Proceeding of 2nd International Science Postgraduate Conference 2014 (ISPC2014)

SECOND ORDER LIMIT LANGUAGE ASSOCIATED WITH TWO RULES

[1*]MUHAMMAD AZRIN AHMAD, [2]NOR HANIZA SARMIN,[3]FONG WAN HENG AND [4]YUHANI YUSOF

[1,2]Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia,81310 UTM Johor Bahru, Johor, Malaysia
[3]Ibnu Sina Institute for Fundamental Science Studies
Universiti Teknologi Malaysia,81310 UTM Johor Bahru, Johor, Malaysia
[4]Faculty of Industrial Science and Technology,
Universiti Malaysia Pahang, 26300 UMP Gambang, Pahang, Malaysia

[1*]azrinahmad1@gmail.com[2]nhs@utm.my, [3]fwh@ibnusina.utm.my, [4]yuhani@ump.edu.my

*Corresponding author

Abstract. The concepts of splicing system involve the study of cut and paste phenomenon of deoxyribonucleic acid (DNA). The splicing language, which is resulted from a splicing system, can be classified as inert persistent language, active persistent language and limit language. As one of the types of splicing language, limit language can biologically be referred as the remaining molecule after the process has attained its equilibrium state. In this paper, the existence of second order limit language is explored strictly using at most two initial strings and two rules. In addition, the actual biological examples are presented. By using the Y-G splicing system, the results obtained are then used to prove the existence of second order limit language in few classes of Y-G splicing system.

Keywords Splicing system; splicing language; second order limit language

1.0 INTRODUCTION

Deoxyribonucleic Acid or known as DNA plays an important role as the hereditary factor in a living cell. It consists of three crucial parts that are sugar, phosphate group and nitrogenous base as shown in Figure 1 [1]. Four types of bases are adenine (A), guanine (G), cytosine (C) and thymine (T), which can later be grouped into purines (A and G), and pyrimidines (C and T). By Watson and Crick complementarity, the possible pairs to exist among those bases are A with T and C with G and vice versa. In reality, DNA is in the form of double helix where two strands are wounded around each other. In this study, DNA is represented as a double stranded DNA (dsDNA) [2]. The dsDNA is formed by single stranded DNA, which is connected by phosphodiester bonds. The hydrogen bond then connects the bases to their respective complement bases [2].

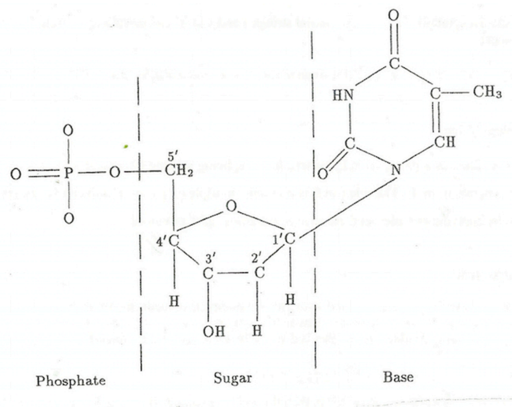

Figure 1 DNA Structure

Head has initiated the study of recombinant behavior on DNA with the existence of appropriate enzymes [3]. By the fact that bases in DNA can be written as alphabet in Formal Language Theory, a mathematical model was developed. Formal Language Theory is known as a branch of theoretical computer science, which studies the set of finite strings of symbols chosen from a prescribed finite set [4]. In his pioneered work, Head has focused on persistent and permanent splicing systems. Since then, more researchers have come out with other splicing models to attain the needs of their research. It is believed that the direction of researchers in developing new models of splicing system has been separated into two ways namely a model based on generation of language and a model to preserve the biological characteristics of splicing system. In 2011, Yusof [5] developed an improved splicing model, which is aligned with Head and Goode-Pixton splicing system named as Yusof-Goode (Y-G) splicing system. It is claimed to work better with the actual DNA biological process.

