DNA Splicing Systems with at Most Two Cutting Sites of a Non-Palindromic Restriction Enzyme

Nurul Izzaty Ismail, Wan Heng Fong* and Nor Haniza Sarmin

Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Johor, Malaysia

*Corresponding author: fwh@utm.my

Abstract The modelling of splicing systems is simulated by the process of cleaving and recombining DNA molecules with the presence of a ligase and restriction enzymes which are biologically called as endodeoxyribonucleases. The molecules resulting from DNA splicing systems are known as splicing languages. Palindrome is a sequence of strings that reads the same forward and backward. In this research, the splicing languages resulting from DNA splicing systems with one non-palindromic restriction enzyme are determined using the notation from Head splicing system. The generalisations of splicing languages for DNA splicing systems involving a cutting site and two non-overlapping cutting sites of one non-palindromic restriction enzyme are presented in the first and second theorems, respectively, which are proved using direct and induction methods. The result from the first theorem shows a trivial string which is the initial DNA molecule; while the second theorem determines a splicing language consisting of a set of resulting DNA molecules from the respective DNA splicing system.

Keywords DNA; splicing system; splicing language; non-palindromic; restriction enzyme.

Mathematics Subject Classification 68Q45, 92B05

1 Introduction

Deoxyribonucleic acid (DNA) splicing systems were introduced by Head [1] in 1987 as a study between formal language theory and molecular biology. The splicing systems are initially known as Head Splicing System, S_H. The systems are simulated by experimental procedures which are inspired by the biological process of recombinant DNA where DNA molecules react with ligase and restriction enzymes through wet lab experiment [2]. In the modelling of splicing systems, the restriction enzymes act as rules to cut and recombine DNA molecules (or strings) to generate further molecules.

The strings resulting from DNA splicing system depict a language known as a splicing language which is analysed using formal language theory. In formal language theory, a language

consists of a set of strings of symbols derived from an alphabet [3]. Some notations in formal languages can be applied in the modelling of DNA splicing system such as λ or 1, +, $\bullet$ and * which indicate the empty string, union, concatenation and star-closure respectively [3]. In splicing system, the splicing language is associated with three sets using formal language theory. The first set consists of double stranded DNA (dsDNA) symbols generated from Watson-Crick complementarity: adenine (A) pairs with thymine (T), while cytosine (C) pairs with guanine (G) [4]. The second set is the set of initial DNA molecules [1]. Lastly, the third set consists of the recombination behaviours allowed by formation rules of enzymatic activities [1]. The rule for the cleavage pattern of restriction enzymes is denoted in term of a triple: left context, crossing and right context [5]. The symbols $\downarrow$ and $\uparrow$ denote the upper and lower cutting sites of the restriction enzymes respectively. In this research, the length of cutting sites must be taken into consideration since overlapped cutting sites of restriction enzymes may affect enzymatic activity [6].

Over the years, several models had been extended and developed like Paun [7], Pixton [8], and Yusof-Goode [9] splicing systems. Besides that, Laun [10] also developed a new model of splicing system known as Goode-Pixton splicing system. Morever, a new extension of splicing systems known as a simple splicing system had also been introduced [10]. The comparison between Head's and these splicing models was studied in [9].

The modelling of DNA splicing systems involving palindromic sequences has been discussed by Head in [1]. Yusof [11] had also studied and mentioned the definition of palindromic string in DNA splicing systems. Palindrome is a sequence of strings that reads the same forward and backward [12]. Palindromic sequence can be recognised in the strings of restriction enzymes. Previously, research on the generalisations of splicing languages resulting from DNA splicing systems has been carried out involving palindromic restriction enzyme [13].

As an extension of the previous research, the splicing languages resulting from the modelling of DNA splicing systems with one non-palindromic restriction enzyme are presented in this paper. The aim of this research is to generalise the splicing languages in DNA splicing systems with at most two cutting sites of a non-palindromic restriction enzyme.

2 Methodology

In this research, DNA splicing systems with one non-palindromic restriction enzyme are modelled using Head's splicing system. The definitions of a splicing system and a splicing language are stated in the following.

