Some restrictions on the existence of second order limit language

Muhammad Azrin Ahmad, Nor Haniza Sarmin, Yuhani Yusof, and Wan Heng Fong

Citation: AIP Conference Proceedings **1682**, 020048 (2015); doi: 10.1063/1.4932457
View online: http://dx.doi.org/10.1063/1.4932457
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1682?ver=pdfcov
Published by the AIP Publishing

Some Restrictions on the Existence of Second Order Limit Language

Muhammad Azrin Ahmad[a], Nor Haniza Sarmin[b], Yuhani Yusof[c] and Wan Heng Fong[d]

[a,b]*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor*
[c]*Faculty of Industrial Sciences and Technology, Universiti Malaysia Pahang, 26300 UMP Gambang, Pahang*
[d]*Ibnu Sina Institute for Fundamental Science Studies, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor*

Abstract. The cut and paste phenomenon on DNA molecules with the presence of restriction enzyme and appropriate ligase has led to the formalism of mathematical modelling of splicing system. A type of splicing system named Yusof-Goode splicing system is used to present the transparent behaviour of the DNA splicing process. The limit language that is defined as the leftover molecules after the system reaches its equilibrium point has been extended to a second order limit language. The non-existence of the second order limit language biologically has lead to this study by using mathematical approach. In this paper, the factors that restrict the formation of the second order limit language are discussed and are presented as lemmas and theorem using Y-G approach. In addition, the discussion focuses on Yusof-Goode splicing system with at most two initial strings and two rules with one cutting site and palindromic crossing site and recognition sites.

Keywords: Y-G splicing system, splicing language, second order limit language
PACS: 87.14.gk, 87.14.ej

INTRODUCTION

Deoxyribonucleic acid (DNA) which contains biological instructions that make organisms unique from one another, plays important roles in living organism such as coding for protein synthesis and also as hereditary agent that passes information from cell to cell [1]. The DNA is made up of three important molecules which are sugar, phosphate and a nitrogenous base. The chemical bases namely adenine (*A*), guanine (*G*), cytosine (*C*) and thymine (*T*) are grouped as purines (*A* with *G*) and also pyrimidines (*C* with *T*). Watson-Crick complementarity concludes that the only possible pairings are *A* with *T*, likewise *C* with *G*. The structure of DNA molecule is given in Figure 1 [2].

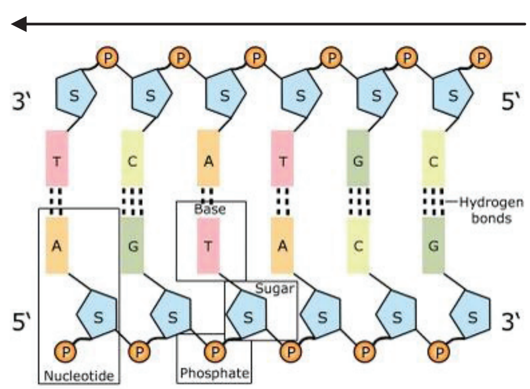

FIGURE 1. The structure of a double-stranded DNA (dsDNA) molecule

A restriction enzyme which has been isolated from bacteria can cut the DNA molecules given appropriate conditions. Interestingly, the DNA molecules which have been separated into two single strands can be 'pasted' (ligated) under the presence of a proper ligase [1]. Head [3] formulated this cut and paste phenomenon as

The 22nd National Symposium on Mathematical Sciences (SKSM22)
AIP Conf. Proc. 1682, 020048-1–020048-7; doi: 10.1063/1.4932457

mathematical modelling of splicing system under formal language theory and the study of informational macromolecules. Basically, the splicing system consists of four parameters which are a set of alphabets A that contains a, c, g and t, a set of initial strings, I that are strings made up from alphabets and finite sets B and C of triples (c, x, d) called patterns. The languages produced from the splicing system are called the splicing languages.

In splicing system, a splicing process is modeled either based on the generation of languages or a splicing model that preserves the biological traits in a splicing system [4]. Hence, Yusof-Goode (Y-G) splicing system which presents the translucent behavior of the DNA biological process is used as a model to formulate our problems throughout this paper. This model also has been used in various study related to DNA splicing languages which can be seen in [5,6].

The splicing languages can be categorized as inert/ active, transient and limit language. Recent study in [7] has renamed previous inert language to inert persistent language and also defines a new language namely active persistent language. These terms are obtained based on the wet-lab works which validate the theorems biologically [4,7,8,9]. Previous study only concern on these types of splicing languages. Therefore, the extension of limit language namely the second order limit language has been introduced and studied in [10,11].

