MATEMATIKA, 2018, Volume 34, Number 1, 59–71
© Penerbit UTM Press. All rights reserved

Generalisations of DNA Splicing Systems with One Palindromic Restriction Enzyme

Wan Heng Fong* **and Nurul Izzaty Ismail**
Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Johor, Malaysia
*Corresponding author: fwh@utm.my

Article history
Received: 18 July 2017
Received in revised form: 1 November 2017
Accepted: 16 January 2018
Published on line: 1 June 2018

Abstract In DNA splicing system, the potential effect of sets of restriction enzymes and a ligase that allow DNA molecules to be cleaved and re-associated to produce further molecules is modelled mathematically. This modelling is done in the framework of formal language theory, in which the nitrogen bases, nucleotides and restriction sites are modelled as alphabets, strings and rules respectively. The molecules resulting from a splicing system is depicted as the splicing language. In this research, the splicing language resulting from DNA splicing systems with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites are generalised as regular expressions.

Keywords DNA; formal language theory; palindromic; restriction enzyme; splicing system.

Mathematics Subject Classification 68Q45; 92B05

1 Introduction

DNA is the molecule that plays the main role in DNA computing. The basic idea in DNA computing is the information-processing capabilities of organic molecules in computers which can be replaced with digital switching primitives [1]. Several models have been carried out using present technology where some components and models are built based on DNA computing. The mathematical model in splicing system is one of the models in DNA computing that generates languages by using formal language theory.

Formal language theory is a branch of theoretical computer science that is devoted to the study of sets of finite strings (called languages) of symbols chosen from a prescribed finite set (called an alphabet) [2]. The fundamental knowledge of formal language theory and some related molecular biological terms related to splicing system are presented in [3]. The language resulting from a splicing system by using formal language theory is called a splicing language. Research on DNA splicing systems with different types of restriction enzymes has been discussed in [4], where the restriction enzymes cut DNA molecules in specific ways based on the cleavage pattern of the enzymes.

2 Literature Review

Deoxyribonucleic Acid (DNA) is a polymer strung together from monomers [1]. DNA has four bases in their nucleotide chain which are two purines, adenine (A) and guanine (G); and two pyrimidines, cytosine (C) and thymine (T). The set of double stranded DNA (dsDNA) is represented by four-symbol alphabets a, g, c, and t where each alphabet stands for [A/T], [G/C], [C/G] and [T/A], respectively.

A new manner of relating formal language theory to the study of informational macromolecules is initiated as splicing system by Head [2] in 1987. In a splicing system, a language is associated with each pair of sets where the first set consists of double stranded DNA molecules and the second set consists of the recombination behaviours allowed by specified classes of enzymatic activities [2].As years passed by, several models in splicing system were developed, namely Paun, Pixton, Goode Pixton and Yusof-Goode splicing system. Paun Splicing System, known as S_{PA}, was introduced by Paun in 1996 [5]. Paun's formalism for splicing systems patterns are pairs which define a binary relation, and that strings are allowed to tie together if they contain patterns which are in this relation [5]. In the same year, Dennis and Pixton [6] introduced Pixton Splicing system, S_{PI}. This splicing system is not really a formal description of the biological cut and paste phenomenon since it is presented as a substitute operation of the splicing operation [7]. Besides that, Laun also developed a new model of splicing system known as Goode-Pixton splicing system [8]. Then, a new extension of splicing system also has been developed from Head and Goode-Pixton splicing system models as a simple splicing system in [8]. Next, a new extension of splicing system known as Yusof-Goode (Y-G) splicing system is presented by Yusof [9] in 2013. It was invented to present the transparent behaviours of the DNA biological process besides the characteristics of the restriction enzyme itself. Furthermore, simple, semi-simple and semi-null splicing systems are introduced in [10].

A splicing language is a language resulting from a splicing system. There are many types of splicing languages such as persistent, strictly locally testable and uniform splicing language which have been discussed in [2, 11, 12]. Besides, simple spicing languages are defined by Laun [8] in 1999. A series of language-theoretic properties of simple splicing languages generated by such systems with finite sets of axioms are investigated in [13, 14].

In wet lab, enzymes are biological catalysts that increase the rate of chemical reactions taking place within living cells without themselves suffering any overall change[15]. In splicing system, DNA molecules mix with a ligase and restriction enzymes which are called as endodeoxyribonucleases that allow the molecules to be cut and recombined [16].

In a splicing system, a restriction enzyme cuts DNA molecules based on the cleavage pattern of the enzyme [1]. Every restriction enzyme consists of a triple known as the cleavage pattern of the enzyme [1]. The cleavage patterns of restriction enzymes can be shown through three ways; $5'$ overhang, $3'$ overhang or blunt ends. The recombinations of DNA molecules are allowed in splicing systems by the cutting sites of enzymes. The symbols ↓ and ↑ refer to the cutting sites by the restriction enzymes. The cutting sites and cleavage patterns of enzymes *Age*I, *Kpn*I and *Dpn*I for $5'$ overhang, $3'$ overhang and blunt ends respectively are illustrated in the following:

$$\textit{Age}\text{I }(a,\ ccgg,\ t):\ \begin{array}{llll} 5'\text{–A} & \downarrow & \text{CCGG} & \\ 3'\text{–T} & & \text{GGCC} & \uparrow \end{array}\ ,$$

$$Kpn\mathrm{I}\ (g,\ gtac,\ c)\text{:}\ \begin{array}{lclcl} 5'-\mathrm{G} & & \mathrm{GTAC} & \downarrow & \mathrm{C}-3' \\ 3'-\mathrm{C} & \uparrow & \mathrm{CATG} & & \mathrm{G}-5' \end{array}, \quad \text{and}$$

$$Dpn\mathrm{I}\ (ga,\ 1,\ tc)\text{:}\ \begin{array}{lcl} 5'-\mathrm{GA} & \downarrow & \mathrm{TC}-3' \\ 3'-\mathrm{CT} & \uparrow & \mathrm{AG}-5' \end{array}.$$

This research is done on DNA splicing system with 5' overhang palindromic restriction enzymes. Palindromic enzyme is one of the enzymes types in dsDNA where reading the single strand 5' to 3' forward matches with 3' to 5' from backward [1]. For instance, the enzyme *Age*I is palindromic since the single strand 5'-ACCGGT-3' is exactly the same with the single strand 3'-TGGCCA-5' from backward. The name and sequence for every restriction enzyme that have been used in this research are taken from [17].