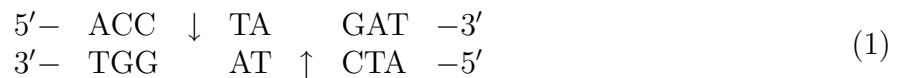**Definition 1** [1] (Splicing System and Splicing Language)
A splicing system $S = (A, I, B, C)$ consists of a finite alphabet A, a finite set I of initial strings in A^*, and finite sets B and C of triples (c, x, d) with c, x and d in A^*. Each such triple in B or C is called a pattern. For each such 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 $L = L(S)$.

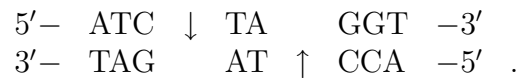Next, an example of a splicing system is given to illustrate the process of cutting and recombining DNA strings.

Example 1

Given a DNA splicing system $S = (A, I, B, C)$ in which $A = \left\{ \dfrac{A}{T}, \dfrac{C}{G}, \dfrac{G}{C}, \dfrac{T}{A} \right\}$ is the set of dsDNA symbols, $I = \left\{ \dfrac{\text{ACCTAGAT}}{\text{TGGATCTA}} \right\}$ is the set consisting of an initial string, set $B = \left\{ \left(\dfrac{C}{G}, \dfrac{TA}{AT}, \dfrac{G}{C} \right) \right\}$ is the set of cleavage pattern for the restriction enzyme *Bfa*I and set C is the empty set.

The initial string with the cutting site of the enzyme *Bfa*I is shown in the following:

$$
\begin{array}{llllll}
5'- & \text{ACC} & \downarrow & \text{TA} & \text{GAT} & -3' \\
3'- & \text{TGG} & & \text{AT} & \uparrow \text{CTA} & -5'
\end{array}
\tag{1}
$$

or written 180 degree wise due to the complementary nature of the two DNA strands [14]

$$
\begin{array}{llllll}
5'- & \text{ATC} & \downarrow & \text{TA} & \text{GGT} & -3' \\
3'- & \text{TAG} & & \text{AT} & \uparrow \text{CCA} & -5' \quad .
\end{array}
$$

When the enzyme *Bfa*I is added to the initial string,

$$
\begin{array}{ll}
5'- & \text{ACC} - 3' \\
3'- & \text{TGGAT} - 5'
\end{array}
$$

combines with

$$
\begin{array}{ll}
5' - \text{TAGGT} & -3' \\
3' - \text{CCA} & -5'
\end{array}
$$

which gives the following string

$$
\begin{array}{lll}
5'- & \text{ACCTAGGT} & -3' \\
3'- & \text{TGGATCCA} & -5' \quad .
\end{array}
\tag{2}
$$

Then,

$$
\begin{array}{lll}
5' - \text{TAGAT} & -3' \\
3' - \text{CTA} & -5'
\end{array}
$$

combines with

$$
\begin{array}{ll}
5'- & \text{ATC} - 3' \\
3'- & \text{TAGAT} - 5'
\end{array}
$$

to give the following string

$$
\begin{array}{lll}
5'- & \text{ATCTAGAT} & -3' \\
3'- & \text{TAGATCTA} & -5' \quad .
\end{array}
\tag{3}
$$

Thus, the set of strings in (1), (2) and (3) depicts the splicing language resulting from this splicing system.

Next, the definition of a palindromic string is presented.

Definition 2 [11] Palindromic String
A string I of a 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.

This research determines the generalisations of the splicing languages from DNA splicing systems with non-palindromic restriction enzyme. For example, the enzyme *Hba*I $\begin{array}{l} 5' - \text{GCCCAA} - 3' \\ 3' - \text{CGGGTT} - 5' \end{array}$ is a non-palindromic restriction enzyme since the upper single strand $5' - \text{GCCCAA} - 3'$ does not match with the lower single strand $3' - \text{CGGGTT} - 5'$ when read from backward.

