

УДК: 577.323:539.196

## ЭНЕРГИЯ ВЗАИМОДЕЙСТВИЯ ЦЕПЕЙ ДНК В ПРОМОТОРАХ E. COLI В СВЯЗИ С ИХ ФУНКЦИОНАЛЬНОЙ АКТИВНОСТЬЮ

Ю.Г.Шкорбатов<sup>1</sup>, А.Ю.Бережной<sup>2</sup>

<sup>1</sup>Харьковский национальный университет им В.Н.Каразина,  
пл. Свободы, 4, Харьков 610077, Украина  
E-mail: Yury.G.Shkorbatov@univer.kharkov.ua

<sup>2</sup>Национальный научный центр Харьковский физико-технический институт,  
ул. Академическая, 1, Харьков, 61108, Украина

Поступило в редакцию 24 февраля 2006 г.

Проведены расчеты энергии взаимодействия между комплементарными нуклеотидами в промоторе. Нами были разработаны методы графического отображения энергетических свойств промоторов. Из полученных данных видно, что распределение энергии по всей длине промотора имеет 3 минимума с центрами на позициях -35, -8 и +7 нуклеотидов относительно стартового участка транскрипции. Наиболее важная и отличающаяся от случайной последовательности область промотора лежит на позиции -8 относительно стартового участка транскрипции. В блоке Прибнова и в -35 последовательности были выделены 4 группы промоторов с различными энергетическими свойствами. В блоке Прибнова для средних показателей в группе обнаружена негативная корреляция между энергией взаимодействия цепей ДНК и активностью (силой) промотора и положительная корреляция между этими показателями в -35 последовательности.

**КЛЮЧЕВЫЕ СЛОВА:** промоторы, эффективность промоторов, нуклеотидный состав, энергия взаимодействия цепей ДНК, физические свойства промоторов.

Бактериальные гены, кодирующие белки, должны по-разному экспрессироваться в ходе клеточного цикла и в связи с широким разнообразием внеклеточных сигналов, регулирующих метаболизм клетки. Инициация транскрипции РНК-полимеразой (РНКП) требует наличия различных регуляторных элементов ДНК, включая наиболее функционально важные участки промоторов. Такие участки промоторов прокариот, в частности, у *Escherichia coli*, для РНК-полимеразного комплекса с  $\sigma^{70}$ , обычно расположены относительно стартового участка транскрипции на участке от -60 до +20. Имеют два наиболее важных участка инициации транскрипции: на позициях -35 и -10 (блок Прибнова). Нуклеотидные последовательности на участках -35 и -10 могут влиять на присоединение РНК полимеразы к промоторной ДНК и формирование РНК-полимеразного комплекса [1].

Основные (коровые) компоненты РНКП *Escherichia coli* могут осуществлять стадию элонгации транскрипции, но только холофермент, содержащий фактор  $\sigma$  осуществляет специфическую инициацию транскрипции. Распознавание промотора холоферментом, содержащим  $\sigma$ -фактор, например,  $\sigma^{70}$ , происходит с помощью взаимодействия  $\sigma$  субъединицы с тремя участками промотора. Проблема взаимосвязи силы промотора с его структурой интенсивно исследовалась в 80-х годах. Гексамер -10 (консенсусная последовательность 5'-ТАТААТ-3') распознается  $\sigma$ -областью 2.3-2.4 [2], дополнительный -10 участок (консенсусная последовательность 5'-ТГТГп-3') распознается  $\sigma$ -областью 3.0 [3], а гексамер -35 (консенсусная последовательность 5'-ТТГАСА-3') -  $\sigma$ -областью 4.2 [4]. С-терминальные области, состоящие из двух  $\alpha$  субъединиц ( $\alpha$ CTDS) в некоторых промоторах взаимодействует со специфическими последовательностями, расположенными выше гексамера -35 [5].

Первым этапом инициации транскрипции обычно является расплетение промоторной ДНК и образование так называемого открытого комплекса. Мутации в области промотора уменьшают скорость образования открытого комплекса транскрипции. Связывание с открытым комплексом происходит за счет  $\sigma$  субъединицы холофермента [6]. Известно, что участок -11 является особенно важным для формирования открытого комплекса РНКП с промотором. При замене нуклеотидных последовательностей в промоторах часто наблюдается сильный эффект уменьшения степени образования открытого комплекса [7].

