Jurnal Teknologi

Full paper

A New Relation of Second Order Limit Language in Simple and Semi-Simple Splicing System

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

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

*Corresponding author: nhs@utm.my

Article history

Received : 31 July 2014
Received in revised form :
23 November 2014
Accepted : 1 December 2014

Graphical abstract

$$L_{2(\text{left pattern, right pattern})}(S) = L(S) = \left\{\alpha b(ab)^k a\beta, \beta'(ab)^k a\beta\right\},$$

Abstract

Splicing system, which is an abstraction of operations on DNA molecules, can be modelled mathematically under the framework of formal language theory and informational macromolecules. The recombinant behavior of the set of double-stranded DNA molecules under the influence of restriction enzyme and ligase further lead to the cut and paste phenomenon in splicing system. The theoretical study of splicing language has contributed to a new type of splicing language known as a second order limit language, which is an extension of limit language. Some types of splicing system can produce second order limit language. Y-G splicing system is chosen among other models to model the DNA splicing process as this model preserves the biological traits and presents the transparent behavior of the DNA splicing process. In this paper, the relation between second order limit language with simple splicing and semi-simple splicing system are presented.

Keywords: Y-G splicing system; Y-G splicing language; second order limit language

Abstrak

Sistem hiris-cantum yang merupakan suatu operasi pemujaradan ke atas molekul-molekul DNA boleh dimodelkan secara matematik di bawah rangka kerja teori bahasa formal dan makromolekul berinformasi. Tingkah laku rekombinan molekul-molekul DNA bebenang dua di bawah pengaruh enzim pembatas dan ligase telah membawa kepada fenomena potong dan tampal di dalam sistem hiris-cantum. Kajian teori bahasa hiris-cantum telah menyumbang kepada jenis bahasa hiris-cantum yang baharu yang dikenali sebagai bahasa batas berperingkat dua yang juga merupakan lanjutan kepada bahasa batas. Beberapa jenis sistem hiris-cantum boleh menghasilkan bahasa batas berperingkat dua. Sistem hiris-cantum Y-G dipilih daripada model-model lain bagi membentuk proses hiris-cantum DNA kerana model ini dapat mengekalkan ciri-ciri biologi dan mewakili tingkah laku telus proses hiris-cantum DNA. Dalam kertas kerja ini, hubungan di antara bahasa batas berperingkat dua dengan sistem hiris-cantum mudah dan sistem hiris-cantum separa-mudah diperkenalkan.

Kata kunci: Sistem hiris-cantum Y-G; bahasa hiris-cantum Y-G; bahasa batas berperingkat dua

1.0 INTRODUCTION

Deoxyribonucleic acid (DNA) acts as a hereditary factor that passes information from parent cell to offspring cell and also as a coding for protein production in living organisms.[1] DNA is a polymer made up of many monomers called deoxyribonucleotides. These deoxyribonucleotides consist of three different components which are sugar, phosphate and nitrogenous base. Moreover, there are four types of bases namely adenine (A), guanine (G), cytosine (C) and thymine (T). In addition, the bases can be grouped as purines (A and G) and pyrimidines (C and T) accordingly. Watson-Crick complementarity[2] stated that the only possible pairings are A with T, C with G and vice versa. Basically the sugar component has five carbon atoms that are numbered from $1'$ to $5'$. The phosphate is attached to the $5'$ carbon while the base is attached to the $1'$ carbon. Meanwhile a hydroxyl group (OH) is attached to the $3'$ carbon.

In 1987, Head[3] pioneered a mathematical modelling of splicing system. A splicing system is a study of the recombinant behavior of restriction enzymes on DNA molecules. The recognition process that determines the cutting site is depicted by the restriction enzyme. The restriction enzyme will clamp at the specific nucleotides sequence known as the crossing site on DNA molecules and the cutting process will take place. Besides that, interdisciplinary knowledge between formal language theory and the study of informational molecules is initiated. In splicing system, a splicing model is either based on the generation of language or a splicing model that preserves the biological traits in a splicing system.

Our focus in this research is on a splicing model that preserves the biological traits of DNA splicing process. Goode[4] and Yusof[5] has studied the mathematical model based on this characteristic. In this paper, Y-G splicing system is chosen since it presents the transparent behavior of the DNA splicing process. The second order limit language was deduced from the *n*-th order limit language.[6] Since then, its existence and properties in DNA splicing system were studied. Besides, the splicing languages resulting from the corresponding splicing systems have been studied through some wet-lab works. [5, 7, 8, 9, 10] In the previous works, researchers investigated on the adult, transient and limit language. In this paper, the relation between the second order limit language with simple splicing, and semi-simple splicing system is presented.

2.0 PRELIMINARIES

In this section, some definitions regarding formal language theory and splicing system which are used in achieving the results in this paper are given.The first three fundamental definitions are related to formal language theory.

Definition 1 [11]: Alphabet, *A*
An alphabet, *A*, is a finite nonempty set of symbols.

Definition 2 [11]: String
A string is a finite sequence of symbols from the alphabet.

Definition 3 [11]: Language, *L*
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*.

The definition of Y-G splicing system that is used throughout this paper is stated below.

Definition 4 [5]: 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.

Besides that, two types of splicing languages that have contributed to the development of second order limit language are the transient and limit language. Experimentally, a splicing language is called transient if a set of strings is eventually used up and disappears in a given system. Besides that, a splicing language is a limit language given that it 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 definition of second order limit language is given in the following.

Definition 5 [6]: Second Order Limit Language
Let the set L_2 of second order limit words of *L* to 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.