3 Methodology

Splicing languages can be denoted using the notation of regular expressions in formal language theory. Formal language theory is a theory on the general characteristics of programming languages in computer science [18]. A formal language consists of a set of finite strings of symbols anda set of alphabets in which the combinations of the symbols are called languages [1]. A set of strings of concatenating zero or more symbols from an alphabet A is denoted as $A*$. The empty string is also contained in the set where the empty string is indicated as λ or 1; while a set of strings of symbols without the empty string is known as A^{+}. The notation of regular expressions consists of a combination of strings of symbols from the alphabet, parentheses, and the operators $+$, $\cdot$ and * [18]. The symbols $+$, $\cdot$ and * denote union, concatenation and star-closure respectively.

Next, the definitions of splicing system and splicing language are stated.

Definition 1 [1] (Splicing System and Splicing Language)
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 to $ucxfq$ and $pexdv$ 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 $L=L(S)$.

Here, a splicing system involving certain DNA molecules and restriction enzyme is discussed. A restriction enzyme cuts DNA molecule in a very specific way and the DNA molecules will paste together with the existence of a ligase [1]. DNA molecules are taken from the sub sequences or pattern in protein or nucleotide sequences which are also called as initial string in splicing systems [19]. An example of a splicing system involving the restriction enzyme *Eco*RI is given in Example 1.

Example 1 Suppose that $S=(A, I, B, C)$ is a splicing system in which

$$A=\left\{ \begin{array}{c} \mathrm{A, C, G, T} \\ \mathrm{T\ G\ C\ A} \end{array} \right\}$$

is the set of DNA alphabets,

$$I = \left\{ \begin{matrix} \text{GAATTCTCTGTAAT} \\ \text{CTTAAGAGACATTA} \end{matrix} \right\}$$

is the set consisting of an initial string of molecules,

$$B = \left\{ \begin{matrix} \text{G, AATT, C} \\ \text{C TTAA A} \end{matrix} \right\}$$

is the set of cleavage pattern for the enzyme *Eco*RI and set C is the null (Ø) set.

The initial string *gaattctctgtaat* with the cutting sites of the enzyme *Eco*RI is shown in the following:

$$\begin{matrix} 5'^{-}\text{G} \downarrow \text{AATT} \quad \text{CTCTGTAAT}^{-}3' \\ 3'^{-}\text{C} \quad \text{TTAA} \uparrow \text{GAGACATTA}^{-}5', \end{matrix}$$

or written 180 degree wise,

$$\begin{matrix} 5'^{-}\text{ATTACAGAG} \downarrow \text{AATT} \quad \text{C}^{-}3' \\ 3'^{-}\text{TAATGTCTC} \quad \text{TTAA} \uparrow \text{G}^{-}5'. \end{matrix}$$

There should be four different molecules since one form is rotated through 180 degrees. When the enzyme *Eco*RI is added to the initial string, the result is as the following:

$$\begin{matrix} 5' - \text{G} - 3' \\ 3' - \text{CTTAA} - 5' \end{matrix}$$

combines with

$$\begin{matrix} 5'\text{–AATTC–}3' \\ \quad\quad 3'\text{–G–}5' \end{matrix}$$

which gives the following molecule

$$\begin{matrix} 5'\text{–GAATTC} - 3' \\ 3'\text{–CTTAAG} - 5' \end{matrix} \quad .$$

Furthermore,

$$\begin{matrix} 5' - \text{AATTCTCTGTAAT–}3' \\ \quad\quad 3' - \text{GAGACATTA–}5' \end{matrix}$$

combines with

$$\begin{matrix} 5'\text{–ATTACAGAG–}3' \\ 3'\text{–TAATGTCTCTTAA} - 5' \end{matrix}$$

gives the following molecule

$$\begin{matrix} 5'\text{–ATTACAGAGAATTCTCTGTAAT} - 3' \\ 3'\text{–TAATGTCTCTTAAGAGACATTA} - 5' \end{matrix} \quad .$$

Thus, the new molecules are shown in the following:

$$\begin{matrix} 5'\text{–GAATTC} - 3' \\ 3'\text{–CTTAAG} - 5' \end{matrix}$$

and

$$\begin{array}{l} 5'\text{–ATTACAGAGAATTCTCTGTAAT}-3' \\ 3'\text{–TAATGTCTCTTAAGAGACATTA}-5' \end{array} \; .$$

Therefore, the splicing languages resulting from this splicing system are:

$$\begin{array}{l} 5'\text{–GAATTCTCTGTAAT}-3' \\ 3'\text{–CTTAAGAGACATTA}-5' \end{array} \; ,$$

$$\begin{array}{l} 5'\text{–ATTACAGAGAATTC}-3' \\ 3'\text{–TAATGTCTCTTAAG}-5' \end{array} \; ,$$

$$\begin{array}{l} 5'\text{–GAATTC}-3' \\ 3'\text{–CTTAAG}-5' \end{array}$$

and

$$\begin{array}{l} 5'\text{–ATTACAGAGAATTCTCTGTAAT}-3' \\ 3'\text{–TAATGTCTCTTAAGAGACATTA}-5' \end{array} \; .$$

4 Results and Discussion

The concept of palindromic restriction enzymes in DNA splicing systems has beenused in this research. Two theorems to generalise DNA splicing systems with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites are presented respectively.

