Original article / Оригинальная статья
https://doi.org/10.47093/2218-7332.2023.907.12

In silico prediction of the transcription factor-enhancer interaction as a first stage of axonal growth regulation

Danil D. Kotelnikov, Ivan A. Sinyakin, Evgeny A. Borodin✉, Tatyana A. Batalova
Amur State Medical Academy
95, Gorky str., Blagoveshchensk, 675001, Russia

Abstract

The development of neurodegenerative diseases is associated with proper neuronal circuit formation, axonal guidance. The DCC receptor (deleted in colorectal cancer / colorectal cancer suppressor) and SHH (sonic hedgehog protein) are among the key regulators of axonal guidance.

Aim. Interaction prediction of specific enhancer regions of *DCC* and *SHH* genes with respectively annotated transcription factors.

Materials and methods. An *in silico* study was performed. The iEnhancer-2L and ES-ARCNN algorithms were selected to estimate enhancer sequence strength. The interaction between transcription factor and enhancer sequence was assessed using the molecular docking method. The enhancer sequence of DCC and SHH protein genes were taken from the NCBI open-source database in FASTA format. Ensembl database was used for enhancer mapping, GeneCards was used for screening and selection of potentially appropriate enhancers and transcription factors associated with these enhancers. The structures of transcription factors as well as their DNA-binding domains were taken from the UniProtKB/Swiss-prot database. An HDOCK scoring function was used as a metric for assessing the possibility of interaction of the target gene transcription factor with associated enhancer sequence.

Results. The results showed that the interactions of transcription factor NANOG with the *DCC* gene enhancer sequence and the interaction of transcription factor CEBPA with the *SHH* gene enhancer sequence predicted by molecular docking method are potentially possible. The iEnhancer-2L and ES-ARCNN algorithms predicted the enhancer sequence of the *SHH* gene as strong one. The enhancer sequence of the *DCC* gene was estimated as strong in the iEnhancer-2L algorithm and as weak in ES-ARCNN. Binding of the DCC gene enhancer sequence to the transcription factor NANOG at 1–206 bp and 686–885 bp sites is the most probable, binding of the SHH gene enhancer sequence to the transcription factor CEBPA at 1–500 bp (HDOCK limitation of 500 bp) is possible.

Conclusion. *In silico* techniques applied in this study demonstrated satisfactory results of predicting the interaction of the transcription factor with the enhancer sequence. Limitations of the current techniques is the lack of consideration of specific transcription factor binding sites. This drawback can be eliminated by implementing an *ab initio* molecular dynamics simulations into the present pipeline.

Keywords: molecular docking; *in silico*; *DCC*; *SHH*; binding site; CEBPA; NANOG; SITECON
MeSH terms:
AXONS
TRANSCRIPTION FACTORS – ANALYSIS
GENE EXPRESSION REGULATION
COMPUTER SIMULATION
For citation: Kotelnikov D.D., Sinyakin I.A., Borodin E.A., Batalova T.A. *In silico* prediction of the transcription factor-enhancer interaction as a first stage of axonal growth regulation. Sechenov Medical Journal. 2023; 14(4): 42–50. Epub ahead of print 29.11.2023. https://doi.org/10.47093/2218-7332.2023.907.12

CONTACT INFORMATION
Evgeny A. Borodin, Dr. of Sci. (Medicine), Professor, Head of the Department of Chemistry, Amur State Medical Academy
Address: 95, Gorky str., Blagoveshchensk, 675001, Russia
Tel.: +7 (962) 284-06-40
E-mail: borodin54@mail.ru

Ethics statements. No ethical approval by the local Ethics Committee was required since the study was not conducted on humans or laboratory animals.
Conflict of interests. The authors declare that there is no conflict of interests.
Financial support. The study was not sponsored (own resources).

Received: 24.08.2023
Accepted: 28.09.2023
Date of publication online: 29.11.2023
Date of publication: 30.11.2023

УДК [611.018.83.008.6:575.1]:004.94

In silico предсказание взаимодействия транскрипционного фактора и энхансера как первого этапа регуляции аксонального роста

Д.Д. Котельников, И.А. Синякин, Е.А. Бородин✉, Т.А. Баталова
ФГБОУ ВО «Амурская государственная медицинская академия» Минздрава России
ул. Горького, д. 95, г. Благовещенск, 675001, Россия

Аннотация

Развитие нейродегенеративных заболеваний ассоциировано с правильным формированием нейронной цепи – аксональным наведением. Среди ключевых регуляторов аксонального наведения – рецептор DCC (deleted in colorectal cancer / colorectal cancer suppressor, супрессор колоректального рака) и белок SHH (sonic hedgehog protein, «сверхзвуковой ёжик»).

Цель. Предсказание взаимодействия определенных энхансерных областей генов *DCC* и *SHH* с аннотированными для них факторами транскрипции.

