

PAPER • OPEN ACCESS

Deoxyribonucleic Acid (DNA) Splicing System from Graph Theoretic Perspective

To cite this article: Nooradelena Mohd Ruslim *et al* 2021 *J. Phys.: Conf. Ser.* **1988** 012081

View the [article online](#) for updates and enhancements.

A promotional banner for the 240th ECS Meeting. The banner features a colorful striped border at the top. On the left, the ECS logo is displayed in a green circle. To its right, the text reads "240th ECS Meeting" in large blue font, followed by "Oct 10-14, 2021, Orlando, Florida" in a smaller black font. Below this, it says "Register early and save up to 20% on registration costs" in bold black text, and "Early registration deadline Sep 13" in a smaller black font. At the bottom left, there is a red "REGISTER NOW" button. On the right side of the banner, there is a photograph of a diverse group of people in professional attire, smiling and clapping, suggesting a successful event or presentation.

ECS **240th ECS Meeting**
Oct 10-14, 2021, Orlando, Florida
**Register early and save
up to 20% on registration costs**
Early registration deadline Sep 13
REGISTER NOW

Deoxyribonucleic Acid (DNA) Splicing System from Graph Theoretic Perspective

Nooradelena Mohd Ruslim, Marta Elizabeth, Yuhani Yusof, Mohd Sham Mohamad and Noraziah Adzhar

Centre for Mathematical Sciences, College of Computing and Applied Sciences
Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan,
Pahang, Malaysia

Email: PSE20003@stdmail ump.edu.my, yuhani@ump.edu.my

Abstract. In 1987, Head introduced the recombinant process of deoxyribonucleic acid (DNA) splicing system through a framework of Formal Language Theory. To express this phenomenon, a comprehensive writing system of the rule sets, proper terms and definitions pioneered by Head. The analysis on the recombinant behavior of double-stranded DNA molecules, characteristics and various type of splicing systems has sparks the researcher to explore more splicing mechanism extensively. The aim of this paper is to give an exhaustive review on the application of graph in DNA splicing system.

1. Introduction

DNA was first discovered in the 1860s by a Swiss chemist named Johann Friedrich Miescher [1]. However, its role in genetic inheritance was not demonstrated until 1943. In 1953, James Watson and Francis Crick determined that the structure of DNA is a double-helix polymer, a spiral consisting of two DNA strands wound around each other [2]. The DNA are molecules which had been made from thousand pairs of nucleotides which consist of a base, phosphate and sugar [3]. Then, the structure of nucleotides in DNA are diversified by their bases namely adenine, guanine, cytosine and thymine. Following the Watson-Crick complementary, the adenine (A) pairs to thymine (T) and guanine (G) pairs to cytosine (C), and vice versa [4]. These pairing can be written as [A/T], [G/C], [C/G] and [T/A] or can be simplified as a , g , c and t respectively.

In DNA splicing system, the DNA molecules are cut and recombined with the existence enzyme to generate new molecules. A mathematical model of splicing system has been developed to depict the scenario by using the concept of formal language theory (FLT). The FLT is a field that is associated with theoretical computer science, discrete mathematics, and information macromolecules. The relationship between formal language theory and the study of informational macromolecules was introduced by Head in 1987. The Head splicing system formed a model consists of a set of alphabets that represents pairing between the nitrogenous bases, a set of rules that represent restriction enzymes and a set of initial strings represent double stranded DNA (dsDNA). The research is devoted to the study of finite sets of finite strings (called languages) of symbols chosen from a prescribed finite set which are derived in the form of alphabets.

Besides, other findings on the definitions, example of splicing system and splicing languages are discovered. The main result in [5] focused on persistent splicing system and splicing language. In addition, some examples of splicing system are provided and defined as null contexts and uniform splicing system. Moreover, the concepts of constant and strictly locally testable (SLT) language were also introduced. Equivalencies between constant and SLT with persistent and uniform splicing



languages were verified. Head proved that, a splicing process performed by a finite set of rules on a certain set of alphabets will generate the class of strictly locally testable languages.