Силу промотора можно определить различными способами. Kajitani и Ishihama с помощью смешанной транскрипционной системы *in vitro* определили два параметра силы промотора: степень образования комплекса между РНКП и промотором и уровень насыщенности образования этого комплекса [8]. Vogel и соавторы определили силу промотора как скорость, с которой открытый комплекс  $RP_0$  of  $RNAP \cdot \sigma^{54}$  (R) образует открытый комплекс в реакции  $R + P \rightleftharpoons RP_0 \rightleftharpoons \dots \rightleftharpoons RP_n$  [9].

Целью данной работы являлось изучение закономерностей и взаимосвязей между функциональной активностью промотора и его структурой и энергетическими свойствами.

### МАТЕРИАЛЫ И МЕТОДЫ

Нами было получено 106 нуклеотидных последовательностей промоторов *Escherichia coli* использующих  $\sigma^{70}$  субъединицу из базы данных Regulon благодаря содействию администратора этой базы данных. Все промоторы были транскрибированы с помощью субъединицы  $\sigma^{70}$ . Данные о силе промоторов были предоставлены Hisanogi Kiryu. Эти данные были взяты из базы данных KEGG EXPRESSION [10,11]. В этой базе данных содержались данные для *Synechocystis* PCC6803, *Bacillus subtilis*, и *Escherichia coli*, полученные японским научным сообществом. Сила промоторов была измерена с помощью метода флуоресцентных меток в полном геноме *E. coli* [12]. Ориентация нуклеотидных последовательностей в геноме определялась как прямая или обратная, в зависимости от положения гена. Ориентация промотора, насколько нам известно, не влияет на функции гена. Нами было проанализировано 106 промоторных последовательностей. В данных рис. 1–3 мы рассматривали только «сильные» промоторы – у которых сила больше единицы. Количество прямых последовательностей равнялось 28, обратных – 34. Кроме того, для сравнения данных мы взяли 30 случайных последовательностей. В результатах, представленных в таблицах и на Рис. 4–7 мы рассматривали все 106 промоторов.

Мы ввели понятие энергии промотора как сумму энергий взаимодействия всех пар нуклеотидов в промоторе по отношению к количеству нуклеотидов. Энергия взаимодействия комплементарных нуклеотидов в паре была взята из работы [13]. Для анализа содержания АТ-пар и энергии взаимодействия пар в промоторе мы применили метод «скользящего окна». Мы суммировали содержание АТ-пар и энергию взаимодействия пар на участке длиной 10 нуклеотидов (длина окна) и оценивали средние значения этих параметров. На каждом последующем этапе анализа окно сдвигалось на одну нуклеотидную пару. На рисунках представлены полученные нами средние значения этих параметров.

### РЕЗУЛЬТАТЫ

Названия промоторов *E. coli* и соответствующие им номера в наших исследованиях представлены в таблице 1.

Распределение энергии взаимодействия пар оснований по длине промотора измерялось методом скользящего окна. Средняя энергия АТ-пары равна  $-29.33$  кДж/моль, GC-пары составляет, по данным Кудришской и Данилова  $-70.35$  кДж/моль [13]. Рассчитанные нами средние данные для промоторов в прямых и обратных последовательностях (количество прямых последовательностей – 28, обратных – 34) представлены на Рис.1. Из Рис.1 видно, что распределение энергии взаимодействия пар оснований в промоторе имеет 3 пика. Первый из них лежит в области между позициями окна  $-40$  и  $-30$  относительно точки инициализации транскрипции (+1), второй – между позициями  $-15$  и  $-10$ , последний – между позициями  $-4$  и  $+10$ . Эти области расположены в наиболее важных участках промотора и соответствуют  $-35$ ,  $-35$  и  $+10$  консенсусным последовательностям. Мы рассчитали величину  $t$ -критерия Стьюдента для различий между энергиями взаимодействия пар в различных положениях промоторной последовательности, который представлен на Рис. 2. Из представленных данных видно, что в интервалах от  $+5$  до  $+15$  и от  $-15$  до  $-5$ , а также в интервале от  $-35$  до  $-20$  отличия в нуклеотидном составе от состава случайной последовательности статистически достоверны.

Данные о силе промоторов представлены на Рис.3. Из этих данных видно, что анализируемые промоторные последовательности значительно различались между собой по силе.

