MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Ruslim *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 20 (2024) 618-625

RESEARCH ARTICLE

The Generalisations of *n*-Cutting Sites Splicing Languages via Yusof-Goode Splicing System using a Non-Palindromic Rule and Crossing Site

Nooradelena Mohd Ruslim[a], Yuhani Yusof[a]*, Mohd Sham Mohamad[a], Mohd Firdaus Abdul-Wahab[b], Mohammad Hassan Mudaber[c]

[a]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia; [b]Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia; [c]Department of Mathematics, Faculty of Natural Sciences, Kabul Education University, 5th District, Afshshar Road, Kabul, Afghanistan

Abstract Yusof-Goode (Y-G) splicing system was introduced in the context of Formal Language Theory. Splicing system is a dry model that presents enzymatic activities between initial strings and restriction enzymes, while splicing language is the generated strings from the splicing system. Splicing language will yield either as new molecules or initial string itself, and can be either in adult or inert, limit or transient languages. In this paper, some mathematical results on generating and generalising the *n*-cutting sites splicing languages are established using a Y-G splicing system consisting of a single pattern of strings with non-palindromic rule and crossing site. Two lemmas are presented to discuss the Y-G splicing system when two and three cutting sites exist in a single pattern of string. Different characteristics concerning the features of left and right contexts are established. A theorem is then proposed based on the lemmas to generalise the *n*-cutting sites splicing languages resulting from a Y-G splicing system with a single pattern of string and a non-palindromic rule when *n*-cutting sites exist in a single pattern of string.

Keywords: Y-G splicing system, non-palindromic, splicing language, crossing site, restriction enzyme.

Introduction

Deoxyribonucleic acid (DNA) recombination is a fundamental biological process that describes a reaction when DNA molecules from two different sources are combined to form a new genetic sequence. This process plays a crucial role in genetic diversity, evolution, and the repair of damaged DNA. Site-specific recombination is a mechanism through which specific DNA sequences called recombination sites are recognized and exchanged. In 1987, Head [1] first introduced the splicing system as a formal characterization of particular enzymatic activity working on DNA molecules under the framework of Formal Language Theory. Theoretical Computer Science and Applied Discrete Mathematics are two branches of Formal Language Theory, where a formal language can be regarded as a collection of strings over a finite alphabet that follow some rules in order to form a language. With the aim of investigating the relationship between the recombination process and formal language theory, Head [1] developed a mathematical splicing system based on the biological perspective of the cutting and pasting process. Head splicing system offers a writing notation on the elements involved in the splicing system, given as $S = (A, I, B, C)$ where A is the set of alphabets of the DNA, I is the initial strings involved in the system, while B and C refers to the rule that is used in the system. After the remarkable work by Head, there have been many explorations in this field. The notion in writing the splicing system has been found in various research such as Paun [2], Goode-Pixton [3], and Yusof-Goode (Y-G) [4] splicing systems.

***For correspondence:**
yuhani@umpsa.edu.my

Received: 24 Nov. 2023
Accepted: 6 May 2024

Among these four types of splicing systems, the difference in one another is mainly on the writing notation of the system. Motivated by the Head splicing system, Yusof [4] introduced a new writing notation. The Y-G splicing system can clearly show the type of rule adopted in the system, either a rule with left cutting, right cutting or both cuttings [4]. Therefore, the Y-G splicing system is chosen to be referred in the lemmas, cases and theorem discussed in this paper, due to the transparent approach of this system. Years after, this field caught interest of many researchers to do extensive studies in bio-molecular. Lim [5] adopted the Y-G splicing system in investigating single-stage splicing languages, while Mudaber [6] deployed the Y-G splicing system in studying two stages splicing languages. Ahmad [7] proposed a new type of language called second-order limit language, also by using a Y-G splicing system. These studies also adopted mathematical approaches in presenting their results; via limit adjacency matrix [8], de Bruijn graph [9], and automata diagram [10]. A comprehensive review [11] on the DNA splicing system from graph theoretic perspective showed that DNA is best presented in graphs, as shown in [12–14].

