, STATISTICA 6.0. Для оценки частоты каждого из нуклеотидов в определенной позиции мы вычисляли средние частоты нуклеотидов каждого типа для всех проанализированных последовательностей. Мы также вычислили средние данные для отрезков по 50 нуклеотидов (Таблицы 1-4 Приложения). Данные по средней энергии взаимодействия комплементарных цепей ДНК представлены на Рис. 7-9.

РЕЗУЛЬТАТЫ

Анализируя суммарное количество А и Т, с одной стороны, и G и C, с другой, можно увидеть, что содержание А + Т увеличено в промоторной зоне от -150 до -1 нуклеотида и снижено в кодирующей части от 1 до 200. Особенно выражено повышение G + C в кодирующей части гена на участках от 100 до 200 нуклеотида (Рис. 1-4 и таблицы 6-9 Приложения).

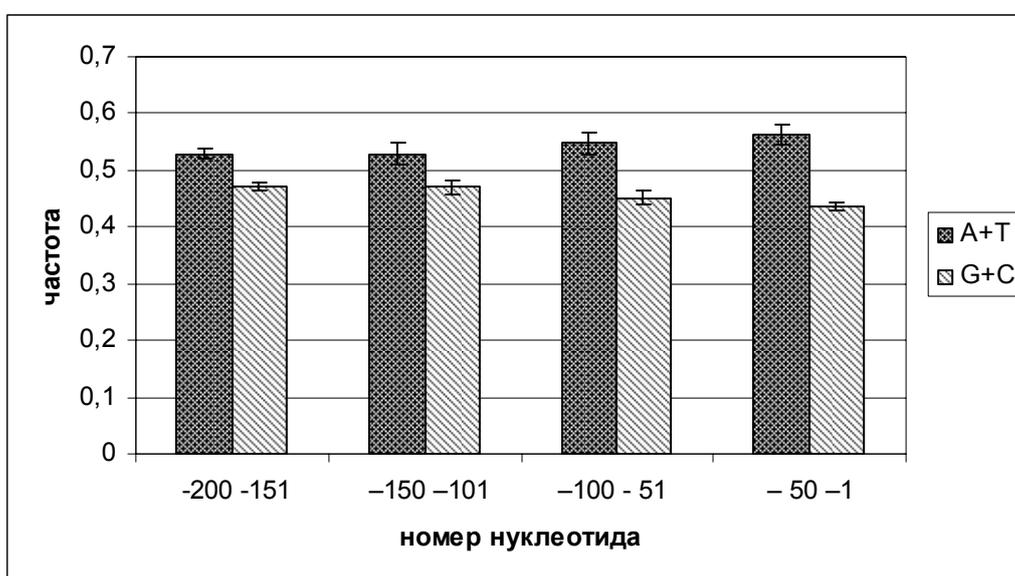


Рис 1. Суммы нуклеотидных частот в 50-нуклеотидных участках в промоторной зоне генов *E. coli* с прямой ориентацией