In the following, the definition of simple splicing and semi-simple splicing system are provided. These definitions have been modified according to the definition of Y-G splicing system.[5]

Definition 6 [12]: Simple Splicing System
Let $S=(A,I,R)$ be a splicing system in which all rules, $r \in R$ have the form $(a,1,1:a,1,1)$, $(1,a,1:1,a,1)$ or $(1,1,a:1,1,a)$ where $a \in A$, then *S* is called a simple splicing system.

Definition 7 [4]: Semi-Simple Splicing System
Let $S=(A,I,R)$ be a splicing system in which *I* and *R* are finite and every rule, $r \in R$ in *S* has the form $(a,1,1:b,1,1)$ where $a,b \in A$. Thus $\sigma=(A,R)$ is called a semi-simple splicing scheme and (A,I,R) a semi-simple splicing system.

Next, the following proposition is given which is needed to prove the main results.

Proposition 1 [5]
Every simple splicing system is a semi-simple splicing system of the form (*A*, *I*, *R*).

In the next section, the relations of some types of splicing systems and language with second order limit language are given in two theorems.

3.0 RESULTS AND DISCUSSION

In this section, the relation between two types of splicing language namely simple splicing and semi-simple splicing system with the second order limit language are discussed. The first theorem shows the existence of a second order limit language in a simple splicing system.

Theorem 1
Let $S=(A,I,R)$ be a Y-G splicing system where $A=\{a,c,g,t\}$ is an alphabet. If a splicing system is simple where $R = (a, 1, 1: a, 1, 1)$ and $a \in A$, then there exists a second order limit language.

Proof
Suppose $S=(A,I,R)$ is a Y-G splicing system. Let *A* be a set of alphabets, $I=\{\alpha baba\beta\}$ is a set of initial strings such that a with a' and b with b' are complement to each other and $\alpha,\beta,\alpha',\beta' \in A^*$ and $R = \{r\}$, where $r=(a,1,1:a,1,1)$, $a \in A$. In the simple splicing system, since the crossing site is in null form, the splicing action has to be considered in two patterns which are, the left pattern and right pattern. The resulted splicing language is shown in the following:

i. The splicing language of left pattern $\{\alpha bababa\beta\} \xrightarrow{R} I \cup \{\alpha ba\beta, \beta' ab\beta, \beta' aba\beta\}$.

ii. The splicing language of right pattern $\{\alpha bababa\beta\}\xrightarrow{R} I\cup\{\alpha ba\beta,\beta' ab\beta,\beta' aba\beta\}$.

When the splicing process occurs once again and splice among the resulted splicing languages from the previous splicing, the following new languages are formed:

$$L_{2(\text{left pattern, right pattern})}(S)=L(S)=\left\{\alpha b(ab)^k a\beta,\beta'(ab)^k a\beta\right\},$$

for $k\geq 2$ and $k\in\mathbb{Z}^+$.

Since the splicing language is the same for both the left and right pattern, then the same second order limit language is produced by both patterns. According to Definition 5, it is shown that $L_2(S)$ is a second order limit language. Hence, the second order limit language, $L_2(S)$ exists in the simple splicing system.

Theorem 2

If a splicing system $S = (A, I, R)$ is semi-simple where $R = (a, 1, 1: b, 1, 1)$ and $a, b\in A$, then there exists a second order limit language.

Proof

By Proposition 1, since there exists a second order limit language in a simple splicing system, then there also exists a second order limit language in a semi-simple splicing system. Hence, Theorem 2 holds. □

■4.0 CONCLUSION

As a conclusion, the relationship between simple splicing and semi-simple splicing system with the second order limit languages are established. The existence of the second order limit language in those two types of splicing system and splicing languages are proved in Theorem 1 and Theorem 2. By using the result in Theorem 1 and Proposition 1, we proved that Theorem 2 holds. The results show that the second order limit language can be extended to specific types of splicing system.

Acknowledgement

The first author is indebted to the Ministry of Education (MOE) Malaysia for his financial funding through MyPhD, MyBrain15 Scholarship. The second and third 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] Tamarin, R.H. 2001. 7th Edition. *Principles of Genetics*. USA: The Mac - Graw Hill Companies.

[2] Paun, G., Rozernberg, G. and Salomaa, A. 1998. *DNA Computing New Computing Paradigm*. Germany: Springer Verlag.

[3] Head, T. 1987. Formal Language Theory and DNA: An Analysis of Generative Capacity of Specific Recombinant Behaviors. *Bulletin of Mathematical Biology*. 49 (6). 737–759.

[4] Goode, T.E. 1999. *Constant and Splicing System*. Ph.D. Thesis. State University of New York.

[5] Yusof, Y. 2011. *Bio Molecular Inspiration in DNA Splicing System*. Ph.D. Thesis. Universiti Teknologi Malaysia.

[6] Ahmad, M.A., Sarmin, N.H., Fong, W.H. and Yusof, Y. 2014. An Extension of First Order Limit Language. *AIP Proceedings*. 1602: 627–631.

[7] Goode, E. and Pixton, D. 2004. Splicing to the Limit. *Lecture Notes in Computer Science*. 2950: 189-201.

[8] Fong, W.H. 2008. *Modelling of Splicing Systems Using Formal Language Theory*. Ph.D. Thesis. Universiti Teknologi Malaysia.

[9] Karimi, F. 2013. *Mathematical Modelling of Persistent Splicing Systems in DNA Computing*. Ph.D. Thesis, Universiti Teknologi Malaysia.

[10] Laun, E. and Reddy, K.J. 1997. Wet Splicing Systems. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* 48: 73–86.

[11] Linz, P. 2006. *An Introduction to Formal Languages and Automata*. USA: Jones and Bartlett Publishers.

[12] Mateescu, A., Paun, Gh., Rozenberg, G. and Salomaa, A. 1998. Simple Splicing System. *Discrete Applied Mathematics* 84: 145–163