Apart from the wet model which conducts laboratory experiments that require time and money, the establishment of a splicing system has sparked more researches in the bio-molecular field. The splicing system is a dry model that illustrates enzymatic activities between DNA initial strings and rules. DNA is a molecule that contains the genetic instructions required for the development and function of all living organisms. It is a double-stranded helical structure, with a nucleotide repeat structure, arranged in a long polymer [15]. Each nucleotide is composed of three components: a pentose sugar, a nitrogenous base, and a phosphate group. The nitrogenous bases are made up of two classes namely purines and pyrimidines. There are two purines that is adenine, A and guanine, G and two pyrimidines that is cytosine, C and thymine, T in DNA structure [16]. The pairing between the bases follows the Watson and Crick complementarity that illustrates the nitrogenous bases pairing where A bonds to T, T bonds to A, C bonds to G and G bonds to C. These bonding can be expressed as A/T, T/A, C/G and G/C or a, c, g, t respectively. Chargaff [17] showed a unique pattern of base ratios, where the A/T and C/G are almost equally likely 1:1. The bonding of the nitrogenous bases will form an antiparallel double-stranded DNA (dsDNA), that is in $5'-3'$ direction of the upper strand, while the lower strand will be in $3'-5'$ direction. For instance, if an upper strand has the sequence of $5'...ATCGTT...3'$, then the lower strand is antiparallel to the upper strand, $3'...TAGCAA...5'$. These sequences can simply be written as dsDNA as shown below:

$$\begin{array}{l} 5'...ATCGTT...3' \\ 3'...TAGCAA...5' \end{array}$$

Other than DNA, a restriction enzyme that involves in a splicing system plays an important role in recognising specific DNA sequences and cuts the DNA at those recognition sites [16]. This enzyme which usually has a length of 4 to 6 base-pairs will cleave the strings into fragments, depending on the features of the restriction enzyme. The sequence of the enzyme can be recognised either in palindromic or non-palindromic. The features of the palindromic or non-palindromic sequence are referred to the sequence whether they read the same in both directions on its upper and lower strands. There exists a certain sequence in restriction enzyme called recognition site. When the recognition site matches the sequence in strings, the strings will cleave at a certain cutting site. The cuts made by restriction enzymes produce DNA fragments with sticky ends or blunt ends, depending on the type of enzyme used. For example, *Hae*III that has four base-pairs is categorised as blunt-end digestion, *Bam*HI that has six base-pairs and *Not*I that has eight base-pairs are sticky-end digestions. The more recognition sites found in a string, the more cutting sites will exist. This is relevant with the aim of this paper which is to generalise the generated splicing languages when *n*-cutting sites exist in a single pattern of strings.

Resulting from enzymatic activities between dsDNA, restriction enzyme, and ligase, is a set of words which is mathematically known as splicing language. The splicing language can be in various types, namely adult or inert language, limit language or transient language [18]. Other than these three types of languages, in 2012, Yusof [4] renamed the limit language to inert persistent language and reclassified the limit language as active persistent language, which denotes to a collection of strings that contain in the limit language and will further splice. The cutting and pasting (i.e. digesting and ligating) process in a splicing system occurs at a specific place in a string. The selected restriction enzyme will cut the dsDNA when the recognition site matches a specific sequence in a string, resulting in sticky or blunt ends [19]. The possibility of pasting among the fragments depends on the features of the crossing site. A palindromic crossing site will give more chances of religation among fragments, in producing an initial string itself and new molecules. Nonetheless, a non-palindromic crossing site will lessen the type and number of generated splicing languages in a certain splicing system. Henceforth, this paper aims to investigate if the *n*-cutting sites splicing languages can be generalised when *n*-cutting sites exists in a

string. It is also intended to see if the generated languages will be in infinitely long molecules. Additionally, by having a non-palindromic rule and crossing site in a splicing system, the possibility of religation between the initial string and its rotational molecules will be investigated.

Some significant definitions pertaining to this study are provided in the next section.