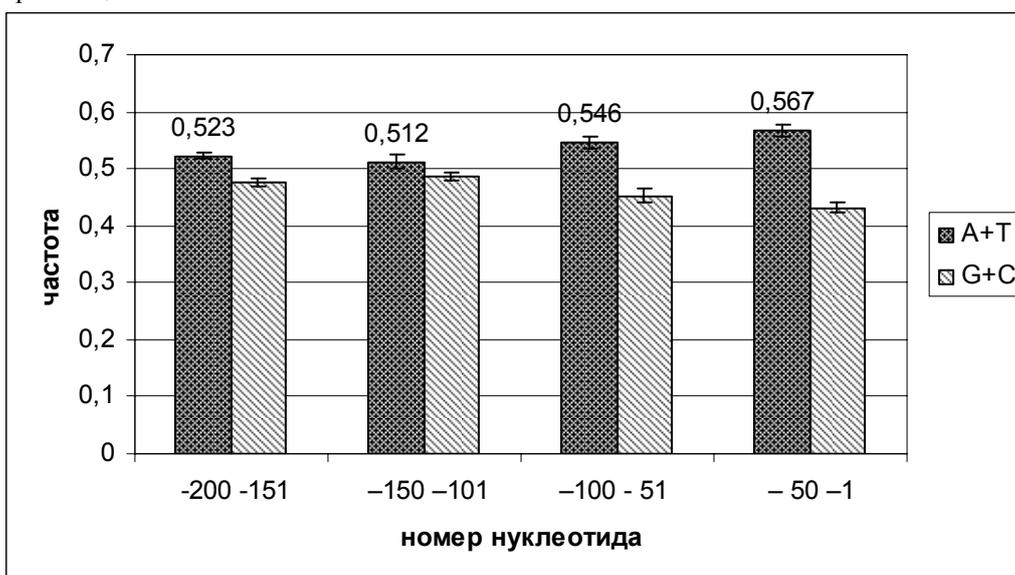


Рис 2. Суммы нуклеотидных частот в 50-нуклеотидных участках в промоторной зоне генов *E. coli* с обратной ориентацией

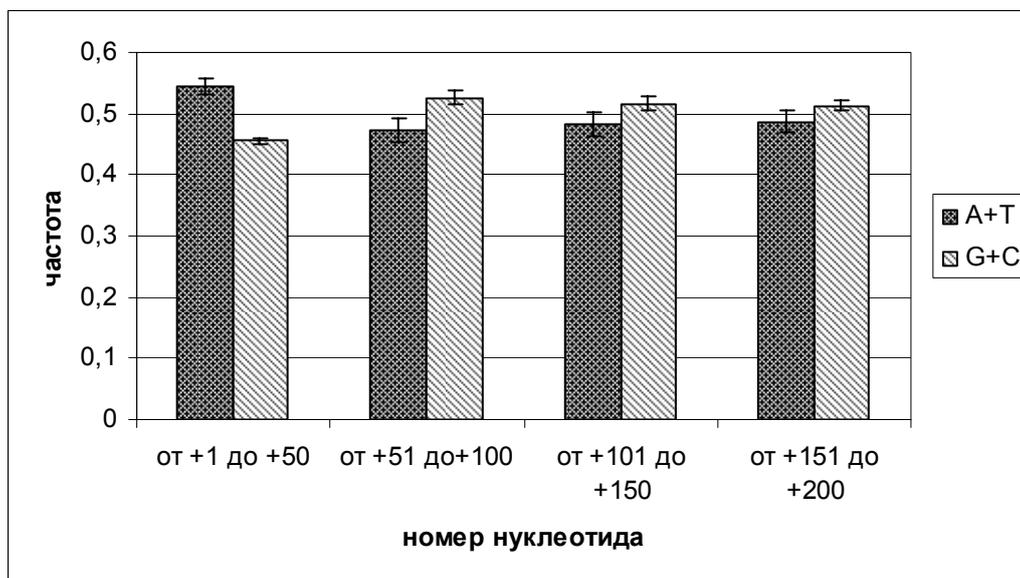


Рис 3. Суммы нуклеотидных частот в 50-нуклеотидных участках в кодирующей части генов *E. coli* с прямой ориентацией