Материалы и методы. Проведено исследование *in silico*. Для оценки силы энхансерной последовательности выбраны алгоритмы iEnhancer-2L и ES-ARCNN. Анализ взаимодействия транскрипционного фактора с энхансерной последовательностью производился с использованием метода молекулярного докинга. Энхансерная последовательность генов белков DCC и SHH взята из открытой базы данных NCBI в FASTA-формате. Для картирования энхансеров использовалась база Ensembl, для отбора потенциальных энхансеров и транскрипционных факторов к ним – GeneCards. Структуры транскрипционных факторов, а также их ДНК-связывающие домены были взяты из базы данных UniProtKB/Swiss-prot. В качестве метрики оценки возможности взаимодействия транскрипционных факторов с целевой энхансерной последовательностью использована оценочная функция (score).

Результаты. Результаты исследования показали, что взаимодействие транскрипционного фактора NANOG с энхансерной последовательностью гена *DCC* и взаимодействие транскрипционного фактора CEBPA с энхансерной последовательностью гена *SHH*, предсказанные путем метода межмолекулярного докинга, являются потенциально возможными. Алгоритмы iEnhancer-2L и ES-ARCNN предсказали энхансерную последовательность гена *SHH* как сильную. Энхансерная последовательность гена *DCC* в алгоритме iEnhancer-2L оценена как сильная, в ES-ARCNN – как слабая. Связывание энхансерной последовательности гена *DCC* с транскрипционным фактором NANOG на промежутках 1–206 bp и 686–885 bp является наиболее вероятным, связывание энхансерной последовательности гена *SHH* с транскрипционным фактором CEBPA на промежутке 1–500 bp (ограничение HDOCK в 500 bp) является возможным.

Заключение. Примененные методики в исследовании *in silico* продемонстрировали удовлетворительные результаты предсказания взаимодействия транскрипционного фактора с энхансерной последовательностью. Ограничением методики является отсутствие учета конкретных сайтов связывания транскрипционных факторов с дезоксирибонуклеиновой кислотой. Этот недостаток может быть устранен включением в пайплайн *ab initio* метода молекулярной динамики.

Ключевые слова: молекулярный докинг; *in silico*; *DCC*; *SHH*; сайт связывания; CEBPA; NANOG; SITECON

Рубрики MeSH:
АКСОНЫ
ТРАНСКРИПЦИИ ФАКТОРЫ – АНАЛИЗ
ГЕННОЙ ЭКСПРЕССИИ РЕГУЛЯЦИЯ
КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Для цитирования: Котельников Д.Д., Синякин И.А., Бородин Е.А., Баталова Т.А. *In silico* предсказание взаимодействия транскрипционного фактора и энхансера как первого этапа регуляции аксонального роста. Сеченовский вестник. 2023; 14(4): 42–50. Публикация онлайн 29.11.2023. https://doi.org/10.47093/2218-7332.2023.907.12

КОНТАКТНАЯ ИНФОРМАЦИЯ:
Бородин Евгений Александрович, д-р мед. наук, профессор, заведующий кафедрой химии ФГБОУ ВО «Амурская государственная медицинская академия» Минздрава России
Адрес: ул. Горького, д. 95, г. Благовещенск, 675001, Россия
Тел.: +7 (962) 284-06-40
E-mail: borodin54@mail.ru

Соответствие принципам этики. Утверждение протокола Локальным этическим комитетом не требовалось, так как исследование не проводилось на людях или лабораторных животных.
Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
Финансирование. Исследование не имело спонсорской поддержки (собственные ресурсы).

Поступила: 24.08.2023
Принята: 28.09.2023
Дата публикации онлайн: 29.11.2023
Дата печати: 30.11.2023

Abbreviations:

Bp – base pairs
DCC – deleted in colorectal cancer / colorectal cancer suppressor
DNA – deoxyribonucleic acid
SHH – sonic hedgehog
TFBS – transcription factor-binding site
ES – enhancer sequence
TF – transcription factor

Machine training techniques have begun to be widely used in the field of biomedicine to search for potential ligands subsequently used as pharmacological targets [1]. Classically, enhancers can be defined as cis-acting deoxyribonucleic acid (DNA) sequences whose function is to increase the transcription (expression) of a gene. Typically, they function independently of orientation and at varying distances from the target-promoter (or promoters) [2]. One of the necessary conditions for the functioning of enhancers is the availability of their transcription factor-binding site (TFBS), that is, the enhancer and promoter itself, as well as the DNA sequence of any gene itself, must be in a decondensed state [3]. The length of a typical enhancer sequence (ES) is approximately 100–1000 bp (base pairs).

Functionally, it is possible to assess the work of an enhancer by using several metrics as the type of transcription factor (TF) expected to bind, the orientation (location) of the enhancer, its affinity; the order, number and distance between multiple TFBSs throughout the enhancer and ultimately the underlying DNA topology, collectively referred to as “enhancer architecture” [4]. The ability of potential control by enhancers as one of the regulating protein expression elements is an important task in modern molecular biology.