After the phenomenon of the Head splicing system in 1987, the concept of splicing system is extensively explored. Paun splicing system which was slightly differed from Head splicing system introduced in 1996 [6]. Meanwhile, in the same year, Pixton splicing system is introduced [7]. This system is presented as a substitution operation of the splicing. Due to this, Pixton splicing system is found to be informal description of the biological recombinant. These two formalisms of splicing systems do not focus on the crossing site thus it does not work biologically, since all restriction enzymes have three main parts, and the crossing site plays the most important role in the DNA recombination process.

The theoretic progress of the splicing systems found in [5-7] has led to the formulation of the most recent development in splicing system, namely Yusof-Goode (Y-G) splicing system. This new writing notation of splicing system is introduced by Yusof in 2012, inspired from the Head splicing system [8]. Compared to Head splicing system, this new writing system presents the translucent behaviour of the DNA biological process which is inspired by the characteristics of the restriction enzyme involved in the model. Yusof *et. al* in [9] compared the four variants of splicing system. It concludes that, Head splicing system does not present the translucent way of splicing process, while the Paun splicing system does not correspond to the cut and paste biological operations as described in Head splicing system. Furthermore, it is not necessarily symmetric and reflective. Meanwhile, the Pixton splicing system is not really a formal description of the biological cut and paste phenomenon since it is presented as a substitution operation of the splicing. As Y-G splicing system is being used throughout this article, the definition will be given in the next section.

In this article, the review on the applications of graph in DNA splicing system are presented including some related definitions.

2. Preliminaries

In this section, some fundamental definitions used in this article are given. The definitions of Y-G splicing system, splicing language, limit language, adult language and graph are stated in this section. the definitions of relevant terms used in this paper are given.

Definition 1 [8]: Y-G Splicing System and Splicing Language

A splicing system $S = (A, I, R)$ consists of a set of alphabets A , a set of initial strings I in A^* and a set of rules, $r \in R$ where $r = (u; x, v: y; x, z)$ for left pattern or $r = (u, x; v: y, x; z)$ for right pattern. However, the notation $r = (u, x, v: y, x, z)$ indicates that both patterns of rules were applied on DNA string. For $s_1 = auxv\beta$ and $s_2 = \gamma yxz\delta$ elements of I , splicing s_1 and s_2 using r produces the initial string I together with $auxz\delta$ and $\gamma yxv\beta$, presented in either other where $\alpha, \beta, \gamma, \delta, u, x, v, y$ and $z \in A^*$ are the free monoids generated by A with the concatenation operation and 1 as the identity element. A language L is a splicing language if there exists a splicing system S for which $L = L(S)$.

As the limit language is explored in this study, thus the definition of the term is presented as follows.

Definition 2 [10]: Limit Language

A limit language is the set of words that are predicted to appear if some amount of each initial molecule is present, and sufficient time has passed for the reaction to reach equilibrium state, regardless of the balance of the reactants in a particular experimental run of the reaction.

The next definition is the adult language, a type of language that also has been discussed in this article.

Definition 3 [11]: Adult Language

Adult language (also called as inert) in a splicing language, are strings in a splicing system which cannot be used for further splicing. Adult molecules show a steady increase in quantity throughout the reaction and do not involved in further interactions with other molecules or enzymes. These molecules therefore lie in the adult language.

Finally, the definition of the main focus in this article is defined.

Definition 4 [12]: Graph

A graph is a pair consisting of two sets where the first set is the non-empty set of vertices and the other set is the set of edges. A graph can be denoted as $G = (V, E)$ where V is the set of vertices and E is the set of edges connecting between vertices.

3. Graph and its application in DNA splicing system

The notion of splicing system was introduced by Head in 1987. This pioneer paper revealed the recombination process of dsDNA molecules when it is reacted with restriction enzymes and forming new molecules with the existence of ligases. In this section, graph splicing system is discussed. Besides, the new generalization on the concept of graph called semigraph is presented, followed by exploration in limit and de Bruijn graphs.