The generalisation of resulting splicing languages in DNA splicing system with one palindromic restriction enzyme and one cutting site is presented in Theorem 1.

Theorem 1 *Given* $S = (A, I, B, C)$ *is a splicing system in which*

$$A = \left\{ \begin{array}{c} \mathrm{A, C, G, T} \\ \mathrm{T \;\; G \;\; C \;\; A} \end{array} \right\}$$

is the set of DNA alphabets,

$$I = \left\{ \begin{array}{c} N_1N_1 \ldots N_1X_1YX_2N_2N_2 \ldots N_2 \\ N_1'N_1' \ldots N_1'X_1'Y'X_2'N_2'N_2' \ldots N_2' \end{array} \right\}$$

is the set consisting of an initial string with one cutting site of a palindromic restriction enzyme

$$\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} .$$

Set

$$B = \left\{ \begin{array}{ccc} X_1 & ,Y, & X_2 \\ X_1' & Y' & X_2' \end{array} \right\}$$

is the set of cleavage pattern for restriction enzyme and set C *is the null (Ø) set, the resulting splicing language is*

$$\left(\begin{array}{c} N_1N_1 \ldots N_1 \\ N_1'N_1' \ldots N_1' \end{array} + \begin{array}{c} N_2'N_2' \ldots N_2' \\ N_2N_2 \ldots N_2 \end{array} \right) \begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} \left(\begin{array}{c} N_2N_2 \ldots N_2 \\ N_2'N_2' \ldots N_2' \end{array} + \begin{array}{c} N_1'N_1' \ldots N_1' \\ N_1N_1 \ldots N_1 \end{array} \right)$$

where $\begin{array}{c} Y \\ Y' \end{array}$ *is the crossing site,* N_1*,* N_2*,* X_1*,* Y *and* X_2 *are complementaries for* N_1'*,* N_2'*,* X_1'*,* Y' *and* X_2' *respectively,* $\begin{array}{c} X_1 \\ X_1' \end{array}$*,* $\begin{array}{c} Y \\ Y' \end{array}$*,* $\begin{array}{c} X_2 \\ X_2' \end{array}$*,* $\begin{array}{c} N_1 \\ N_1' \end{array}$ *and* $\begin{array}{c} N_2 \\ N_2' \end{array}$ *denote arbitrary DNA alphabet(s), and*

$$\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} \notin \left\{ \begin{array}{c} N_1N_1\ldots N_1 \\ N_1'N_1'\ldots N_1' \end{array}, \begin{array}{c} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array} \right\}.$$

Proof Suppose the restriction enzyme is palindromic, so the base sequence of enzyme reads the same backwards and forwards:

$$\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} = \begin{array}{c} X_2'Y'X_1' \\ X_2YX_1 \end{array}.$$

Then $\begin{array}{c} X_1 \\ X_1' \end{array} = \begin{array}{c} X_2' \\ X_2 \end{array}$, $\begin{array}{c} Y \\ Y' \end{array} = \begin{array}{c} Y' \\ Y \end{array}$ and $\begin{array}{c} X_2 \\ X_2' \end{array} = \begin{array}{c} X_1' \\ X_1 \end{array}$.

The initial string $\begin{array}{c} N_1N_1\ldots N_1X_1YX_2N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1'Y'X_2'N_2'N_2'\ldots N_2' \end{array}$ with the cutting site of the enzyme $\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array}$ is shown in the following:

$$\begin{array}{lclcl} N_1N_1\ldots N_1X_1 & \downarrow & Y & & X_2N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1' & & Y' & \uparrow & X_2'N_2'N_2'\ldots N_2' \end{array}. \tag{1}$$

Hence, the initial string is cut into two parts:

$$\begin{array}{ll} N_1N_1\ldots N_1X_1 & \\ N_1'N_1'\ldots N_1'X_1' & Y' \end{array} \tag{2}$$

and

$$\begin{array}{l} YX_2N_2N_2\ldots N_2 \\ \quad X_2'N_2'N_2'\ldots N_2' \end{array}. \tag{3}$$

Besides that, the initial string $\begin{array}{c} N_1N_1\ldots N_1X_1YX_2N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1'Y'X_2'N_2'N_2'\ldots N_2' \end{array}$ can read 180 degree wise as

$$\begin{array}{c} N_2'N_2'\ldots N_2'X_2'Y'X_1'N_1'N_1'\ldots N_1' \\ N_2N_2\ldots N_2X_2YX_1N_1N_1\ldots N_1 \end{array}. \tag{4}$$

Since $\begin{array}{c} X_1 \\ X_1' \end{array} = \begin{array}{c} X_2' \\ X_2 \end{array}$, $\begin{array}{c} Y \\ Y' \end{array} = \begin{array}{c} Y' \\ Y \end{array}$ and $\begin{array}{c} X_2 \\ X_2' \end{array} = \begin{array}{c} X_1' \\ X_1 \end{array}$, then (4) becomes

$$\begin{array}{lclcl} N_2'N_2'\ldots N_2'X_1 & \downarrow & Y & & X_2N_1'N_1'\ldots N_1' \\ N_2N_2\ldots N_2X_1' & & Y' & \uparrow & X_2'N_1N_1\ldots N_1 \end{array} \tag{5}$$

with the cutting site of the enzyme $\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array}$. The results of cutting (5) into two parts are

$$\begin{array}{ll} N_2'N_2'\ldots N_2'X_1 & \\ N_2N_2\ldots N_2X_1' & Y' \end{array} \tag{6}$$

and

$$\begin{array}{c} YX_2N_1'N_1'\dots N_1' \\ X_2'N_1N_1\dots N_1 \end{array} . \tag{7}$$