Preliminaries

This section provides some key definitions related to this study. In this paper, motivated by the notion of rule writing, the Yusof-Goode splicing system is preferred in generating the *n*-cutting sites splicing languages. Hence, its definition is first given as follows:

Definition 1 [4]: Yusof-Goode (Y-G) Splicing System and Splicing Language
Let $S=(A,I,R)$ be a Y-G splicing system. The splicing system consists of a set of alphabets $A=\{a,c,g,t\}$, an initial string $I\in A^*$, and a set of rules $r\in R$. If the rule is given by $r=(m,y,n:s,y,t)$, for two initial strings, $I_1=\gamma myn\delta$ and $I_2=\alpha syt\beta$, where $I_1,I_2\in I$, then $\gamma myt\beta$ and $\alpha syn\delta$ are the new languages generated from the system besides its initial strings, where $\gamma,\delta,\alpha,\beta,m,n,s,t,y\in A^*$ are free monoids generated by A with the concatenation process and 1 as the identity element. Therefore, a set of words, L is known as a splicing language if there exists a splicing system S for which $L=L(S)$.

Next, an important characteristic on elements in splicing system is known as palindromic.

Definition 2 [4]: Palindromic
A sequence in dsDNA strand is said to be palindromic if the sequence of the upper strand from left to right is similar to the sequence of the lower strand from right to left.

The palindromic feature is common in the DNA splicing process, which can occur either in the dsDNA strand, the restriction enzyme or the crossing site, and plays a vital role in the pasting process.

Restriction enzyme namely *Acc*65I has a palindromic sequence, $\begin{matrix}5'...GGTACC...3'\\3'...CCATGG...5'\end{matrix}$ shows a similar reading for its upper and lower strand, reads from left to right and right to left, respectively. Conversely, the non-palindromic feature is recognised when the sequence of the upper strand from left to right varies from the sequence of the lower strand from right to left. For instance, *Aci*I has a non-palindromic sequence, $\begin{matrix}5'...CCGC...3'\\3'...GGCG...5'\end{matrix}$, showing a different reading from left to right of its upper strand, and from right to left, of its lower strand. Later, the features of non-palindromic rule and crossing site can be seen in the rule presented in the next section of this paper.

When the enzymatic activities occur in a splicing system, the recombination of DNA molecules will take part, producing either new DNA molecules or the initial string itself. Finally, the outcome of the splicing process can be seen in a set of words called splicing language. The generated splicing languages can be in several types, one of which is called transient language. The definition of transient language is as given in Definition 3.

Definition 3 [4]: Transient Language
A set of language which will finally be consumed and vanished is called transient language.

The results and discussion are presented in the following section.

Results and Discussion

The Yusof-Goode (Y-G) splicing system, which comprises of a single pattern of strings with a non-palindromic rule and crossing site is discussed in this section. Based on the restriction enzyme features presented in this study, two lemmas and a theorem are established. Biologically, the lemmas will predict if the features of the restriction enzyme will affect the formation of splicing languages when ligases take

part in the recombination process after the restriction enzyme cuts the string. There are a few possible types of the generated splicing languages to be explored in this study, which can be either in adult or inert language, limit language or transient language. The Y-G splicing system is deliberated in two cases namely Case 1 which deliberates a distinct left and right context of a rule, and Case 2 for a unique left and right context of a rule. Both cases are obtainable in Lemma 1 and Lemma 2, which are further discussed below.

Single Pattern of Strings Containing Two Cutting Sites with a Non-palindromic Rule and Crossing Site

Aiming to investigate the splicing language from a Y-G splicing system containing of a single pattern of strings that contain two cutting sites, the rule in the system is chosen to be non-palindromic with non-palindromic crossing site. Thus, Lemma 1 presents a Y-G splicing system with certain criteria set on the rule.

Lemma 1

Let $S=(A,I,R)$ *be a Y-G splicing system. The initial string,* I *is comprises of set of alphabets* $A=\{a,c,g,t\}$*, with two cutting sites exist. If the system contains a non-palindromic rule,* $r\in R$ *with non-palindromic crossing site, then n-cutting sites splicing languages is generated in* $\gamma-\delta$ *sequence. With the increasing number of repetitions of the middle segment in the languages,* k *where* $k\geq 0$ *and* $k\in Z^{+}$*, the Y-G splicing system will produce infinitely long molecules in the type of transient language.*

Proof.