A central challenge in understanding gene regulation remains to explain how specific sets of genes are selected for an expression during a cell growth, differentiation, or in response to environmental cues. Primary goal: to determine how a fixed-size genome establishes a huge range of different developmental programs. All information required for regulatory functions mediated by enhancers and promoters is encoded in the DNA sequence through their unique combination of modules [5]. That is, each individual module binds one or more TFs, thus performing one of the functions of the entire regulatory element. In addition, certain modules can serve as central switching units, reacting accordingly to input data from other modules of the same element [5].

Structural and functional studies of TFs have shown that they are modular proteins having distinct regions dedicated to different functions: a DNA-binding domain that directs the protein to a specific DNA site, a multimerization domain that allows the accumulate

homo- or hetero-multimers, and an effector domain that can modulate the rate of transcription (activation or repression) [5]. The modular nature of TFs along with the modular architecture of enhancers and promoters provides the basis for a combinatorial mode of gene expression; and unlimited ability to mix and match enhancer and protein modules suggests that there may be an infinite number of unique gene expression programs embedded in genomes of relatively limited size. Enhancers and promoters are assumed to be at the endpoints of signal transmission pathways that modify TFs. Thus, a specific gene is expressed only if the cell perceives and equally interprets specific signals [5].

Axonal guidance, along with cell migration and synaptogenesis, is one of the key processes required for the proper formation of a neural circuit. It is regulated by a wide range of signaling cascades occurring both in neurons themselves and in other cells, including neuroglia [6].

With technical improvements in imaging of axonal guidance processes both in vivo and in vitro, defects in axonal transport have been shown to be associated with the development of neurodegenerative diseases. These defects are caused by genetic mutations leading to a lack of binding of motor proteins (kinesin, dynein) or dysfunction or development of microtubule instability [6].

Among the regulators of axonal guidance it is possible to distinguish the DCC receptor (deleted in colorectal cancer / colorectal cancer suppressor) that is constitutively expressed on the surface of axons. The binding of netrin-1 to DCC causes chemoattraction [7]. Netrin-1 itself is secreted by cells of the plate of the spinal cord fundus, diffuses into the extracellular matrix and creates a gradient that attracts growing commissural axons to the ventral midline of the spinal cord. Various signal-guiding receptors on the surface of the neuron's growth cone are constantly in contact with the environment, interacting with appropriate signals released by target cells, thereby allowing the axon to move correctly along a precise trajectory among many possible routes [8]. According to some data, dysfunction of the DCC receptor together with netrin-1 contributes to optic nerve hypoplasia [9].

The SHH protein (sonic hedgehog protein) can be identified as an intracellular regulator of axonal direction and growth. It plays a significant role in mediating axon guidance of commissural neurons in the developing spinal cord through a non-canonical transcription-independent pathway, interacting with the Boc receptor and a member of the receptor family combined with G-protein, the SMO receptor (Smoothened receptor). Activation of Src family kinase through SMO by SHH guides the axons of commissural neurons, playing a key role in the entire process of growth and guidance [10, 11].

The goal of the work is to predict the interaction of certain enhancer regions of the DCC and SHH genes with transcription factors annotated for them.

MATERIALS AND METHODS

Data preparation

The netrin receptor DCC and the SHH protein, which indirectly participate in the process of directed axon growth, were selected as target-proteins.

The ES of the DCC and SHH protein genes was taken from the NCBI open database (National Center for Biotechnology, US National Center for Biotechnology Information) in FASTA format[1,2]. The Ensembl database was used to map enhancers, and Gene Cards was used to select potential enhancers and TFs for them.

The structures of transcription factors as well as their DNA-binding domains were obtained from the UniProtKB/Swiss-prot database.

Finding of enhancer sequences and predicting their strength

To perform the experiment, sequences were selected whose activity, according to Ensembl, was observed in cells of neural origin; for the DCC and SHH genes, stellate neuroglial astrocytes cells were the common cell population. Other important selection criteria were the presence of an ES in both the Ensembl database and the GeneHencer database (presented on the GeneCards website), its close location relative to the gene itself (high GeneHencer Gene Association Score) and the presence of the desired gene as an annotated target (for example, DCC).

Algorithms for enhancer sequence strength assessment

To assess the strength of ES, two algorithms with relatively high prediction accuracy and easy to use were selected.

The first algorithm chosen was iEnhancer-2L[3] [12]. Its input data is the putative ES in FASTA format, and the output is a list of substrings obtained by fragmenting the original sequence with a reading frame 200 bp long (model enhancer length) and a step of 1 bp, as well as the strength of the resulting substring. Using iEnhancer-2L, the strength of the entire sequence was analyzed and if

[1] https://www.ncbi.nlm.nih.gov/nuccore/NC_000007.14?report=fasta&from=155815219&to=155815723&strand=true (access date: 01.06.2023).

[2] https://www.ncbi.nlm.nih.gov/nuccore/NC_000018.10?report=fasta&from=52310401&to=52313800 (access date: 01.06.2023).