When the enzyme $\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array}$ is added to the initial string, (2) combines with (7) gives the following string

$$\begin{array}{c} N_1N_1\dots N_1X_1YX_2N_1'N_1'\dots N_1' \\ N_1'N_1'\dots N_1'X_1'Y'X_2'N_1N_1\dots N_1 \end{array} . \tag{8}$$

Furthermore, (3) joins with (6) gives the following string

$$\begin{array}{c} N_2'N_2'\dots N_2'X_1YX_2N_2N_2\dots N_2 \\ N_2N_2\dots N_2X_1'Y'X_2'N_2'N_2'\dots N_2' \end{array} . \tag{9}$$

From (1), (5), (8) and (9), the resulting splicing language is

$$\left(\begin{array}{c} N_1N_1\dots N_1 \\ N_1'N_1'\dots N_1' \end{array} + \begin{array}{c} N_2'N_2'\dots N_2' \\ N_2N_2\dots N_2 \end{array} \right) \begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} \left(\begin{array}{c} N_2N_2\dots N_2 \\ N_2'N_2'\dots N_2' \end{array} + \begin{array}{c} N_1'N_1'\dots N_1' \\ N_1N_1\dots N_1 \end{array} \right) . \qquad \square$$

Theorem 2 is presented for generalisation of resulting splicing languages in DNA splicing system with one palindromic restriction enzyme and two non-overlapping cutting sites.

Theorem 2 *Given* $S = (A, I, B, C)$ *is a splicing system in which*

$$A = \left\{ \begin{array}{c} \mathrm{A, C, G, T} \\ \mathrm{T\ \ G\ \ C\ \ A} \end{array} \right\}$$

is the set of DNA alphabets,

$$I = \left\{ \begin{array}{c} N_1N_1\dots N_1X_1\,Y\,X_2M\,M\dots MX_1\,Y\,X_2\,N_2N_2\dots N_2 \\ N_1'N_1'\dots N_1'X_1'Y'X_2'M'M'\dots M'X_1'Y'X_2'N_2'N_2'\dots N_2' \end{array} \right\}$$

is the set consisting of an initial string with two non-overlapping cutting sites of a palindromic restriction enzyme $\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array}$*, set* $B = \left\{ \begin{array}{ccc} X_1 & ,Y, & X_2 \\ X_1' & Y' & X_2' \end{array} \right\}$ *is the set of cleavage pattern for restriction enzyme and set* C *is the null* $(\emptyset)$ *set, the resulting string is*

$$\left(\begin{array}{c} N_1N_1\dots N_1 \\ N_1'N_1'\dots N_1' \end{array} + \begin{array}{c} N_2'N_2'\dots N_2' \\ N_2N_2\dots N_2 \end{array} \right) \begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} \left(\left(\begin{array}{c} M\,M\dots M \\ M'M'\dots M' \end{array} + \begin{array}{c} M'M'\dots M' \\ M\,M\dots M \end{array} \right) \begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} \right)^{n-1} \left(\begin{array}{c} N_2N_2\dots N_2 \\ N_2'N_2'\dots N_2' \end{array} + \begin{array}{c} N_1'N_1'\dots N_1' \\ N_1N_1\dots N_1 \end{array} \right)$$

where $\begin{array}{c} Y \\ Y' \end{array}$ *is the crossing site,* N_1*,* M*,* N_2*,* X_1*,* Y *and* X_2 *are complementaries for* N_1'*,* M'*,* N_2'*,* X_1'*,* Y' *and* X_2' *respectively,* $\begin{array}{c} X_1 \\ X_1' \end{array}$*,* $\begin{array}{c} Y \\ Y' \end{array}$*,* $\begin{array}{c} X_2 \\ X_2' \end{array}$*,* $\begin{array}{c} N_1 \\ N_1' \end{array}$*,* $\begin{array}{c} M \\ M' \end{array}$ *and* $\begin{array}{c} N_2 \\ N_2' \end{array}$ *denote arbitrary DNA alphabet(s), and* $\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} \notin \left\{ \begin{array}{c} N_1N_1\dots N_1 \\ N_1'N_1'\dots N_1' \end{array}, \begin{array}{c} M\,M\dots M \\ M'M'\dots M' \end{array}, \begin{array}{c} N_2N_2\dots N_2 \\ N_2'N_2'\dots N_2' \end{array} \right\}$.

Proof Suppose the restriction enzyme is palindromic, so the base sequence of enzyme reads the same backwards and forwards:

$$\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix} = \begin{matrix} X_2'Y'X_1' \\ X_2YX_1 \end{matrix} \quad .$$

Then $\begin{matrix} X_1 \\ X_1' \end{matrix} = \begin{matrix} X_2' \\ X_2 \end{matrix}$, $\begin{matrix} Y \\ Y' \end{matrix} = \begin{matrix} Y' \\ Y \end{matrix}$ and $\begin{matrix} X_2 \\ X_2' \end{matrix} = \begin{matrix} X_1' \\ X_1 \end{matrix}$.