Мы исследовали вопрос о возможности разбиения на группы всех анализируемых промоторов по распределению энергии взаимодействия нуклеотидов в блоке Прибнова. С помощью кластерного анализа, используя метод определения евклидовых расстояний, мы рассортировали все 106 промоторов на четыре кластера (здесь мы рассматривали также и «слабые» промоторы). Полученные данные представлены в Табл. 1. Все кластеры сильно различаются по распределению энергии взаимодействия оснований комплементарных нуклеотидов в целом промоторе (81 нуклеотид), см. Табл. 1. Для каждой из групп промоторов мы определили средние величины энергии взаимодействия нуклеотидов в блоке Прибнова и среднюю величину силы промоторов. В Табл. 2. представлены величины энергии взаимодействия цепей ДНК в блоке Прибнова и в  $-35$  последовательности для каждой из групп промоторов. Из наших данных видно отсутствие линейной зависимости между энергией взаимодействия цепей в промоторе и силой промотора. Однако, согласно нашим данным, наиболее корректно описывает эту зависимость  $y = 4,34 \cdot 10^{-27} \exp(6,78 \cdot 10^{-2}x)$ . Формулы регрессионной зависимости и коэффициенты корреляции зависимости силы промотора от энергии взаимодействия цепей ДНК в блоке Прибнова и представлены в Табл. 3.

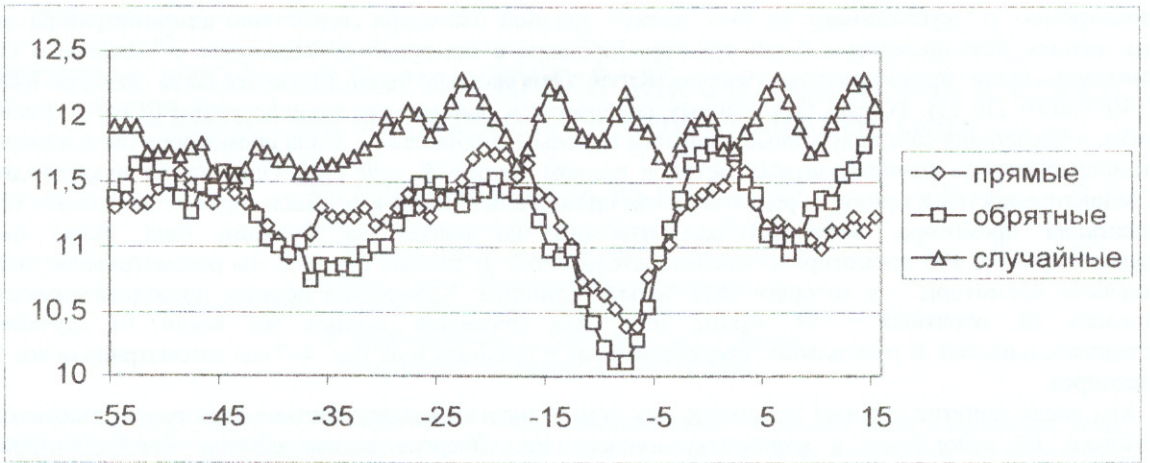


Рис. 1. Распределение энергии взаимодействия комплементарных пар в промоторных и случайных последовательностях.

