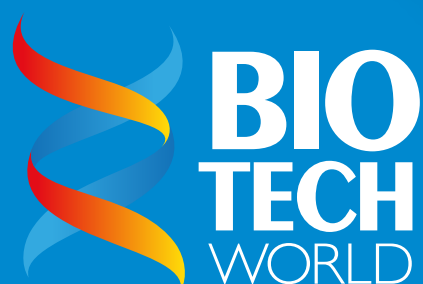


МАТЕРИАЛЫ КОНГРЕССА  
CONGRESS PROCEEDINGS



МЕЖДУНАРОДНЫЙ КОНГРЕСС  
**БИОТЕХНОЛОГИЯ:  
СОСТОЯНИЕ  
И ПЕРСПЕКТИВЫ  
РАЗВИТИЯ**

INTERNATIONAL CONGRESS  
**BIOTECHNOLOGY:  
STATE OF THE ART  
AND PERSPECTIVES**

**25 - 27 ФЕВРАЛЯ 2019**  
МОСКВА, ГОСТИНЫЙ ДВОР,  
ИЛЬИНКА, 4

**25 - 27 FEBRUARY 2019**  
ILYNKA 4, GOSTINY DVOR,  
MOSCOW

MACHINE LEARNING FOR MICROBIOTA ANALYSIS: INTERINDIVIDUAL VARIABILITY OF THE RESPONSE TO DIETARY INTERVENTION, Klimentko N. S., Popenko A. S., Alexeev D. G., Tyakht A. V.....	350
12. МЕТОД МОРФОМЕТРИИ КОЛОСА ПШЕНИЦЫ НА ОСНОВЕ АНАЛИЗА ИЗОБРАЖЕНИЙ, Комышев Е.Г., Генаев М.А., Афонников Д.А. ....	351
THE METHOD OF WHEAT SPIKE MORPHOMETRY BASED ON IMAGE ANALYSIS, Komyshev E.G., Genaev M.A., Afonnikov D.A. ....	352
13. МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ, ПОСТРОЕННЫЕ НА ИНФОРМАЦИИ О ПОСЛЕДОВАТЕЛЬНОСТЯХ И СТРУКТУРЕ, ДЛЯ РАСПОЗНАВАНИЯ СТРУКТУР СТЕБЕЛЬ-ПЕТЛЯ НА 3'-КОНЦАХ ТРАНСПОЗОНОВ L1 И ALU В ГЕНОМЕ ЧЕЛОВЕКА, А. Заикин, А. Шеин, М. Попцова .....	353
SEQUENCE-BASED AND STRUCTURE-BASED MACHINE-LEARNING MODELS FOR RECOGNITION OF 3'-END L1 AND ALU STEM-LOOPS IN HUMAN GENOME, A. Zaikin, A. Shein, M. Poptsova .....	354
14. МОДЕЛИРОВАНИЕ ЭВОЛЮЦИИ МЕТАБОЛИЗМА ПРОКАРИОТ В ПРОСТРАНСТВЕННО-ГЕТЕРОГЕННЫХ СРЕДАХ БЕЗ И ПРИ ФАГОВОЙ ИНФЕКЦИИ, Матушкин Ю.Г., Клименко А.И., Мустафин З.С., Лашин С.А. ....	355
MODELING THE EVOLUTION OF METABOLISM OF PROKARYOTES IN SPATIALLY HETEROGENEOUS ENVIRONMENTS, WITH AND WITHOUT PHAGE INFECTION, Matushkin Yu.G., Klimentko A.I., Mustafin Z.S., Lashin S.A. ....	356
15. НЕЙРОСЕТЕВАЯ МОДЕЛЬ СИГНАЛЬНОГО ПУТИ RAGE–NF-KB, Васильев П.М., Спасов А.А., Яналиева Л.Р., Кочетков А.Н., Ворфоломеева В.В., Клочков В.Г., Аппазова Д.Т.....	357
NEURAL NETWORK MODEL OF THE RAGE–NF-KB SIGNALING PATHWAY, Vassiliev P.M., Spasov A.A., Yanaliev L.R., Kochetkov A.N., Vorfolomeeva V.V., Klochkov V.G., Appazova D.T. ....	357
16. ОСОБЕННОСТИ СВЯЗЫВАНИЯ miRNA С mRNA ГЕНОВ ИМЕЮЩИХ НУКЛЕОТИДНЫЕ ПОВТОРЫ, Белкожаев А.М., Ниязова Р.Е. ....	358
FEATURES OF miRNA BINDING WITH mRNA OF GENES HAVING NUCLEOTIDE REPEATS, Belkozhaev A.M., Niyazova R.E.....	359
