

PERSPECTIVES ON ADDING NEW LETTERS TO GENETIC ALPHABET

LAMIA ALGHANNAM¹

¹The Catholic University of America

Email: La.alghannam@gmail.com

Received: 20 Jan 2016, Revised and Accepted: 20 Mar 2016

ABSTRACT

Objective: Today, DNA-based technologies are limited in the range, efficiency and accuracy of their application by the four bases in the structure of DNA. Development of DNA molecules with a higher number of bases could contribute to the resolution of the problem. Moreover, the addition of new letters to the genetic alphabet could be used for the treatment of disorders and development of DNA molecules with extended functionality. The aim of the research was to analyse the current abilities and future perspectives of expanded alphabet with respect to the mentioned problems.

Methods: The research questions were addressed by the research of the articles present in NCBI and Google Scholar databases. With the application of developed key terms, such as "expanded gene code", "expanded genetic alphabet" and "genetic alphabet and medicine", and certain inclusion and exclusion criteria, six articles published between 2006 and 2016 were selected for analysis.

Results: It was identified that most of the applications of DNA with six letters possible for today use are associated with the expansion in the functional abilities of the modern DNA-based methods. Some of the researchers show the higher binding ability and affinity of the artificial DNA aptamers and suggest their application for treatment and increased functions of DNA. However, still a great number of other applications are suggested for the future. They include the production of proteins and enzymes with new qualities, making DNA molecule a molecular probe for tumour detection and a number of other applications.

Conclusion: Addition of new letters to the genetic alphabet can be a powerful tool for improvement of diagnostic technologies used today. However, more research in this field is still needed for wider application and development of new treatment approaches.

Keywords: Expanded alphabet, Personalized medicine, Pharmaceuticals, cancer, Gene technology, Molecular beacon

© 2016 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

Since its discovery and analysis, the DNA molecule has been considered as an ideal one because it was a first chemical compound, capable of replication and self-assembly in the limited extend. It also could store the information about the organization of other molecules inside the cell and pass this information to executors in the role of ribosomes via the production of complementary RNA molecules. However, further analysis of DNA and attempts of its application in medical diagnosis showed that the molecule was not as ideal as suggested. In addition, DNA sequencing technology showed significant problems when short fragments of DNA sequences had to be assembled into the entire molecule. The non-ideal nature of DNA was explained by the low density of information encoded by the four bases [1]. It is considered that the number of nitrogen bases, present in the natural DNA molecules, is limited to four because of the DNA origin from the RNA molecules, developed in the RNA world. RNA molecules were less stable than DNA and had to perform replication on their own. The presence of a high number of different units, which in turn had to have some conservative organization, within the first RNA could result in the increased risks for improper pairing. Thus, the number of bases was probably limited to four, and this feature was passed to the DNA molecules, which evolved later [2].

Nevertheless, the high stability of the modern DNA molecules and protein assistance during the replication process did not result in the evolution of DNA towards the formation of a higher number of nitrogen bases. The presence of only four bases does not create significant problems for the modern life, however significantly limits the application of DNA-based technology. False-positive as well as false-negative results can be identified with the application of four-base DNA. In order to resolve this problem, scientists incorporated the new synthetic bases into the structure of DNA. Several molecules, which can par with each other based on the complementarity principle, were found to fit the B-structure of DNA. They were referred to as new letters of the DNA alphabet. These nucleotides have been successfully introduced into synthetic DNA molecules, which were then efficiently replicated in the living cells [3].

It means that the artificially designed nucleotide pairs can be used for the development of DNA molecules having the higher density of information and recognized by the natural biological molecules. It leads to a suggestion about the possible wider range of application of the DNA molecules, having six or more bases instead of four.

The perspectives of extended DNA alphabet are associated with the possibility of improvement of DNA-based technologies, treatment of incurable diseases and assignment of new functions for the DNA molecules. These questions were addressed in the current research of the potential application of synthetic nucleotides in the structure of natural DNA.

MATERIALS AND METHODS

NCBI and Google Scholar data bases were searched for the key terms, which included "expanded gene code", "expanded genetic alphabet" and "genetic alphabet and medicine". In the case of both databases, the primary research articles published within the period of 2006-2016 were selected. Many of the articles found by the system in response to the listed key terms were excluded from the research. Some of the excluded ones addressed the expansion of the repeats presents on the human genome, in the case of Huntington's disease for example. Others studied the effects of gene code expansion not via the addition of new nucleotides, but via the suppression of the stop codons, making them encode for the non-standard amino acid. Finally, the research articles, where the search for the new nucleotides was performed only, or possibility of unusual nucleotides recognition by cellular enzymes was studied, were excluded as well. Only the articles, where the practical application of the additional nucleotides or DNA molecules with expanded alphabet, were incorporated in the research. Six articles, which met the inclusion criteria, were selected for analysis in the current study.