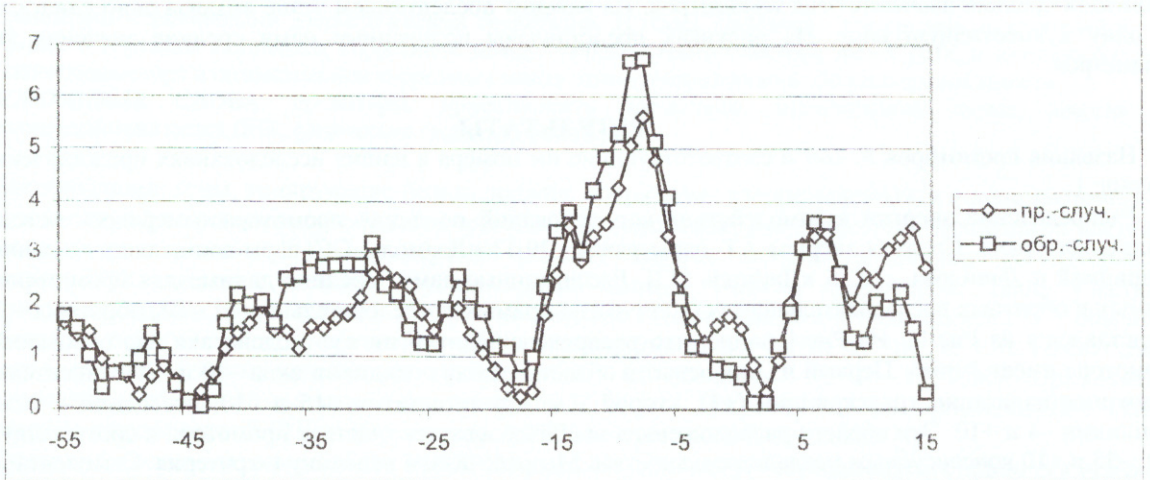


Рис. 2. Величина  $t$ -критерия Стьюдента для характеристики различий между консенсусными последовательностями промоторов (в прямых и обратных последовательностях) и случайной нуклеотидной последовательностью.

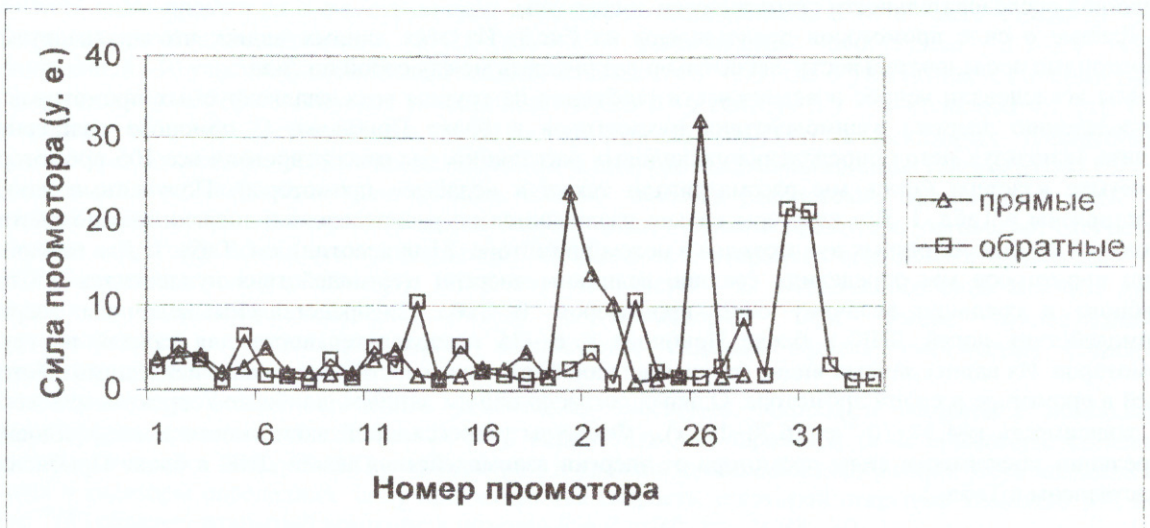


Рис. 3. Сила анализируемых промоторов в прямых и обратных последовательностях.

Энергия взаимодействия цепей ДНК в промоторах *E. coli*...

Табл. 1. Энергетические свойства промоторов и их сила в группах по степени сходства блоков Прибнова.

Номер кластера	Количество промоторов в кластере	Средняя энергия взаимодействия цепей ДНК в промоторе для данного кластера (Ккал/моль).	Средняя энергия взаимодействия цепей ДНК в блоке Прибнова для данного кластера (Ккал/моль).	Средняя сила промотора для кластера
1	2	890,1	71,4	0,62
2	10	903,8	61,6	1,54
3	59	915,1	54,7	3,21
4	40	904,5	51,4	3,00

Табл. 2. Энергетические свойства промоторов и их сила в группах по степени сходства -35 последовательности.

Номер кластера	Количество промоторов в кластере	Средняя энергия взаимодействия цепей ДНК в промоторе для данного кластера (Ккал/моль).	Средняя энергия взаимодействия цепей ДНК в -35 последовательности для данного кластера (Ккал/моль).	Средняя сила промоторов для кластера
1	2	890,1	61,6	1,12
2	7	898,5	54,6	1,20
3	72	906,4	62,3	2,97
4	25	917,5	71,4	3,47

Табл. 3. Формулы регрессионной зависимости силы промоторов от энергии взаимодействия цепей ДНК в различных участках промотора и в целом промоторе.