17. ПОИСК НЕКАНОНИЧЕСКИХ СТРУКТУР ДНК КАК НУКЛЕОСОМНЫХ БАРЬЕРОВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ, Теванян Э.А., Попцова М.С.....	360
SEARCHING FOR NON-B-DNA STRUCTURES AS NUCLEOSOME BARRIERS WITH MACHINE LEARNING METHODS, E. A. Tevanyan, M. S. Poptsova .....	361
18. ПОИСК НОВЫХ ГЕНОВ В «СКРЫТОЙ» ЧАСТИ ТРАНСКРИПТОМОВ СЕЛЬСКОХОЗЯЙСТВЕННЫХ РАСТЕНИЙ, М.А. Генаев, Н.А. Шмаков, З.С. Мустафин, А.М. Мухин, Д.К. Константинов, А.В. Дорошков, С.А. Лашин, Д.А. Афонников .....	362
SEARCH FOR NEW GENES IN THE “HIDDEN” PART OF AGRICULTURAL PLANT TRANSCRIPTOMES, M.A. Genaev, N.A. Shmakov, Z.S. Mustafin, A.M. Mukhin, D.K. Konstantinov, A.V. Doroshkov, S.A. Lashin, D.A. Afonnikov .....	363
19. ПОЛУЧЕНИЕ НОВОГО БИОЛОГИЧЕСКОГО ЗНАНИЯ ИЗ РЕЗУЛЬТАТОВ ПОЛНОГЕНОМНЫХ ИССЛЕДОВАНИЙ АССОЦИАЦИЙ С ПОМОЩЬЮ ПЛАТФОРМЫ GWAS-MAP, Т.И.Шашкова, Д.Д.Горев, Я.А.Цепилов, Е.Д.Пахомов, А.А. Торгашева, П.Джоши, Ю.С.Аульченко .....	364
OBTAINING NEW BIOLOGICAL KNOWLEDGE FROM THE RESULTS OF GENOME-WIDE ASSOCIATION STUDIES USING THE GWAS-MAP PLATFORM, T.Shashkova, D.Gorev, Y.Tsepilov, E.Pakhomov, A.Torgasheva, P.Joshi, Y.Aulchenko.....	367
20. РАЗРАБОТКА МЕТОДОВ АНАЛИЗА «БОЛЬШИХ ДАННЫХ» ДЛЯ ПОИСКА НОВЫХ АНТИ-ВИЧ СОЕДИНЕНИЙ, П.И.Савосина, Л.А.Столбов, Д.С.Дружилловский, Д.А.Филимонов, В.В.Поройков .....	371
DEVELOPMENT OF METHODS FOR BIG-DATA ANALYSIS TO DISCOVER NEW ANTI-HIV COMPOUNDS, P.Savosina, L.Stolbov, D.Druzhilovskiy, D.Filimonov, V.Poroikov .....	372
21. РАСПРЕДЕЛЕНИЕ СИСТЕМ РЕСТРИКЦИИ-МОДИФИКАЦИИ В ОБРАЗЦАХ ИЗ МИРОВОГО ОКЕАНА, М. Хачатурян, И. Русинов, А. Ершова .....	374
DISTRIBUTION OF RESTRICTION-MODIFICATION SYSTEMS GENES IN SAMPLES OF THE WORLD OCEAN, M. Khachaturyan, I. Rusinov, A. Ershova .....	375
22. РОЛЬ АМИЛОИДОГЕНЕЗА В ЗАПАСАНИИ БЕЛКА В СЕМЕНАХ РАСТЕНИЙ, Нижников А.А., Белоусов М.В., Белоусова М.Е., Косолапова А.О., Штарк О.Ю., Антоненц К.С. ....	376
THE ROLE OF AMYLOIDOGENESIS IN THE PROTEIN STORAGE IN PLANT SEEDS, Nizhnikov A.A., Belousov M.V., Belousova M.E., Kosolapova A.O., Stark O.Yu., Antonets K.S. ....	376
23. САЙТЫ СВЯЗЫВАНИЯ miRNA С mRNA ГЕНОВ СЕМЕЙСТВА МЫВ ЖИВОТНЫХ, Мырзабекова М.О. ....	377

