Abstract. Graph has many applications both in mathematics and various other disciplines especially to describe model problem. Graph can be used to describe the process of splicing system in DNA molecule. Splicing system is a mathematical model of the system in the action of restriction enzyme in DNA molecule. Splicing language is a language constructed by splicing system. In this paper we discussed about 2 cut splicing and 4 cut splicing on DNA molecule. It had been recognized that at the time of splicing, the graph takes the semi graph structure and it generates the words which leads to the Parikh concepts. Since a DNA which consist of \((A, C, G, T)\) can be seen as a word, thus the modelling can be done within the framework of formal language theory. After the splicing process, the main element of the DNA molecule is mapped into the binary alphabet \((a, b)\). The language formed from the binary alphabet is mapped into a matrix form.

1. Introduction
Splicing system is a mathematical model of the system in restriction actions on DNA molecules. The concept of splicing systems is introduced as a tool to represent string restructuring that occurs during linear polymer deviation in the presence of specific enzyme activity [1].

DNA or deoxyribonucleic acid is a molecule that carries the genetic information in living cell. DNA sequences are three-dimensional objects are more easily described using graphs [2]. So that the main structure of the DNA molecule \((A, G, T, C)\) can be mapped into the binary alphabet \((a, b)\). The Parikh matrix is a type of upper triangular matrix. The Parikh matrix results from string mapping formed from the binary alphabet into the matrix form introduced by Mateescu [3].

2. Theoretical Model
2.1. Genetic
DNA or deoxyribonucleic acid is a molecule that carries the genetic information in living cell. Those are sugar (deoxyribose sugar), a phosphate, and a base. There are four type of base; adenine, thymine, cytosine, and guanine, abbreviated by \(A\), \(T\), \(C\), and \(G\) respectively.

Two single strands of DNA molecule can bind together to form a double stranded DNA molecule. The two strands hold together by (hydrogen) bonds between their bases. These bonds hold between complementarity’s bases. That is between \(A\) in one strand and \(T\) in the other, and between \(C\) in one strand and \(G\) in the other.

2.2. Restriction enzyme
Restriction enzymes recognize specific and continuous subpart, called restriction sites, of a double stranded DNA molecule and cut it at precisely specifiable positions. Each enzyme has a series of 4-6
specific base pairs contained in a DNA strand. Each restriction enzyme has a different and specific cutting recognition site especially in the nitrogen base. There are two kinds of restriction endonuclease enzymes which produce blunt ends and sticky or cohesive ends. There are two kinds of sticky ends, first the 5'-end which results in cutting results extending at the 5'-end and 3'-end which results in cutting lengthwise at the 3'-end.

| Enzyme | Restriction sites | Enzyme | Restriction sites |
|--------|-------------------|--------|-------------------|
| TaqI   | T CGA             | HindIII| A AGCTTT          |
| RsaI   | GT AC             | PvuI   | CAG CTG           |
| HhaI   | GCG C             | PvuII  | GGTAC C           |
| BsaRI  | GCG CC            | KpnI   | GGTAC C           |
| EcoRI  | G AATTTC          | BclI   | TGG CCA           |
| EcoRV  | GAT ATCC          | Smal   | CCC GGG           |
| BamHI  | G GATCC           | BssHI  | G CGGCC           |
| PacI   | A CATGT           | PstI   | CGTA G            |
| Acc65I | G GTACC           | MfeI   | C ATTAG           |
| MluI   | A CGCGGT          | AvrII  | C TAGGG           |
| BsrAI  | GCTAG C           | BspHI  | T CATGA           |

2.3. Context-free Grammar

A context-free Grammar \( G = (V_N, V_T, P, S) \) where \( V_N \) is a finite set of variables or nonterminal, \( S \) is a special element of \( V_N \) called the starting symbol, \( V_T \) is a finite set of terminals, and \( P \) is a finite set of production. Each production is of the form \( a \rightarrow b \) with \( a \) in \( V_N \) and \( b \) in \( (V_N \cup V_T)^* \).

A grammar \( G \) is a set of rules for constructing a language. The constructed language called the language build by the grammar, is denoted by \( L(G) \).

The language on n-cut spliced semi graph is generated by following context-free grammar in the Greiback normal form,

\[
G_{4593} \rightarrow G_{45} , G_{45} \rightarrow G_{53} G_{45} G_{54} | G_{54} , G_{54} \rightarrow G_{38}
\]

2.4. Graph

Definition 2.1: Suppose \( G \) is a graph, when splicing \( G \), a new vertices is obtained which is called semi-vertices denoted by \( V' \), where \( |V'| = p' \).

