RESEARCH ARTICLE | JULY 10 2014

Second order limit language in variants of splicing system

Muhammad Azrin Ahmad; Nor Haniza Sarmin; Fong Wan Heng; Yuhani Yusof

AIP Conf. Proc. 1605, 639–643 (2014)
https://doi.org/10.1063/1.4887664

Second Order Limit Language in Variants of Splicing System

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

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

Abstract. The cutting and pasting processes that occur in DNA molecules have led to the formulation of splicing system. Since then, there are few models used to model the splicing system. The splicing language, which is the product of splicing system, can be categorized into two, namely the adult and limit language. In this research, limit language is extended to the second order limit language. Few problems are approached which lead to the formation of second order limit language which is then analyzed using various types of splicing system.

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

INTRODUCTION

Deoxyribonucleic acid (DNA) is an important cell located inside every living organism [1]. The DNA molecules are made up from thousand of pairs of nucleotides which consist of a base, sugar, and phosphate. The base where the information in DNA stored as a form of code is built up from four types of chemical bases: adenine (*A*), guanine (*G*), cytosine (*C*) and thymine (*T*). By Watson-Crick complementarity [2], the only possible pairing out of those bases are adenine with thymine (*A* - *T*), guanine with cytosine (*C* - *G*) and vice versa. In 1987, Head [3] introduced the formal presentation in presenting the recombination of DNA molecules. As time evolved, many researches made numerous works regarding splicing system which then contribute to the formation of enhanced or extended version of splicing system such as Paun, Pixton, Goode-Pixton (G-P) and Yusof-Goode (Y-G) splicing system [4, 5, 6, 7].

The rising of model of splicing system was tremendously found after the remarkable work pioneered by Head. One of them is Paun splicing system that is recognized as a powerful tool in the formalism of splicing system because the rule in his system can deal with either infinite or finite form. Then, Pixton splicing system was introduced which is not representing the biological cut and paste process in the DNA molecules. The factor β which appeared in Paun model, acts as an intervening effect after the rule which is not more than just a substitution operation of the splicing [10]. Due to this limitation, we will omit this splicing system. When the Pixton splicing system was found mistake by many researches, the following Goode-Pixton splicing system was introduced as a substitute model to have a better representation in explaining the cutting and pasting process in DNA molecules. As the study concerned, rules in splicing system that represent restriction enzymes can be categorized into three types of pattern which are 5*'* overhang, blunt end and 3*'* overhang. These patterns later can be grouped as left pattern and right pattern. Based on the given model, the chosen restriction enzymes are difficult to recognize either they fall under left pattern or right pattern.

In the recent years, Yusof in [7], introduced a new model extended from Head's and Goode-Pixton's version of splicing system which is the Y-G splicing system. This version of splicing system presents the transparent behaviour of the DNA biological process. It is claimed so because the splicing model itself is inspired by the characteristics of the restriction enzymes contained in the model. In obtaining the splicing languages, the role of symmetric and reflexive rules are quite important. It affects the formation of splicing languages since the splicing process is up to two stages in the splicing systems.

From those types of splicing languages, it can be seen that they are categorized into two categories. The first one is a splicing model based on generation of language while the other one is a splicing model that preserves the biological traits in splicing system. Second order limit language is more to biological characteristics from splicing system. The output of splicing system, namely splicing language can be categorized into three types: adult or inert, transient and limit language.

Proceedings of the 21st National Symposium on Mathematical Sciences (SKSM21)
AIP Conf. Proc. 1605, 639-643 (2014); doi: 10.1063/1.4887664

Yusof in [8] had re-defined few terms for the types of splicing languages resulted with inert persistent language from inert language and also contributed to a new type of splicing language called active persistent language. The extension of limit language, the n^{th} order limit language has been discussed in [9]. Based on [8, 9, 11], their interest of study is limited only to limit language. The researchers also have made laboratory work through wet-lab to verify their theories related to limit language. In this paper, the methods on recognizing second order limit language are introduced. The motivation to study second order limit language rather than limit language is that it requires two stage of splicing for them to exist. One cannot simply choose normal biological example to obtain second order limit language. Some properties of initial language and rules in the splicing system should be monitored to ensure the existence of second order limit language. Other than that, the appearance of second order limit language in variants of splicing system is studied in the types of splicing system.

In the next section, some preliminaries results that will be used in this study are presented as definition.

PRELIMINARIES

In this section, the definitions of Head and Y-G splicing system are given. In addition, the definitions related to splicing languages are also presented.