the AutoDock Vina program, an ensemble docking of these compounds into sites 66 of the 3D models of 22 valid target proteins was performed, and the minimum docking energy  $\Delta E$  was determined. A training set was formed on the levels of activity and affinity of 183 known RAGE inhibitors in relation to 22 target proteins of the RAGE–NF- $\kappa$ B signaling pathway, which includes 4026  $\Delta E$  values. Using the method of artificial multilayer perceptron neural networks with the “bottleneck” architecture, the RAGE–NF- $\kappa$ B signaling pathway model was built using Statistica Neural Networks software. Using the ROCK analysis method, the accuracy of the obtained model was estimated, which for a high level of RAGE inhibitory activity was 90% ( $p < 0.001$ ). The model makes it possible to predict the experimental level of RAGE inhibitory activity of chemical compounds according to calculated values of their affinity for 22 target proteins of the RAGE–NF- $\kappa$ B signaling pathway. The resulting model will be used for the targeted search for highly active systemic multifunctional multitarget RAGE-inhibitors for the treatment of patches in diabetes mellitus and Alzheimer’s disease.

This work was carried out with the financial support of the Russian Foundation for Basic Research (project 18-015-00499).

УДК 577.21

## ОСОБЕННОСТИ СВЯЗЫВАНИЯ miRNA С mRNA ГЕНОВ ИМЕЮЩИХ НУКЛЕОТИДНЫЕ ПОВТОРЫ

Белкожаев А.М., Ниязова Р.Е.

НИИ Проблем биологии и биотехнологии, КазНУ им. аль-Фараби, Алматы, Казахстан  
 050038, Алматы, Проспект аль-Фараби, д. 71, корп. 6  
 e-mail: [Ayaz\\_jarkent@mail.ru](mailto:Ayaz_jarkent@mail.ru)

Нуклеотидные повторы в 5’UTR, CDS и 3’UTR имеют важное биологическое значение, поэтому необходимо изучить особенности взаимодействия miRNA с mRNA генов, имеющих нуклеотидные повторы.

**Ключевые слова:** miRNA, mRNA, нуклеотидные повторы, ген

Гены имеющие нуклеотидные повторы являются одной из основных причин развития наследственных заболеваний человека [1]. В данной работе изучались особенности связывания miRNA с mRNA генов, имеющих нуклеотидные повторы. Нуклеотидные последовательности mRNA 95 генов человека были получены из Genbank (<http://www.ncbi.nlm.nih.gov>), и 2567 miRNA были взяты из miRBase (<http://mirbase.org>). Поиск генов мишеней для miRNA проводили, используя программу MirTarget [2]. Сайты связывания miRNA с mRNA отбирали с отношением  $\Delta G/\Delta G_m$  равным 86% и более. Из них 73 сайтов связывания расположены в CDS, 28 сайтов в 5’UTR и 194 сайтов в 3’UTR. Результаты исследования показали, что из 95 генов, имеющих нуклеотидные повторы, в mRNA только 56 генов найдены 295 сайтов связывания miRNA. В mRNA гена *HTT* имелись сайты связывания miRNA с величиной  $\Delta G/\Delta G_m$ , выше 90%. Нами установлены miRNA, связывающиеся с высокой эффективностью с mRNA гена *HTT* (CAG повторы). Полученные результаты показывают, что 16 генов с нуклеотидными повторами, являются мишенями двух и более miRNA: *ADRB3* (miR-466, miR-6845-5p) динуклеотидные повторы GU, *GHR* (miR-1273a, miR-1273c, miR-1273f) динуклеотидные повторы UC, *IGF2R* (miR-4307, miR-466, miR-4734) тринуклеотидные AAG и динуклеотидные повторы UG, *LDLR* (miR-1303, miR-6751-5p) динуклеотидные повторы AC, *TGFB1* (miR-4651, miR-6824-5p, miR-6877-5p, miR-877-3p) пентануклеотидные GCCCC и тринуклеотидные повторы GCC, *TNFRSF18* (miR-4298, miR-8073) тринуклеотидные повторы UCC, *VSNL1* (miR-574-5p, miR-7111-3p, miR-877-3p) динуклеотидные повторы AC, *XRCC1* (miR-4763-3p, miR-574-5p) динуклеотидные повторы AC, *HLA-A* (miR-4786-3p, miR-548ah-3p) тетрануклеотидные повторы GGGC, *HTT* (miR-1200, miR-185-5p, miR-203a, miR-4722-3p, miR-5008-5p, miR-5193, miR-6072, miR-6722-3p, miR-6746-5p, miR-6748-5p, miR-6778-5p, miR-1322) динуклеотидные UG и тринуклеотидные повторы CAG, *ATXN2* (miR-6089, miR-466) динуклеотидные повторы GU, *EP400* (miR-6089, miR-6874-3p, miR-6882-3p) тринуклеотидные повторы CAG, *FAM157B* (miR-1273g-3p, miR-3162-3p, miR-3193) тринуклеотидные повторы CUC, *IRF2BPL* (miR-5091, miR-7110-5p) динуклеотидные повторы UC, *IRS1* (miR-4655-3p, miR-466) динуклеотидные повторы GU, *KIAA2018* (miR-1277-5p, miR-574-5p) динуклеотидные повторы AU. В результате нами были отобраны шесть miRNA, которые имеют сайты связывания с высокой свободной энергией связывания в mRNA двух и более генов, имеющих нуклеотидные повторы: miR-466 (*ADCYAP1R*, *ADRB3*, *FGF9*, *IGF2R*, *SEMA6D*, *IRS1*, *NCOA3*), miR-574-5p (*DMD*, *KLF7*, *VSNL1*, *XRCC1*, *AFF3*, *ARID3B*, *KIAA2018*), miR-1273a (*CYP4F3*, *GHR*, *KCNN3*, *MPRIIP*), miR-1273f (*GHR*, *MEF2A*), miR-1273g-3p (*AFF3*, *FAM157B*,