[3] http://bliulab.net/iEnhancer-2L// (access date: 01.06.2023).

the prediction output indicated that all 200 bp substrings were strong, then the sequence itself was considered to be strong. When obtaining heterogeneous data: alternating strong, weak and non-functional enhancer regions, it was assumed that the sequence structure is not completely involved in binding to transcription factors, but only to potentially strong ones. The computational capabilities of iEnhancer-2L are limited when analyzing very long sequences, so the EP for the DCC gene (3399 bp) was divided into several substrings of 500 bp in length. These substrings were analyzed through iEnhancer-2L, and those of them being strong were selected as output data. This helps optimize the time allotted for docking a given EP with a transcription factor.

The second algorithm for analyzing the strength of EP was ES-ARCNN [13]; its input data is also the ES in FASTA format, and its output is the strength of the entire ES. According to a study conducted by T. Zhang and coauthors [13], the relative prediction accuracies of iEnhancer-2L and ES-ARCNN are 60.5 and 65.5%, respectively.

Molecular docking

To analyze the interaction of TF with ES, the molecular docking method was chosen. This is one of the modern methods in computational biology, the essence of which is to predict the optimal relative position of two biomolecules, ensuring their stable binding. The process itself represents the generation of potentially possible conformations/orientations of the ligand in the protein binding site [14].

HDOCK was chosen as the docking algorithm for our work[4]. It differs from its typical analogues, since it takes a DNA sequence in FASTA format as a ligand, from which it builds a structural model; the second binding substrate is the TF itself, which is loaded in pdb format. Additional settings included annotated on UniProtKB TF DNA-binding sites, if available [15]. The interpretation of binding metrics was carried out in accordance with the rules specified on the HDOCK server. Score is the result of calculating the ITScorePP or ITScorePR evaluation functions and is the authors' own development, and the confidence score serves as an empirically derived docking evaluation function and is calculated using the formula:

$$\text{Confidence score} = 1.0 / [1.0 + e^{0.02*(\text{Docking Score}+150)}]$$

It is important to clarify that the evaluation metric was not equated with the actual binding energy, since the former was not optimized basing on experimental data.

Analysis of the enhancer sequence for the presence of a transcription factor binding site

Selected areas of the ES were analyzed for the presence of corresponding TFBS using the Find TFBS with SITECON and Find TFBS with matrices modules in the free bioinformatics software Unipro UGENE[5] [16]. These modules match DNA sequences with an existing SITECON library, as well as with positional weight/frequency matrices. The selected ESs were loaded in FASTA format into Unipro UGENE, then a TFBS search was carried out: for CEBPA (CCAAT/enhancer-binding protein alpha, CCAAT/enhancer-binding protein alpha) SITECON[6] modules and weight matrices JASPAR ID: MA0102 were used. 2 (technically, a positional frequency matrix was used – a file in pfm format), for NANOG (homeobox protein NANOG) only positional weight matrices taken from the HOCOMOCO[7] database (data taken from quality A) and adapted to the structure used by Unipro UGENE were used (file in pwm format) [17]. The numbers used are calculated based on empirical data from SELEX (SITECON for CEBPA) as well as ChIP-Seq (matrix for CEBPA and NANOG). As output data, marked areas of potential TFBS were obtained, among which those with the highest score metric were taken into account.

Visual interpretation of the results was carried out using PyMol ver. 2.5.4 (Schrödinger, LLC, USA).

RESULTS

The obtained TFBS results are presented in Table 1.

The used models of TF, ES, its localization, length and predicted strength are summarized in Table 2.

The study of ES using different algorithms showed heterogeneous results: on the one hand, both iEnhancer-2L and ES-ARCNN predicted the ES of the SHH gene was assessed as strong, on the other hand, the ES of the DCC gene was assessed by iEnhancer-2L as strong, and by ES-ARCNN as weak.

The results of the binding of ES genes with the transcription factors indicated in Table 2 are shown in Table 3.

In accordance with the data obtained in Table 3, it can be confirmed that the binding of the ES gene DCC to the NANOG TF in the intervals 1–206 bp and 686–885 bp is most likely, binding of the SHH gene ES to the CEBPA TF in the range of 1–500 bp (HDOCK limit of 500 bp) is potentially possible.

[4] http://hdock.phys.hust.edu.cn/ (access date: 01.06.2023).
[5] http://ugene.net/ru/ (access date: 01.06.2023).
[6] http://wwwmgs.bionet.nsc.ru/mgs/programs/sitecon/tutorial.html (access date: 01.06.2023).
[7] https://hocomoco11.autosome.org/ (access date: 01.06.2023).

The figure demonstrates a visual interpretation of the results.