The initial string, $\begin{matrix} N_1N_1\ldots N_1X_1\,Y\,X_2MM\ldots MX_1\,Y\,X_2\,N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1'Y'X_2'M'M'\ldots M'X_1'Y'X_2'N_2'N_2'\ldots N_2' \end{matrix}$ with the cutting site of the enzyme $\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix}$ is shown in the following:

$$\begin{array}{lcccccl} N_1N_1\ldots N_1X_1 & \downarrow & Y & & X_2MM\ldots MX_1 & Y & X_2N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1' & & Y' & \uparrow & X_2'M'M'\ldots M'X_1' & Y' & X_2'N_2'N_2'\ldots N_2' \end{array} \tag{10}$$

for the first cutting site and

$$\begin{array}{lcccccl} N_1N_1\ldots N_1X_1 & Y & X_2MM\ldots MX_1 & \downarrow & Y & & X_2N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1' & Y' & X_2'M'M'\ldots M'X_1' & & Y' & \downarrow & X_2'N_2'N_2'\ldots N_2' \end{array} \tag{11}$$

for the second cutting site. The initial string can be written 180 degree wise as

$$\begin{matrix} N_2'N_2'\ldots N_2'X_2'Y'X_1'M'M'\ldots M'X_2'Y'X_1'N_1'N_1'\ldots N_1' \\ N_2N_2\ldots N_2X_2\,Y\,X_1MM\ldots MX_2\,Y\,X_1N_1N_1\ldots N_1 \end{matrix} \quad . \tag{12}$$

Since $\begin{matrix} X_1 \\ X_1' \end{matrix} = \begin{matrix} X_2' \\ X_2 \end{matrix}$, $\begin{matrix} Y \\ Y' \end{matrix} = \begin{matrix} Y' \\ Y \end{matrix}$ and $\begin{matrix} X_2 \\ X_2' \end{matrix} = \begin{matrix} X_1' \\ X_1 \end{matrix}$, then the first and second cutting sites of (12) are shown respectively in the following:

$$\begin{array}{lcccccl} N_2'N_2'\ldots N_2'X_1 & \downarrow & Y & & X_2M'M'\ldots M'X_1 & Y & X_2N_1'N_1'\ldots N_1' \\ N_2N_2\ldots N_2X_1' & & Y' & \uparrow & X_2'MM\ldots MX_1' & Y' & X_2'N_1N_1\ldots N_1 \end{array} \tag{13}$$

and

$$\begin{array}{lcccccl} N_2'N_2'\ldots N_2'X_1 & Y & X_2M'M'\ldots M'X_1 & \downarrow & Y & & X_2N_1'N_1'\ldots N_1' \\ N_2N_2\ldots N_2X_1' & Y' & X_2'MM\ldots MX_1' & & Y' & \uparrow & X_2'N_1N_1\ldots N_1 \end{array} \quad . \tag{14}$$

When the enzyme $\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix}$ is added to the initial string, (10) combines with (11) gives

$$\begin{array}{lcccl} N_1N_1\ldots N_1X_1 & \downarrow & Y & & X_2N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1' & & Y' & \uparrow & X_2'N_2'N_2'\ldots N_2' \end{array} \quad . \tag{15}$$

Furthermore, (13) combines with (14) gives the new string

$$\begin{array}{lcccl} N_2'N_2'\ldots N_2'X_1 & \downarrow & Y & & X_2N_1'N_1'\ldots N_1' \\ N_2N_2\ldots N_2X_1' & & Y' & \uparrow & X_2'N_1N_1\ldots N_1 \end{array} \quad . \tag{16}$$

The results of the combination of (10) with (13) and (14) are shown in the following:

$$\begin{array}{c} N_1N_1\ldots N_1X_1Y'X_1'M'M'\ldots M'X_1{\downarrow}YX_2N_1'N_1'\ldots N_1' \\ N_1'N_1'\ldots N_1'X_1'YX_1MM\ldots MX_1'Y'{\uparrow}X_2'N_1N_1\ldots N_1 \end{array} \quad , \tag{17}$$

$$\begin{array}{c} N_2'N_2'\ldots N_2'X_1\,Y\,X_2M\,M\ldots M\,X_1\downarrow Y\quad X_2N_2N_2\ldots N_2 \\ N_2N_2\ldots N_2X_1'Y'X_2'M'M'\ldots M'X_1'\quad Y'\uparrow X_2'N_2'N_2'\ldots N_2' \end{array} \quad , \tag{18}$$

$$\begin{array}{c} N_1N_1\ldots N_1X_1YX_2N_1'N_1'\ldots N_1' \\ N_1'N_1'\ldots N_1'X_1'Y'X_2'N_1N_1\ldots N_1 \end{array} \quad . \tag{19}$$

The results of the combination of (11) with (13) and (14) are:

$$\begin{array}{c} N_2'N_2'\ldots N_2'X_1YX_2N_2N_2\ldots N_2 \\ N_2N_2\ldots N_2X_1'Y'X_2'N_2'N_2'\ldots N_2' \end{array} \quad , \tag{20}$$

$$\begin{array}{llllllll} N_2'N_2'\ldots N_2'X_1 & \downarrow & Y & & X_2M'M'\ldots M'X_1 & Y & X_2N_2N_2\ldots N_2 & \\ N_2N_2\ldots N_2X_1' & & Y' & \uparrow & X_2'MM\ldots MX_1' & Y' & X_2'N_2'N_2'\ldots N_2' & , \end{array} \tag{21}$$

$$\begin{array}{llllllll} N_1N_1\ldots N_1X_1 & \downarrow & Y & & X_2\,M\,M\ldots M\,X_1 & Y & X_2N_1'N_1'\ldots N_1' & \\ N_1'N_1'\ldots N_1'X_1' & & Y' & \uparrow & X_2'M'M'\ldots M'X_1' & Y' & X_2'N_1N_1\ldots N_1 & . \end{array} \tag{22}$$

Moreover, when (15) and (16) combine with (21) and (22) respectively, the other new strings arise:

$$\begin{array}{c} N_1N_1\ldots N_1X_1YX_2M'M'\ldots M'X_1YX_2N_2N_2\ldots N_2 \\ N_1'N_1'\ldots N_1'X_1'Y'X_2'M\,M\ldots MX_1'Y'X_2'N_2'N_2'\ldots N_2' \end{array} \quad ,$$

$$\begin{array}{c} N_2'N_2'\ldots N_2'X_1\,Y\,X_2\,M\,M\ldots M\,X_1YX_2N_1'N_1'\ldots N_1' \\ N_2N_2\ldots N_2X_1'Y'X_2'M'M'\ldots M'X_1'Y'X_2'N_1'N_1\ldots N_1 \end{array} \quad .$$

By using induction, this theorem can be proved. For $n = 1$, the strings are stated in (15), (16), (19) and (20).