*KCNN3, KIAA2018, MEF2A*), miR-6089 (*TGFB1, ATXN2, EP400*). Гены *CYP4F3, GHR, KCNN3, MPRIP, GHR, MEF2A, AFF3, FAM157B, KCNN3, KIAA2018, MEF2A* являются мишенями для семейства miR-1273. Для некоторых miRNA обнаружены полисайты: для miR-466 – девять сайтов в mRNA гена *ADCYAP1R1*, 11 сайтов в mRNA гена *ADRB3*, восемь сайтов в mRNA гена *FGF9*, семь сайтов в mRNA гена *IGF2R*, 12 сайтов в mRNA гена *SEMA6D*, шесть сайтов в mRNA гена *IRS1* и четыре сайта в mRNA гена *NCOA3*; для miR-574-5p – шесть сайтов в mRNA гена *DMD*, три сайта в mRNA гена *KLF7*, 13 сайтов в mRNA гена *VSNL1*, 11 сайтов в mRNA гена *XRCC1*, семь сайтов в mRNA гена *AFF3*, три сайта в mRNA гена *ARID3B*, девять сайтов в mRNA гена *KIAA2018*.

Литература:

1. Сиянова Е.Ю., Миркин С.М. Экспансия тринуклеотидных повторов // Молекулярная биология. – 2001. - №2. – С. 208-223.
2. Ivashchenko A., Berillo O., Pyrkova A., Niyazova R., Atambayeva S. MiR-3960 binding sites with mRNA of human genes // Bioinformation. – 2014. - 10(7). – P. 423-427

UDC 577.21

## FEATURES OF miRNA BINDING WITH mRNA OF GENES HAVING NUCLEOTIDE REPEATS

**Belkozhayev A.M., Niyazova R.E.**

Research Institute of Problems of Biology and Biotechnology, al-Farabi Kazakh National University, Almaty, Kazakhstan  
050038, Almaty, Prospect al-Farabi, 71, corp.6  
e-mail: [Ayaz\\_jarkent@mail.ru](mailto:Ayaz_jarkent@mail.ru)

Nucleotide repeats in 5'UTR, CDS and 3'UTR have an important biological significance, therefore it is necessary to study the features of the interaction of miRNA with mRNA genes having nucleotide repeats.