Assume $S=(A,I,R)$ is a Y-G splicing system containing a single string with a single rule. Two cases will be considered where Case 1 presents a splicing system with distinct left and right context in its rule, while in Case 2, a unique left and right context of rule are considered.

Case 1: Distinct Left and Right Context of Rule

Assume r is a non-palindromic rule with distinct left and right contexts, i.e. $r_1=(p,mn,q:p,mn,q)$. Let $s_1=\gamma pmnqpmnq\delta$ be an initial string containing two cutting sites, such that $[p/p'],[q/q'],[m/m']$ and $[n/n']$ for $\forall\gamma,\delta,p,q,m,n\in A^{*}$. Since the crossing site is non-palindromic, it follows that $m\neq n'$ and $n\neq m'$, and are not complemented to each other. When the rule is acted on the given initial string, then the string will cleave as follows:

$$s_{1_0}=\begin{matrix}\gamma & p^{\blacktriangledown} & m & n & q & p^{\blacktriangledown} & m & n & q & \delta\\ \gamma' & p' & m' & n'_{\blacktriangle} & q' & p' & m' & n'_{\blacktriangle} & q' & \delta'\end{matrix} \tag{1}$$

The dsDNA given in (1) can be alternated by 180°, giving:

$$s_{1_{180}}=\begin{matrix}\delta' & q'^{\blacktriangledown} & n' & m' & p' & q'^{\blacktriangledown} & n' & m' & p' & \gamma'\\ \delta & q & n & m_{\blacktriangle} & p & q & n & m_{\blacktriangle} & p & \gamma\end{matrix} \tag{2}$$

Owing to the non-palindromic feature of the crossing site, the $5'-$overhangs in (1) and (2) have no chance to religate with each other. Thus, the religation occurs among fragments from s_{1_0} or $s_{1_{180}}$ only, producing splicing languages in $\gamma-\delta$ sequence. Depending on the number of k repeats in the splicing language, different length of molecules will be generated, either new or the initial string itself. The generated splicing languages, $L(S)$ from this splicing system is as shown in (3):

$$\{\gamma pmnq\delta,\gamma pmnqpmnq\delta,\gamma pmnqpmnqpmnq\delta,\gamma pmnqpmnqpmnqpmnq\delta,\dots\} \tag{3}$$

Accordingly, the generated splicing languages with $k\geq 0$ and $k\in Z^{+}$ can be generalised as follows:

$$s_{Y-G_1} = \{\gamma p(mnqp)^k mnq\delta\} \quad (4)$$

Case 2: Unique Left and Right Context of Rule

Let $r_2 = (p, mn, p : p, mn, p)$ be a rule with a unique left and right context with $s_2 = \gamma pmnppmnp\delta$ is an initial string containing two cutting sites given that $[p / p']$,$[m / m']$ and $[n / n']$ for $\forall \gamma, \delta, p, m, n \in A^*$. Given the crossing site is non-palindromic, thus $m \neq n'$ and $n \neq m'$, and not complemented to each other. When the initial string and the rule reacted, the string will cleave as follows:

$$s_{2_0} = \begin{matrix} \gamma & p^{\blacktriangledown} & m & n & p & p^{\blacktriangledown} & m & n & p & \delta \\ \gamma' & p' & m' & n'_{\blacktriangle} & p' & p' & m' & n'_{\blacktriangle} & p' & \delta' \end{matrix} \quad (5)$$

The dsDNA given in (5) can be alternated by 180°, giving:

$$s_{2_{180}} = \begin{matrix} \delta' & p'^{\blacktriangledown} & n' & m' & p' & p'^{\blacktriangledown} & n' & m' & p' & \gamma' \\ \delta & p & n & m_{\blacktriangle} & p & p & n & m_{\blacktriangle} & p & \gamma \end{matrix} \quad (6)$$

