MJFAS MALAYSIAN JOURNAL OF FUNDAMENTAL AND APPLIED SCIENCES
PRINT ISSN: 2289-5981 | ONLINE ISSN: 2289-599X

RESEARCH ARTICLE

Computation of splicing languages from DNA splicing system with one palindromic restriction enzyme

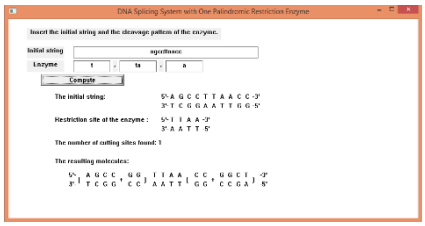
Nurul Izzaty Ismail, Wan Heng Fong*, Nor Haniza Sarmin

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

* *Corresponding author: fwh@utm.my*

Article history
Submitted 25 October 2017
Revised 29 January 2018
Accepted 9 April 2018
Published Online 4 June 2018

Graphical abstract



Abstract

In DNA splicing system, the potential effects of sets of restriction enzymes and a ligase that allow DNA molecules to be cleaved and reassociated to produce further molecules are studied. A splicing language depicts the molecules resulting from a splicing system. In this research, a C++ programming code for DNA splicing system with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites is developed. A graphical user interface, GUI is then designed to allow the user to insert the initial DNA string and restriction enzymes to generate the splicing languages which are the result of the computation of the C++ programming. This interface displays the resulting splicing languages, which depict the results from in vitro experiments of the respective splicing system. The results from this research simplify the lenghty manual computation of the resulting splicing languages of DNA splicing systems with one palindromic restriction enzyme.

Keywords: DNA, palindromic, restriction enzyme, splicing system, C++ Visual Programming

INTRODUCTION

DNA is a polymer strung together from monomers which are known as deoxyribonucleotides [1]. Nucleotides are linked together in many ways such as the base of one nucleotide can interact with the base of the other to form a hydrogen bond, which is classified as a weak bond [1]. The hydrogen bond has a restriction on the base pairing that are: adenine (A) and thymine (T) are paired together, and cytosine (C) and guanine (G) are paired together. This pairings hence make up the double stranded DNA (dsDNA) which are represented by four symbols *a*, *g*, *c*, and *t* where each alphabet stands for [A/T], [G/C], [C/G] and [T/A], respectively. DNA plays an important part in DNA computing. In DNA computing, the information-processing capabilities of DNA molecule in every living organism can be used in computers for the purpose of digital switching primitives [1]. Splicing system have been modelled mathematically where some components and models are built based on DNA computing.

A splicing system is initiated by Head [2] in 1987, which is described as a new manner of relating formal language theory to the study of informational macromolecules. Later, the concept of splicing systems were extended into many models like Paun [3], Pixton [4], Goode Pixton [5] and Yusof-Goode splicing systems [6]. In a splicing system, the set of the DNA molecules resulted from splicing operations with the effect of enzymatic activities is depicted as splicing language [2]. DNA molecules are taken from the sub sequences or pattern in protein or nucleotide sequences which are also called as initial strings in splicing systems [7].

Formal language theory is a study devoted from the field of theoretical computer science and discrete mathematics [8]. In formal laguage theory, a language consists of a set of strings of symbols derived from an alphabet [8]. Some notations in formal languages are λ or 1, +, · and * that indicate the empty string, union, concatenation and star-closure respectively. A^* denotes a set of strings concatenating zero or more symbols from an alphabet A while A^+ is a set of strings of symbols without the empty string. The modelling of splicing languages generated by splicing systems is done by using formal language theory where the DNA bases, sequences of nucleotides and restriction enzymes are modelled as alphabets, strings and rules respectively.