3 Results and Discussion

In this research, the splicing languages from DNA splicing systems involving one non-palindromic restriction enzyme, which acts as the rule, are generalised using Head splicing system and presented as theorems. Some examples are provided to determine the splicing languages from DNA splicing systems with one non-palindromic restriction enzyme using the theorems. The generalisation of splicing languages resulting from DNA splicing systems with one cutting site of a non-palindromic restriction enzyme is presented in Theorem 1.

Theorem 1   *Given S = (A, I, B, C) is a DNA splicing system in which* $A = \left\{ \dfrac{A}{T}, \dfrac{C}{G}, \dfrac{G}{C}, \dfrac{T}{A} \right\}$ *is the set of dsDNA symbols,* $I = \left\{ \begin{array}{cccccc} N_1 N_1 ... N_1 & X_1 & Y & X_2 & N_2 N_2 ... N_2 \\ N_1' N_1' ... N_1' & X_1' & Y' & X_2' & N_2' N_2' ... N_2' \end{array} \right\}$ *is the set consisting of an initial string with one cutting site of a non-palindromic rule* $\begin{array}{ccc} X_1 & Y & X_2 \\ X_1' & Y' & X_2' \end{array}$ *where* $\dfrac{N_1}{N_1'}$, $\dfrac{X_1}{X_1'}$, $\dfrac{Y}{Y'}$, $\dfrac{X_2}{X_2'}$ *and* $\dfrac{N_2}{N_2'}$ *are variables used to denote any arbitrary dsDNA and* N_1', X_1', Y', X_2' *and* N_2' *are complementaries for* N_1, X_1, Y, X_2 *and* N_2 *respectively, set* $B = \left\{ \left(\dfrac{X_1}{X_1'}, \dfrac{Y}{Y'}, \dfrac{X_2}{X_2'} \right) \right\}$ *is the set of the restriction site where* $\dfrac{Y}{Y'}$ *is the crossing and set C is the empty set, the resulting splicing language is*

$$\begin{array}{ccccc} N_1 N_1 ... N_1 & X_1 & Y & X_2 & N_2 N_2 ... N_2 \\ N_1' N_1' ... N_1' & X_1' & Y' & X_2' & N_2' N_2' ... N_2' \end{array}$$

where $\left\{ \begin{array}{ccc} X_1 & Y & X_2 \\ X_1' & Y' & X_2' \end{array}, \begin{array}{ccc} X_2' & Y' & X_1' \\ X_2 & Y & X_1 \end{array} \right\} \notin \left\{ \begin{array}{cc} N_1 N_1 ... N_1 & N_2 N_2 ... N_2 \\ N_1' N_1' ... N_1' & N_2' N_2' ... N_2' \end{array} \right\}$ *which indicates no other cutting site is present in strings* $\begin{array}{c} N_1 N_1 ... N_1 \\ N_1' N_1' ... N_1' \end{array}$ *and* $\begin{array}{c} N_2 N_2 ... N_2 \\ N_2' N_2' ... N_2' \end{array}$.

Proof   Let the rule be not palindromic, so the base sequence of the site is not the same backward and forward

$$\begin{array}{ccc} X_1 & Y & X_2 \\ X_1' & Y' & X_2' \end{array} \neq \begin{array}{ccc} X_2' & Y' & X_1' \\ X_2 & Y & X_1 \end{array}.$$

Then $\dfrac{X_1}{X_1'} \neq \dfrac{X_2'}{X_2}$, $\dfrac{Y}{Y'} \neq \dfrac{Y'}{Y}$ and $\dfrac{X_2}{X_2'} \neq \dfrac{X_1'}{X_1}$.

The initial string $\dfrac{N_1 N_1 ... N_1 \quad X_1 \quad Y \quad X_2 \quad N_2 N_2 ... N_2}{N_1' N_1' ... N_1' \quad X_1' \quad Y' \quad X_2' \quad N_2' N_2' ... N_2'}$ with the cutting site $\dfrac{X_1 \quad Y \quad X_2}{X_1' \quad Y' \quad X_2'}$ is shown in the following:

$$\begin{matrix} N_1 N_1 ... N_1 & X_1 & \downarrow & Y & & X_2 & N_2 N_2 ... N_2 \\ N_1' N_1' ... N_1' & X_1' & & Y' & \uparrow & X_2' & N_2' N_2' ... N_2' \end{matrix} \tag{4}$$