Due to the non-palindromic feature of the crossing site, the $5'-$overhangs in (5) and (6) cannot religate with each other. Thus, the religation occurs among fragments in s_{2_0} or $s_{2_{180}}$, producing splicing languages in $\gamma - \delta$ sequence only. The different lengths of new molecules depend on the number of k repeats in the splicing language. This splicing system produce a set of splicing languages, $L(S)$ given as:

$$\{\gamma pmnp\delta, \gamma pmnppmnp\delta, \gamma pmnppmnppmnp\delta, \gamma pmnppmnppmnppmnp\delta, \dots\} \quad (7)$$

Consequently, for $k \geq 0$ and $k \in Z^+$, the set of splicing languages in (7) can be generalised as follows:

$$s_{Y-G_2} = \{\gamma p(mnpp)^k mnp\delta\} \quad (8)$$

In the following section, the splicing system contains a single pattern of strings with three cutting sites, a non-palindromic rule and crossing site is presented.

Single Pattern of Strings Containing Three Cutting Sites with a Non-palindromic Rule and Crossing Site

Next, to examine the splicing language from a Y-G splicing system comprising of a single pattern of strings that contain three cutting sites, the rule in the system is chosen to be non-palindromic with non-palindromic crossing site. Accordingly, Lemma 2 presents a Y-G splicing system with certain criteria set on the rule.

Lemma 2

Let $S = (A, I, R)$ be a Y-G splicing system. The initial string, I is comprises of set of alphabets $A = \{a, c, g, t\}$, with three cutting sites exist. If the system contains a non-palindromic rule, $r \in R$ with non-palindromic crossing site, then n-cutting sites splicing languages is generated in $\gamma - \delta$ sequence. With the increasing number of repetitions of the middle segment in the languages, k where $k \geq 0$ and $k \in Z^+$, the Y-G splicing system will produce infinitely long molecules in the type of transient language.

Proof.

The splicing process of this lemma is as shown in Lemma 1, which is when 3-cutting sites exist in a string. Hence for Case 1 and Case 2, the generalised *n*-cutting sites splicing languages for $k \geq 0$ and $k \in Z^+$ can be specified as given in (4) and (8).

Seemingly, from Lemma 1 and Lemma 2, the generalised splicing languages for Case 1 and Case 2 are alike. Consequently, both Lemma 1 and Lemma 2 lead to the establishment of Theorem 1, that is the

generalisation of the *n*-cutting sites splicing languages when *n*-cutting sites exist in a string.

Theorem 1

Given Y-G splicing system, $S = (A, I, R)$ *with* $I = \{s\}$ *as an initial string containing set of alphabets* $A = \{a, c, g, t\}$ *with n-cutting sites,* $R = \{r\}$ *is a rule with non-palindromic characteristic rule in its sequence and its crossing site, then n-cutting sites splicing languages can be generalised as follows:*

for distinct left and right context:

$$s_{Y-G_1} = \{\gamma p(mnqp)^k mnq\delta\}$$

for unique left and right context:

$$s_{Y-G_2} = \{\gamma p(mnpp)^k mnp\delta\}$$

where $k \geq 0$ and $k \in Z^+$.

Proof.

This theorem will be proved by using mathematical induction. For a given Y-G splicing system, the generalised *n*-cutting sites splicing languages for distinct left and right context is $\{\gamma p(mnqp)^k mnq\delta\}$, while for unique left and right context is $\{\gamma p(mnpp)^k mnp\delta\}$.

In Case 1 (distinct left and right context), for $k = 0$,

$$s_{Y-G_1} = \gamma pmnq\delta$$

Hence, it is true when the middle segment of the generalised splicing languages, $(mnqp)^k$ is not exist i.e. when $k = 0$.