RESULTS

Among the six research articles most addressed the problems of DNA application for the improvement of DNA-based technologies. They included improvement of the DNA sequencing technology with respect to single-stranded DNA assembly [1], of DNA detection by

means of PCR with fluorescent probes [4], including virus detection and differentiation [5], and development of molecular beacons [6]. Prevalence of articles concerned with such problems can be associated with the fact that current understanding of the DNA with additional letters of the alphabet can allow only this approach. One of the articles was concerned with the development of DNA with an increased range of functions. Only the higher binding ability of new DNA aptamers and thus, possibility of regulation of functions of some proteins, is considered as a possible extended function of DNA [7]. Finally, the possibility of earlier cancer diagnosis was suggested in one of the articles [8] based on the higher affinity of binding and wider range of targets for DNA molecules, which have six bases instead of four. At the same time, the broader application of DNA molecules with an enhanced number of letters in the alphabet is suggested to be possible in future.

DISCUSSION

The currently possible applications of DNA aptamers with six nucleotides are mainly concerned with the improvement of the currently existing DNA-based methods and can potentially be used for the improvement of gene technology in future. Today, assembly of single-stranded nucleotide sequences is possible for the limited number of fragments because of the low information density reported for the natural DNA molecules. Application of DNA molecules with six nucleotides was shown to result in efficient assembly of single stranded fragments into the entire molecule. Then, the additional nucleotides were removed during PCR reaction with the application of only natural nucleotides and polymerase with reduced proofreading ability. Based on tautomer formation process, the enzyme substituted the pairs of synthetic nucleotides with the AT pairs. The incorporation of the obtained sequence into the plasmid and finally, its delivery into bacterial cell resulted in the normal production of protein, providing resistance to kanamycin. The control treatment, in which the DNA fragments with four bases only were allowed to assemble into the single strand, showed much lower efficiency. Hairpin formation and non-canonised folding were the main obstacles for DNAs with low information density [1].

The described outcome shows that six letters in DNA alphabet can contribute to the improvement of DNA technology when vector should be formed with the incorporation of the target sequence. It is also suggested that in future there will be no requirement for the removal of synthetic bases because the new codons will be developed to encode for the new amino acids. As a result, the wider range of some useful proteins will be obtained [9]. Some of the proteins will have the unique properties allowing their interaction

with the particular targets, probably responsible for disorder development, or participation in the processes, which are important in the progression of the particular disorders. Today, a number of disorders are untreatable because of the inability of the pharmaceutical industry to produce the required drug or inability to design the protein, which would participate in the treatment because of the reduced stability of the protein molecule containing the required amino acids in the certain position. Expanded gene code and DNA alphabet would allow producing the proteins or enzymes with the desired properties due to the presence of non-specific amino acids. Such perspective is considered an important tool for the treatment of cancer, which is poorly managed today [8].

Finally, the self-assembling DNA fragments can be important for genome sequencing because they would make it faster and less expensive. The concept of personalized medicine is rather popular today because of the ability of each patient to get the medical care needed by him or her with respect to genome specificities. In addition, the medical forecast can be made for each patient and thus, development of disorders, to which the patient is susceptible, can be prevented. The problem with personalized medicine consists of the high price and relatively low speed of DNA sequencing affordable by the patients. With the application of additional nucleotides, the problem can be resolved because assembly of the sequenced fragments will take less time [9].

It would be reasonable to mention that development of the high-affinity DNA aptamers can also be a highly efficient tool in personalized medicine. The presence of the additional letters in DNA sequence was shown to be associated with the increased strength of DNA binding to certain proteins [7]. Such binding can be associated with the inability of the protein to perform its functions and in the case of infectious nature of the protein, in the case of prion for example, the progression of the disorder can be reduced. Thus, the addition of the new letters can allow the development of new functions by DNA, application of this molecule as a pharmaceutical, for example. Such possibility becomes even more challenging if the presence of hydrophobic synthetic nucleotides is taken into consideration. All the natural nitrogen bases are polar, and the presence of non-polar ones allows binding of a wider range of molecules [7]. It is supposed that DNA molecules will be used for the detection of cancer cells at the earliest stages of their development in the nearest future [8]. There is a working scheme for the selection of DNA molecules, capable of binding to different markers on the tumour cells (fig. 1.). The DNA sequences capable of binding the cancer cells to undergo positive selection while those binding normal cells undergo negative selection.