Semi-Edges

Definition 2.2: Let \( G \) be a graph when splicing \( G \), we obtain new edges by decomposition of edges which are called as semi–edges denoted by \( E' \), where \( |E'| = q' \).

Bipartite Semi graphs

Definition 2.3: \( G \) is bipartite if its vertex set \( V \) can be partitioned into sets \( \{V_1, V_2\} \) such that both \( V_1 \) and \( V_2 \) are independent.

2.4.1. Euler’s Polyhedral formula for Splicing System

The Euler’s polyhedral formula for DNA spliced semi graph is \( P - (q + q') + r = k + 1 \), where \( P \) is the total number of vertices and semi vertices in the semi graph, \( q \) and \( q' \) are the number of edges and semi edges, \( r \) is the number of faces of semi graph and \( k \) is the number of components.

2.4.2. Parikh Matrices

Let \( \Sigma = \{a_1, a_2, a_3, ..., a_k\} \) with \( a_1 < a_2 < a_3 < \cdots < a_k \) are regular a alphabet. \( \Psi_k : \Sigma^* \rightarrow M_{k+1} \) is morphism defined by: \( \Psi_k(\lambda) = I_{k+1} \) and \( f \Psi_k(a_q) = \)
\[(m_{ij})_{1 \leq i, j \leq (k+1)}, \text{ then for each } 1 \leq i \leq (k + 1), \ m_{i,i} = 1, \ m_{q,q+1} = 1, \ all \ other \ elements \ of \ the \ matrix \ \Psi_k(a_q) \ are \ 0.\]

For a word \(w = a_1a_2a_3 \ldots a_m, \ a_j \in \sum\) for \(1 \leq j \leq m\) we have

\[\Psi_k(w) = \Psi_k(a_1)\Psi_k(a_2)\Psi_k(a_3) \ldots \Psi_k(a_m).\]

2.5. Norm of Parikh Matrices

A function \(\| \cdot \|: \mathbb{R}^{m,n} \rightarrow \mathbb{R}\) is called a matrix norm on \(\mathbb{R}^{m,n}\) if for all \(A, B \in \mathbb{R}^{m,n}\) and all \(\alpha \in \mathbb{R}\)

i) \(\| A \| \geq 0\) with equality if and only if \(A = 0\) (positivity),

ii) \(\| \alpha A \| = |\alpha|\| A \|\) (homogeneity),

iii) \(\| A + B \| \leq \| A \| + \| B \|\) (subadditivity),

iv) \(\| AB \| \leq \| A \| \| B \|\) (submultiplicativity).

Theorem 2.5: For \(A \in \mathbb{R}^{m,n}\) we have

i) \(\| A \|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|\) (max column sum)

ii) \(\| A \|_{\infty} = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|\) (max row sum)

3. Result

3.1. 2-Cut splicing on DNA molecule with strand 5'-end and 3'-end of sticky ends overhang

If the double strand DNA molecule’s restriction site is 4 bases in length and produce lengthwise cuts in the 5'-end and 3'-end is described in graph form according to the graph splicing scheme that has been given will produce the following graph.

Figure 4.1: Representation Sticky Ends Overhang 3'-end (Left) and Overhang 5'-end (Right) with L=4

Because the Euler's Polyhedral formula \(P - (q + q') + r = 16 (6 + 8) + 1 = 2 + 1 = 3 = k + 1\) is fulfilled, Figure 4.1 is 2-cut splicing. Based on the 2-cut splicing in Figure 4.1, two semigraphs are obtained, as follows

Figure 4.2: 2-cut spliced semi graph
Let the weight of each edge of a spliced semi graph as ‘b’ then the weight of each semi edge becomes \( \frac{b}{2} \). After 2–Cut splicing so we have spliced semi graphs in the form of two bipartite semi graphs. Each one has four semi edges and three edges. Therefore let the language be \( \left\{ a^{n}b^{n} \right\} \), Therefore the language obtained is \( L = \left\{ \frac{b}{2} \right\} b^{3} = \left\{ a^{n}b^{n} \right\} a = \frac{b}{2} \).