3.1 Graph splicing system

DNA sequences are found to be in three-dimensional objects in a three-dimensional space. Problem arises when this sequence is discussed in one-dimensional space. Thus, to overcome the complexity of dsDNA structure that is originally in three-dimensional structure and involved in a complex biochemical process of three-dimensional macromolecules in the three-dimensional space strings, Freund [13] and [14] suggested graph as a more appropriate approach to model the splicing process by introducing graph splicing system. The model of graph splicing system illustrates DNA splicing process where the vertices represent DNA molecules. The splicing rules that represent the restriction enzymes are used in this splicing scheme. Regular graph splicing system on graphs discussed in this article exhibited relations with splicing systems on strings. From the examples illustrated in [13], it shows that every DNA splicing system can be expressed through a suitable graph splicing scheme. In this new proposed model, it described the splicing systems on strings within a uniform framework. It is evident that, regular graph splicing schemes works on linear graphs corresponds to the splicing systems that works on strings and can also be implemented on circular strings. Hence, regular graphs splicing schemes build up a uniform framework for splicing schemes on strings and circular strings. From examples discussed in [13], it shows that graph splicing system can produce graph languages representing non-context free string languages, in which the length growth of the linear graphs derivable with a graph splicing system can only be linear. This shows that graph splicing system is able to capture non-context free features. This itinerary graph splicing systems will lead to the investigation in different variants of non-regular graph splicing schemes working on strings and circular strings in the future.

3.2 Semigraph

In year 2000, generalization on the concept of graph is made when a new notion called semigraph is introduced [15]. Semigraph holds various properties through its structure in describing fundamental operations in graph theory. Then, in 2011, the exploration on the graph splicing system continues with work by [16]. The main objective of this work is to correlate the graph splicing system in [13] with semigraphs introduced in [15]. They introduced semigraph folding technique for DNA splicing system and proved that any n -cut spliced semigraph for any value of n except 1 can be folded onto an edge and two semiedges at the maximum of four semigraph folding. The cutting is conducted for 1-cut and 2-cut and it is shown that the Euler's Polyhedral formula for splicing system is satisfied. This is consistent with the theorem proposed and proved by induction on n -cut splicing. After the application of n -cut splicing, the semigraph is then transformed into bipartite semigraph. This is relevant to the theorem proposed, where for every n -cut splicing will form two bipartite semigraphs. Thus, with the application of semigraph, the resulting language can be seen clearly.

In 2019 and 2020, the work on semigraph continued by Aisah [17] and Razak [18] respectively. In [17], the author discussed on 2-cut splicing and 4-cut splicing. Through the splicing process, the reaction of restriction enzyme on DNA molecules in splicing system can be visualized in term of semigraph

structure. This process led to generation of new words. The language generated from the splicing process in the form of DNA molecules, (a, g, c, t) is recorded into the binary alphabet (a, b) , which then is mapped into a matrix form. The 2-cut splicing had been carried out on 3' overhang and 5' overhang sticky ends. These two cutting fulfilled the Euler's Polyhedral formula and produced two semigraphs, which then had been presented in two bipartite semigraphs. The generated language on n -cut spliced semigraph followed context-free grammar in the Greiback normal form is mapped into an upper triangular matrix called Parikh matrix. This also holds for 4-cut splicing.

In contrast with research by Aisah, Razak in his article suggested that, a type of semigraph called n -cut spliced semigraph can be obtained. The investigation focused on the application of folding technique on n -cut spliced semigraph, to form a new semigraph. In the article, the process of folding technique is observed for $n = 1, 2$ and 3. The idea of n -cut splicing was initiated to illustrate the cleavage site of splicing system with verdict of n -cut splicing will generate two components of n -cut spliced semigraphs by considering $n + 2$ bonds to be cut [16]. By using vertex mapping, it is observed that the maximum number of folding for every n -cut spliced semigraph is $n+1$ times. This can also be shown by applying a proposed algorithm in [18].