Последовательность	Формула регрессионной зависимости силы промотора от энергии взаимодействия в указанной области	Множественный коэффициент корреляции
Блок Прибнова	$y = 280 \exp(-8,5 \cdot 10^{-2}x)$	0,909
Весь промотор (группировка по Блоку Прибнова)	$Y = -93,4 + 0,106x$	0,883
-35 последовательность	$y = -6,59 + 0,141x$	0,802
Весь промотор (группировка по -35 последовательности)	$y = -85,2 + 9,68 \cdot 10^{-2}x$	0,935

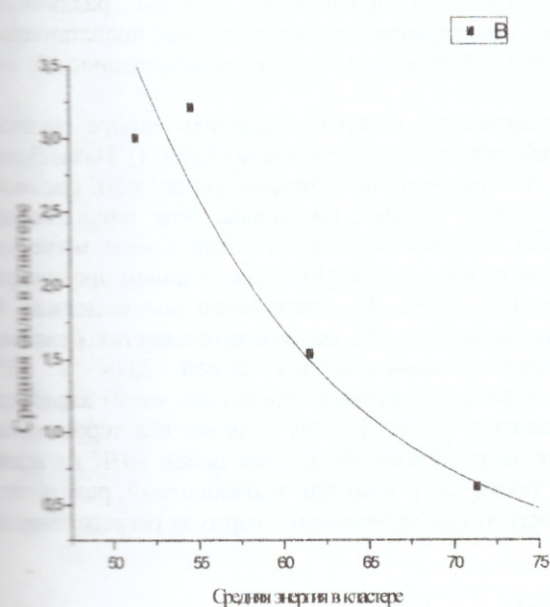


Рис. 4. Зависимость силы промотора от энергии взаимодействия цепей ДНК в блоке Прибнова для кластеров из последовательностей блока Прибнова.

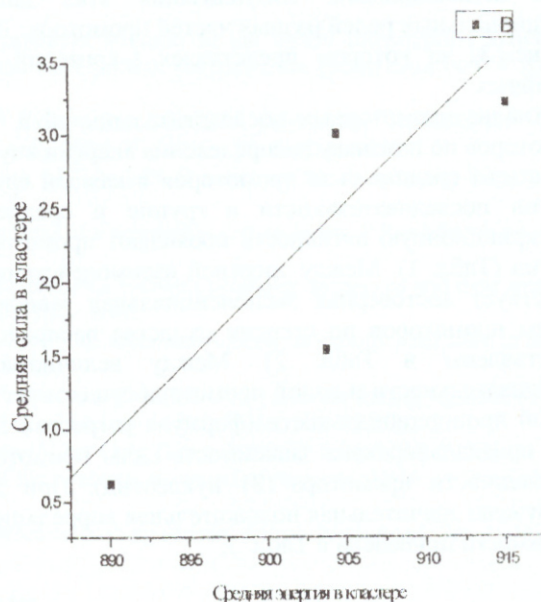


Рис. 5. Зависимость силы промотора от энергии взаимодействия цепей ДНК во всем промоторе для кластеров из последовательностей блока Прибнова.

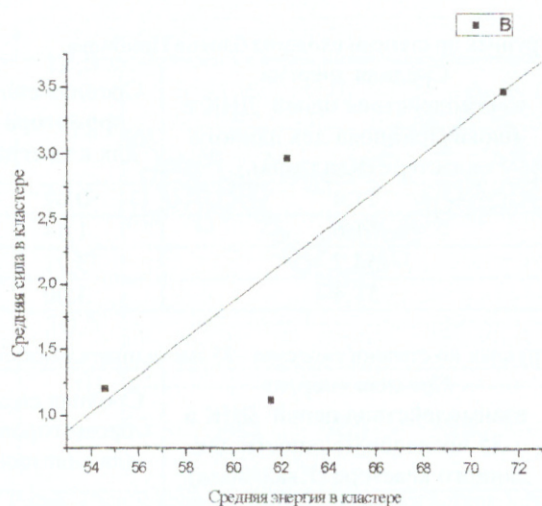


Рис. 6. Зависимость силы промотора от энергии взаимодействия цепей ДНК в  $-35$  последовательности для кластеров из  $-35$  последовательностей.