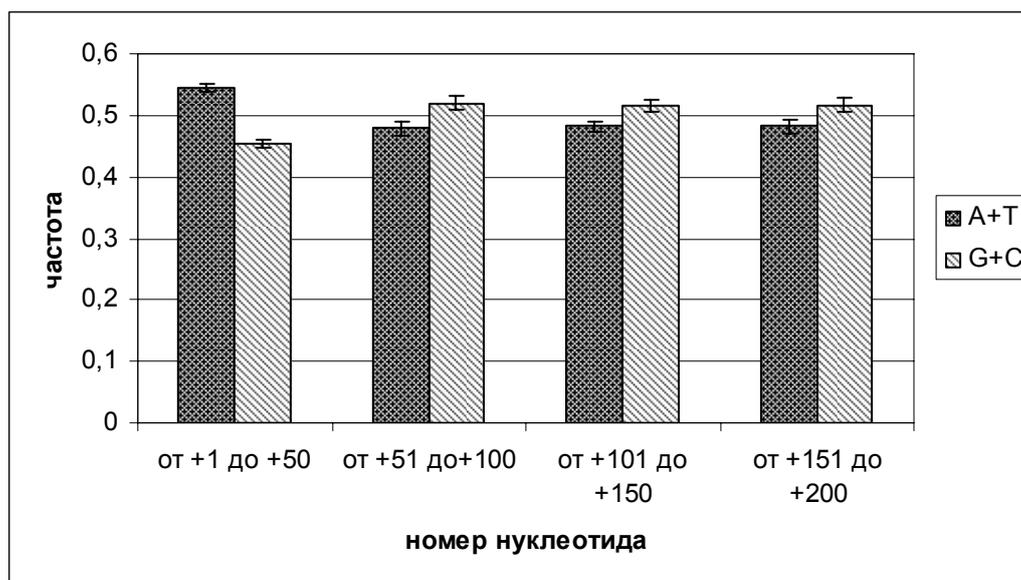


Рис 4. Суммы нуклеотидных частот в 50-нуклеотидных участках в кодирующей части генов *E. coli* с обратной ориентацией

Мы проанализировали энергию взаимодействия нуклеотидов в комплементарных цепях ДНК (Рис. 6-7 и табл. 5 "Приложения"). Как показано на рис. 5, распределение энергии вдоль ДНК в генах с прямой и обратной ориентацией имеет некоторые общие особенности. Общее уменьшение энергии наблюдается вблизи начала транскрипции: от -150 до +80 пары нуклеотидов. Минимальная энергия отмечена в позициях -100, -70, -40, +5, +20, +30. Однако гены в обратной ориентации имеют дополнительные минимумы в -140, -120, -25 и +40. Минимум в -85 перемещается относительно к минимуму в -80 в генах с прямой ориентацией. В целом, энергия взаимодействия цепей ДНК больше в области промотора, чем в кодирующей части гена (Рис. 6). Важно отметить, что энергия взаимодействия комплементарных цепей ДНК снижена не

только в области промотора, но и в начале кодирующей части гена. Увеличение энергии взаимодействия ДНК в кодирующей части гена наблюдается после 100-ой позиции (Рис. 7). Энергия взаимодействия цепей ДНК, в среднем, меньше в обратных, чем в прямых последовательностях. Отличия между прямыми и обратными последовательностями статистически достоверны в области от -150 до -101 и от 1 до 51 (Рис. 6, 7).