Next, let $n = k$

$$\left(\begin{array}{c} N_1N_1\ldots N_1 \\ N_1'N_1'\ldots N_1' \end{array} + \begin{array}{c} N_2'N_2'\ldots N_2' \\ N_2N_2\ldots N_2 \end{array}\right) \begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array} \left(\left(\begin{array}{c} M\,M\ldots M \\ M'M'\ldots M' \end{array} + \begin{array}{c} M'M'\ldots M' \\ M\,M\ldots M \end{array}\right) \begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array}\right)^{k-1}$$
$$\left(\begin{array}{c} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array} + \begin{array}{c} N_1'N_1'\ldots N_1' \\ N_1N_1\ldots N_1 \end{array}\right),\ k\epsilon Z^+ \tag{23}$$

The following strings are among the strings in (23):

$$\begin{array}{ll} N_1N_1\ldots N_1 & X_1\downarrow Y\quad X_2 \\ N_1'N_1'\ldots N_1' & X_1'\quad Y'\uparrow X_2' \end{array} \left(\begin{array}{ll} M\,M\ldots M & X_1YX_2 \\ M'M'\ldots M' & X_1'Y'X_2' \end{array}\right)^{k-1} \begin{array}{l} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array} \quad , \tag{24}$$

$$\begin{array}{ll} N_1N_1\ldots N_1 & X_1\downarrow Y\quad X_2 \\ N_1'N_1'\ldots N_1' & X_1'Y'\uparrow X_2' \end{array} \left(\begin{array}{ll} M\,M\ldots M & X_1YX_2 \\ M'M'\ldots M' & X_1'Y'X_2' \end{array}\right)^{k-1} \begin{array}{l} N_1'N_1'\ldots N_1' \\ N_1N_1\ldots N_1 \end{array} \quad , \tag{25}$$

$$\begin{array}{lll} N_1N_1\ldots N_1 & X_1\downarrow Y & X_2 \\ N_1'N_1'\ldots N_1' & X_1' \ \ Y'\uparrow & X_2' \end{array} \begin{pmatrix} M'M'\ldots M' & X_1YX_2 \\ M\,M\ \ldots M & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array}, \tag{26}$$

$$\begin{array}{lll} N_1N_1\ldots N_1 & X_1\downarrow Y & X_2 \\ N_1'N_1'\ldots N_1' & X_1' \ \ Y'\uparrow & X_2' \end{array} \begin{pmatrix} M'M'\ldots M' & X_1YX_2 \\ M\,M\ \ldots M & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_1'N_1'\ldots N_1' \\ N_1N_1\ldots N_1 \end{array}, \tag{27}$$

$$\begin{array}{lll} N_2'N_2'\ldots N_2' & X_1\downarrow Y & X_2 \\ N_2N_2\ldots N_2 & X_1'Y'\uparrow & X_2' \end{array} \begin{pmatrix} M\,M\ldots\ M & X_1YX_2 \\ M'M'\ldots M' & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array}, \tag{28}$$

$$\begin{array}{lll} N_2'N_2'\ldots N_2' & X_1\downarrow Y & X_2 \\ N_2N_2\ldots N_2 & X_1'Y'\uparrow & X_2' \end{array} \begin{pmatrix} M\,M\ldots\ M & X_1YX_2 \\ M'M'\ldots M' & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_1'N_1'\ldots N_1' \\ N_1N_1\ldots N_1 \end{array}, \tag{29}$$

$$\begin{array}{lll} N_2'N_2'\ldots N_2' & X_1\downarrow Y & X_2 \\ N_2N_2\ldots N_2 & X_1' \ \ Y'\uparrow & X_2' \end{array} \begin{pmatrix} M'M'\ldots M' & X_1YX_2 \\ M\,M\ \ldots M & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array}, \tag{30}$$

$$\begin{array}{lll} N_2'N_2'\ldots N_2' & X_1\downarrow Y & X_2 \\ N_2N_2\ldots N_2 & X_1'Y'\uparrow & X_2' \end{array} \begin{pmatrix} M'M'\ldots M' & X_1YX_2 \\ M\,M\ \ldots M & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_1'N_1'\ldots N_1' \\ N_1N_1\ldots N_1 \end{array}. \tag{31}$$

Then, the above resulting strings combine with (11), (17), (18) and (14) when the enzyme $\begin{array}{c} X_1YX_2 \\ X_1'Y'X_2' \end{array}$ is added. Hence, the other recombinations can be shown through four cases.

Case 1: The string (11) combines with (24) and (25) which produces new strings

$$\begin{array}{lllll} N_1N_1\ldots N_1X_1 & Y & X_2\,M\,M\ldots M\,X_1 & Y & X_2 \\ N_1'N_1'\ldots N_1'X_1' & Y' & X_2'M'M'\ldots M'X_1' & Y' & X_2' \end{array} \begin{pmatrix} M\,M\ldots\ M & X_1YX_2 \\ M'M'\ldots M' & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array} \tag{32}$$

and

$$\begin{array}{lllll} N_1N_1\ldots N_1X_1 & Y & X_2\,M\,M\ldots M\,X_1 & Y & X_2 \\ N_1'N_1'\ldots N_1'X_1' & Y' & X_2'M'M'\ldots M'X_1' & Y' & X_2' \end{array} \begin{pmatrix} M\,M\ldots\ M & X_1YX_2 \\ M'M'\ldots M' & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_1'N_1'\ldots N_1' \\ N_1N_1\ldots N_1 \end{array}. \tag{33}$$