In a splicing system, the mixture of DNA molecules with a ligase and restriction enzymes are called as endodeoxyribonucleases that allow the molecules to be cut and recombined [9]. The restriction enzymes cut DNA molecules in specific ways based on the cleavage pattern of the enzymes which is known as a triple [2]. In the modelling of splicing system, the triple is denoted as a rule for restricton enzyme which consists of left context, crossing and right context [2]. DNA molecules can be cleaved differently through three types of cutting sites generated by the restriction enzymes such as $5'$ overhang, $3'$ overhang or blunt ends. The symbols $\downarrow$ and $\uparrow$ refer to the cutting sites by the restriction enzymes. The cutting sites and cleavage patterns of enzymes for $5'$ overhang, $3'$ overhang and blunt ends are shown through some enzymes *Hpa*II, *Hha*I and *Alu*I respectively in the following:

*Hpa*II (*c*, *cg*, g): $5'-C\downarrow CG\quad G-3'$ / $3'-G\quad GC\uparrow C-5'$,

*Hha*I (*g*, *cg*, *c*): 5'-G CG↓C-3' / 3'-C↑GC G-5', and

*Alu*I (*ag*, 1, *ct*): 5'-AG↓CT-3' / 3'-TC↑GA-5'.

The concept of DNA splicing system with $5'$ overhang palindromic restriction enzymes is used in this research. Palindrome is a sequence that reads the same forward and backward [10]. In DNA splicing systems, DNA molecule can read in two ways: forward and backward. A sequence of DNA molecule which is exactly the same both ways is

called as a palindromic string. The definition of a palindromic string is stated in the following:

Definition 1 [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.

Palindromic sequences can also be recognised in the sequence of restriction enzymes [12]. For example, the enzyme *Mse*I $5' - \text{TTAA} - 3'$ $3' - \text{AATT} - 5'$ is a palindromic restriction enzyme since the upper single strand of base sequence of enzyme *Mse*I matches with the lower single strand when read from backward. Research on DNA splicing systems involving palindromic sequences has been studied in [2,13,14, 15, 16, 17,]. The name and sequence for every restriction enzyme that are used in this research are taken from [18].

Research on DNA splicing systems is normally done using one or two restriction enzymes in experiments. As this research on the generalisation of splicing languages resulting from DNA splicing system is still in its preliminary stage, the modelling of DNA splicing system with the presence of one palindromic restriction enzyme is discussed in this paper. An algorithm is designed and implemented in C++ visual programming to compute the resulting splicing strings of a DNA splicing system with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites. A graphical user interface, GUI is also developed to display the resulting splicing languages generated from the respective DNA splicing system.

RESEARCH METHODOLOGY

In this section, the definition of a splicing system and some generalisations of DNA splicing systems with the presence of one palindromic restriction enzyme are presented. In this research, Head splicing system is used to generalise the resulting splicing languages. The definitions of Head splicing system and a splicing language are stated next.

Definition 2 [2] 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 $L = L(S)$.

The resulting splicing language generated by the splicing system above are demonstrated in the following: the left part of string *ucxdv* combines with the right part of string *pexfq* which produces the string *ucxfq*. The other combination between the right part of string *ucxdv* with the left part of string *pexfq* results in the string *pexdv*.

Next, Theorem 1 is used for the generalisation of resulting splicing languages generated from DNA splicing system with one palindromic restriction enzyme involving one cutting site.

Theorem 1 [19]

Given $S = (A, I, B, C)$ is a splicing system in which $A = \begin{Bmatrix} \text{A,C,G,T} \\ \text{T G C A} \end{Bmatrix}$ is the set of dsDNA symbols, $I = \begin{Bmatrix} N_1N_1 \ \dots N_1X_1YX_2N_2N_2 \dots N_2 \\ N_1'N_1' \dots N_1'X_1'Y'X_2'N_2'N_2' \dots N_2' \end{Bmatrix}$ is the set consisting of an initial string with a cutting site of a palindromic restriction enzyme $\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix}$, set $B = \begin{Bmatrix} X_1, Y, X_2 \\ X_1'Y'X_2' \end{Bmatrix}$ is the set of cleavage pattern for restriction enzyme and set C is the empty (Ø) set, the resulting splicing language consists the strings of the form

$$\begin{pmatrix} N_1N_1 \dots N_1 \\ N_1'N_1' \dots N_1' \end{pmatrix} + \begin{matrix} N_2'N_2' \dots N_2' \\ N_2N_2 \dots N_2 \end{matrix} \begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix} \begin{pmatrix} N_2N_2 \dots N_2 \\ N_2'N_2' \dots N_2' \end{pmatrix} + \begin{matrix} N_1'N_1' \dots N_1' \\ N_1N_1 \dots N_1 \end{matrix}$$

where $\begin{matrix} N_1 & X_1 & Y & X_2 \\ N_1', & X_1', & Y' & X_2' \end{matrix}$ and $\begin{matrix} N_2 \\ N_2' \end{matrix}$ denote arbitrary dsDNA symbol(s), N_1', X_1', Y', X_2' and N_2' are complementaries for N_1, X_1, Y, X_2 and N_2 respectively, $\begin{matrix} Y \\ Y' \end{matrix}$ is the crossing, and $\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix} \notin \begin{Bmatrix} N_1N_1 \ \dots N_1 \ N_2N_2 \dots N_2 \\ N_1'N_1' \dots N_1', N_2'N_2' \dots N_2' \end{Bmatrix}$.