DISCUSSION

Based on our results obtained using a combination of several in silico methods, it can be confirmed that the interaction of NANOG with the ES of the DCC gene and the interaction of CEBPA with the ES of the SHH gene is potentially possible and may make for interesting further research. Since our work proposes an original research protocol, and the score and confidence score metrics are tied to a specific docking algorithm (HDOCK), comparison of the data obtained with the results of other researchers is limited. Nevertheless, the molecular docking method has already established itself as an independent method. Thus, in a study by P. Giri et al. [18], using the *in silico* docking method, patterns (AtMAPK3P) in *Arabidopsis thaliana* were identified. Out of the 131 transcription factors studied, only MYB 41 showed interaction with AtMAPK3P. Several novel MYB-interacting proteins have also been reported using minimal sequence motif searches as well

Table 1. **Transcription factor binding site data of *DCC* and *SHH* genes with target enhancer sequence**
Таблица 1. **Данные сайта связывания транскрипционного фактора энхансерной последовательности генов *DCC* и *SHH***

Parameter / Параметр	Gene / Ген		
	***DCC* (1–206 bp)**	***DCC* (686–885 bp)**	***SHH***
Transcription factor / Транскрипционный фактор	Homeobox protein NANOG		CCAAT/enhancer-binding protein alpha CEBPA
LOGO (ChIP-Seq based)			
Obtained TFBS / Найденный TFBS	**PWM**: GCTTTTCACATGACTGA (+) (131–147 bp) GTCAGTCATGTGAAAAG (–) (132–148 bp)	**PWM**: GGCATAGTTATGTCATC (+) (802–818 bp) GGCAGAAAGATTCTGAG (–) (824–840 bp)	**SITECON** : GGGGATTTCCCAAACCGGCCAAGC (+) (243–266 bp) CCGGTTTGGGAAATCCCCGCAGTC (–) (237–260 bp) **JASPAR**: TTTCCCAAA (+) (248–256 bp) TTGGGAAAT (–) (247–255 bp)
Score / Оценка	**PWM**: 68.6% (+), 72.8% (–)	**PWM**: 74% (+), 68.5% (–)	**SITECON**: 75.4% (+), 73.6 (–) **JASPAR**: 81% (+), 86% (–)

Table 2. **Summary of enhancer sequences of *DCC* and *SHH* genes**
Таблица 2. **Итоговые данные энхансерной последовательности генов *DCC* и *SHH***

Protein / Белок, UniProtKB ID	Gene / Ген, Cards ID	Localization / Локализация	Length / Длина	Power / Сила	Transcription factor / Транскрипционный фактор, UniProtKB URL
SHH/ Q15465	SHH (homo sapiens)/ GC07M155799	Chromosome 7: 155815219 – 155815723	505 bp	Strong / Сильная (iEnhancer-2L, ES-ARCNN)	CCAAT/enhancer-binding protein alpha (CEBPA)[8]
DCC/ P43146	DCC (homo sapiens)/ GC18P052340	Chromosome 18: 52310401 – 52313800	3399 bp	Weak / Слабая (ES-ARCNN) Strong / Сильная (**1–206**, 10–209, 14–219, 25–229, **686–885**, 689–888)	Homeobox protein NANOG[9]

Note: those regions of enhancer sequences of *DCC* gene that were used in this study are highlighted in bold.
Примечание: жирным шрифтом выделены те участки энхансерной последовательности гена *DCC*, которые использовались в данном исследовании.

[8] https://www.uniprot.org/uniprotkb/P49715/entry (access date: 01.06.2023)
[9] https://www.uniprot.org/uniprotkb/Q9H9S0/entry (access date: 01.06.2023).

Table 3. **HDOCK docking results**
Таблица 3. **Результаты докинга HDOCK**

Gene / Ген	Enhancer sequence / Энхансерная последовательность	Transcription factor / Транскрипционный фактор	Docking results / Результаты докинга	
			Score / Оценка	Confidence Score / Оценка достоверности
SHH	1–500 bp	CEBPA	–189.40	0.6874
DCC	1–206 bp	NANOG	–244.80	0.8694
	686–885 bp		–242.36	0.8638