The splicing language that is the end product of a splicing system can produce several types of languages, including inert or adult language, transient and limit language. Yusof [6] re-defined the term inert language to inert persistent language and also found a new type of splicing language namely the active persistent language. In this study, the work is devoted to the extension of limit language [7], which is called the second order limit language. The case where at least two initial strings and two rules that lead to the formation of second order limit language is investigated. The results obtained are then used to prove the existence of second order limit language in the few classes of Y-G splicing system.

2.0 PRELIMINARIES

In this section, the definition of Y-G splicing system, transient, limit language and second order limit language are given. Besides that, the definition of strictly locally testable is also presented.

Definition 1 [5]: Y-G Splicing System. If $r \in R$, where $r = (u, x, v\text{: } y, x, z)$ and $s_1 = \alpha uxv\beta$ and $s_2 = \gamma yxz\delta$ are elements of I, then splicing s_1 and s_2 using r produces the initial string I together with $\alpha uxv\beta$ and $\gamma yxz\delta$ presented in either order where $\alpha, \beta, \gamma, \delta, u, x, v, y$ and $z \in A^*$ are the free monoid generated by A with the concatenation operation and 1 as the identity element.

Definition 2 [7]: Transient Language. A splicing language is called transient if a set of strings is eventually used up and disappears in a given system.

Definition 3 [7]: Limit Language. A limit language is the set of words that are predicted to appear if some amount of each initial molecule is present, and sufficient time has passed for the reaction to reach equilibrium state, regardless of the balance of the reactants in a particular experimental run of the reaction.

Definition 4: Second Order Limit Language. Let L_1be the set of second order limit words of L, the set L_2 of second order limit words of L to be the set of first order limits of L_1. We obtain L_2 from L_1 by deleting words that are transient in L_1.

Definition 5 [8]: Simple Splicing System. Let $S = (A, I, R)$ be a splicing system in which all rules in R have the form $(a, 1, 1: a, 1, 1)$ or $(1, a, 1: 1, a, 1)$ or $(1, 1, a: 1, 1, a)$ where $a \in A$. Then S is called a simple splicing system.

Definition 6 [9]: Strictly Locally Testable (SLT). A language is strictly locally testable if there is a positive integer k for which every factor of L of length k is a constant.

In the next section, an actual biological example of second order limit language in the case of an initial string with two rules is provided. After that, a case of second order limit language in the case of two initial string with two rules is discussed.

3.0 INTERPRETATION OF SECOND ORDER LIMIT LANGUAGE IN THE ACTUAL CASE

Firstly, an example of second order limit language with an initial string with two rules in the actual biological process is given. Two restriction enzymes are chosen from New England Biolabs Catalogue [10]. The existence of second order limit language in the case of two initial strings with two rules is presented in the following.

Example 1. Let $S=(A,I,R)$ be a Y-G splicing system consisting of two restriction enzymes, namely *Acc*65I and BsrGI, where $A=\{a,c,g,t\}$, $I=\{ggtacctctagctgtaca\}$ such that $R=\{(r_1 : r_2)\}$ where $r_1=(g,gtac,c)$ and $r_2=(t,gtac,a)$. When splicing occurs, the following splicing languages are generated:

$$ggtacctctagctgtaca \xrightarrow{r_1,r_2} I \cup \begin{Bmatrix} ggtacc, tgtacagctagaggtacctctagctgtaca, \\ ggtacctctagctgtacagctagaggtacc, tgtaca, \\ ggtaca, ggtacctctagctgtacctctagctgtaca, \\ ggtacagctagaggtacc, tgtacctctagctgtaca \end{Bmatrix}.$$

Based on the rules used above, when the resulted splicing languages are being spliced again, one new splicing languages is obtained which is listed below.

$$\{ggtacctctagctgtacctctagctgtacctctagctgtaca\}.$$

In the next example, a case where second order limit language with two initial string and two rules is illustrated.