Hence, the initial string is cut into two parts:

$$\begin{matrix} N_1 N_1 ... N_1 & X_1 & \\ N_1' N_1' ... N_1' & X_1' & Y' \end{matrix}$$

And

$$\begin{matrix} Y X_2 & N_2 N_2 ... N_2 \\ X_2' & N_2' N_2' ... N_2' \end{matrix}$$

Besides that, the initial string $\dfrac{N_1 N_1 ... N_1 \quad X_1 \quad Y \quad X_2 \quad N_2 N_2 ... N_2}{N_1' N_1' ... N_1' \quad X_1' \quad Y' \quad X_2' \quad N_2' N_2' ... N_2'}$ can be written 180 degree wise as

$$\begin{matrix} N_2' N_2' ... N_2' & X_2' & Y' & X_1' & N_1' N_1' ... N_1' \\ N_2 N_2 ... N_2 & X_2 & Y & X_1 & N_1 N_1 ... N_1 \end{matrix} \tag{5}$$

Since $\dfrac{X_1}{X_1'} \neq \dfrac{X_2'}{X_2}$, $\dfrac{Y}{Y'} \neq \dfrac{Y'}{Y}$ and $\dfrac{X_2}{X_2'} \neq \dfrac{X_1'}{X_1}$, then there is no cutting site in (5) and no new resulting string.

From (4) and (5), the resulting splicing language is the initial string:

$$\begin{matrix} N_1 N_1 ... N_1 & X_1 & Y & X_2 & N_2 N_2 ... N_2 \\ N_1' N_1' ... N_1' & X_1' & Y' & X_2' & N_2' N_2' ... N_2' \end{matrix}$$

Thus, Theorem 1 is proved. $\qquad\square$

Example 2 shows the splicing language obtained from a DNA splicing system with one cutting site of a non-palindromic restriction enzyme namely *Aci*I.

Example 2

Given a DNA splicing system $S = (A, I, B, C)$ where $I = \left\{ \begin{matrix} \text{TTCCCGCGAC} \\ \text{AAGGGCGCTG} \end{matrix} \right\}$ is the set of initial string, set $B = \left\{ \begin{pmatrix} \text{C} & \text{CG} & \text{C} \\ \text{G} & \text{GC} & \text{G} \end{pmatrix} \right\}$ is the set of cleavage pattern for the enzyme *Aci*I and set C is the empty set.

The enzyme *Aci*I, $\begin{matrix} 5'- & \text{CCGC} & -3' \\ 3'- & \text{GGCG} & -5' \end{matrix}$ is a non-palindromic restriction enzyme since the base sequence of enzyme *Aci*I do not read the same forwards and backwards.

The initial string $\begin{matrix} 5'- & TTCC & \downarrow & CG & & CGAC & -3' \\ 3'- & AAGG & & GC & \uparrow & GCTG & -5' \end{matrix}$ has one cutting site of the enzyme *Aci*I. So, by using Theorem 1, the resulting splicing language is the initial string

$$\begin{matrix} 5'- & TTCCCGCGAC & -3' \\ 3'- & AAGGGCGCTG & -5' \end{matrix} \quad .$$

Next, the splicing languages from DNA splicing systems with two non-overlapping cutting sites of a non-palindromic restrictions enzyme are generalised and given as Theorem 2.