Next, assume $k = t$ is true for some $t \in Z^+$ i.e.

$$s_{Y-G_1} = \{\gamma p(mnqp)^t mnq\delta\}$$

Then, need to show it is true when $k = t + 1$,

$$\begin{aligned} s_{Y-G_1} &= \gamma p(mnqp)^t mnq\delta + \gamma p(mnqp)mnq\delta \\ &= \gamma p(mnqp)^t (mnqp)mnq\delta \\ &= \gamma p(mnqp)^{t+1} mnq\delta \end{aligned}$$

Therefore, it is true that the generalised *n*-cutting sites splicing languages for distinct left and right context is given as:

$$s_{Y-G_1} = \{\gamma p(mnqp)^k mnq\delta\}$$

Next, consider for Case 2 (unique left and right context), for $k = 0$,

$$s_{Y-G_2} = \gamma pmnp\delta$$

Hence, it is true when the middle segment of the generalised splicing languages, $(mnpp)^k$ is not exist i.e. when $k = 0$.

Next, assume $k = t$ is true for some $t \in Z^+$ i.e.

$$s_{Y-G_2} = \{\gamma p(mnpp)^t mnp\delta\}$$

Then, need to show it is true when $k = t+1$,

$$s_{Y-G_2} = \gamma p(mnpp)^t mnp\delta + \gamma p(mnpp)mnp\delta$$
$$= \gamma p(mnpp)^t (mnpp)mnp\delta$$
$$= \gamma p(mnpp)^{t+1} mnp\delta$$

Therefore, it is true that the generalised *n*-cutting sites splicing languages for unique left and right context is given as:

$$s_{Y-G_2} = \{\gamma p(mnpp)^k mnp\delta\}$$

Hence, the theorem is proved. □

Some discussion will focus on two aspects: the effect of having a non-palindromic rule and crossing site in the Y-G splicing system, and the type of the resulting splicing language. Lemma 1 is to show the produced splicing languages when two cutting sites exist in a string, while Lemma 2 is when three cutting sites exist in a string. The proving of the lemmas is given in Case 1 and Case 2. A distinct left and right context of restriction enzyme is used in Case 1, whereas in Case 2, a unique left and right context is measured. Thus, from Lemma 1 and Lemma 2, a theorem is proposed. Mathematical induction method is used to prove the theorem. It is observed that, from both lemmas, the generated splicing languages can be generalised for both Case 1 and Case 2. This implies that, as stated in Theorem 1, a generalised splicing language can be formed, as given in equation (4) and (8) for Case 1 and Case 2, respectively.

From the cases discussed in Lemma 1 and Lemma 2, it is observed that, the splicing languages is generated only in $\gamma - \delta$ sequence, which is contradict with findings in [20] that used a palindromic rule with palindromic crossing site. This is due to the non-palindromic features of the crossing site, that does not allow religation between sticky ends from the initial string and its rotational molecules. Thus, in the reaction between initial strings and non-palindromic rule and crossing site, the features of left and right context have no effect on the generated splicing languages. Next, in the context of type of splicing language, for Case 1 and Case 2 discussed in Lemma 1 and Lemma 2, the generated splicing languages are in transient language, for all $k \geq 0$ and $k \in Z^+$. Hence, as shown in Theorem 1, for any number of cutting sites exist in a single pattern of strings, the generalised *n*-cutting sites splicing languages will be in the type of transient language. Additionally, as k increases, the Y-G splicing system will produce infinitely long molecules.

Conclusions

The generalisation of *n*-cutting sites splicing languages for a Y-G splicing system is shown. The Y-G splicing system offered in this paper which is presented in two lemmas and a theorem verified that the splicing languages can be generalised. In this paper, the restriction enzyme is restricted to a non-palindromic rule and crossing site; to see the pattern of languages and how the features of the rule affect the generated languages. Lemma 1 shows a Y-G splicing system with a single pattern of strings with two cutting sites, while Lemma 2 shows a system with a single pattern of strings with three cutting sites. Two cases demonstrated the proving in Lemma 1 and Lemma 2, assuming a different feature of the context of restriction enzyme is measured. From these lemmas, a theorem is suggested. The finding of this study shows the generalisation of *n*-cutting sites splicing languages when *n*-cutting sites exist in a string. Moreover, it is discovered that the properties of the restriction enzyme in terms of its left and right contexts have no effect on the generated splicing languages because it only appears in $\gamma - \delta$ sequence. In the future, the generalisation of splicing languages can be investigated by taking other restriction enzyme characteristics into account.