Case 2: The string (17) combines with (26) and (27) produces new strings

$$\begin{array}{lllll} N_1N_1\ldots N_1X_1 & Y & X_2M'M'\ldots M'X_1 & Y & X_2 \\ N_1'N_1'\ldots N_1'X_1' & Y' & X_2'M\,M\ldots\ M\,X_1' & Y' & X_2' \end{array} \begin{pmatrix} M'M'\ldots M' & X_1YX_2 \\ M\,M\ \ldots M & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array} \tag{34}$$

and

$$\begin{array}{lllll} N_1N_1\ldots N_1X_1 & Y & X_2M'M'\ldots M'X_1 & Y & X_2 \\ N_1'N_1'\ldots N_1'X_1' & Y' & X_2'M\,M\ldots\ M\,X_1' & Y' & X_2' \end{array} \begin{pmatrix} M'M'\ldots M' & X_1Y?{-}_2 \\ M\,M\ \ldots M & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_1'N_1'\ldots N_1' \\ N_1N_1\ldots N_1 \end{array}. \tag{35}$$

Case 3: The string (18) combines with (28) and (29) produces new strings

$$\begin{array}{llll} N_2'N_2'\ldots N_2'X_1 & Y & X_2\,M\,M\ldots M\,X_1 & Y\,X_2 \\ N_2N_2\ldots N_2X_1' & Y' & X_2'M'M'\ldots M'X_1' & Y'X_2' \end{array} \begin{pmatrix} M\,M\ldots\ M & X_1YX_2 \\ M'M'\ldots M' & X_1'Y'X_2' \end{pmatrix}^{k-1} \begin{array}{l} N_2N_2\ldots N_2 \\ N_2'N_2'\ldots N_2' \end{array} \tag{36}$$

and

$$\begin{array}{cccccc} N'_2N'_2\ldots N'_2X_1 & Y & X_2\,M\,M\ldots M\,X_1 & Y & X_2 \\ N_2N_2\ldots N_2X'_1 & Y' & X'_2M'M'\ldots M'X'_1 & Y' & X'_2 \end{array} \left(\begin{array}{cc} M\,M\ldots M & X_1YX_2 \\ M'M'\ldots M' & X'_1Y'X'_2 \end{array}\right)^{k-1} \begin{array}{c} N'_1N'_1\ldots N'_1 \\ N_1N_1\ldots N_1 \end{array}. \tag{37}$$

Case 4: The string (14) combines with (30) and (31) produces new strings

$$\begin{array}{cccccc} N'_2N'_2\ldots N'_2X_1 & Y & X_2M'M'\ldots M'X_1 & Y & X_2 \\ N_2N_2\ldots N_2X'_1 & Y' & X'_2M\,M\ldots M\,X'_1 & Y' & X'_2 \end{array} \left(\begin{array}{cc} M'M'\ldots M' & X_1YX_2 \\ M\,M\ldots M & X'_1Y'X'_2 \end{array}\right)^{k-1} \begin{array}{c} N_2N_2\ldots N_2 \\ N'_2N'_2\ldots N'_2 \end{array} \tag{38}$$

and

$$\begin{array}{cccccc} N'_2N'_2\ldots N'_2X_1 & Y & X_2M'M'\ldots M'X_1 & Y & X_2 \\ N_2N_2\ldots N_2X'_1 & Y' & X'_2M\,M\ldots M\,X'_1 & Y' & X'_2 \end{array} \left(\begin{array}{cc} M'M'\ldots M' & X_1YX_2 \\ M\,M\ldots M & X'_1Y'X'_2 \end{array}\right)^{k-1} \begin{array}{c} N'_1N'_1\ldots N'_1 \\ N_1N_1\ldots N_1 \end{array}. \tag{39}$$

By simplifying strings (32), (33), (34), (35), (36), (37), (38) and (39), the resulting strings are:

$$\begin{array}{cc} N_1N_1\ldots N_1 & X_1YX_2 \\ N'_1N'_1\ldots N'_1 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M\,M\ldots M & X_1YX_2 \\ M'M'\ldots M' & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N_2N_2\ldots N_2 \\ N'_2N'_2\ldots N'_2 \end{array},$$

$$\begin{array}{cc} N_1N_1\ldots N_1 & X_1YX_2 \\ N'_1N'_1\ldots N'_1 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M\,M\ldots M & X_1YX_2 \\ M'M'\ldots M' & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N'_1N'_1\ldots N'_1 \\ N_1N_1\ldots N_1 \end{array},$$

$$\begin{array}{cc} N_1N_1\ldots N_1 & X_1YX_2 \\ N'_1N'_1\ldots N'_1 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M'M'\ldots M' & X_1YX_2 \\ M\,M\ldots M & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N_2N_2\ldots N_2 \\ N'_2N'_2\ldots N'_2 \end{array},$$

$$\begin{array}{cc} N_1N_1\ldots N_1 & X_1YX_2 \\ N'_1N'_1\ldots N'_1 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M'M'\ldots M' & X_1YX_2 \\ M\,M\ldots M & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N'_1N'_1\ldots N'_1 \\ N_1N_1\ldots N_1 \end{array},$$

$$\begin{array}{cc} N'_2N'_2\ldots N'_2 & X_1YX_2 \\ N_2N_2\ldots N_2 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M\,M\ldots M & X_1YX_2 \\ M'M'\ldots M' & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N_2N_2\ldots N_2 \\ N'_2N'_2\ldots N'_2 \end{array},$$

$$\begin{array}{cc} N'_2N'_2\ldots N'_2 & X_1YX_2 \\ N_2N_2\ldots N_2 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M\,M\ldots M & X_1YX_2 \\ M'M'\ldots M' & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N'_1N'_1\ldots N'_1 \\ N_1N_1\ldots N_1 \end{array},$$

$$\begin{array}{cc} N'_2N'_2\ldots N'_2 & X_1YX_2 \\ N_2N_2\ldots N_2 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M'M'\ldots M' & X_1YX_2 \\ M\,M\ldots M & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N_2N_2\ldots N_2 \\ N'_2N'_2\ldots N'_2 \end{array},$$

$$\begin{array}{cc} N'_2N'_2\ldots N'_2 & X_1YX_2 \\ N_2N_2\ldots N_2 & X'_1Y'X'_2 \end{array} \left(\begin{array}{cc} M'M'\ldots M' & X_1YX_2 \\ M\,M\ldots M & X'_1Y'X'_2 \end{array}\right)^{(k+1)-1} \begin{array}{c} N'_1N'_1\ldots N'_1 \\ N_1N_1\ldots N_1 \end{array}.$$