Theorem 2   *Given $S = (A, I, B, C)$ is a DNA splicing system in which $A = \left\{ \dfrac{A}{T}, \dfrac{C}{G}, \dfrac{G}{C}, \dfrac{T}{A} \right\}$ is the set of dsDNA symbols, $I = \left\{ \dfrac{N_1 N_1 ... N_1\; X_1\; Y\; X_2\; M M ... M\quad X_1\; Y\; X_2\; N_2 N_2 ... N_2}{N_1' N_1' ... N_1'\; X_1'\; Y'\; X_2'\; M' M' .... M'\; X_1'\; Y'\; X_2'\; N_2' N_2' ... N_2'} \right\}$ is the set consisting of an initial string with two non-overlapping cutting sites of a non-palindromic rule $\dfrac{X_1\; Y\; X_2}{X_1'\; Y'\; X_2'}$ where $\dfrac{N_1}{N_1'}, \dfrac{X_1}{X_1'}, \dfrac{Y}{Y'}, \dfrac{X_2}{X_2'}, \dfrac{M}{M'}$ and $\dfrac{N_2}{N_2'}$ are variables used to denote any arbitrary dsDNA and N_1', X_1', Y', X_2', M' and N_2' are complementaries for $N_1, X_1, Y, X_2,$ M and N_2 respectively, set $B = \left\{ \left(\dfrac{X_1}{X_1'}, \dfrac{Y}{Y'}, \dfrac{X_2}{X_2'} \right) \right\}$ is the set of the restriction site where $\dfrac{Y}{Y'}$ is the crossing and set C is the empty set, the resulting splicing language consists of strings of the form*

$$\begin{matrix} N_1 N_1 ... N_1 & X_1 & Y & X_2 \\ N_1' N_1' ... N_1' & X_1' & Y' & X_2' \end{matrix} \left(\begin{matrix} M M ... M & X_1 & Y & X_2 \\ M' M' .... M' & X_1' & Y' & X_2' \end{matrix} \right)^{n-1} \begin{matrix} N_2 N_2 ... N_2 \\ N_2' N_2' ... N_2' \end{matrix} \tag{6}$$

where $\left\{ \dfrac{X_1\; Y\; X_2}{X_1'\; Y'\; X_2'}, \dfrac{X_2'\; Y'\; X_1'}{X_2\; Y\; X_1} \right\} \notin \left\{ \dfrac{N_1 N_1 ... N_1}{N_1' N_1' ... N_1'}, \dfrac{M M ... M}{M' M' .... M'}, \dfrac{N_2 N_2 ... N_2}{N_2' N_2' ... N_2'} \right\}$ which indicates no other cutting site is present in strings $\dfrac{N_1 N_1 ... N_1}{N_1' N_1' ... N_1'}, \dfrac{M M ... M}{M' M' .... M'}$ and $\dfrac{N_2 N_2 ... N_2}{N_2' N_2' ... N_2'} \quad .$

Proof   Let the rule be not palindromic, so the base sequence of the site is not the same backward and forward

$$\begin{matrix} X_1 & Y & X_2 \\ X_1' & Y' & X_2' \end{matrix} \neq \begin{matrix} X_2' & Y' & X_1' \\ X_2 & Y & X_1 \end{matrix} \quad .$$

Then $\dfrac{X_1}{X_1'} \neq \dfrac{X_2'}{X_2}, \dfrac{Y}{Y'} \neq \dfrac{Y'}{Y}$ and $\dfrac{X_2}{X_2'} \neq \dfrac{X_1'}{X_1} \quad .$

The initial string, $\begin{matrix} N_1 N_1 ... N_1 & X_1 & Y & X_2 & M M ... M & X_1 & Y & X_2 & N_2 N_2 ... N_2 \\ N_1' N_1' ... N_1' & X_1' & Y' & X_2' & M' M' .... M' & X_1' & Y' & X_2' & N_2' N_2' ... N_2' \end{matrix}$ with the

cutting site $\dfrac{X_1 Y X_2}{X_1' Y' X_2'}$ is shown in the following:

$$\begin{matrix} N_1 N_1 ... N_1 & X_1 & \downarrow & Y & & X_2 & M M ... M & X_1 & Y & X_2 & N_2 N_2 ... N_2 \\ N_1' N_1' ... N_1' & X_1' & & Y' & \uparrow & X_2' & M' M' .... M' & X_1' & Y' & X_2' & N_2' N_2' ... N_2' \end{matrix} \tag{7}$$

for the first cutting site and

$$
\begin{array}{ccccccccccc}
N_1N_1...N_1 & X_1 & Y & X_2 & M\,M\,...M & X_1 & \downarrow & Y & X_2 & N_2N_2...N_2 \\
N_1'N_1'...N_1' & X_1' & Y' & X_2' & M'M'....M' & X_1' & & Y' & \uparrow & X_2' & N_2'N_2'...N_2'
\end{array}
\tag{8}
$$

for the second cutting site.