In these two articles, the number of cutting that is the n -cut are differed. In [17], the cutting is focused for $n = 2$ and 4, while in [18], the n -cut is for $n = 1, 2$ and 3. The representation of the cutting is in the form of bipartite graph, for 5' overhang and 3' overhang sticky ends molecules in [17], while [18] illustrated examples for $n = 1, 2$ and 3 by considering the 5' overhang sticky end on n -cut spliced semigraph. From these two articles, the similarity is found in the number of cutting, 2-cut. The words generated leads to the Parikh's concept, where the main structure of DNA molecules can be mapped into binary alphabet (a, b) . The Parikh matrix known as an upper triangular matrix is resulted from string mapping which formed from the binary alphabet into matrix form. With Parikh matrix, the semigraph procedure can finally generated 35 conceivable strings (languages) [17]. Though, the exploration using folding technique gave the number of resulted semigraph which later can be extended in finding the number of resulted languages. The search may also be extended by considering other n -cut splicing since the restriction enzyme that recognize a specific site in dsDNA molecules has a length up to 6 base pairs [17].

The application of n -cut and bipartite semigraph can be seen in [16] and [17]. From these two articles, the researchers considered the number of cutting, n is up to 4. It is shown that, for $n = 2$, the number of bipartite semigraph is equivalent to 2 for both articles. This proved the theorem stated in [16]. However, Aisah had extended the cutting by considering $n = 4$. With 4-cut splicing, the semigraph is also resulted in the form of two bipartite semigraph. Besides, Razak has applied the same n -cut splicing by using the folding technique where the n -cut spliced semigraph is used to study the concept of graph folding. The exploration on semigraph can be extended with other number of cutting by considering other properties of rules.

3.3 Limit and de Bruijn graph

Other than the utilization of semigraph in DNA splicing system, the concept of limit graph is extensively explored in this research. The introduction of limit graph which shown the sequence of splicing process had been initiated in 2004. Earlier in 2004, the modelling of limit graph at an equilibrium state when the experiment has run to the completion was explored [10]. Also, the notion of limit language is first defined in this work and can be termed as the limit molecules of the system, that are left at an equilibrium state and corresponds to the formal language defined by the splicing system.

The concept of limit graph is further extending with theorems, examples and wet-lab experiments in order to identify the behaviour of splicing language. The pioneer framework of Formal Language Theory introduced by Head [5], led to the existence of other splicing system models, namely Paun [6], Goode-Pixton [7] and Y-G [8], which inspired by the molecular biological process of DNA splicing. Hence, the objective to focus on the translucent DNA splicing process which is particularly in generating languages by adopting the concept of limit graph had been explored in [8]. It is theoretically analyzed the non-uniqueness of limit language which has been defined in [10]. The non-uniqueness of limit

language is shown when Yusof introduced active persistent language [8]. This extended language is a set of strings that participate in further splicing and also contained in the limit language. Through theorem, Yusof presented the characteristics of the limit graph which involves limit language. It is found that, to have an active persistent language in limit graph, all vertices that represent strings in the language must have 'in' and 'out' edges between the vertices in the same component of a graph.

Additionally, to overcome the complexity of the limit graph with regards to the increment of the number of vertices and edges which associated to the increase of number of splicing languages generated, the limit adjacency matrix is applied [19]. Lim in [19] had extended the study in [8] by introducing the concept of single stage limit language. The limit graph concepts that applied in limit adjacency matrix is utilized to determine the behaviour of single-stage splicing language. This method is found to be more appropriate and effective tool for investigating the complexity of dsDNA recombination process and determining the limit language. Also, in 2015, Lim had developed a software programming by using Visual Basic program. This GUI is to determine the behaviour of single-stage splicing language wherein the output was found to validate the outcomes of wet lab experiment [20].

