UDC 519.6

10.23947/2587-8999-2019-1-1-29-34

Computer modelling of primers search in the DNA chain***O. Y. Kiryanova**, L.U. Akhmetzianova, B.R. Kuluev, I. M. Gubaydullin, A.V. Chemeris**

Ufa State Petroleum Technological University, Ufa, Russia

Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences, Ufa, Russia

Polymerase chain reaction (PCR) is one of the most common experimental methods for solving DNA analysis problems. The possibility of PCR experiment conduction and its success are vastly depend on oligonucleotide structures. Oligonucleotide primers are important component of any PCR, and therefore, there are a number of requirements for their design. In this regard, it is essential to provide computer analysis for the primer selection. In current paper a new approach is proposed for a specific primer design which is based on Boyer-Moore search algorithm. Computer software is developed for computer-aided primer design, which noticeably simplifies the pre-experiment phase and improves PCR results.

Keywords: polymerase chain reaction, primer design, Boyer-Moore algorithm, computer analysis.

Introduction. Oligonucleotide primers determine specificity, efficiency and possibility of PCR reaction in the presence of all other components. The specificity of PCR is based on the formation of complementary complexes between a matrix and primers – short synthetic oligonucleotides with length from 10 to 30 bases. Each primer is complementary to one of the two chains of the double-stranded matrix and limits the beginning and the end of the amplified region. Reaction temperature depends on a particular composition of primers. Therefore, there are a number of requirements for the selection of nucleotide sequences and their lengths in primers in a view of problem being solved and object of experiment.

There are several well-known software solutions for the computer-aided primer design [1-5]. However, it is not always possible to implement a specific search, adjust parameters and initial conditions in case of using the aforementioned computer programs. Therefore, a new software is developed that allows primer design with the ability of introducing more stringent conditions on the desired primers and their location in a particular genome. It allows repeatedly perform computer analysis in several variations in order to determine the most favorable PCR conditions.

The object of research. Polymerase chain reaction (PCR) is one of the most common experimental method for solving DNA analysis problems. This method of gene diagnostics is widely used in various fields of biology and medicine. The essence of this analysis consists in a

* The reported study was funded by RFBR according to the research projects № 17-44-020120.

** E-mail: olga.kiryanova27@gmail.com.

multiple increase the amount of specific fragments by using special enzymes that repeatedly copy them for particular genomes.

PCR is performed in three stages. The first stage is denaturation: the divergence of two DNA chains (temperature 94-96°C). The second stage is annealing. Reaction temperature is reduced after the chains divergence so that the primers can bind to the one-chained matrix. Then the replication occurs (the synthesis of affiliated molecule DNA) where the primer is used as a priming [6].

Oligonucleotide primers (artificially created oligonucleotides that search for the desired DNA fragment) are important component due to the specificity of the PCR reaction, as well as its possibility occurs in the presence of all other components [7]. Thus, it is important to conduct a preliminary computer analysis for a further direct experimental research. There is a number of software solutions implementing design of oligonucleotide primers [8]. However, such software does not allow more detailed design of primers (for example, a specific composition of a primer, narrow PCR temperature range). Hence, there is no opportunity to solve small routine subtasks.

One of the computer-aided tasks is searching and determination of short sequences localization (primers) up to nucleotide. It is necessary in order to determine the possible places for primers annealing and their number. In fact, the task is similar to the searching for a word (10-20 characters length) in some text (up to 1 billion characters). In this case, the alphabet has the following letters determine four nucleotides: A (adenine), G (guanine), C (cytosine), and T (thymine). In its turn the molecule consists of nucleotides mentioned above. It is important not only to locate the «words» but also the number of its occurrences.

Searching short sequences has two destinations. The first one is for random PCR. It is important to know how many times the primer occurs in entire DNA chain and at what position it is located. In this case, the considered length of short sequences, as a rule, is from 8 to 25 nucleotides. The more often the desired site occurs, the higher the likelihood of a successful experiment. Figure 1 shows the search pattern.

Fig. 1. Searching for primers (random PCR).

The second application is a searching for the annealing sites of slightly longer primers which should occur on the average of every 16 million nucleotides. A subsequent sequence of nucleotides after

each site is of interest assuming that three rather than four nucleotides are taken in the amplification reaction. The termination of chain (completion of the synthesis) occurs on the missing one. In this case, we are interested not only in position of the primer, but also in the adjacent DNA segments. It is worth noting, the forward and reverse primers are considered as equivalent in this case. That is why both primers are used to search for the annealing site. The scheme of this search is presented on Figure 2. Therefore, the proposed approach is take into account only such places of annealing where specific nucleotide is not met on the sufficiently long distance (for example, guanine G as shown on Figure 2). Moreover, it is important to know the length of sections, the total molecular mass of their constituent nucleotides.