Definition 1 [3]: Head Splicing System

A splicing system $S = (A, I, B, C)$ consists of a finite alphabet A, a finite set I of initial strings in A^*, and finite sets B and C of triples (c, x, d) with c, x and d in A^*. Each such triple in B or C is called a pattern. For each such triple the string *cxd* is called a site and the string x is called a crossing. Patterns in B are called left patterns and patterns in C are called right patterns. The language $L = L(S)$ generated by S consists of the strings in I and all strings that can be obtained by adjoining the words *ucxfq* and *pexdv* to L whenever *ucxdv* and *pexfq* are in L and (c, x, d) and (e, x, f) are patterns of the same hand. A language L is a splicing language if there exists a splicing system S for which $L = L(S)$.

Definition 2 [7]: Y-G Splicing System

If $r \in R$, where $r = (u,x,v\!:\!y,x,z)$ 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 $\alpha uxz\beta$ and $\gamma yxz\delta$, presented in either order where $\alpha,\beta,\gamma,\delta,u,x,v,y$ and $z \in A^*$.

In the following, the definitions of transient language and limit language that contribute to the understanding of second order limit language are presented. In addition, definition of n^{th} order limit language is given which is then deduced to form the definition of second order limit language.

Definition 3 [9]: Transient Language

A splicing language is called transient if a set of strings is eventually used up and disappears in a given system.

The other types of splicing language, which is limit language, is given below.

Definition 4 [9]: 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 its equilibrium state, regardless of the balance of the reactants in a particular experimental run of the reaction.

The extension of limit language to the n^{th} order is described in the following definition.

Definition 5 [9]: n^{th} Order Limit Language

Let L_{n-1} be the set of second order limit words of L, the set L_n of n^{th} order limit words of L to be the set of first order limit word of L_{n-1}. We obtain L_n from L_{n-1} by deleting the words that are transient in L_{n-1}.

The definition of second order limit language that is deduced from n^{th} order limit language is defined as follows.

Definition 6: Second Order Limit Language

Let L_1 be 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.

After the definition of second order limit language is presented, the methods of recognizing second order limit language in splicing system are given as two theorems. Then two examples are illustrated to explain the existence of second order limit language in two different categories of splicing system: Paun and Y-G splicing system.

MAIN RESULTS

In this section, we introduced two methods to recognize second order limit language. Those methods are presented as two theorems.

Theorem 1
If the combination of two splicing languages of first stage splicing under the given rule has different lengths from those two splicing languages of first stage splicing, then the second order limit language is identified and exists.

Proof
Suppose $S = (A, I, R)$ is a splicing system where A is an alphabet, I is the initial language and the rule R possesses symmetric and reflexive properties. Regardless of the number of rules and initial language involved, the following languages are produced. Let say, $L(S)$ be the set of languages generated from the splicing system,

$$L(S) = \{a, c, \dots , z\}.$$

For each string, they can be written as $a = a_1aa_2$, $c = c_1cc_2$ and up to $z = z_1zz_2$ where we are expected the length for each string to be different, the length of produced language are varies. According to the splicing rule in which the first half of the cut string are pasted to second half of another string, the new string which is the hybrid of both length of the strings are also having different length from their parent string. It is shown, that the new language, we denote as *a'*, is formed where this new string, the result of cut and paste from string a and c, has different length from the original one, denoted as $a' = a_1acc_2$. Hence, second order limit language is identified and existed. □

Theorem 2
If the resulted splicing language that is derived from splicing language of first stage splicing has different language from the resulted splicing language, then it is second order limit language.

Proof
Using the proof of Theorem 1, the language obtained which is identified of having different length from the parent strings are also considered as a new splicing language from the first one (the previous splicing language), therefore it is the second order limit language. □

By definition of second order limit language, it is quite hard to compute second order limit language if the model used the same restriction enzyme that has been used in the normal situation. Some alteration in the model should be made to ensure the existence of second order limit language. This idea can be seen in the examples provided in the following examples. First and foremost, to give a better explanation in explaining the existence of second order limit language in these two different category models of splicing system, the same restriction enzyme from actual biological case from [11] is used. Example 1 shows the splicing language formed in Paun splicing system while in Example 2, the example of second order limit language in Y-G splicing system is discussed. As Y-G splicing system is a modification from Head and Goode-Pixton splicing system, the Example 2 is also considered explaining second order limit language in Head and Goode-Pixton splicing system.

Example 1

Let $\sigma = (A, R)$ be a Paun splicing system where A is an alphabet and $R \subseteq cccgcttaa\#cg\$c\#cgc$ is a set of splicing rules, for #, $ two special symbols not in A. An initial string $I = \{\alpha cccgcttaacg\beta\}$ in A^* can be spliced via rule r in R to produce the following language:

$$\sigma(I) = \{\alpha cccgcttaacgcttaacg\beta\}.$$

As symmetric and reflexive properties do not hold automatically we assumed the example above is not symmetric and reflexive and by Paun splicing system, second order limit language cannot be determined. It can be seen clearly that Paun splicing system due to the symmetric and reflexive properties that do not hold automatically is not having a right mechanism in explaining the formation of second order limit language which can be detected at second stage of splicing.