In [7], the non-existence of second order limit language is shown when the splicing process involves two initial strings and two rules for one cutting site. Thus, the characteristics of non-second order limit language are important in this study. In this paper, the factors that restricted the formation of the second order limit language in Y-G splicing system are investigated according to the crossing site and recognition site of a set of rules. This study focuses a few cases on Y-G splicing system up to two initial strings and two rules.

PRELIMINARIES

In this section, some fundamental definitions on formal language theory are presented. Besides, the definition of Y-G splicing system and second order limit language are also given.

In the biological point of view, the set of alphabet consists of elements of a, c, g, t representing [A/T], [C/G], [G/C] and [T/A] respectively. The string represents the DNA sequences. In addition, a language can be biologically described as a set of DNA molecules that is produced after the initial molecules of dsDNA are spliced with the restriction enzyme and pasted with an appropriate ligase. The first three definitions related to formal language theory are given as follows.

Definition 1 [12]: Alphabet
An alphabet, A, is a finite, nonempty set of symbols.

Definition 2 [12]: String
A string is a finite sequence of symbols from the alphabet.

Definition 3 [12]: Language
A set of strings all of which are chosen from some A^*, where A is a particular alphabet, is called a language.

Note that, A^* denotes the set of all strings over an alphabet A which is obtained by concatenating zero or more symbols from A.

In the following, the definition of Y-G splicing system will be given. Yusof in [7] introduced a new notation for writing rules in a splicing system and a new extension of splicing system extracted from Head's and Goode-Pixton's splicing system.

Definition 4 [7]: Yusof-Goode (Y-G) splicing system
A splicing system $S=(A,I,R)$ consists of a set of alphabets A, a set of initial strings I in A^* and a set of rules, $r \in R$ where $r = (u, x, v: y, x, z)$. For $s_1 = \alpha uxv\beta$ and $s_2 = \gamma yxz\delta$ elements of I, splicing s_1 and s_2 using r produces the initial string I together with $\alpha uxz\delta$ and $\gamma yxv\beta$, 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.

The transient and limit language has been introduced in [13] through a biological experiment. Eventually, transient language is a set of strings that disappear and are used up through the DNA splicing process. In addition, the limit language is the set of existing strings after the system has reached its equilibrium state, regardless of the balance of the reactants in a particular experimental run of the reaction. The definition of second order limit language that has been deduced from n^{th} order limit language [13] is given as follows.

Definition 5 [10]: Second Order Limit Language
Let the set L_2 of second order limit words of L be the set of first order limit words of L_1. We obtain L_2 from L_1 by deleting words that are transient in L_1.

The definition of palindromic which can be applied to strings, crossing sites and also the recognition sites of rules is given in the following definition.

Definition 5 [7]: Palindromic
A string I of dsDNA is said to be palindromic if the sequence from the left to the right side of the upper single strand is equal to the sequence from the right to the left side of the lower single strand.

In the next section, the restrictions that affect the formation of second order limit language in Y-G splicing system are discussed and presented as lemmas and theorem.

MAIN RESULTS

In this section, four lemmas and a theorem on the non-existence of second order limit language in Y-G splicing system is introduced. The theorem is restricted to a set of initial strings with one cutting site and a set of rules where either its crossing site or its recognition site is palindromic.

Lemma 1
Let $S=(A,I,R)$ be a Y-G splicing system where A is a set of alphabets, with one initial string, $I=\{s\}$ and a rule, $R=\{r\}$ involved. Then, no second order limit language exists.

Proof
Case 1: Suppose that the recognition site of the rule is palindromic such that $I=\{\alpha abab\beta\}$ and $r=(a;ba,b:a;ba,b)$ where a and b are complement to each other and $\alpha,\beta,a,b\in A^*$. Therefore, the splicing language is generated as follows:

$$\{\alpha abab\beta\}\xrightarrow{r} I\cup\{\alpha ababa\alpha',\beta' abab\beta\}.$$

So, $L(S)=I\cup\{\alpha ababa\alpha',\beta' abab\beta\}$. By splicing those strings in the generated splicing language under the same rule for the second time, only the same splicing language is generated.

$$I\cup\{\alpha ababa\alpha',\beta' abab\beta\}\xrightarrow{r} I\cup\{\alpha ababa\alpha',\beta' abab\beta\}.$$

Since no distinct splicing language is generated [11], then no second order limit language is generated.