The single stage proposed by [19] is extended. Two stages splicing languages which is the union of stage 1 and stage 2 splicing languages is then presented in Yusof-Goode (Y-G) notation [21]. The application in graph theory for two stages splicing languages is presented in the de Bruijn graph. The connection between mathematics and DNA molecular by presenting the connection as de Bruijn graph in which every edge of the de Bruijn graph gives a k -mers composition of DNA molecules and also each path of the de Bruijn graph gives a DNA sequence and vice-versa is shown [22]. The de Bruijn graphs that represent the DNA strands are constructed according to the crossing sites of restriction enzymes used in the study. Two cases were considered; two rules with palindromic disjointed crossing sites and two rules with palindromic identical crossing sites. Besides, the suggested model also can predict the persistency, permanency, non-persistency and non-permanency of two stages DNA splicing languages. In reality, there are mutually relations between paths of the de Bruijn graph and DNA sequences, so that every edge of the graph presents a k -mer composition of the recombinant DNA molecules as well as every path of the graph gives a unique sequence of the DNA and vice-versa.

4. Result and discussion

The application of graph in predicting and performing the splicing languages namely active persistent language, single stage limit language and two stages splicing languages are discussed in this section.

4.1 Connection between graph splicing system and de Bruijn graph

This section discusses the relationship between DNA sequence and graph. It is evidenced that graph is a suitable method to represent the sequence of DNA which is found to be in three-dimensional structure [13]. Furthermore, the introduction of graph splicing system helps to illustrate DNA splicing process by representing the DNA molecules as vertices. Freund in the article exhibited relations between regular graph splicing system and splicing system on strings. The proposed model described splicing system on linear string as well as on circular strings within a uniform framework. This model particularly focuses on DNA molecules that is the initial string. With the reaction of splicing rule on DNA molecules, the graph will split into two parts according to certain cutting depends on the recognition site found in the restriction enzymes. Later, the graph will recombine at the identical crossing site, producing new molecules. However, in the de Bruijn graph proposed by [22], the composition of graph is based on the generated splicing languages. It is to see the connection between mathematics and DNA molecular. In the proposed two cases, splicing languages are generated via Y-G approach. The splicing process taking into consideration the number of cutting site in initial strings, crossing and context factor of restriction enzymes. From this model, the type of splicing system namely persistent and permanent can be recognized. The de Bruijn graph will construct all sets of stage one and stage two splicing languages as the two stages splicing languages examined in this article is the union of stage one and stage two splicing languages. Nonetheless, challenging task awaits when more considerations should be taken into account.

Variants of graph splicing scheme that cannot be interpreted as linear or circular graphs is something to investigate while the characteristics of restriction enzymes can be included in both studies, for example by using molecular example with non-palindromic rules, same or different crossing sites and different number of initial strings involved in the process.

4.2 The development of semigraph

In this paper, the new generalization of graph called semigraph is discussed. This model is extensively explored by researchers with different number of cutting and different technique in representing the cutting process. Thus, to see the connection between [17] and [18], this article uses an example taken from [17] that is for $n = 4$. The semigraph for 4-cut splicing is then presented in folding technique applied in [18], to see if the proposition holds for 4-cut splicing.

Example 1 [17]

With the existence of certain enzyme, a dsDNA molecule that produce 5'-overhang is used to illustrate the 4-cut splicing. It is described in graph form according to the graph splicing system and produce the following graph.

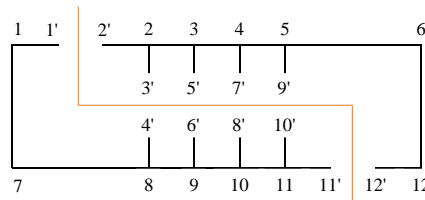


Figure 1. Spliced Semigraph for 4-cut (SSG_4).

Adapting folding technique by [18] in this example, the 4-cut splicing by [17] is transformed into spliced semigraph.