Lastly, the generalisation of resulting splicing strings generated from DNA splicing system with one palindromic restriction enzyme involving two non-overlapping cutting sites is presented in Theorem 2.

Theorem 2 [19]

Given $S = (A, I, B, C)$ is a splicing system in which $A = \begin{Bmatrix} \text{A,C,G,T} \\ \text{T G C A} \end{Bmatrix}$ is the set of dsDNA symbols, $I = \begin{Bmatrix} N_1N_1 \ \dots N_1X_1 \ Y \ X_2M \ M \dots MX_1 \ Y \ X_2 \ N_2N_2 \dots N_2 \\ N_1'N_1' \dots N_1'X_1'Y'X_2'M'M' \dots M'X_1'Y'X_2'N_2'N_2' \dots N_2' \end{Bmatrix}$ is the set consisting of an initial string with two non-overlapping cutting sites of a palindromic restriction enzyme $\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix}$, set $B = \begin{Bmatrix} X_1, Y, X_2 \\ X_1'Y'X_2' \end{Bmatrix}$ is the set of cleavage pattern for restriction enzyme and set C is the empty (Ø) set, the resulting splicing language consists the strings of the form

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

where $n \in \mathbb{Z}^+$, $\begin{matrix} N_1 & X_1 & Y & X_2 & M \\ N_1', & X_1', & Y' & X_2', & M' \end{matrix}$ and $\begin{matrix} N_2 \\ N_2' \end{matrix}$ denote arbitrary dsDNA symbol(s), 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, $\begin{matrix} Y \\ Y' \end{matrix}$ is the crossing, and $\begin{matrix} X_1YX_2 \\ X_1'Y'X_2' \end{matrix} \notin \begin{Bmatrix} N_1N_1 \ \dots N_1 \ M \ M \dots M \ N_2N_2 \dots N_2 \\ N_1'N_1' \dots N_1'', M'M' \dots M'' N_2'N_2' \dots N_2' \end{Bmatrix}$.

RESULTS AND DISCUSSION

A C++ programming code is designed to compute the resulting splicing strings of DNA splicing system with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites. The generalisations of resulting splicing languages from DNA splicing systems with one palindromic restriction enzyme from Theorem 1 and 2 are used in this coding to compute the results. Then, a user-friendly GUI is developed for the purpose of displaying the resulting splicing languages in place of the time-consuming manual calculations. The procedure and design of generating the algorithm of the C++ programming code are illustrated through four steps:

Step 1: User is required to insert an initial string and a cleavage pattern of restriction enzyme by using the four dsDNA symbols, *a*, *c*, *g*, and *t*. The first step is to convert the initial string and the restriction enzyme into the double stranded DNA (dsDNA) sequence.

For example, the dsDNA sequence of the initial string *agcccgtgg* is

$5' - \text{AGCCCGTGG} - 3'$
$3' - \text{TCGGGCACC} - 5'$

Step 2: Calculate the number of cutting sites in the initial string. The C++ programming code is designed for the DNA splicing system with one restriction enzyme for one or two (non-overlapping) cutting sites to proceed to the next step. The flowchart of designing the coding for calculating and displaying the number of cutting sites is demonstrated in Fig. 1.

For instance, the initial string *ggccgctat* has one cutting site of the enzyme *Aci*I with the cleavage pattern{*c*,*cg*,*c*}:

$$5' - \text{GGC} \downarrow \text{CG} \quad \text{CTAT} - 3'$$
$$3' - \text{CCG} \quad \text{GC} \uparrow \text{GATA} - 5'$$

Step 3: Determine if the enzyme is palindromic, i.e., it is read the same forward and backward. Also, the length of restriction enzyme and the

crossing sites of the enzyme must be even in order to be a palindromic enzyme.