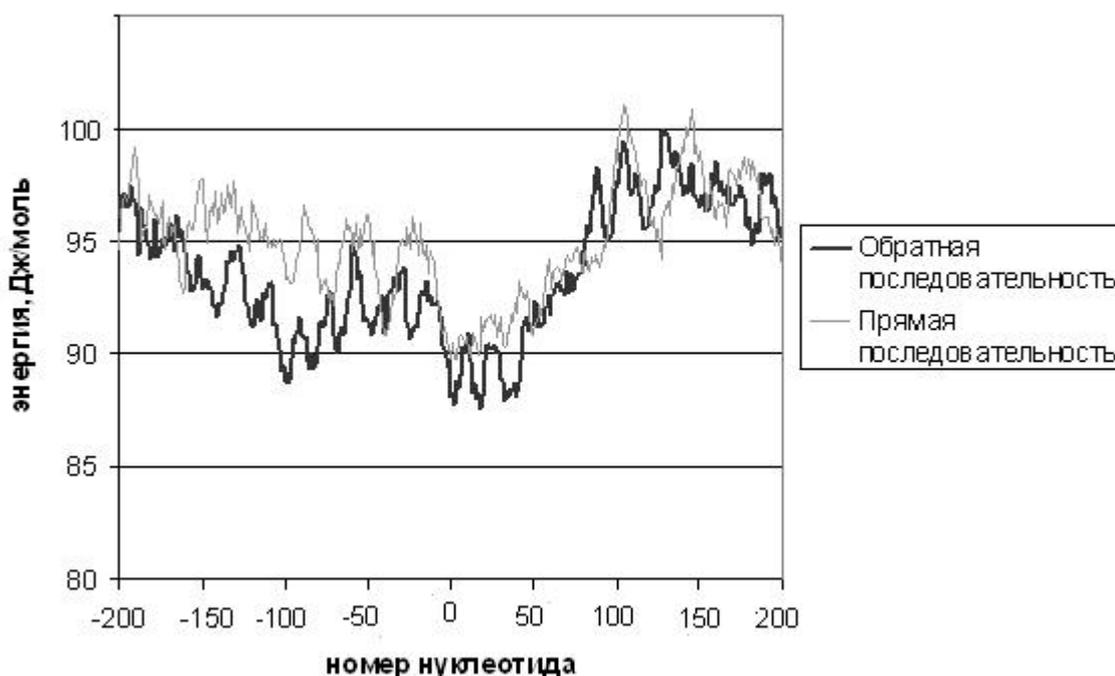


Рис. 5. Энергия взаимодействия (по абсолютной величине) 3 комплементарных цепей ДНК в генах с различной ориентацией

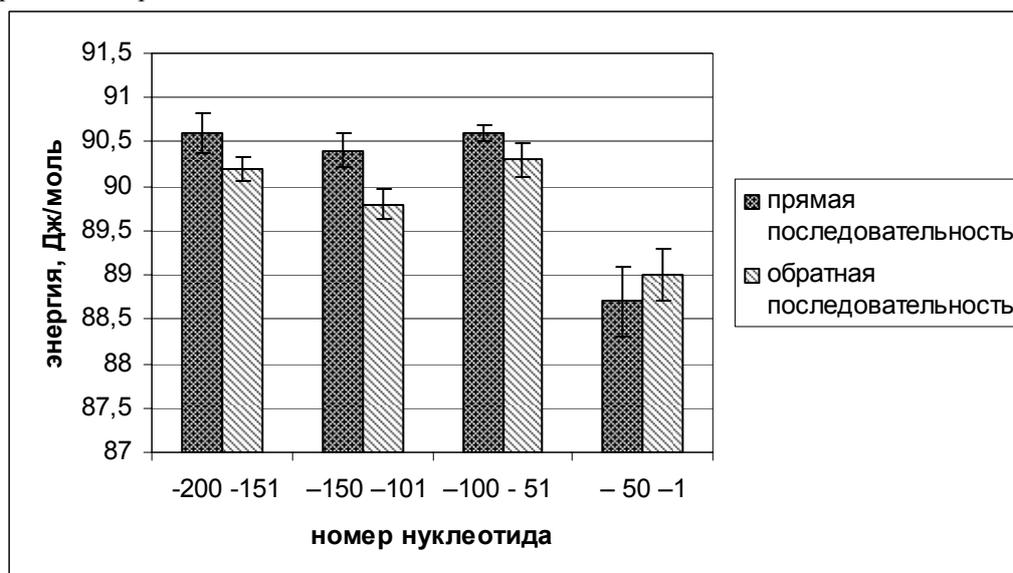


Рис. 6. Энергия взаимодействия (по абсолютной величине) коомплементарных нуклеотидов в промоторной зоне генов *E. coli* (прямая и обратная ориентация)