When the rule $\dfrac{X_1 Y X_2}{X_1' Y' X_2'}$ is added to the initial string, (7) combines with (8) which gives

$$
\begin{array}{ccccc}
N_1N_1...N_1 & X_1 & Y & X_2 & N_2N_2...N_2 \\
N_1'N_1'...N_1' & X_1' & Y' & X_2' & N_2'N_2'...N_2'
\end{array}
\tag{9}
$$

and

$$
\begin{array}{ccccccccccccccc}
N_1N_1...N_1 & X_1 & Y & X_2 & M\,M\,...M & X_1 & Y & X_2 & M\,M\,...M & X_1 & Y & X_2 & N_2N_2...N_2 \\
N_1'N_1'...N_1' & X_1' & Y' & X_2' & M'M'....M' & X_1' & Y' & X_2' & M'M'....M' & X_1' & Y' & X_2' & N_2'N_2'...N_2'
\end{array}.
$$

The initial string can be written 180 degree wise as

$$
\begin{array}{ccccccccc}
N_2'N_2'...N_2' & X_2' & Y' & X_1' & M'M'....M' & X_2' & Y' & X_1' & N_1'N_1'...N_1' \\
N_2N_2...N_2 & X_2 & Y & X_1 & M\,M\,...M & X_2 & Y & X_1 & N_1N_1...N_1
\end{array}.
\tag{10}
$$

Since $\dfrac{X_1}{X_1'} \neq \dfrac{X_2'}{X_2}$, $\dfrac{Y}{Y'} \neq \dfrac{Y'}{Y}$ and $\dfrac{X_2}{X_2'} \neq \dfrac{X_1'}{X_1}$, then there is no cutting site in (10)

By using induction, this theorem can be proved. Let string (6) be a statement

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

For $n = 1$, $P(1)$ is true since it is stated in (9). Next, let $n = k \in \mathbf{Z}^+$. Thus,

$$
P(k): \begin{array}{c}
N_1N_1...N_1\ X_1\ \downarrow\ Y \qquad X_2 \\
N_1'N_1'...N_1'\ X_1' \qquad Y'\ \uparrow\ X_2'
\end{array}
\left(\begin{array}{c}
M\,M\,...M\ \ X_1\ Y\ X_2 \\
M'M'....M'\ X_1'\ Y'\ X_2'
\end{array} \right)^{k-1}
\begin{array}{c}
N_2N_2...N_2 \\
N_2'N_2'...N_2'
\end{array}.
\tag{11}
$$

Then, the string (8) combines with (11) which produces new string

$$
\begin{array}{c}
N_1N_1...N_1\ X_1\ Y\ X_2\ M\,M\,...M\ \ X_1\ Y\ X_2 \\
N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ X_1'\ Y'\ X_2'
\end{array}
\left(\begin{array}{c}
M\,M\,...M\ \ X_1\ Y\ X_2 \\
M'M'....M'\ X_1'\ Y'\ X_2'
\end{array} \right)^{k-1}
\begin{array}{c}
N_2N_2...N_2 \\
N_2'N_2'...N_2'
\end{array}.
\tag{12}
$$

By simplifying string (12), the resulting strings are:

$$
\begin{array}{c}
N_1N_1...N_1\ X_1\ Y\ X_2 \\
N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'
\end{array}
\left(\begin{array}{c}
M\,M\,...M\ \ X_1\ Y\ X_2 \\
M'M'....M'\ X_1'\ Y'\ X_2'
\end{array} \right)^{(k+1)-1}
\begin{array}{c}
N_2N_2...N_2 \\
N_2'N_2'...N_2'
\end{array}.
\tag{13}
$$

Thus, $P(k+1)$ is true since it depicts string (13) when $n = k + 1$.

Hence, Theorem 2 is proved. $\qquad\qquad\square$

Example 3 presents the splicing language from a DNA splicing system with two non-overlapping cutting sites of a non-palindromic restrictions enzyme namely *Bss*SI.