Figure 4.3: Representation of 2-cut spliced bipartite semi graph

Let \( G = (V_{N}, V_{T}, P, S) \) be the grammar defined on the languages on edges, where \( V_{N} = \{ S \}, V_{T} = \{ a, b \}, P = \{ S \rightarrow (Aa, b), A \rightarrow a \} \), the set of productions P is given from the language \( L(G) = \left\{ a^{n}b^{n} \right\} a = \frac{b}{2} \).

Since the language generated on the splicing graph such as \( L = L(G) = \left\{ a^{n}b^{n} \right\} a = \frac{b}{2} \), than \( L \) is splicing language. Based on the \( L \), we obtained \( V_{T} = \{ a, b \} \) with \( a < b \). The whole set of strings formed from the initial string on the \( L \) is \( V_{T}^{*} \) as much \( \frac{71}{413} = 35 \) strings.

The Parikh matrix mapping \( \Psi_{2} \) is a mapping from \( V_{T}^{*} \) to \( M_{3} \) with

\[
\Psi_{2}(a) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Psi_{2}(b) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}
\]

\( \Psi_{2}(ab) = \Psi_{2}(a)\Psi_{2}(b), \ a, b \in V_{T}^{*} \), where multiplication of matrices is the operation on the right side of this equation.

**Table 2.** Norm of Parikh Matrices for the 1-Cut splicing system words

| Word       | \( \Psi(w_{i}) \) | \( ||\Psi(w_{i})||_{1} \) | \( ||\Psi(w_{i})||_{\infty} \) |
|------------|-------------------|-------------------------|-------------------------|
| \( w_{1} = aaaaabbb \)  | \( \begin{bmatrix} 1 & 4 & 12 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \) | 16 | 17 |
| \( w_{2} = aaababb \)  | \( \begin{bmatrix} 1 & 4 & 11 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \) | 15 | 16 |
| \( w_{3} = aababb \)  | \( \begin{bmatrix} 1 & 4 & 10 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \) | 14 | 15 |
| \( w_{4} = aaabbaa \)  | \( \begin{bmatrix} 1 & 4 & 9 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \) | 13 | 14 |
| \( w_{5} = abaaabb \)  | \( \begin{bmatrix} 1 & 4 & 8 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \) | 12 | 13 |
Norm of Parikh matrices produced on 2-cut splicing in Table 2 has a range $4 \leq \|\Psi(w)\|_1 \leq 16$ and $5 \leq \|\Psi(w)\|_\infty \leq 17$ and $\|\Psi(w)\|_\infty = \|\Psi(w)\| + 1$.

3.2. 4-Cut splicing on DNA molecule with strand 5’-end and 3’-end of sticky ends overhang

If a 6-base-length, double-stranded DNA restriction site that produces longitudinal pieces at 5’-end and 3’-end is describe in graph form according to the splicing chart scheme that has been given, it will produce the following graph.

Consider $L = 6$, the number of set of combinations required for 4-Cut splicing. Here $P = 24$, $q = 10$, $q' = 12$, $r = 1$, $k = 2$. Then $P - (q + q') + r = 24 - (10 + 12) + 1 = 2 + 1 = 3 = k + 1$ therefore is satisfied in 2-Cut splicing, as described below.
Let the weight of each edge of a spliced semi graph as ‘b’ then the weight of each semi edge becomes \( \frac{b}{g^3 + 9} \). After 4–Cut splicing so we have spliced semi graphs in the form of two bipartite semi graphs. Each one has six semi edges and five edges. Therefore let the language be \( \left( \frac{b}{g^3 + 9} \right)^6 b^5 \). Therefore the language obtained is \( \left( \frac{b}{g^3 + 9} \right)^6 \left( \frac{b}{g^3 + 9} \right)^5 \left( \frac{b}{g^3 + 9} \right)^3 \).

The language of the grammar of \( \left( \frac{b}{g^3 + 9} \right)^6 \left( \frac{b}{g^3 + 9} \right)^5 \left( \frac{b}{g^3 + 9} \right)^3 \) obtained from the set of result P, where \( V_n = \{S\} \), \( V_r = \{a, b\} \), \( P = \{S \rightarrow (A5b, Aab), A \rightarrow a\} \), is \( L(G) = \left\{ a^6b^5 | a = \frac{b}{g^3 + 9} \right\} \).

Since the language generated on the splicing graph such as \( L = L(G) = \left\{ a^6b^5 | a = \frac{b}{g^3 + 9} \right\} \), then \( L \) is splicing language. Based on the \( L \), we obtained \( V_r = \{a, b\} \) with \( a < b \). The whole set of strings formed from the initial string on the \( L \) is \( V_r^* \) as much \( 0.51 = 642 \) strings.