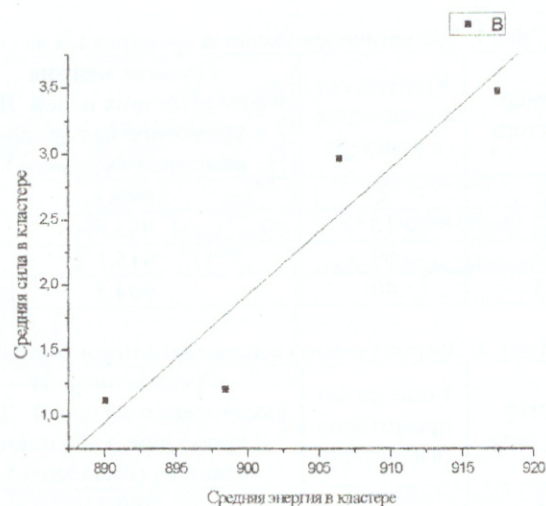


Рис. 7. Зависимость силы промотора от энергии взаимодействия цепей ДНК во всем промоторе для кластеров из  $-35$  последовательностей.

### ОБСУЖДЕНИЕ

Давно известно, что для начала транскрипции нити ДНК должны быть расплавлены в области промотора. Многие регуляторы транскрипции активны на этапах, приводящих к плавлению нитей ДНК. Специфическое взаимодействие между определенными областями ДНК и  $70\text{-S}$  субъединицей РНКП приводит к разделению пар оснований между нитями ДНК в области  $-10$  промотора и к освобождению кодирующей нити ДНК для синтеза РНК. В этом процессе холофермент сначала присоединяется к промотору для образования закрытого комплекса, а затем открывает область ДНК примерно между позициями  $-11$  и  $+3$ . Коровая часть РНКП может участвовать в реакции плавления ДНК [16].

Из наших результатов видно, что энергии пар в промоторах значительно понижаются в областях между  $-5$  и  $-15$  и между  $-25$  и  $-40$ , а также в области от  $+10$  до  $+5$  (Рис. 1). Это связано с повышением концентрации АТ-пар в этих областях (Рис. 2). Уменьшение энергии пар нуклеотидов делает эти области более легкоплавкими. Визуализация этих данных объясняет физические основы различных функциональных ролей разных частей промотора. Данные, подтверждающие эти различия, представлены на Рис. 3, на котором представлен  $t$ -критерий отличия исследовавшихся последовательностей от случайных.

Анализ промоторных последовательностей в блоке Прибнова позволил выделить четыре группы промоторов по признаку распределения энергии внутри данной последовательности (Табл. 1). Нами была определена средняя сила промоторов в каждой группе. Мы рассчитали величины корреляции средней энергии последовательности в группе и средней силой в группе. Оказалось, что наибольшую транскрипционную активность проявляют промоторы, для которых характерно минимальное значение энергии (Табл. 1). Между энергией взаимодействия нуклеотидов в блоке Прибнова и силой промотора существует достоверная экспоненциальная зависимость (см. Табл. 3). Аналогично мы выделили 4 группы промоторов по степени сходства распределения энергии в  $-35$  последовательностях (данные представлены в Табл. 2). Между величиной энергии взаимодействия цепей ДНК в  $-35$  последовательности и силой промотора существует достоверная корреляция, однако она имеет характер прямой пропорциональности (формула регрессии приведена в Табл. 3). Для этих же кластеров нами была проанализирована зависимость силы промотора от энергии взаимодействия цепей ДНК на всей протяженности промотора (81 нуклеотид). При этом также был получен неожиданный результат: обнаружена значительная положительная корреляция между этими величинами, формула регрессионной зависимости приведена в Табл. 3.

### ВЫВОДЫ

Механизмы регуляции транскрипции на уровне инициации очень сложны, но, несмотря на это, мы предположили, что на уровне структуры промотора могут быть найдены некоторые простые закономерности, объясняющие различия в силе промоторов. Из наших данных видно, что распределение энергии взаимодействия пар оснований по длине промотора имеет три минимума с центрами на