For example, the enzyme *Age*I with the cleavage pattern {*a*, *ccgg*, *t*} is palindromic since the base sequence of the enzyme $\begin{matrix}5'-\text{CCGC}-3'\\3'-\text{GGCG}-5'\end{matrix}$ reads the same forward and backward and the length of the crossing, *ccgg* is four which is even.

Next, the enzyme *Aci*I with the cleavage pattern {*c*,*cg*,*c*} is not palindromic since the upper single strand of the enzyme $5'-\text{CCGC}-3'$ is not exactly the same with the lower single strand of the enzyme $3'-\text{GGCG}-5'$ when read from backward.

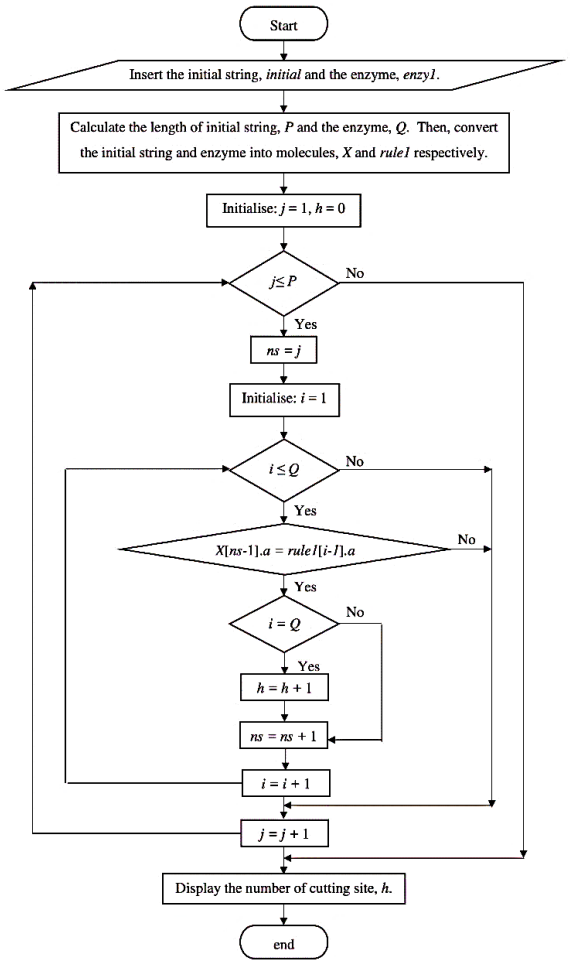


Fig. 1 Flowchart of calculating the number of cutting sites.

Step 4: Generate the resulting splicing strings for one palindromic restriction enzyme in the DNA splicing system for one and two (non-overlapping) cutting sites by using Theorems 1 and 2 respectively.

The resulting languages generated from DNA splicing system with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites are presented in Example 1 and Example 2 respectively.

Example 1

Given a splicing system $S = (A, I, B, C)$ where $I = \{agccttaacc\}$ is the set of initial string, set $B = \{t, ta, a\}$ is the set of cleavage pattern for the enzyme *Mse*I and set C is the empty ($\emptyset$) set.

Solution

The enzyme *Mse*I $\begin{matrix}5'-\text{TTAA}-3'\\3'-\text{AATT}-5'\end{matrix}$ is palindromic since the base sequence of enzyme *Mse*I can read the same forward and backward. So, $\frac{Y}{Y'}=\frac{Y'}{Y}$, $\frac{X_1}{X_1'}=\frac{X_2'}{X_2}$ and $\frac{X_2}{X_2'}=\frac{X_1'}{X_1}$ which are the strings $\begin{matrix}\text{TA}\\\text{AT}\end{matrix}$ (or the crossing), $\begin{matrix}\text{T}\\\text{A}\end{matrix}$ and $\begin{matrix}\text{A}\\\text{T}\end{matrix}$ respectively.

The initial string $\begin{matrix}5'-\text{AGCCT}\downarrow\text{TA}\quad\text{ACC}-3'\\3'-\text{TCGGA}\quad\text{AT}\uparrow\text{TGG}-5'\end{matrix}$ has one cutting site of the enzyme *Mse*I with the cleavage pattern {*t*, *ta*, *a*}.