Fig. 2. Search for primers and annealing sites in the nucleotide sequence

Results of research. In order to solve the above problems, an algorithm was developed for searching short primers in the DNA chain. The proposed approach was based on the Boyer-Moore algorithm [9-10] with the additional conditions for the choice of primers. This algorithm allows to find the inclusion of specific fragments of sequences in the DNA chain up to nucleotide, as well as to determine the composition and size of the amplicons. During the search, a selection of primers was produced taking into account the stated requirements.

	A	B	C	D		A	B	C	D
1	#GGATCTTT	AAAGATCC	length of amplicon		1	#GGATCTTT	AAAGATCC	length of amplicon	
2	39883835	39884052	217		2	39883835	39884052	217	
3	55375264	55375548	284		3	55375264	55375548	284	
4	29569657	29569969	312		4	29569657	29569969	312	
5	38393029	38393375	346		5	38393029	38393375	346	
6	49519668	49520023	355		6	49519668	49520023	355	
7	41540764	41541163	399		7	41540764	41541163	399	
8	8231987	8232448	461		8	8231987	8232448	461	
9	37414390	37414862	472		9	37414390	37414862	472	
10	56234084	56234565	481		10	56234084	56234565	481	
11					11				

(a)

(b)

Fig. 3. The output of the program

- (a) search sequence GGATCTTT (reverse primer AAAGATCC) for the analysis of random PCR
 (b) search for GGATCTTTAC sequence (reverse primer GTAAAGATCC) to detect annealing sites

The software was developed using Python 3.5 language and BioPython library [11]. This library contains tools for calculations in the field of computational biology and bioinformatics. In addition, library tools allow to work with files in fasta-format (text format for nucleotide or polypeptide sequences, in which

nucleotides or amino acids are indicated using single-letter codes) [12]. All calculations were done for the model objects (chromosomes of Arabidopsis). The program output is presented on Figure 3.

Conclusion. The proposed software allows varying the size of a primer, the length of an amplicon. Moreover the developed computer-aided system could change the conditions for the annealing site (such as size and nucleotide composition).

The results obtained allow to predict the conditions for experimental PCR. On the base of the algorithm, a computer program was developed that allows computer-aided primer design. On the base of performed computer analysis, it was revealed that it is inexpedient to carry out experimental studies of PCR diagnostics since there are a small number of sites containing the required primers or they do not exist at all. The received results simplifies and optimizes the work of geneticists and experimenters are providing PCR experiments. If the size of the amplicons expected from computer analysis and their number are known, we can them on the gel electrophoresis in the form of bands during the PCR. In case of the absence of the desired sites in the genome under study the successful conduct of a full-scale experiment is unlikely.

References

1. Cheng-Hong Yang, Yu-Huei Cheng, Li-Yeh Chuang, Hsueh-Wei Chang, Specific PCR product primer design using memetic algorithm: *Biotechnology Progress* 25(3) – 2009. – P 745-753.
2. Konwar K., Mandoiu I., Russell A., Shvartsman A., Approximation Algorithms for Minimum PCR Primer Set Selection with Amplification Length and Uniqueness Constraints, *Proceedings of the 3th Asia-Pacific Bioinformatic conference (APBC)*, Imperial College Press – 2005. – P 41-45.
3. Yu-Huei Cheng, Estimation of Teaching-Learning-Based Optimization Primer Design Using Regression Analysis for Different Melting Temperature Calculations: *IEEE Transactions on NanoBioscience* 14(1) – 2015. – P. 3-12.
4. Yung-Fu Chen, Rung-Ching Chen, Yung-Kuan Chan, Reo-Hao Pan, You-Cheng Hseu, Elong Lin, Design of multiplex PCR primers using heuristic algorithm for sequential deletion applications: *Computational biology and chemistry*, 33 – 2009. – P 181-188.
5. Li-Yeh Chuang, Yu-Huei Cheng, Chang-Hsuan Ho, Specific primer design for the polymerase chain reaction: *Biotechnology Letters*, 35(10) – 2013. – P. 1541-1549.
6. Kleppe K., Ohtsuka E., Kleppe R., Molineux I., Khorana H.G. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. – *Mol. Biol. Bd.* – 2002. – Vol. 56 – P. 341 – 364.
7. Glik B., Pasternak Dzh. *Molekulyarnaya biotekhnologiya. Principy i primeneniye.* — M.: Mir, 2002. — 589 p.
8. Chemeris D.A., Kiryanova O.Y., Gubaidullin I.M., Chemeris A.V. Dizajn prajmerov dlya polimeraznoj cepnoj reakcii (kratkij obzor komp'yuternyh programm i baz dannyh) – *Biomika.* – 2016. – T. 8. – №3. – P. 215-238.
9. Knuth D.E., Morris (Jr) J.H., Pratt V.R. Fast pattern matching in strings — *SIAM Journal on Computing.* – 1977. – Vol. 6(1) – P. 323-350.
10. Boyer R. S., Moore J. S. A fast string searching algorithm, *Carom. ACM* 20, (10) – 1977. – P. 262–272.
11. Biopython [Electronic resource]. – Access mode: <https://biopython.org/>
12. Pearson W.R., Lipman D.J. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* – 1988. – Vol. 85(8). – 2444 p.