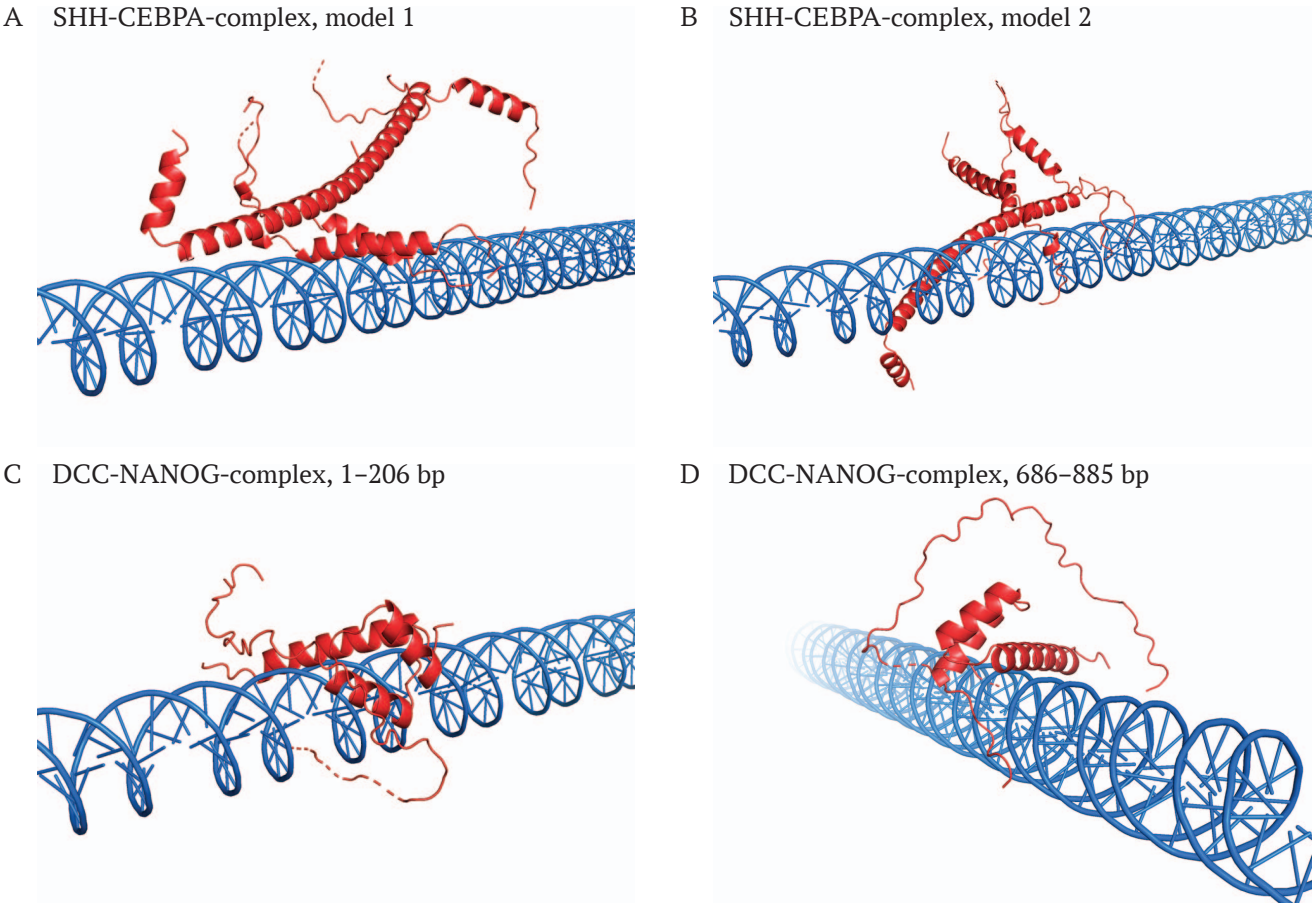

FIG. Visualization of successful docking conformations.
РИС. Визуализация успешных результатов межмолекулярного докинга.

as docking techniques, which need to be confirmed by in vitro kinase analyses [18]. This piece of research, along with ours, demonstrates the successful application of a bioinformatics approach at the initial stages of an experiment as a good predictive modeling.

In the study of domestic scientists A.M. Andrianova et al. [19] using virtual screening and molecular modeling methods, six potential peptidomimetics of the cross-reactive anti-HIV-1 (human immunodeficiency virus) neutralizing antibody N6 were identified, capable of imitating the pharmacophoric properties of this immunoglobulin through specific and effective interactions with the CD4-binding site of the gp120 protein of the virus envelope HIV [19]. It has been shown that a key role in the binding of these compounds to the gp120 protein is played by van der Waals interactions with conservative residues of the Phe43 cavity of the glycoprotein, which are critical for the attachment of HIV-1 to the CD4 cell receptor, as well as a hydrogen bond with the Asp-368gp120 residue, the formation of which increases the chemical affinity without activation of an undesirable allosteric effect [19]. Based on the results obtained, the authors concluded that the identified compounds can be considered as promising candidates for detailed experimental studies with a view to their further use in the development of new antiviral drugs.

Thus, it can be confirmed that the *in silico* study of A.M. Andrianova et al. [19], so as our study, was also successful in terms of identifying core domains as candidates for pharmacological molecules.

It is also worth mentioning that the developed protocol is not perfect for several reasons: the method does not take into account specific TF binding sites, i.e. docking was carried out "blindly"; the intermolecular docking method itself is "static" in nature, since it shows one of the possible conformations of the DNA-protein complex (which may be unstable); score and confidence score values are not universal.

The previously mentioned disadvantages can be eliminated by including the *ab initio* molecular dynamics method in the pipeline. Molecular dynamics simulation uses force fields being parametric equations describing the components for various forces (tensile, van der Waals, etc.) acting between atoms within and between molecules [20] and is a computational method used for the stability of conformations and dynamics of various biomolecules, including proteins and nucleic acids [21].