Thus, by using Theorem 1, the resulting language is

$$\begin{matrix}5'-\\3'-\end{matrix}\left(\begin{matrix}\text{AGCC}\\\text{TCGG}\end{matrix}+\begin{matrix}\text{GG}\\\text{CC}\end{matrix}\right)\begin{matrix}\text{TTAA}\\\text{AATT}\end{matrix}\left(\begin{matrix}\text{CC}\\\text{GG}\end{matrix}+\begin{matrix}\text{GGCT}\\\text{CCGA}\end{matrix}\right)\begin{matrix}-3'\\-5''\end{matrix}$$

Example 2

Given a splicing system $S = (A, I, B, C)$ where $I = \{atcggcgcaagcgcctt\}$ is the set of initial string, set $B = \{g, cg, c\}$ is the set of cleavage pattern for the enzyme *Sci*NI and set C is the empty ($\emptyset$) set.

Solution

The enzyme *Sci*NI $\begin{matrix}5'-\text{GCGC}-3'\\3'-\text{CGCG}-5'\end{matrix}$ is palindromic since the base sequence of enzyme *Sci*NI can read the same forward and backward. So, $\frac{Y}{Y'}=\frac{Y'}{Y}$, $\frac{X_1}{X_1'}=\frac{X_2'}{X_2}$ and $\frac{X_2}{X_2'}=\frac{X_1'}{X_1}$ which are the strings $\begin{matrix}\text{CG}\\\text{GC}\end{matrix}$ (or the crossing), $\begin{matrix}\text{G}\\\text{C}\end{matrix}$ and $\begin{matrix}\text{C}\\\text{G}\end{matrix}$ respectively.

The initial string has two non-overlapping cutting sites of the enzyme *Mse*I with the cleavage pattern {*g*, *cg*, *c*}:

$$\begin{matrix}5'-\text{ATCGG}\downarrow\text{CG}\quad\text{CAAG}\downarrow\text{CG}\quad\text{CCTT}-5'\\3'-\text{TAGCC}\quad\text{GC}\uparrow\text{GTTC}\quad\text{GC}\uparrow\text{GGAA}-3''\end{matrix}$$

Therefore, by using Theorem 2, the resulting language is

$$\begin{matrix}5'-\\3'-\end{matrix}\left(\begin{matrix}\text{ATCG}\\\text{TAGC}\end{matrix}+\begin{matrix}\text{AAG}\\\text{TTA}\end{matrix}\right)\begin{matrix}\text{GCGC}\\\text{GCGC}\end{matrix}\left(\left(\begin{matrix}\text{AA}\\\text{TT}\end{matrix}+\begin{matrix}\text{TT}\\\text{AA}\end{matrix}\right)\right)^{n-1}\left(\begin{matrix}\text{CTT}\\\text{GAA}\end{matrix}+\begin{matrix}\text{CGAT}\\\text{GCTA}\end{matrix}\right)\begin{matrix}-3'\\-5'\end{matrix}$$

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

The overall process of computing the resulting strings is explained in the flowchart in Fig. 2.

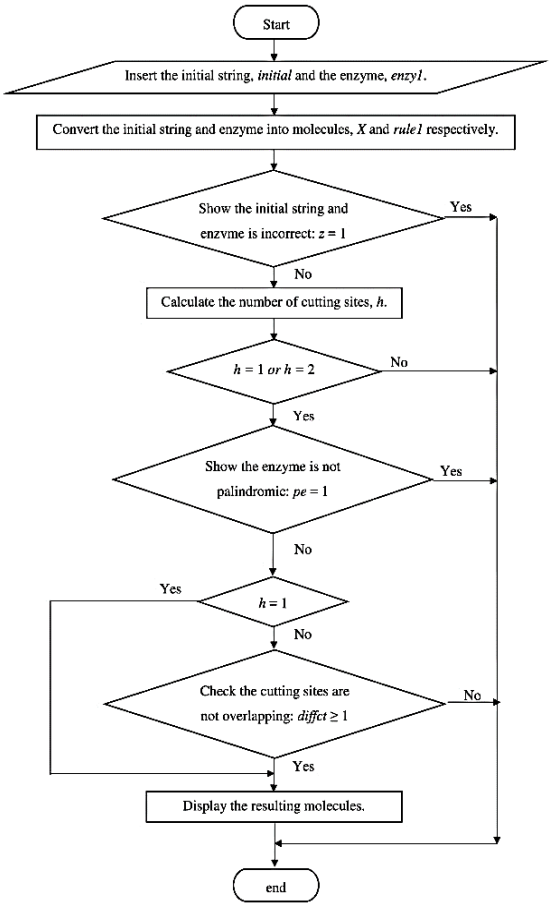


Fig. 2 Flowchart of computing the resulting splicing languages.