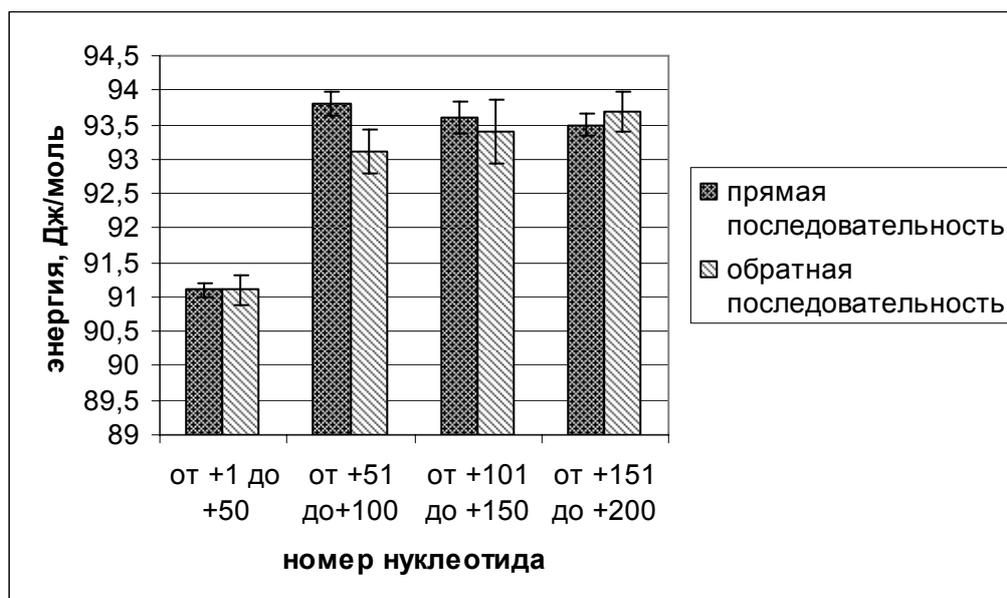


Рис 7. Энергия взаимодействия (по абсолютной величине) коомплементарных нуклеотидов в кодирующей зоне генов *E. coli* (прямая и обратная ориентация гена)

ОБСУЖДЕНИЕ

Подробный анализ нуклеотидных последовательностей *E. coli* позволил выявить следующее. Согласно [3] относительное содержание GC пар в ДНК *E. coli* K12 равняется 50.8%. Мы наблюдали превышение содержания G над другими нуклеотидами в участке от -200 до -51 в генах с прямой ориентацией и от -150 до -51 в генах с обратной ориентацией, а также превышение содержания C над другими нуклеотидами в регионе от -50 до -1 как для генов с прямой, так и с обратной ориентацией, что хорошо согласуется с данными работы [3] (Табл. 1-4. Приложение).

Ранее было показано, что лидирующая цепь ДНК *E. Coli* обогащается G и T, а отстающая цепь обогащается в C и A [6]. Интересно, что геном различных бактерий имеет ту же асимметрию: G - относительно больше, чем C в лидирующей цепи кодирующей части гена. Эта закономерность менее выражена в отстающей цепи, которая характеризуется большим содержанием T [14]. Мы не обнаружили значительных отличий в содержании нуклеотидов между генами в прямой и обратной ориентации. Но мы обнаружили, что энергия взаимодействия цепей ДНК ниже в генах с обратной ориентацией, что означает, что гены с прямой ориентацией относительно обогащены G и C. Общее уменьшение энергии взаимодействия цепей ДНК наблюдается в регионе от -150 до +80. Этот участок взаимодействует с разными регуляторными белками и мы считаем, что обнаруженные нами участки с минимальной энергией взаимодействия цепей ДНК соответствуют сайтам ДНК-белковых взаимодействий. Расплетание ДНК облегчено в участках с низкой энергией взаимодействия цепей ДНК. Расплетание ДНК необходимо, поскольку этот процесс облегчает специфическое узнавание определенных сайтов ДНК в ходе взаимодействия белок-ДНК. Эти участки совпадают с известными последовательностями нуклеотидов как в промоторной зоне, например, -40, +5, +20, а также нами обнаружены новые сайты -100, -70, +20, +30. Мы считаем, что в этих позициях расположены места взаимодействия между ДНК и регуляторными белками.