Premature use of such a powerful method without prior testing would be irrational, since such simulations are computationally (using GPUs or graphics processors) and financially expensive (renting a server for computing). At present, our group is already studying a new method and developing an updated pipeline that will complement and, possibly, correct the results obtained by docking.

CONCLUSION

In silico prediction of the interaction of TF with ES demonstrated satisfactory results and achieved the stated goal. Despite the fact that the applied methods are highly inaccurate and do not allow any precise conclusions to be drawn, the study showed that the use of such methods can serve as a useful tool for screening the potential interaction of TFs with a relevant DNA sequence (with an enhancer, promoter, etc.).

AUTHOR CONTRIBUTIONS

Danil D. Kotelnikov made a major contribution to the development of the concept and design of the study, as well as the preparation of illustrations. Ivan A. Sinyakin contributed to the search for literary sources, writing and editing the text. Evgeny A. Borodin and Tatyana A. Batalova supervised the process of statistical processing, writing and editing the text. All authors approved the final version of the article and are ready to take responsibility for all aspects of the submitted publication.

ВКЛАД АВТОРОВ

Д.Д. Котельников внес основной вклад в разработку концепции и дизайна исследования, а также подготовку иллюстраций. И.А. Синякин внес вклад в поиск литературных источников, написание и редактирование текста. Е.А. Бородин и Т.А. Баталова руководили процессом статистической обработки, написания и редактирования статьи. Все авторы одобрили окончательный вариант статьи и готовы взять на себя ответственность за все аспекты представленной публикации.

REFERENCES / ЛИТЕРАТУРА

1. *Borodin E.A., Chupalov A.P., Timkin P.D., et al.* Selection of potential ligands for TRPM8 using deep neural networks and intermolecular docking. Bûlleten' fiziologii i patologii dyhaniâ = Bulletin Physiology and Pathology of Respiration. 2021; (80): 26–33 (In Russian). https://doi.org/10.36604/1998-5029-2021-80-26-33. EDN: MUFKIQ / *Бородин Е.А., Чупалов А.П., Тимкин П.Д. и др.* Подбор потенциальных лигандов к TRPM8 с помощью глубоких нейронных сетей и межмолекулярного докинга. Бюллетень физиологии и патологии дыхания. 2021; (80): 26–33. https://doi.org/10.36604/1998-5029-2021-80-26-33. EDN: MUFKIQ
2. *Jing F., Zhang S.W., Zhang S.* Prediction of enhancer-promoter interactions using the cross-cell type information and domain adversarial neural network. BMC Bioinformatics. 2020; 21(1): 507. https://doi.org/10.1186/s12859-020-03844-4. PMID: 33160328
3. *Catarino R.R., Stark A.* Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. 2018; 32(3–4): 202–223. https://doi.org/10.1101/gad.310367.117. PMID: 29491135
4. *Panigrahi A., O'Malley B.W.* Mechanisms of enhancer action: the known and the unknown. Genome Biol. 2021; 22(1): 108. https://doi.org/10.1186/s13059-021-02322-1. PMID: 33858480
5. *Broeders M., Herrero-Hernandez P., Ernst M.P.T., et al.* Sharpening the Molecular Scissors: advances in gene-editing technology. iScience. 2020; 23(1): 100789. https://doi.org/10.1016/j.isci.2019.100789. Epub 2019 Dec 19. PMID: 31901636
6. *Russell S.A., Bashaw G.J.* Axon guidance pathways and the control of gene expression. Dev Dyn. 2018; 247(4): 571–580. https://doi.org/10.1002/dvdy.24609. Epub 2018 Jan 5. PMID: 29226467
7. *Leonardo E.D., Hinck L., Masu M., et al.* Vertebrate homologues of C. elegans UNC-5 are candidate netrin receptors. Nature. 1997; 386(6627): 833–838. https://doi.org/10.1038/386833a0. PMID: 9126742
8. *Finci L., Zhang Y., Meijers R., Wang J.-H.* Signaling mechanism of the netrin-1 receptor DCC in axon guidance. Prog Biophys Mol Biol. 2015; 118(3): 153–160. https://doi.org/10.1016/j.pbiomolbio.2015.04.001. Epub 2015 Apr 14. PMID: 25881791
9. *Deiner M.S., Kennedy T.E., Fazeli A., et al.* Netrin-1 and DCC mediate axon guidance locally at the optic disc: loss of function leads to optic nerve hypoplasia. Neuron. 1997; 19(3): 575–589. https://doi.org/10.1016/s0896-6273(00)80373-6. PMID: 9331350
10. *Yam P.T., Langlois S.D., Morin S., Charron F.* Sonic hedgehog guides axons through a noncanonical, Src-family-kinase-dependent signaling pathway. Neuron. 2009 May 14; 62(3): 349–362. https://doi.org/10.1016/j.neuron.2009.03.022. PMID: 19447091
11. *Lepelletier L., Langlois S.D., Kent C.B., et al.* Sonic Hedgehog guides axons via Zipcode binding protein 1-mediated local translation. J Neurosci. 2017 Feb 15; 37(7): 1685–1695. https://doi.org/10.1523/JNEUROSCI.3016-16.2016. Epub 2017 Jan 10. PMID: 28073938
12. *Liu B., Fang L., Long R., et al.* iEnhancer-2L: a two-layer predictor for identifying enhancers and their strength by pseudo k-tuple nucleotide composition. Bioinformatics. 2016; 32(3): 362–369. https://doi.org/10.1093/bioinformatics/btv604. Epub 2015 Oct 17. PMID: 26476782
13. *Zhang T.H., Flores M., Huang Y.* ES-ARCNN: Predicting enhancer strength by using data augmentation and residual