Example 2

Let $S = (A, I, R)$ be a Y-G splicing system consisting of two restriction enzymes namely *Smu*I and *Aci*I, where $A = \{a, c, g, t\}$, $I = \{\alpha cccgcttaacg\beta\}$ such that $\alpha, \beta \in A^*$, and $R = \{(r_1 : r_2)\}$ where $r_1 = (cccgcttaa; cg, 1)$ and $r_2 = (c; cg, c)$. When splicing occurs, the following splicing languages are generated:

$$(\alpha cccgcttaacg\beta) \xrightarrow{r_1, r_2} \{\alpha cccgcttaacg\beta,\ \alpha cccgcttaacgttaagcggg\alpha',\ \beta' cg\beta,\ \alpha cccggg\alpha',\ \beta' cgttaagcgcttaacg\beta,$$
$$\alpha cccgcttaacgcttaacg\beta,\ \alpha cccg\beta,\ \alpha cccgcttaacggg\alpha',\ \beta' cgttaaggcg\beta\}.$$

Based on the rules stated above, when the resulted splicing languages are being spliced again, other new splicing languages are obtained which are listed below:

$$(\alpha cccgcttaacgttaagcggg\alpha', \alpha cccgcttaacgcttaacg\beta) \xrightarrow{r_1, r_2} \alpha cccgcttaacgcttaacgcttaacg\beta.$$

$$(\alpha cccgcttaacgcttaacg\beta, \alpha cccgcttaacgttaagcggg\alpha') \xrightarrow{r_1, r_2} \alpha cccgcttaacgcttaacgttaagcggg\alpha'.$$

$$(\alpha cccgcttaacgttaagcggg\alpha', \alpha cccgcttaacggg\alpha') \xrightarrow{r_1, r_2} \alpha cccgcttaacgcttaacggg\alpha'.$$

It can be seen that when $\alpha cccgcttaacgttaagcggg\alpha'$ and $\alpha cccgcttaacgcttaacg\beta$ are spliced together, they are actually being used up to form this $\alpha cccgcttaacgcttaacgcttaacg\beta$ which are the limit words. According to the definition, transient words that are being used up are deleted to form second order limit language. From the examples above, we arrive at the following conjecture.

Conjecture 1

If the splicing system contains second order limit language, then the preferable splicing model is Y-G splicing system.

CONCLUSION

In this paper, the methods to recognize second order limit language are presented in two theorems. Some justifications of types of splicing system to the existence of second order limit language are discussed. Two examples are discussed to illustrate the justification where Y-G splicing system is more preferable to study second order limit language compare to other variants of splicing systems.

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), University Teknologi Malaysia (UTM) for the financial funding through UTM Research University Fund Vote No. 07J41.

REFERENCES

1. R. H. Tamarin, *Principles of Genetics*, 7th. ed. USA: The Mac-Graw Hill Companies, 2001.
2. I. E. Alcamo, *DNA Technology The Awesome Skill*, 2nd. ed. USA: Harcourt/Academic Press, 2001.
3. T. Head, *Bulletin of Mathematical Biology* **49**, 737-759 (1987).
4. Gh. Paun, *Discrete Applied Mathematics* **70**, 57-79 (1996).
5. D. Pixton, Discrete Applied Mathematics **69**, 101-124 (1996).
6. T. E. G Laun, “Constants and Splicing Systems”, Ph.D. Thesis, State University of New York, 1999.
7. Y. Yusof, N. H. Sarmin, T. E. Goode, W. H. Fong, *International Journal of Modern Physics: Conference Series* **9**, 271-277 (2012).
8. Y. Yusof, "DNA Splicing System Inspired by Bio Molecular Operations", Ph.D. Thesis, Universiti Teknologi Malaysia, 2011.
9. T. E. Goode, D. Pixton, *Lecture Notes in Computer Science* **2950**, 189 – 201 (2004).
10. P. Bonizzoni, C. Ferretti, G. Mauri, R. Zizza, *Information Processing Letters* **79**, 255-259 (2001).
11. F. Karimi, “Mathematical Modelling of Persistent Splicing System in DNA Computing”, Ph.D. Thesis, Universiti Teknologi Malaysia, 2013.
12. Research Biolabs Sdn. Bhd. *New England Biolabs 2011-12 Catalogue & Technical Reference*. USA: Catalogue. 2011.