ВЫВОДЫ

Мы показали, что, в соответствии с ранее опубликованными данными, содержание нуклеотидов в кодирующем участке гена значительно отличается от содержания нуклеотидов в зоне близкой к промотору: в частности, увеличено содержание гуанина и цитозина. Мы обнаружили минимумы энергии взаимодействия комплементарных цепей ДНК в промоторной и в кодирующей областях генов *E. coli* вблизи следующих позиций: -100, -70, -40, +5, +20, +30. Эти участки частично совпадают с известными консенсусными последовательностями нуклеотидов. Общее уменьшение энергии взаимодействия цепей ДНК наблюдается в регионе от -150 до +80. Гены в обратной ориентации имеют дополнительные минимумы в позициях: -140, -120, -25 и +40. Кроме того, в генах с обратной ориентацией имеется минимум энергии взаимодействия цепей ДНК в области -85, в генах с прямой ориентацией аналогичный минимум имеется в области -80. Кроме того, в генах с обратной ориентацией энергия взаимодействия цепей ДНК, в среднем, значительно ниже, чем таковая в генах с прямой ориентацией, в районах от -150 до -100 и от 1 до 50.

ЛИТЕРАТУРА

1. Lewin B. Genes VIII. Upper Saddle River, New Jersey, Pearson Education Inc., 2004. 1002 p.
2. Morton R.A., Morton B.R. Separating the effects of mutation and selection in producing DNA skew in bacterial chromosomes. *BMC Genomics* 2007; 8: 369.
3. Lobry J.R., Sueoka N. Asymmetric directional mutation pressures in bacteria. *Genome Biol.* 2002; 3(10): research0058.1–research0058.14.
4. Mrazek J., Karlin S. Strand compositional asymmetry in bacterial and large viral genomes. *Proc. Natl. Acad. Sci. USA.* 1998; 178: 3720-3725.
5. Rocha C. The replication-related organization of bacterial genomes. *Microbiology* 150 2004; 220: 1609-1627.
6. Tillier E.R., Collins R.A. Replication orientation affects the rate and direction of bacterial gene evolution. *J Mol. Evol.* 2000; 51(5): 459-63.
7. Burland V., Plunkett G., Daniels D.L., Blattner F.R. DNA sequence and analysis of 136 kb of the *Escherichia coli* genome: organizational symmetry around the origin of replication // *Genomics.* 1993. V. 16: 551–561.
8. Ogasawara N., Yoshikawa H. Genes and their organization in the replication origin region of the bacterial chromosome // *Mol. Microbiol.* 1992. V. 6. P. 629–634.
9. Kunst, F., Ogasawara, N., Moszer, I. & 148 other authors. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis* // *Nature* 1997. V. 390. P.249–256.
10. Blattner F.R., Plunkett G., Bloch C.A. & 14 other authors. The complete genome sequence of *Escherichia coli* K-12 // *Science.* 1997. V. 240. P. 1453-1461.
11. Dervyn E., Ehrlich S.D. Two essential DNA polymerases at the bacterial replication fork. *Science.* 2001. V. 294. P. 1716-1719.
12. Yuzhakov, A., Turner, J. & O'Donnell, M. Replisome assembly reveals the basis for asymmetric function in leading and lagging strand replication // *Cell.* 1996. V. 86. P. 877–886.
13. Sponer J., Jurecka P., Hobza P. Accurate interaction energies of hydrogen-bonded nucleic acid Base Pairs // *J. AM. CHEM. SOC.* 2004. V. 126. P. 10142-10151.
14. Rocha E.P., Danchin A., Viari A. Universal replication biases in bacteria // *Mol. Microbiol.* 1999. V.32. P. 11–16/

ПРИЛОЖЕНИЕ

Табл. 1. Нуклеотидные частоты в 50-ти нуклеотидных участках в промоторной зоне генов *E. Coli* с прямой ориентацией

Нуклеотиды: от - до	Нуклеотидные частоты			
	С	G	T	A
-151:-200	0,2±0,009	0,236±0,02	0,259±0,02	0,304±0,007
-101:-150	0,223±0,007	0,228±0,006	0,281±0,007	0,266±0,007
-51:-100	0,224±0,006	0,246±0,006	0,28±0,006	0,248±0,006
-1: -50	0,225±0,007	0,244±0,005	0,276±0,006	0,253±0,007

Табл. 2. Нуклеотидные частоты в 50-ти нуклеотидных участках в промоторной зоне генов *E. Coli* с обратной ориентацией