Case 2: Suppose that the crossing site of the rule is palindromic such that $I=\{\alpha abaa\beta\}$ and $r=(a;ba,a:a;ba,a)$. Therefore, the splicing language is generated as follows:

$$\{\alpha abaa\beta\}\xrightarrow{r} I\cup\{\alpha ababa\alpha',\beta' bbaa\beta\}.$$

Again, no second order limit language is generated since splicing among the strings of the generated splicing language will regenerate the same splicing language. □

Lemma 2

Let $S=(A,I,R)$ be a Y-G splicing system where A is a set of alphabets with one initial string, $I=\{\alpha bbab\beta\}$ and two rules such that $R=\{r_1,r_2\}$, $r_1=(b;ba,b:b;ba,b)$ and $r_2=(a;ba,b:a;ba,b)$. Given that, a and b are complement to each other and $\alpha,\beta,a,b\in A^*$. Then, no second order limit language exists.

Proof

Since the initial string only has one cutting site, we set the crossing site of the first rule to be palindromic and the recognition site of the second rule to be palindromic such that $I=\{\alpha bbab\beta\}$ and $r_1=(b;ba,b:a;ba,b)$. Therefore, the splicing language is generated as follows:

$$\{\alpha bbab\beta\}\xrightarrow{r_1} I\cup\{\alpha bbaa\alpha',\beta' abab\beta\}.$$

The first rule is applied in the first splicing process since there is only one recognition site r_1. In the second splicing process, both rules r_1, r_2 are applied. But then no second order limit language is generated since the same splicing languages are generated after the second splicing process. □

Lemma 3

Let $S=(A,I,R)$ be a Y-G splicing system where A is a set of alphabets with two initial strings, $I=\{s_1,s_2\}$ and one rule, $R=\{r\}$. Then, no second order limit language exists.

Proof

Case 1: Suppose that the recognition site of the rule is palindromic such that $I=\{\alpha abab\beta,\gamma abab\delta\}$ and $r=(a;ba,b:a;ba,b)$, where a and b are complement to each other and $\alpha,\beta,\gamma,\delta,a,b\in A^*$. Therefore, the splicing language is generated as follows:

$$\{\alpha abab\beta,\gamma abab\delta\}\xrightarrow{r} I\cup\left\{\begin{matrix}\alpha abab\alpha',\beta' abab\beta,\gamma abab\gamma',\delta' abab\delta,\\ \alpha abab\delta,\gamma abab\beta,\alpha abab\gamma',\delta' abab\beta\end{matrix}\right\}.$$

By splicing those strings of the generated splicing language under the same rule for the second splicing process, only the same splicing language is generated. Since no new splicing language is generated, then no second order limit language is generated.

Case 2: Suppose that the crossing site of the rule is palindromic such that $I=\{\alpha abaa\beta,\gamma abaa\delta\}$ and $r=(a;ba,a:a;ba,a)$. The splicing language is generated as follows:

$$\{\alpha abaa\beta,\gamma abaa\delta\}\xrightarrow{r} I\cup\left\{\begin{matrix}\alpha abab\alpha',\beta' bbaa\beta,\gamma abab\gamma',\delta' bbaa\delta,\\ \alpha abaa\delta,\gamma abaa\beta,\alpha abab\gamma',\delta' bbaa\beta\end{matrix}\right\}.$$

Again, no second order limit language is generated since splicing among the generated splicing language will regenerate the same splicing language. □

Lemma 4

Let $S=(A,I,R)$ be a Y-G splicing system where A is a set of alphabets with two initial strings, $I=\{s_1,s_2\}$ and two rules, $R=\{r_1,r_2\}$. Then, no second order limit language exists.

Proof

Case 1: Suppose that the recognition site of the rule is palindromic such that $I=\{\alpha abab\beta,\gamma bbaa\delta\}$ and $R=\{r_1,r_2\}$ where $r_1=(a;ba,b:a;ba,b)$ and $r_2=(b;ba,a:b;ba,a)$, a and b are complement to each other and $\alpha,\beta,\gamma,\delta,a,b\in A^*$. Therefore, the splicing language is generated as follows:

$$\{\alpha abab\beta,\gamma bbaa\delta\}\xrightarrow{r_1,r_2} I\cup\left\{\begin{matrix}\alpha abab\alpha',\beta' abab\beta,\gamma bbaa\gamma',\delta' bbaa\delta,\\ \alpha abaa\delta,\gamma bbab\beta,\alpha abaa\gamma',\delta' bbab\beta\end{matrix}\right\}.$$

No new splicing language is generated when the splicing process takes place for the second time. Hence, no second order limit language is generated.