Authors:

Kiryanova Olga Yurevna, Ufa State Petroleum Technological University (1 Kosmonavtov St., Ufa, Russian Federation).

Akhmetzianova Liana Ulfatovna, Ufa State Petroleum Technological University (1 Kosmonavtov St., Ufa, Russian Federation).

Kuluev Bulat Razyapovich, Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences, (71 Prospekt Oktyabrya, 450054, Ufa, Russian Federation), Doctor of Science in Biology.

Gubaydullin Irek Marsovich, Institute of Petrochemistry and Catalysis of the Russian Academy of Sciences (141 Oktyabrya avenue, Ufa, Russian Federation), Ufa State Petroleum Technological University (1 Kosmonavtov St., Ufa, Russian Federation), Doctor of Science in Physics and Maths, Associate professor.

Chemeris Aleksey Viktorovich, Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences (71 Prospekt Oktyabrya, 450054, Ufa, Russian Federation), Doctor of Science in Biology, Associate professor

УДК 519.6

10.23947/2587-8999-2019-1-1-29-34

Компьютерное моделирование поиска праймеров в цепи ДНК***О.Ю. Кирьянова**, Л.У. Ахметзянова, Б.Р. Кулуев, И.М. Губайдуллин, А.В. Чемерис**

Уфимский государственный нефтяной технический университет, Уфа, Россия
Институт биохимии и генетики Уфимского федерального исследовательского центра Российской академии наук

Полимеразная цепная реакция (ПЦР) является одним из самых распространенных экспериментальных методов при решении задач анализа ДНК. Олигонуклеотидные праймеры являются важной составляющей любой ПЦР, и поэтому существует ряд требований к их дизайну. От данных структур зависит успешность и возможность проведения эксперимента в целом. В связи с этим появилась необходимость проведения компьютерного анализа подбора праймеров. Разработан алгоритм на основе алгоритма поиска Бойера-Мура для специфичного дизайна праймеров. В настоящей работе представлены две вариации поиска праймеров. На основе алгоритма разработана программа, позволяющая проводить компьютерный дизайн праймеров перед непосредственным экспериментальным проведением ПЦР. Что значительно упрощает проведение натурного эксперимента. На данный момент расчеты проведены для модельных объектов (хромосом арабидопсиса).

Ключевые слова: полимеразная цепная реакция, дизайн праймеров, алгоритм Бойера-Мура, компьютерный анализ.

Авторы:

Кирьянова Ольга Юрьевна, Уфимский государственный нефтяной технический университет (450062, Уфа, ул. Космонавтов, д. 1).

Ахметзянова Лиана Ульфатовна, Уфимский государственный нефтяной технический университет (450062, Уфа, ул. Космонавтов, д. 1).

Кулуев Булат Разяпович, Институт биохимии и генетики Уфимского федерального исследовательского центра Российской академии наук (450054, Уфа, проспект Октября, д. 71), доктор биологических наук.

Губайдуллин Ирек Марсович, Институт нефтехимии и катализа РАН (450075 Уфа, Проспект Октября, д. 141), Уфимский государственный нефтяной технический университет (450062, Уфа, ул. Космонавтов, д. 1), доктор физико-математических наук, профессор.

Чемерис Алексей Викторович, Институт биохимии и генетики Уфимского федерального исследовательского центра Российской академии наук (450054, Уфа, проспект Октября, д. 71), доктор биологических наук, профессор.

* Работа выполнена при частичной поддержке гранта РФФИ № 17-44-020120.

** E-mail: olga.kiryanova27@gmail.com.