Нуклеотиды: от - до	Нуклеотидные частоты			
	С	G	T	A
-151:-200	0,213±0,007	0,218±0,01	0,262±0,009	0,305±0,008
-101:-150	0,232±0,006	0,220±0,006	0,281±0,007	0,264±0,008
-51:-100	0,244±0,006	0,243±0,007	0,272±0,008	0,239±0,007
-1: -50	0,244±0,007	0,232±0,006	0,259±0,008	0,263±0,006

Табл. 3. Нуклеотидные частоты в 50-ти нуклеотидных участках начала транскрипции генов *E. Coli* с прямой ориентацией

Нуклеотиды: от - до	Нуклеотидные частоты			
	С	G	T	A
1: 50	0,237±0,01	0,248±0,011	0,269±0,013	0,244±0,008
51: 100	0,236±0,009	0,247±0,011	0,266±0,012	0,249±0,006
101: 150	0,224±0,008	0,248±0,011	0,268±0,013	0,258±0,007
151: 200	0,283±0,02	0,261±0,019	0,246±0,019	0,209±0,01

Табл. 4. Нуклеотидные частоты в 50-ти нуклеотидных участках начала транскрипции генов *E. Coli* с обратной ориентацией

Нуклеотиды: от - до	Нуклеотидные частоты			
	С	G	T	A
1: 50	0,246±0,011	0,236±0,012	0,264±0,011	0,253±0,008
51: 100	0,242±0,008	0,241±0,011	0,256±0,012	0,259±0,007
101: 150	0,234±0,009	0,245±0,012	0,262±0,011	0,258±0,006
151: 200	0,273±0,018	0,271±0,019	0,219±0,019	0,235±0,012

Табл. 5.

Энергия кодирующей и промоторной зон ген *E. coli* (прямая и обратная ориентация гена)

Промоторная зона

Нуклеотиды: от - до	Энергия взаимодействия ДНК	
	Прямая	Обратная
-200: -151	48,4±0,12	48,5±0,07
-150: -101	48,1±0,10	49,0±0,09
-100: -51	47,5±0,05	47,5±0,10
-50: -1	46,5±0,21	46,5±0,19

Кодирующая зона

Нуклеотиды: от - до	Энергия взаимодействия ДНК	
	Прямая	Обратная
1: 50	47,9±0,06	47,0±0,08
51: 100	49,4±0,09	49,4±0,14
101: 150	51,6±0,13	51,4±0,08
151: 200	51,0±0,09	50,9±0,06

Табл. 6

Суммы нуклеотидных частот в 50-нуклеотидных участках в промоторной зоне генов *E. coli* с прямой ориентацией

Нуклеотиды: от - до	A+T	G+C
-200: -151	0,529±0,01	0,470±0,007
-150: -101	0,528±0,019	0,471±0,012
-100: -51	0,547±0,019	0,452±0,011
-50: -1	0,563±0,019	0,436±0,008

Табл. 7

Суммы нуклеотидных частот в 50-нуклеотидных участках в промоторной зоне генов *E. coli* с обратной ориентацией

Нуклеотиды: от - до	A+T	G+C
-200: -151	0,523±0,006	0,477±0,007
-150: -101	0,512±0,012	0,488±0,012
-100: -51	0,546±0,011	0,453±0,011
-50: -1	0,567±0,009	0,432±0,010

Табл. 8

Суммы нуклеотидных частот в 50-нуклеотидных участках в кодирующей части генов *E. coli* с прямой ориентацией

Нуклеотиды: от - до	A+T	G+C
1: 50	0,544±0,012	0,455±0,006
51: 100	0,473±0,019	0,526±0,011
101: 150	0,483±0,019	0,516±0,011
151: 200	0,486±0,018	0,513±0,008

Табл. 9

Суммы нуклеотидных частот в 50-нуклеотидных участках в кодирующей части генов *E. coli* с обратной ориентацией

Нуклеотиды: от - до	A+T	G+C
1: 50	0,544±0,007	0,455±0,007
51: 100	0,479±0,011	0,520±0,011
101: 150	0,483±0,011	0,516±0,011
151: 200	0,482±0,008	0,517±0,011