Case 2: Suppose that the crossing site of the rule is palindromic such that $I=\{\alpha abaa\beta,\gamma bbab\delta\}$ and $R=(r_1,r_2)$ where $r_1=(a;ba,a:a;ba,a)$ and $r_2=(b;ba,b:b;ba,b)$. Therefore, the splicing language is generated as follows:

$$\{\alpha abaa\beta,\gamma bbab\delta\}\xrightarrow{r_1,r_2} I\cup\left\{\begin{matrix}\alpha abab\alpha',\beta' bbaa\beta,\gamma bbaa\gamma',\delta' abab\delta,\\ \alpha abab\delta,\gamma bbaa\beta,\alpha abaa\gamma',\delta' abaa\beta\end{matrix}\right\}.$$

Again, no second order limit language is generated.

Case 3: Given $I=\{\alpha abaa\beta,\gamma bbab\delta\}$ where the crossing site of the first rule is palindromic, while the recognition site of the second rule is palindromic such that $R=\{r_1,r_2\}$ where $r_1=(a;ba,a:a;ba,a)$ and $r_2=(a;ba,b:a;ba,b)$. Therefore, the splicing language is generated as follows:

$$\{\alpha abaa\beta,\gamma abab\delta\}\xrightarrow{r_1,r_2} I\cup\left\{\begin{matrix}\alpha abab\alpha',\beta' bbaa\beta,\gamma abaa\gamma',\delta' abab\delta,\\ \alpha abab\delta,\gamma abaa\beta,\alpha abab\gamma',\delta' abaa\beta\end{matrix}\right\}.$$

When splicing those strings of the generated splicing language using the same rules for the second time, no new splicing language is generated. Hence, no second order limit language is generated. □

From Lemma 1 to Lemma 4, we arrive at the following theorem.

Theorem 1

Let $S=(A,I,R)$ be a Y-G splicing system where A is a set of alphabets, I is a set of initial strings with only one cutting site and R is a set of rules where its crossing site or recognition site is palindromic, then there is no second order limit language.

Proof

From Lemma 1 to Lemma 4, Y-G splicing systems involving one initial string with one rule, one initial string with two rules, two initial strings with one rule and two initial strings with two rules are considered respectively, whereby the cases where either the recognition site or the crossing site of the set of rules are palindromic, are also considered. Since, the lemmas do not produce second order limit language, therefore, it has been proven that Y-G splicing systems involving at most two initial strings with one cutting site and two rules do not produce second order limit language. Thus, the proof is complete. □

EXAMPLE OF DNA RECOMBINATION INVOLVING SECOND ORDER LIMIT LANGUAGE OF Y-G SPLICING SYSTEM

In an experiment, a test tube contains DNA templates, restriction enzymes and an appropriate ligase. The DNA templates which are chosen from enterobacteria phage lambda digested with *Hind*III from New England Biolabs are then generated for many copies in PCR. The restriction enzymes from [12] are supplied together with the suitable buffer for robust production. By adding an appropriate ligase to the solution containing DNA templates and the restriction enzymes, new molecules will be formed.

In the following example, two different restriction enzymes namely *Hpa*II and *HinP*1I which are supplied together with CutSmart™ buffer that works best for both restriction enzyme. Note that, the recognition site of the rules are palindromic.

Example 1
Let $S=(A,I,R)$ be a Y-G splicing system consisting of a set of alphabets, $A=\{a,c,g,t\}$, a set of initial strings, $I=\{\alpha ccgg\beta, \gamma gcgc\delta\}$ such that α with β, γ with δ, α' with β', γ' with δ' are complement to each other where $\alpha, \beta, \gamma, \delta, \alpha', \beta', \gamma', \delta' \in A^*$ and $R=\{r_1, r_2\}$ is a set of rules where $r_1=(c;cg,g:c;cg,g)$ and $r_2=(g;cg,c:g;cg,c)$. When splicing occurs, the following splicing language is generated:

$$\{\alpha ccgg\beta, \gamma gcgc\delta\} \xrightarrow{r_1, r_2} I \cup \left\{\begin{matrix}\alpha ccgg\alpha', \beta' ccgg\beta, \gamma gcgc\gamma', \\ \delta' gcgc\delta, \alpha ccgc\delta, \gamma gcgg\beta, \alpha ccgc\gamma', \delta' gcgg\beta\end{matrix}\right\}.$$

Again, this Y-G splicing system does not generate second order limit language since no new string is generated when the splicing process occurs once again.