Conflicts of Interest

All authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

The authors would like to express their appreciation to Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2021/STG06/UMP/02/1 (University reference RDU210107) and Universiti Malaysia Pahang Al-Sultan Abdullah under UMP Fundamental Research Grant RDU220344.

References

[1] Head, T. (1987). Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors. *Bulletin of Mathematical Biology*, *49*(6), 737-759.
[2] Paun, G. (1996). On the splicing operation. *Discrete Applied Mathematics*, *70*, 57-79.
[3] Laun, G. (1999). *Constants and splicing system.* Binghamton University, Binghamton.
[4] Yusof, Y. (2012). DNA splicing system inspired by bio molecular operations. Universiti Teknologi Malaysia, Skudai.
[5] Lim, W. H. (2015). Single stage DNA splicing system via Yusof-Goode approach. Universiti Malaysia Pahang, Kuantan.
[6] Mudaber, M. H. (2015). Persistency and permanency of two stages splicing languages based in DNA recombination process by using Yusof-Goode (Y-G) approach. Universiti Malaysia Pahang, Kuantan.
[7] Ahmad, M. A. (2016). Second order limit language and its properties in Yusof-Goode splicing system. Universiti Teknologi Malaysia, Skudai.
[8] Lim, W. L., Yusof, Y., Rosli, N. & Mudaber, M. H. (2015). Modelling the behavior of single stage splicing language: A Yusof-Goode computational approach. *Jurnal Teknologi*, *73*(1), 135-138.
[9] Mudaber, M. H., Yusof, Y., Mohamad, M. S., Ramli, A. N. M. & Lim, W. L. (2016). Modelling of two stages DNA splicing languages on de Bruijn graph. *Jurnal Teknologi*, *78*(1), 73-78.
[10] Khairuddin, M. S. H., Ahmad, M. A. & Adzhar, N. (2021). Application of automata theory on *n*-th order limit language. *Review of International Geographical Education*, *11*(4), 817-824.
[11] Ruslim, N. M., Elizabeth, M., Yusof, Y., Mohamad, M. S. & Adzhar, N. (2021). Deoxyribonucleic acid (DNA) splicing system from graph theoretic perspective. *Journal of Physics: Conference Series, 1988, 012081.*
[12] Razak, M. N. S. A., Fong, W. H. & Sarmin, N. H. (2021). Graph splicing rules with cycle graph and its complement on complete graphs. *Journal of Physics: Conference Series, 1988, 012067.*
[13] Aisah, I., Jayanti, P. R. E. & Supriatna, A. K. (2019). 2-cut splicing and 4-cut splicing on DNA molecule. *IOP Conference Series: Materials Science and Engineering, 567, 012018.*
[14] Razak, M. N. S. A., Fong, W. H. & Sarmin, N. H. (2020). Folding technique on *n*-Cut spliced semigraph in splicing system. *AIP Conference Proceedings, 2266, 060011.*
[15] Bergtrom, G. (2016). Cell and molecular biology: What we know & how we found it: Section 8.3, DNA Structure (2nd ed). USA: UWM Digital Commons 2016.
[16] Russell, P. J. (2010). iGenetics a molecular approach: Section 10, Recombinant DNA Technology (3rd ed). London: B Cummings 2010.
[17] Chargaff, E. (1968). What Really is DNA? *Progress in Nucleic Acid Research and Molecular Biology*, *8*, 297-333.
[18] Goode, E. & Pixton, D. (2004). Splicing to the Limit. Aspects of Molecular Computing, Lecture Notes Computer Science, 189-201.
[19] Pray, I.A. (2008). Restriction enzymes. *Nature Education*, *1*, 38.
[20] Ruslim, N. M., Yusof, Y. & Adzhar, N. (2022). *n*-Cutting site of DNA splicing language for single string and palindromic rule. *Journal of Physics: Conference Series, 2157, 012033.*