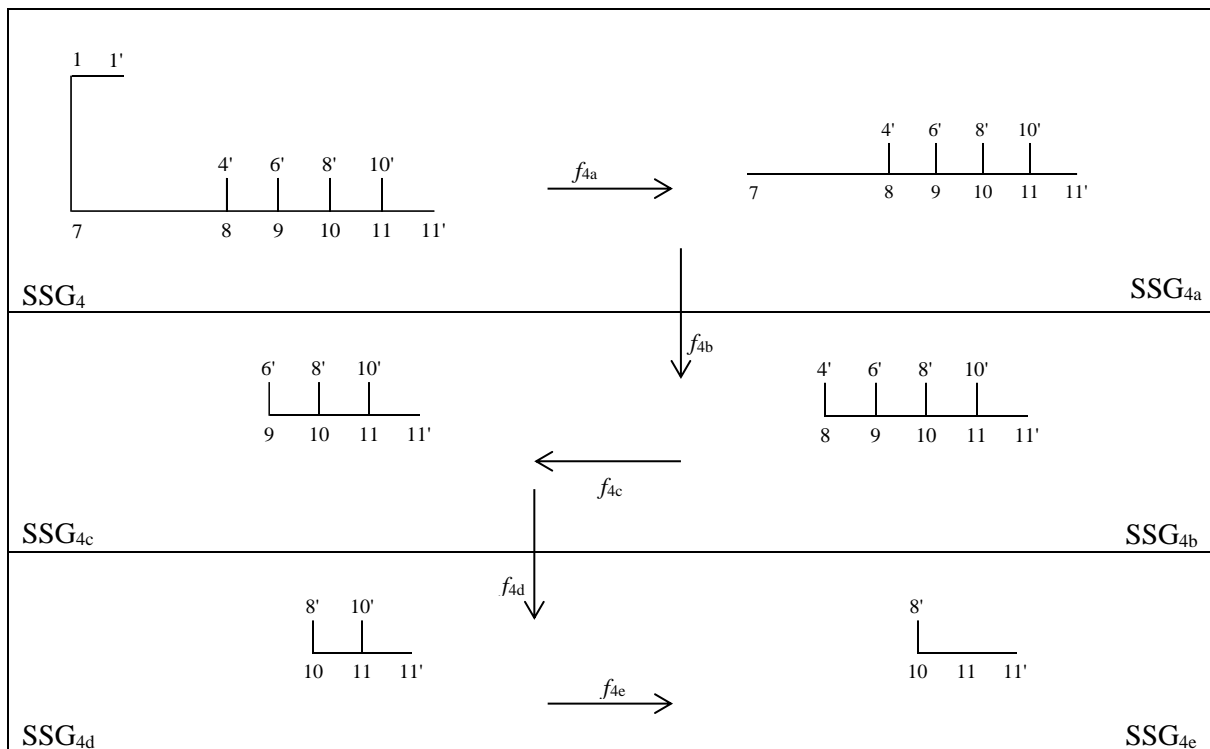


Figure 2. Folding of SSG_4 into a new subgraph SSG_{4e} through $f_{4a}, f_{4b}, f_{4c}, f_{4d}$ and f_{4e} .

From this example, the bipartite semigraph shown in [17] for 4-cut is transformed in graph folding using folding technique on 4-cut in [18]. The resulted semigraph, SSG_{4c} is in the form $\eta(S) = (2,1,2,2)$ since it consists two vertices, one edge, two semivertices and two semiedges. This example holds the proposition stated in [18], that, for $n \geq 2$, the semigraph will fold $n + 1$ times. A complete splicing process is discussed in [17]. The semigraph shows the cutting process of dsDNA with the reaction of restriction enzyme on specific cutting site. With the existence of ligase, the molecules can recombine. Parikh's concept is utilized to see the resulting language from the process.

However, the process of graph folding is not extended to the generation of possible splicing languages in [18]. Thus, the connection between the folding technique and the generation of splicing language is unseen. The study may be extended to complete the splicing process. In these, the discussion involved one string with one cutting site. Further investigation should be conducted by considering more than one initial string or more than one rule with different length of cutting site.