Example 2. Let $S=(A,I,R)$ be a Y-G splicing system. Assume $u\alpha\mu\beta v, p\alpha\mu\beta q$ are two strings in $I \in A^*$ with $R=\{(\alpha,\mu,\beta:\gamma,\mu,\delta)\}$. Only the first rule is used in the first stage of splicing. Thus the obtained strings in $L(S)$ are

$$I \cup \{u\alpha\mu\beta q, p\alpha\mu\beta v\}.$$

As we have to consider the rotation of 180^0, assume that the following is one of the splicing languages formed in the first splicing:

$$u\gamma\mu\delta v.$$

Now, when the second splicing occurs among the first splicing language, the second order limit language is formed through this splicing:

$$\{u\alpha\mu\beta q, u\gamma\mu\delta v\} \xrightarrow{r_1, r_2} u\alpha\mu\delta v.$$

Therefore, $u\alpha\mu\beta q, u\gamma\mu\delta v$ are being used up to form $u\alpha\mu\delta v$. Hence, there exists a second order limit language.

In the next section, three conjectures are presented to show the relation of non strictly locally testable language and simple splicing language with second order limit language.

4.0 RESULTS & DISCUSSION

In the next section, three conjectures are presented to show the relation of non strictly locally testable language and simple splicing language with second order limit language and also a case where the possibility of second order limit language exist in a splicing system. The proof is still ongoing.

Conjecture 1. If a splicing system is not strictly locally testable, then there exists a second order limit language.

Conjecture 2. If a splicing system is simple, then there exists a second order limit language.

Conjecture 3. If a splicing system contains an initial string which consists of two recognition sites of two rules, then there exists a second order limit language.

ACKNOWLEDGMENTS

The first author would like to thank the Ministry of Education (MOE) Malaysia for his financial funding through MyBrain15, MyPhD scholarship. The second and third authors would like to thank the Ministry of Education (MOE) Malaysia and Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the financial funding through UTM Research Universiti Fund Vote No. 07J41.

REFERENCES

[1] R. H. Tamarin, *Principles of Genetics*, 7th. ed. USA: The Mac-Graw Hill Companies, 2001.

[2] G. Paun, G. Rozenberg and A. Salomaa, *DNA Computing New Computing Paradigm*, Germany: Springer-Verlag Berlin Heidelberg New York, 1998.

[3] T. Head, Formal Language Theory and DNA: An Analysis of Generative Capacity of Specific Recombinant Behaviors.*Bulletin of Mathematical Biology* 49:737-759, 1987.

[4] P. Linz, *An Introduction to Formal Languages and Automata*, USA: Jones and Barlett Publisher, 2006.

[5] Y. Yusof, N. H. Sarmin, M. Mahmud, T. E. Goode and W. H. Fong, “An Extension of DNA Splicing System” in 6th International Conference on Bio-Inspired Computing: Theories and Applications, Pulau Pinang, Malaysia, 2011, pp. 246-248.

[6] Y. Yusof, "DNA Splicing System Inspired by Bio Molecular Operations", Ph.D. Thesis, Universiti Teknologi Malaysia, 2011.

[7] E. Goode and D. Pixton, “Splicing to the Limit,” in *Aspects of Molecular Computing,* edited by Janoska, N., Paun, Gh. And Rozenberg, G. (eds), Germany: Springer-Verlag Berlin Heidelberg New York, 2004, pp. 189-201.

[8] Y. Yusof, N. H. Sarmin, T. E. Goode, W. H. Fong, Hierarchy of Certain Types of DNA Splicing Systems.*International Journal of Modern Physics: Conference Series,* 49: 271-277 (2012).

[9] W. H. Fong, "Modelling of Splicing System using Formal Language Theory", Ph.D. Thesis, Universiti Teknologi Malaysia, 2008.

[10] Research Biolabs Sdn. Bhd. *New England Biolabs 2011-12 Catalogue & Technical Reference*. USA: Catalogue. 2011.