convolutional neural network. Anal Biochem. 2021; 618: 114120. https://doi.org/10.1016/j.ab.2021.114120. Epub 2021 Jan 31. PMID: 33535061
14. *Salmaso V., Moro S.* Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. Front Pharmacol. 2018 Aug 22; 9: 923. https://doi.org/10.3389/fphar.2018.00923. PMID: 30186166
15. *Yan Y., Zhang D., Zhou P., et al.* HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res. 2017; 45(W1): W365–W373. https://doi.org/10.1093/nar/gkx407. PMID: 28521030
16. *Okonechnikov K., Golosova O., Fursov M., et al.* Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012; 28(8): 1166–1167. https://doi.org/10.1093/bioinformatics/bts091. Epub 2012 Feb 24. PMID: 22368248
17. *Kulakovskiy I.V., Vorontsov I.E., Yevshin I.S., et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. 2018; 46(D1): D252–D259. https://doi.org/10.1093/nar/gkx1106. PMID: 29140464
18. *Giri P., Kumar A., Taj G.* In silico-prediction of downstream MYB interacting partners of MAPK3 in Arabidopsis. Bioinformation. 2014; 10(12): 721–725. https://doi.org/10.6026/97320630010721. PMID: 25670873
19. *Andrianov A.M., Nikolaev G.I., Kornoushenko Y.V., et al. In silico* Identification of high-affinity ligands of the Hiv-1 Gp120 protein, potential peptidomimetics of neutralizing antibody N6. Math. Biol. Bioinf. 2019; 14(2): 430–449 (In Russian). https://doi.org/10.17537/2019.14.430. EDN: XCKXRV / *Андрианов А.М. , Николаев Г.И., Корноушенко Ю.В. и др. In silico* идентификация высокоаффинных лигандов белка gp120 ВИЧ-1 – потенциальных пептидомиметиков нейтрализующего антитела N6. Математическая биология и биоинформатика 2019; 14(2): 430–449. https://doi.org/10.17537/2019.14.430. EDN: XCKXRV
20. *De Vivo M., Masetti M., Bottegoni G., Cavalli A.* Role of molecular dynamics and related methods in drug discovery. Journal of Medicinal Chemistry. 2016; 59(9): 4035–4061. https://doi.org/10.1021/acs.jmedchem.5b01684. PMID: 26807648
21. *Yoo J.* On the stability of protein–DNA complexes in molecular dynamics simulations using the CUFIX corrections. J. Korean Phys. Soc. 2021; 78: 461–466. https://doi.org/10.1007/s40042-021-00063-9

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Danil D. Kotelnikov, Student, Amur State Medical Academy.
ORCID: https://orcid.org/0009-0003-5159-5796

Котельников Данил Дмитриевич, студент ФГБОУ ВО «Амурская государственная медицинская академия» Минздрава России.
ORCID: https://orcid.org/0009-0003-5159-5796

Ivan A. Sinyakin, Student, Amur State Medical Academy.
ORCID: https://orcid.org/0000-0001-8680-2298

Синякин Иван Алексеевич, студент ФГБОУ ВО «Амурская государственная медицинская академия» Минздрава России.
ORCID: https://orcid.org/0000-0001-8680-2298

Evgeny A. Borodin✉**,** Dr. of Sci. (Medicine), Professor, Head of the Department of Chemistry, Amur State Medical Academy.
ORCID: https://orcid.org/0000-0002-0983-4541

Бородин Евгений Александрович✉**,** д-р. мед. наук, профессор, заведующий кафедрой химии ФГБОУ ВО «Амурская государственная медицинская академия» Минздрава России.
ORCID: https://orcid.org/0000-0002-0983-4541

Tatyana A. Batalova, Dr. of Sci. (Biology), Associate Professor, Head of the Department of Physiology and Pathophysiology, Amur State Medical Academy.
ORCID: https://orcid.org/0000-0002-0983-4541

Баталова Татьяна Анатольевна, д-р биол. наук, доцент, заведующая кафедрой физиологии и патофизиологии ФГБОУ ВО «Амурская государственная медицинская академия» Минздрава России.
ORCID: https://orcid.org/0000-0002-0983-4541

✉ Corresponding author / Автор, ответственный за переписку