Therefore, the resulting splicing language can be summarised where $n = k + 1$:

$$\left(\begin{array}{c} N_1N_1\ldots N_1 \\ N'_1N'_1\ldots N'_1 \end{array} + \begin{array}{c} N'_2N'_2\ldots N'_2 \\ N_2N_2\ldots N_2 \end{array}\right) \begin{array}{c} X_1YX_2 \\ X'_1Y'X'_2 \end{array}$$

$$\left(\left(\begin{array}{c} M\,M\ldots M \\ M'M'\ldots M' \end{array} + \begin{array}{c} M'M'\ldots M' \\ M\,M\ldots M \end{array}\right) \begin{array}{c} X_1YX_2 \\ X'_1Y'X'_2 \end{array}\right)^{(k+1)-1}$$

$$\left(\begin{array}{c} N_2N_2\ldots N_2 \\ N'_2N'_2\ldots N'_2 \end{array} + \begin{array}{c} N'_1N'_1\ldots N'_1 \\ N_1N_1\ldots N_1 \end{array}\right), \; k\epsilon Z^+.$$

Hence, Theorem 2 is proved. □

5 Conclusion

In this research, generalisations of DNA splicing systems with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites are presented in Theorem 1 and Theorem 2 respectively. These theorems are proved by using direct method and induction method respectively. The generalisations hence display the resulting languages of splicing systems as regular expressions. In other words, splicing languages from DNA splicing system with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites can be obtained without having to manually compute them when different initial strings and restriction enzymes are used.

Acknowledgement

The first author would like to thank the Ministry of Higher Education (MOHE) and Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the financial funding through Research University Grant Vote No. 13H18. The second author would also like to thank Perbadanan Insurans Deposit Malaysia Scholarship for funding her study at UTM.

References

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

[2] Head, T. Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors. *Bulletin of Mathematical Biology*. 1987. 49(6): 737-759.

[3] Sarmin, N. H., Ahmad, M. A. and Fong, W. H. On the mathematical modeling of splicing systems in DNA computing. *Proceedings on the 7th SEAMS UGM International Conference on Mathematics and Its Applications 2015*. August 18-21, 2015. Yogyakarta, Indonesia. 2015. 119-125.

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

[5] Paun, G. On the splicing operation. *Discrete Applied Mathematics*. 1996. 70(1): 57-79.

[6] Pixton, D. Regularity of splicing languages. *Discrete Applied Mathematics*. 1996. 69(1): 101-124.

[7] Bonizzoni, P., Ferretti, C., Mauri, G. and Zizza, R. Separating some splicing models. *Information Processing Letters*. 2001. 79(6): 255-259.

[8] Laun, E. G. *Constant and Splicing Systems*. Ph.D. Thesis. State University of New York at Binghamton; 1999.

[9] Yusof, Y., Sarmin, N. H., Fong,W. H., Goode, T. E. and Ahmad, M. A. An analysis of four variants of splicing system. *Proceeding of the 20th National Symposium on Mathematical Sciences - Research in Mathematical Sciences: A Catalyst for Creativity and Innovation,*

SKSM 2012. December 18-20, 2012. Putrajaya, Malaysia: AIP Conference Proceedings. 2013. 888-895.

[10] Yusof, Y., Sarmin, N. H., Goode, T. E., Mahmud, M. and Fong, W. H. Hierarchy of certain types of DNA splicing systems. I*nternational Journal of Modern Physics: Conference Series*. 2012. 9(1). 271-277.

[11] Yusof, Y., Sarmin, N. H., Goode, T. E., Mahmud, M. and Fong, W. H. An extension of DNA splicing system. *Proceeding of the 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications*. September 27-29, 2011. Penang, Malaysia: IEEE. 2011. 246-248.

[12] Karimi, F., Sarmin, N. H. and Fong, W. H. The characterizations of different splicing systems. *International Journal of Modern Physics: Conference Series*. 2012. 9(1): 89-94.

[13] Mateescu, A, Paun, G., Rozenbrg, G. and Salomaa, A. Simple splicing systems. *Discrete Applied Mathematics*. 1998. 84(1): 145-163.

[14] Yusof, Y., Sarmin, N. H., Fong, W. H. and Karimi, F. Some relations on different types of splicing systems. *Journal of Fundamental Sciences*. 2010. 6(2): 143-147.

[15] Palmer, T. and Bonner, P. L. *Enzymes: Biochemistry, Biotechnology, Clinical Chemistry*. 2nd. ed. UK: Woodhead Publishing Limited. 2007.

[16] Kim, S. M. Computational modelling for genetic splicing systems. *SIAM J. Comput*. 1997. 26(5): 1284-1309.

[17] New England Biolabs, *Enzyme Finder*. Available from: <https://www.neb.com/tools-and-resources/interactive-tools/enzyme-finder>. [11 June 2017].

[18] Linz, P. *An Introduction of Formal Language and Automata*. 4th. ed. USA: John and Barlett Publisher. 2006.

[19] National Center for Biotechnology Information, *Expressed Sequence Tag*. Available from: <https://www.ncbi.nlm.nih.gov/nucest>. [11 June 2017].