**Key words:** miRNA, mRNA, nucleotide repeats, gene

Genes with nucleotide repeats are one of the main causes of the development of human hereditary diseases [1]. In this paper, the features of miRNA binding to mRNA genes with nucleotide repeats were studied. The nucleotide sequences of the mRNA 95 human genes were obtained from Genbank (<http://www.ncbi.nlm.nih.gov>), and 2567 miRNAs were taken from miRBase (<http://mirbase.org>). Search for miRNA target genes was performed using the MirTarget program [2]. The miRNA binding sites for mRNA were selected with an  $\Delta G/\Delta G_m$  ratio of 86% or more. Of these, 73 binding sites are located in CDS, 28 sites in 5'UTR and 194 sites in 3'UTR. The results of the study showed that of the 95 genes having nucleotide repeats, only 56 genes in the mRNA are 295 miRNA binding sites. In the mRNA of the *HTT* gene, there were miRNA binding sites with  $\Delta G/\Delta G_m$ , above 90%. We have established miRNAs that bind with high efficiency to the mRNA of the *HTT* gene (CAG repeats). The results show that 16 genes with nucleotide repeats are targets of two or more miRNAs: *ADRB3* (miR-466, miR-6845-5p) dinucleotide repeats GU, *GHR* (miR-1273a, miR-1273c, miR-1273f) dinucleotide repeats UC, *IGF2R* (miR-4307, miR-466, miR-4734) trinucleotide AAG and dinucleotide repeats UG, *LDLR* (miR-1303, miR-6751-5p) dinucleotide repeats AC, *TGFB1* (miR-4651, miR-6824-5p, miR-6877-5p, miR-877-3p) pentanucleotide GCCCC and trinucleotide repeats GCC, *TNFRSF18* (miR-4298, miR-8073) trinucleotide repeats UCC, *VSNL1* (miR-574-5p, miR-7111-3p, miR-877-3p) dinucleotide repeats AC, *XRCC1* (miR-4763-3p, miR-574-5p) dinucleotide repeats AC, *HLA-A* (miR-4786-3p, miR-548ah-3p) tetranucleotide repeats GGGC, *HTT* (miR-1200, miR-185-5p, miR-203a, miR-4722-3p, miR-5008-5p, miR-5193, miR-6072, miR-6722-3p, miR-6746-5p, miR-6748-5p, miR-6778-5p, miR-1322) dinucleotide UG and trinucleotide repeats of CAG, *ATXN2* (miR-6089, miR-466) dinucleotide repeats GU, *EP400* (miR-6089, miR-6874-3p, miR-6882-3p) trinucleotide repeats CAG, *FAM157B* (miR-1273g-3p, miR-3162-3p, miR-3193) trinucleotide repeats CUC, *IRF2BPL* (miR-5091, miR-7110-5p) dinucleotide repeats UC, *IRS1* (miR-4655-3p, miR-466) dinucleotide repeats GU, *KIAA2018* (miR-1277-5p, miR-574-5p) dinucleotide repeats AU. As a result, we selected six miRNAs that have binding sites with high free binding energy in mRNA of two or more genes with nucleotide repeats: miR-466 (*ADCYAP1R, ADRB3, FGF9, IGF2R, SEMA6D, IRS1, NCOA3*), miR-574 -5p (*DMD, KLF7, VSNL1, XRCC1, AFF3, ARID3B, KIAA2018*), miR-1273a (*CYP4F3, GHR, KCNN3, MPRIP*), miR-1273f (*GHR, MEF2A*), miR-1273g-3p (*AFF3, FAM157B, KCNN3, KIAA2018, MEF2A*), miR-6089 (*TGFB1, ATXN2, EP400*). Genes *CYP4F3, GHR, KCNN3, MPRIP, GHR, MEF2A, AFF3, FAM157B, KCNN3, KIAA2018, MEF2A* are targets for the miR-1273 family. For some miRNA polysites were found: for miR-466, nine sites in the mRNA *ADCYAP1R1* gene, 11 sites in the mRNA *ADRB3* gene, eight sites in the mRNA *FGF9* gene, seven sites in the

mRNA *IGF2R* gene, 12 sites in the mRNA *SEMA6D* gene, six sites in the mRNA the *IRS1* gene and four sites in the mRNA *NCOA3* gene; for miR-574-5p, six sites in the mRNA *DMD* gene, three sites in the mRNA *KLF7* gene, 13 sites in the mRNA *VSNL1* gene, 11 sites in the mRNA *XRCC1* gene, seven sites in the mRNA *AFF3* gene, three sites in the mRNA *ARID3B* gene, nine sites in the mRNA *KIAA2018* gene.