The Parikh matrix mapping \( \Psi_{g^3 + 9} \) is a mapping from \( V_r^* \) to \( M_3 \) with \( \Psi_{g^3 + 9}(a) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \), \( \Psi_{g^3 + 9}(b) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \).

\( \Psi_{g^3 + 9}(ab) = \Psi_{g^3 + 9}(a)\Psi_{g^3 + 9}(b) \), \( a, b \in V_r^* \) where multiplication of matrices is the operation on the right side of this equation.

Norm of Parikh matrices produced on 4-cut splicing has a range of \( 6 \leq ||\Psi(w)||_1 \leq 36 \) and \( 7 \leq ||\Psi(w)||_\infty \leq 37 \) and \( ||\Psi(w)||_\infty = ||\Psi(w)|| + 1 \).

Max column sum norm of Parikh matrix obtained by calculating the number of all entries in the third column of each matrix produced by the string where the first entry is the number of \( ab \) in the string, the second entry is the number of letters \( b \) in the string, and the third entry is 1. And max row sum norm of Parikh matrix obtained by calculating the number of all entries in the first row of each matrix produced by the string where the first entry is 1, the second entry is the number of letters \( a \) in the string, and the third entry is the number of \( ab \) in the string. The number of \( a \) in each string is \( n +
2, the number of $b$ is $n + 1$, and the number of substrings $ab$ is $0 \leq |ab| \leq |a||b|$, with $n$ the number of splicing ($n \geq 1$).

4. Conclusion

DNA molecules hold the characterization of the semi graph at the time of splicing. By using formal language and with the help of Algebra software, we can see the amount of splicing on DNA molecule from the norm value of the Parikh matrix. The result of norm of Parikh matrices is $n + 2 \leq ||\Psi(w)||_1 \leq (n + 2)^2$ where $n$ the number of splicing ($n \geq 1$) and $||\Psi(w)||_\infty = ||\Psi(w)||_1 + 1$.

Acknowledgments

The authors thank to the University of Padjadjaran, due to the funding of this work through to the scheme of Academic Leadership Grand (ALG) with contract number 2297/UN6.D/KS/2018.

References

[1] Head, T. Splicing Representations of Strictly Locally Testable Languages. Disc. Applied Math. vol. 87. Binghamton, USA: 139-147; 1998.
[2] Freund, R. Splicing Systems on Graphs, Proc. Intelligence in Neural and Biological Systems, IEEE Press, 189-194; 1995.
[3] Lawson, Mark V. 2009. Finite Automata. Heriot-Watt University, Edinburgh.
[4] Adleman, L.M. Computing with DNA. Scientific American. USA: 54-61; August 1998.
[5] Dinu, L. P. AS garro-A low complexity distance for DNA words fundamental Informatics 73(3), 361 372, 2006.
[6] Frisco, P. Theory of Molecular Computing Splicing and Membrane System. Italy; 1970.
[7] Head, T. Formal Language Theory and DNA: An analysis of the generative capacity of recombinant behaviors. Bulletin of Mathematical Biology, 49, 735-759; 1987.
[8] Jeyabharathi, S., Angayarkanni, M., Sinthanaiselvi, S., Anusha, R. Bipartite Semigraph Structure in DNA Splicing System, International Journal of Computer Application, Vol. VI, E-ISSN 2250-1797; 2012.
[9] Jeyabharathi, S., Sinthanaiselvi, S. Parikh Representation for Splicing System. Proceedings of the International Conference on Applied Mathematics and Theoretical Computer Science. 978-93-82338-35-2. 2013.
[10] Jeyabharathi, S., Thiagarajan, K., Sinthanaiselvi, S. Norms of Parikh Matrices on DNA Splicing System, International Journal of Advanced Engineering Technology, Vol. VII, E-ISSN 0976-3945; 2016.
[11] Laufer, H.B. Discrete Mathematics and Applied Modern Algebra. USA: PWS Publisher; 1984.
[12] Leon, S.J. Linear Algebra with Applications, Ninth Edition. Pearson Education, Inc. United States of America. 2015.
[13] Subramanian, K.G., Huey, A.M., Nagar, A.K. On Parikh Matrices. International Journal of Foundations of Computer Science Vol. 20, No. 2, 211-219; 2009.
[14] Available from: https://www.addgene.org/mol-bio-reference/restriction-enzymes/ (cited August 14, 2018).