DISCUSSION

In the lemmas and theorem above, the initial string has only one cutting site and we have shown in all possible cases that the production of second order limit language is impossible. This is due to the right and left context of the rule and also the condition where only one cutting site exists in the initial string. The recognition site of the rule that is palindromic always lead to molecules with the same sequence as of the initial string except the arrangement of α and β. The left and right context of the rule with fixed alphabet will not produce a variety of strings hence not a second order limit language . In the example that has been provided, the second order limit language exists when the initial language has two cutting sites.

CONCLUSION

In summary, from Lemma 1 to Lemma 4, we consider the cases of one initial string with one rule, one initial string with two rules, two initial strings with one rule and two initial strings with two rules that do not generate second order limit language. The theorem is derived from those four lemmas which conclude that splicing system with at most two initial strings, two rules and one cutting site do not produce second order limit language. The example given illustrates the existence of non-second order limit language from the biological point of view.

ACKNOWLEDGMENTS

The first author is indebted to the Ministry of Education (MOE) Malaysia for his financial funding through MyBrain15 (MyPhD) scholarship. The second and fourth authors would like to thank MOE and Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the financial funding through UTM Research University Fund Vote No. 08H45.

REFERENCES

1. R. H. Tamarin, *Principles of Genetics*, 7th. ed. USA: The Mac-Graw Hill Companies, 2001, pp. 17-38.
2. I. E. Alcamo, *DNA Technology The Awesome Skill*, 2nd. ed. USA: Harcourt/Academic Press, 2001, pp. 21-38.
3. T. Head, *Bulletin of Mathematical Biology* **49**, 737 – 759 (1987).
4. F. Karimi, "Mathematical Modelling of Persistent Splicing System in DNA Computing", Ph.D. Thesis, Universiti Teknologi Malaysia, 2013.
5. M. H. Mudabber, Y. Yusof and M. S. Mohamad, "Some Relations between Two Stages DNA Splicing Languages" in *Proceedings of The 3rd International Conference on Mathematical Sciences*, edited by W. Z. W. Zin et al., AIP Conference Proceedings 1602, American Institute of Physics, Melville, NY, 2014, pp. 254-259.
6. M. H. Mudaber, Y. Yusof and M. S. Mohamad, "Some Sufficient Conditions for Persistency and Permanency of Two Stages DNA Splicing Languages via Yusof-Goode Approach" in *Proceedings of the 21st National Symposium on Mathematical Sciences (SKSM21)*, edited by M. T. Ismail et al, AIP Conference Proceedings 1605, American Institute of Physics, Melville, NY, 2014, pp. 591-595.
7. Y. Yusof, "DNA Splicing System Inspired by Bio Molecular Operations", Ph.D. Thesis, Universiti Teknologi Malaysia, 2011.
8. E. Laun and K. J. Reddy, "Wet Splicing System," in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, edited by H. Rubin and D. H. Wood, USA: American Mathematical Society, 1999, pp. 73 – 83.
9. W. H. Fong, "Modelling of Splicing System using Formal Language Theory", Ph.D. Thesis, Universiti Teknologi Malaysia, 2008.
10. M. A. Ahmad, N. H. Sarmin, W. H. Fong and Y. Yusof, "An Extension of First Order Limit Language" in *Proceedings of The 3rd International Conference on Mathematical Sciences*, edited by W. Z. W. Zin et al., AIP Conference Proceedings 1602, American Institute of Physics, Melville, NY, 2014, pp. 627-631.
11. A. Ahmad, N. H. Sarmin, W. H. Fong and Y. Yusof, "Second Order Limit Language in Variants of Splicing System" in *Proceedings of the 21st National Symposium on Mathematical Sciences (SKSM21)*, edited by M. T. Ismail et al, AIP Conference Proceedings 1605, American Institute of Physics, Melville, NY, 2014, pp. 639-643.
12. P. Linz, *An Introduction to Formal Languages and Automata*, USA: Jones and Barlett Publisher, 2006, pp. 16-19.
13. E. Goode and D. Pixton, "Splicing to the Limit," in *Aspects of Molecular Computing*, edited by N. Janoska, Gh. Paun and G. Rozenberg, New York: Springer Berlin Heidelberg, 2004, pp. 189 – 201.