позициях -35, -8 и +7 относительно начала транскрипции (см. Рис. 1). Полученные результаты не зависят от ориентации промотора в геноме и являются сходными для прямых и обратных последовательностей. Наиболее важная и наиболее отличающаяся от случайной последовательности область промотора расположена возле позиции -8 (см. Рис. 2). Мы обнаружили несколько групп промоторов с различными энергетическими свойствами. Из полученных данных видно, что средняя сила промотора в кластере обратно пропорциональна средней энергии кластера для участков -10 (блок Прибылова). Между тем наблюдается прямая пропорциональность между средней силой промотора в кластере и средней энергии всей последовательности промотора. Напротив, для кластеров в -35 последовательности энергия взаимодействия цепей ДНК в этих последовательностях и во всем промоторе прямо пропорциональна его силе, что, на наш взгляд, является нетривиальным результатом. В дальнейшем планируется провести подробный анализ энергетических свойств различных частей промотора и предпромоторной области и в начале кодирующей части гена в связи с его функциональной активностью.

*Благодарность.* Авторы благодарят доктора Hisanori Kiryu за предоставление данных о силе и нуклеотидной последовательности ряда промоторов.

#### СПИСОК ЛИТЕРАТУРЫ

- Babb K., McAlister J.D., Miller J.C., Stevenson B. Molecular Characterization of *Borrelia burgdorferi* erp Promoter/Operator Elements. // Journal of Bacteriology. 2004. V. 186. No 9. P. 2745-2756.
- Burr T., Mitchell J., Minchin S., Busby, S. DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: A systematic study. // Nucleic Acids Res. 2000. V. 28. P. 1864-1870.
- Murakami K.S., Masuda S., Campbell E.A., Muzzin O., Darst S.A. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. // Science. 2002. V. 296. P. 1285-1290.
- Campbell E.A., Muzzin O., Chlenov, M., Sun J.L., Olson C.A., Weinman O., Trester-Zedlitz M.L., Darst S.A. Structure of the bacterial RNA polymerase promoter specificity  $\sigma$  subunit. // Mol. Cell. 2002. V. 9. P. 527-539.
- Gourse R.L., Ross W., Gaal T. UPs and downs in bacterial transcription initiation: Role of the  $\alpha$  subunit of RNA polymerase in promoter recognition. // Mol. Microbiol. 2000. V. 37. P. 687-695.
- Guo Y., Gralla J.D. Promoter opening via a DNA fork junction binding activity. Proc. Natl. Acad. Sci. // USA. 1998. V. 95. No. 20. P. 11655-11660.
- Roberts C.W., Roberts J.W. Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase. // Cell. 1996. V. 86. No. 3. P. 495-501.
- Kajitani M., and Ishihama A. Determination of the promoter strength in the mixed transcription system. II. Promoters of ribosomal RNA, ribosomal protein S1 and recA protein operons from *Escherichia coli*. // Nucleic Acids Research. 1983. V. 11. No. 12. P.3873-3888.
- Vogel S.K., Schulz A., Rippe K. Binding affinity of *Escherichia coli* RNA polymerase-<sup>54</sup>holoenzyme for the *glnAp2*, *nifH* and *nifL* promoters. // Nucleic Acids Research. 2002. V. 30. No. 18. P. 4094-4101.
- Kanehisa M., Goto S., Kawashima, S., Kuno Y., Hattori M. The KEGG resources for deciphering the genome// Nucleic Acids Res. 2004. V. 32. D277-D280.
- Mori H., Isono K., Horiuchi T., Miki T. Functional genomics of *Escherichia coli* in Japan, // Res Microbiol. 2000. 151. P. 121-128.
- Hisanori Kiryu, Taku Oshima, Kiyoshi Asai. Extracting relations between promoter sequences and their strengths from microarray data // Bioinformatics. 2005. V. 21. No7. 1062-1068.
- Kadritskaya, Z.G., Danilov, V.I. Quantum mechanical study of bases interactions in various associates in atomic dipole approximation // J. Theor. Biol. 1976. V. 59. 301-318.
- Spassky A., Kirkegaard K., Buc H. Changes in the DNA structure of the lac UV5 promoter during formation of an open complex with *Escherichia coli* RNA polymerase // Biochemistry. 1985. V. 24. No11. P.2723-31.
- Gralla J.D. Activation and repression of *E. coli* promoters // Curr Opin Genet. 1996. V. 6. No5. P.526-530.
- Roberts C.W., Roberts J.W. Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase // Cell, 86. 1996. No3. P:495-501.
- Guo Y., Gralla J.D. Promoter opening via a DNA fork junction binding activity // Proc. Natl. Acad. Sci. 1998. V.95. No20. P.1655-11660.