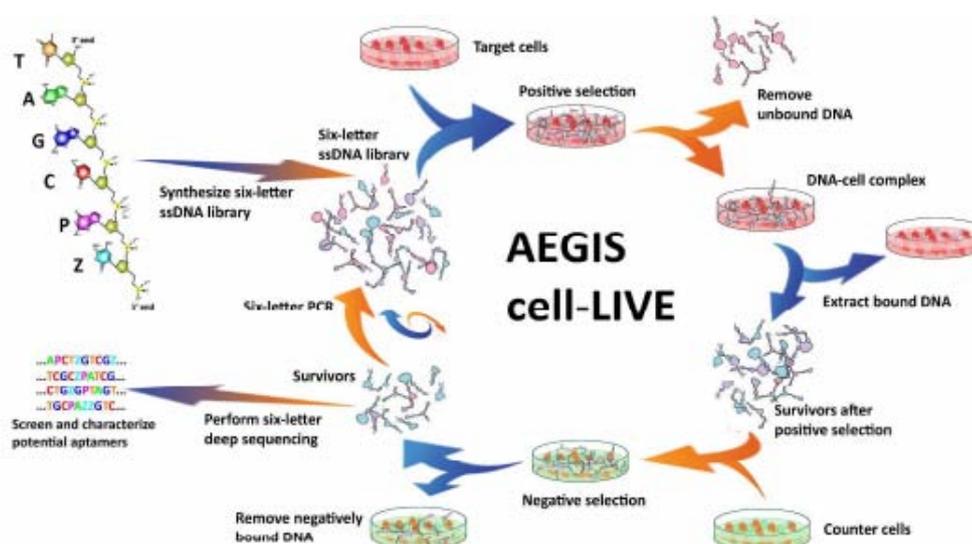


Fig. 1: Selection of the six-base DNA molecules, capable of binding different markers on the surface of cancer cells [8]

The other potential application of six-letter DNA molecules today is the development of molecular beacons [6] and identification of the specific DNA or RNA fragments with the help of fluorescent probes and/or during the PCR reactions [4]. Today, the problems with these methods consist of the possibility of false positive and false-negative results, as well as noise. Noise may come from the nonspecific pairing and failure of primers to interact with the DNA. In the case of four-base system the number of possible combinations is much lower than in the case of the six-base system. It means that the possibility of non-specific pairing between the fluorescent probe and studied DNA is much lower when six letters are used. Thus, false-positive results can be excluded. The reason of false-negative results is often due to requirement in special conditions for hybridization, which can easily be interrupted, as well as possibility of low purity of the sample. The presence of nucleases or protein can prevent the binding of target DNA and probe. However, when six bases are used, the common nucleases may fail to digest DNA with the additional letters and the affinity between DNA probe and studied DNA will be higher [6]. The application of the six-base DNA in nucleic acid assays will contribute to higher efficiency and accuracy of the methods, which in turn will lead to the possibility of better understanding of some cellular processes and finally, treatment of the related disease.

The similar problem appears when related viruses should be identified with the help of PCR method. Identification of the genetic material of virus is one of the most efficient and fast methods of diagnosis of viral infection. Sometimes, it can be extremely important to determine, which pathogen is affecting the patient. If the virus has several related germs, it can be difficult to distinguish between them. In addition, studying several samples, the researcher may allow cross-contamination. The multicode-plx system was developed to resolve this issue. It is a complex technology, which includes the multiplexed PCR, expanded DNA alphabet, and microsphere flow cytometry. With the use of six nucleotides, the ability to differentiate between viruses increased, allowing more accurate medication prescription and development of treatment strategy.

Therefore, the addition of new letters to the genetic alphabet can be an extremely powerful tool for the management of the incurable

diseases and application of DNA for different diagnostic and treatment purposes. Some of the benefits of this technology can be used today while more research is required for the other aspects of six-base DNA applications.

CONFLICT OF INTERESTS

Declare none

REFERENCES

1. Merritt K, Bradley K, Hutter D, Matsuura M, Rowold D, Benner S. Autonomous assembly of synthetic oligonucleotides built from an expanded DNA alphabet. Total synthesis of a gene encoding kanamycin resistance. *Beilstein J Org Chem* 2014;10:2348-60.
2. Georgiadis M, Singh I, Kellett W, Hoshika S, Benner S, Richards N. Structural basis for a six nucleotide genetic alphabet. *J Am Chem Soc* 2015;137:6947-55.
3. Chaput J. Replicating an expanded genetic alphabet in cells. *Chem Biochem* 2014;15:1869-71.
4. Yang Z, Durante M, Glushakova L, Sharma N, Leal N, Bradley K, *et al.* Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *Anal Chem* 2013;85:4705-12.
5. Nolte F, Marshall D, Rasberry C, Schivelbein S, Banks G, Storch G, *et al.* Multicode-PLx system for multiplexed detection of seventeen respiratory viruses. *J Clin Microbiol* 2007;45:2779-86.
6. Sheng P, Yang Z, Kim Y, Wu Y, Tan W, Benner S. Design of a novel molecular beacon: modification of the stem with artificially genetic alphabet. *Chem Commun* 2008;7:5128-30.
7. Kimoto M, Yamashige R, Matsunaga K, Yokoyama S, Hirao I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat Biotechnol* 2013;31:453-7.
8. Zhang L, Yang Z, Sefah K, Bradley K, Hoshika S, Kim M, *et al.* Evolution of functional six-nucleotide DNA. *J Am Chem Soc* 2015;137:6734-7.
9. Hirao I, Kimoto M. Unnatural base pair systems toward the expansion of the genetic alphabet in the central dogma. *Proc Jpn Acad Ser B* 2012;88:345-67.