4.3 Continuation in the study of limit language

Besides semigraph, a number of researchers also explored limit graph in DNA splicing system [8], [23]. This method is meant to identify the behaviour of splicing language when the characteristics of limit language is translated into limit graph. However, with the complexity of splicing language gained from the splicing process, new study has been carried out using limit adjacency matrix. This approach is found to be appropriate in determining the limit language [19]. In [19], the concept of single stage limit language is introduced. The application of directed graph in [8] clearly show the existence of active persistent language. Thus, Example 2 is presented to show the presence of active persistent language introduced by [8] in single stage limit language proposed in [19].

Example 2

Let $S = (A, I, R)$ be a Y-G splicing system consists of an initial string, $I = \alpha gatc\beta$ and palindromic rule $(g; at, c: g; at, c)$. Therefore, the splicing language generated by using Y-G approach is presented as follows:

$$\alpha gatc\beta, \alpha gatc\beta', \beta' gatc\beta$$

These three types of splicing language can be presented as a directed graph as shown in figure 3.

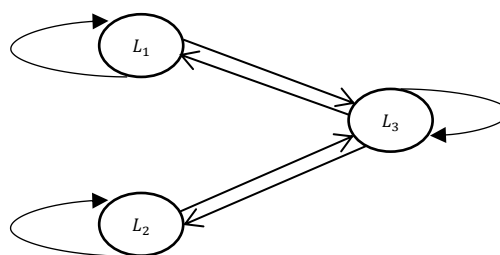


Figure 3. Directed graph of generated splicing language for palindromic rules.

As presented in figure 3, the vertices L_1, L_2 and L_3 represent $\alpha gatc\beta', \beta' gatc\beta$ and $\alpha gatc\beta$ respectively. This figure shows that all three types of splicing language have 'in' and 'out' edges between the vertices. According to [8], those vertices are active persistent language, which lie in limit language. Hence, splicing system with one initial string and one palindromic rule with palindromic cutting site generate three types of splicing languages, $\{\alpha gatc\beta, \alpha gatc\beta', \beta' gatc\beta\}$, which satisfies theorem stated in [19]. The existence of active persistent language is as illustrated in figure 3. Vertices L_1, L_2 and L_3 has 'in' and 'out' edges between the vertices which suggests that L_1, L_2 and L_3 are active persistent language.

However, in some circumstances discussed in [8], there exist vertices that have ‘in’ and ‘out’ edges which are classified as transient language. Thus, it is clear that, the characteristics of vertices in directed graph is a suitable method to show the existence of active persistent language and classifying the vertices as either limit or transient languages.

Limitation in presenting the limit language in directed graph for more than one initial string brought the research in [19] in the modelling of limit adjacency matrix. The limit adjacency matrix clearly shows the existence of inert persistence language, that is an adult language that has been renamed by [8]. Both inert and active persistent languages are subset of single stage limit language [19]. Thus, these two types of languages together with transient language exist in the limit adjacency matrix with its own properties in the matrix. Therefore, it is found that, matrix approach gives a better representation of complex DNA splicing process.

5. Conclusion

In this review article, a brief introduction on the structure of DNA, history of FLT introduced by Head and the development in splicing systems are presented. In addition, some definitions which are related to the focus of this article are provided. Furthermore, a detail discussion on the graph and its application in DNA splicing system is conferred. To see the development of graph application in DNA splicing system, several approach, namely graph splicing system, semigraph, limit graph, limit adjacency matrix and de Bruijn graph has been discussed. A massive development in this field from the biological aspect is well supported with the connection of mathematical modelling in it. Laboratory experiment conducted in determining the splicing languages is both money and time consuming, thus mathematics plays its vital role in the prediction of generated languages [8], [20]. The analysis of DNA splicing behaviours or results can be predicted through mathematical modelling as an alternative to wet lab experiment which are time consuming and costly. However, with application of mathematical modelling, it is shown that, this field is rapidly developed. This article also discussed the connection and continuity between the graph models in which the deficiencies can be covered in future work. Review on other type of graph or any other suitable methods applied in DNA splicing system can be conducted in the near future.