References:

1. Siyanova E.Y., Mirkin S.M. Expansion of trinucleotide repeats // *Molecular Biology*. - 2001. - №2. - p. 208-223.
2. Ivashchenko A., Berillo O., Pyrkova A., Niyazova R., Atambayeva S. MiR-3960 binding sites with mRNA of human genes // *Bioinformatics*. - 2014. - 10 (7). - P. 423-427

## ПОИСК НЕКАНОНИЧЕСКИХ СТРУКТУР ДНК КАК НУКЛЕОСОМНЫХ БАРЬЕРОВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Теванян Э.А., Попцова М.С.

Лаборатория биоинформатики, Департамент больших данных и информационного поиска, факультет компьютерных наук, Национальный исследовательский университет «Высшая школа экономики», Россия, 125319, г. Москва, Кочновский проезд, д.3  
 e-mail: [etevanian@hse.ru](mailto:etevanian@hse.ru); [mpoptsova@hse.ru](mailto:mpoptsova@hse.ru)

Мы обучили алгоритм случайного леса для распознавания паттернов взаимного расположения нуклеосом и вторичных структур ДНК, которые могут служить барьерами для нуклеосом, в геноме мыши. Мы показали, что среди четырех типов рассмотренных структур (Z-ДНК, H-ДНК, G-квадруплексов и участков SIDD) лучшее качество модели достигается для G-квадруплексов и H-ДНК.

**Ключевые слова:** вторичные структуры ДНК, G-квадруплексы, H-ДНК, Z-ДНК, нуклеосомные барьеры, расположение нуклеосом, методы машинного обучения, случайный лес

Геномы обладают большим потенциалом образования вторичных структур ДНК, которые воздействуют на различные геномные процессы, включая транскрипцию. Одним из механизмов регуляции транскрипции является регуляция расположения нуклеосом. Хотя нуклеосомы наматываются только на каноническую, так называемую, В-форму, другие, отличные от В-формы, структуры ДНК могут конкурировать с нуклеосомами за расположение в геноме или служить барьерами, разделяющими нуклеосомные массивы.

В данной работе мы использовали данные перманганат/S1 нуклеаз-футпринтинга [1], который позволяет определить потенциальные сайты образования вторичных структур ДНК на участках расплетенной ДНК. Для определения расположения нуклеосом мы использовали данные MNase-seq [1]. Компьютерная аннотация генома мыши вторичными структурами ДНК производилась с помощью следующих программ и алгоритмов: Z-ДНК – Zhunt [2], H-ДНК - Inverted Repeats Finder [3], G-квадруплексы – QuadParser [4], участки SIDD – алгоритм из [5]. В результате в геноме мыши существует потенциально 250-420 тысяч сайтов образования структур каждого типа, в то время как число вторичных структур ДНК, обогащенных участками расплетенной ДНК, которые были обнаружены при помощи метода перманганат/S1 нуклеаз-футпринтинга, составляет 4-8% от числа структур, предсказанных с помощью компьютерных программ.

Для анализа взаимного расположения нуклеосом и вторичных структур ДНК мы рассмотрели области в 500 п.о., центрированных на структуре. Анализ нуклеосомных профилей вокруг структур ДНК выявил три типа паттернов: 1) структура окружена нуклеосомой с двух сторон, 2) структура расположена только с одной стороны нуклеосомы и 3) на участке нет нуклеосом.

Используя статистику динуклеотидов и триплетов, мы построили модель машинного обучения (классификатор по алгоритму случайного леса), распознающую принадлежность региона к одному из паттернов. Качество модели по метрике ROC-AUC достигло 86% для G-квадруплексов, 79% для H-ДНК, 73% для SIDD и 63% для Z-ДНК.

Модель может быть улучшена добавлением новых характеристик, в числе которых физические и химические свойства ДНК последовательности, такие, как энтальпия, энтропия, энергия Гиббса, гидрофильность, а также структурные свойства спирали (Shift, Rall, Slide, Rise, Tild, Bend), доступные в базе данных DiProDB [6]. Возможно, добавление новых характеристик и не улучшит качество модели, но даст возможность определить, какие структурные свойства ДНК играют важную роль в классификации паттернов.