Example 3

Given a splicing system $S = (A, I, B, C)$ where $I = \left\{ \begin{array}{l} \text{CTCACGAGTTCACGAGGA} \\ \text{GAGTGCTCAAGTGCTCCT} \end{array} \right\}$ is the set of initial string, set $B = \left\{ \left(\begin{array}{ccc} \text{C} & \text{ACGA} & \text{G} \\ \text{G} & \text{TGCT} & \text{C} \end{array} \right) \right\}$ is the set of cleavage pattern for the enzyme *Bss*SI and set C is the empty set.

The enzyme *Bss*SI $\begin{array}{lll} 5'- & \text{CACGAG} & -3' \\ 3'- & \text{GTGCTC} & -5' \end{array}$ is a non-palindromic restriction enzyme since the base sequence of the enzyme *Bss*SI does not read the same forwards and backwards.

The initial string $\begin{array}{lllllllll} 5'- & \text{CTC} & \downarrow & \text{ACGA} & \text{GTTC} & \downarrow & \text{ACGA} & \text{GGA} & -3' \\ 3'- & \text{GAG} & & \text{TGCT} & \uparrow \text{ CAAG} & & \text{TGCT} & \uparrow \text{ CCT} & -5' \end{array}$ has two non-overlapping cutting sites of the enzyme *Bss*SI $\begin{array}{lll} 5'- & \text{CACGAG} & -3' \\ 3'- & \text{GTGCTC} & -5' \end{array}$. Therefore, by using Theorem 2, the resulting splicing language is

$$\begin{array}{llll} 5'- & \text{CTCACGAG} & \left(\text{TTCACGAG} \right)^{n-1} & \text{GA} & -3' \\ 3'- & \text{GAGTGCTC} & \left(\text{AAGTGCTC} \right) & \text{CT} & -5' \end{array}$$

where $n \in \mathbb{Z}^+$.

4 Conclusion

In this research, the generalisations of splicing languages in DNA splicing systems involving one and two non-overlapping cutting sites of one non-palindromic restriction enzyme are given as Theorem 1 and 2 respectively. The results from this research contribute to the evolution of idea in DNA splicing systems where the resulting DNA molecules can be determined without actual lab experiments. For further research, splicing languages from DNA splicing systems for more than one non-palindromic restriction enzyme with same and different crossings can also be generalised.

Acknowledgement

References

[1] Head, T. Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors. *B. Math. Biol.* 1987. 49(6): 737-759.

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

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

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

[5] Head, T. Splicing representations of strictly locally testable languages. *Discrete. Appl. Math.* 1998. 87(1): 139-147.

[6] Horton, R. M., Ho, S. N., Pullen, J. K., Hunt, H. D., Cai, Z. and Pease, L. R. [23] Gene splicing by overlap extension. In Wu, R. (Ed.). *Methods in enzymology*: Academic Press. 1995. 363-372.

[7] Pun, G. On the splicing operation. *Discrete. Appl. Math.* 1996. 70(1): 57-79.

[8] Pixton, D. Regularity of splicing languages. *Discrete. Appl. Math.* 1996. 69(1-2): 101-124.

[9] Yusof, Y., Sarmin, N. H., Fong, W. H., Goode, T. E. and Ahmad, M. A. An analysis of four variants of splicing system. In *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. Melville, NY: AIP Conference Proceedings. 2013. 888-895.

[10] Laun, T. E. G. *Constants and Splicing Systems.* Ph.D. Thesis. State University of New York. 1999.

[11] Yusof, Y. *DNA Splicing System Inspired by Bio Molecular Operations.* Ph.D. Thesis. Universiti Teknologi Malaysia. 2012.

[12] Tomohiro, I., Inenaga, S. and Takeda, M. Palindrome pattern matching. *Theor. Comput. Sci.* 2013. 483: 162-170.

[13] Fong, W. H. and Ismail, N. I. Generalisations of DNA splicing systems with one palindromic restriction enzyme. *MATEMATIKA.* 2018. 34(1): 59-71.

[14] Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M. and Losick, R. *Molecular Biology of the Gene.* 7th ed. New York: Cold Spring Harbor Laboratory Press. 2014.