Acknowledgement

The authors would like to express their special thanks to University College of Yayasan Pahang (UCYP) for supporting this research.

References

- [1] Dahm R 2005 Friedrich Miescher and the discovery of DNA *Developmental Biology* **278** 274–288
- [2] Yusof Y, Sarmin N H, Goode T E, Mahmud M and Fong W H 2011 An extension of DNA splicing system *Proceedings of 6th International Conference on the Bio-Inspired Computing: Theories and Applications (BIC-TA)* pp 246–248
- [3] Amos M, Paun G and Rozenberg G 2002 Topics in the theory of DNA computing *Theoretical Computer Science* **287**(1) 3–38
- [4] Tamarin R H 2001 *Principle of Genetics* USA (The McGraw-Hill Companies) chapter 9 p 218
- [5] 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
- [6] Păun G 1996 On the splicing operation *Discrete Applied Mathematics* **70**(1) 57–79
- [7] Pixton D 1996 Regularity of Splicing Languages *Discrete Applied Mathematics* **69**(95) 101–124
- [8] Yusof Y 2012 DNA Splicing System Inspired by Bio Molecular Operations (Universiti Teknologi Malaysia) chapter 3-6 pp 25-83
- [9] Yusof Y, Sarmin N H, Fong W H, Goode T E and Ahmad M A 2013 An Analysis of four variants of splicing system *AIP Conference Proceedings* **1522** pp 888-895
- [10] Goode T E and Pixton D 2004 Splicing to the Limit *Aspects of Molecular Computing* **2950** pp

- 189-201
- [11] Sarmin N H and Fong W H 2006 Mathematical Modelling of Splicing Systems *Proceedings of the 1st International Conference on Natural Resources Engineering & Technology* pp 524–527
 - [12] Rosen K H 2012 *Discrete Mathematics and Its Applications* (McGraw-Hill Companies) chapter 10 p 641
 - [13] Freund R 1995 Splicing Systems on Graphs *Proceedings of the 1st International Symposium on Intelligence in Neural and Biological Systems* **5**(7) pp 189–194
 - [14] Jeganathan L and Rama R 2007 *Graph Splicing System* pp 1-12
 - [15] Sampathkumar E 2000 Semigraphs and Their Applications *Report on the DST Project*
 - [16] JeyaBharathi S, Padmashree J, Selvi S S and Thiagarajan K 2011 Semigraph Structure on DNA Splicing System *Sixth International Conference on Bio-Inspired Computing: Theories and Applications* **V** pp 182-187
 - [17] Aisah I, Jayanti P R E and Supriatna A K 2019 2-cut Splicing and 4-cut Splicing on DNA Molecule *IOP Conference Series: Materials Science and Engineering* **567** pp 1-8
 - [18] Razak M N S A, Fong W H and Sarmin N H 2020 Folding Technique on n -cut Spliced Semigraph in Splicing System *AIP Conference Proceedings* **2266** pp 1–10
 - [19] Lim W L 2015 Single-Stage DNA Splicing System via Yusof-Goode Approach (Universiti Malaysia Pahang) chapter 3-5 pp 22-82
 - [20] Lim W L, Yusof Y, Rosli N and Mudaber M H 2015 Modelling the Behaviour of Single Stage Splicing Language : A Yusof Goode Computational Approach *Jurnal Teknologi* **73**(1) 135–138
 - [21] Mudaber M H, Yusof Y and Mohamad M S 2014 Some Relations between Two Stages DNA Splicing Languages *Proceedings of the 3rd International Conference on Mathematical Sciences* **1602** pp 254–259
 - [22] Mudaber M H, Yusof Y, Mohamad M S, Ramli A N M and Lim W L 2016 Modelling of Two Stages DNA Splicing Languages on de Bruijn Graph *Jurnal Teknologi* **78**(1) 73-78
 - [23] Yusof Y, Lim W L, Goode T E, Sarmin N H, Fong W H and Wahab M F A Molecular aspects of DNA splicing system *AIP Conference Proceedings* **1660** pp 1–8