Then, the GUI is developed to initiate execution of an application program for a DNA splicing system with one palindromic restriction enzyme using C++ visual programming. Users can use the GUI by entering the initial string and the cleavage pattern of any 5' overhang restriction enzymes. The procedure of using the interface is explained in the following.

The default GUI for DNA splicing system with one palindromic restriction enzyme is illustrated in Fig. 3.

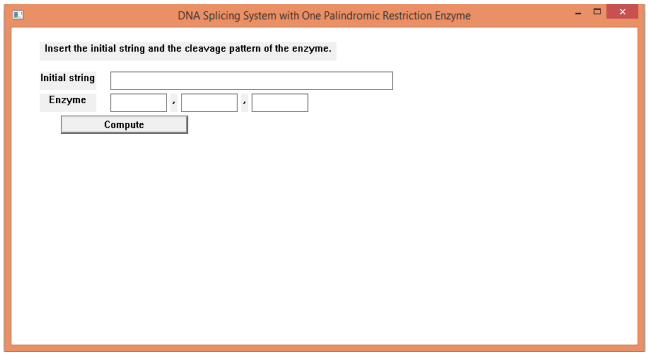


Fig. 3 The default graphical user interface.

Next, users need to insert the initial string and enzyme to generate the output. The resulting splicing languages of DNA splicing systems with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites are shown by clicking the button 'Compute' in the interface.

In Fig. 4, the output for the DNA splicing system involving one cutting site as discussed in Example 1 is presented using the interface where the initial string is *agccttaacc* and the cleavage pattern of the enzyme *Mse*I is $\{t, ta, a\}$, where the resulting molecules $\begin{matrix}5'-\\3'-\end{matrix}\binom{\text{AGCC}}{\text{TCGG}}+\binom{\text{GG}}{\text{CC}}\text{TTAA}\binom{\text{CC}}{\text{GG}}+\binom{\text{GGCT}}{\text{CCGA}}\begin{matrix}-3'\\-5'\end{matrix}$ depict the splicing language corresponding to the respective DNA splicing system.

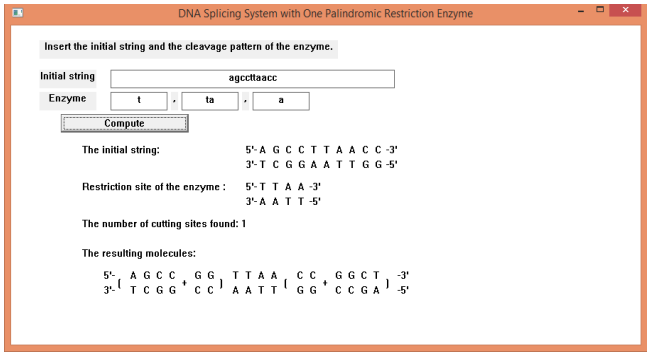


Fig. 4 The output for DNA splicing system involving one cutting site.

Fig. 5 shows the output for the DNA splicing system mentioned in Example 2 involving two non-overlapping cutting sites with the initial string, *atcggcgcaagcgcctt* and the cleavage pattern of the enzyme *Sci*NI, $\{g, cg, c\}$, where the resulting molecules $\begin{matrix}5'-\\3'-\end{matrix}\binom{\text{ATCG}}{\text{TAGC}}+\binom{\text{AAG}}{\text{TTA}}\text{GCGC}\left(\binom{\text{AA}}{\text{TT}}+\binom{\text{TT}}{\text{AA}}\right)^{n-1}\binom{\text{CTT}}{\text{GAA}}+\binom{\text{CGAT}}{\text{GCTA}}\begin{matrix}-3'\\-5'\end{matrix}$, $n \in \mathbb{Z}^+$depict the splicing language corresponding to the respective DNA splicing system.

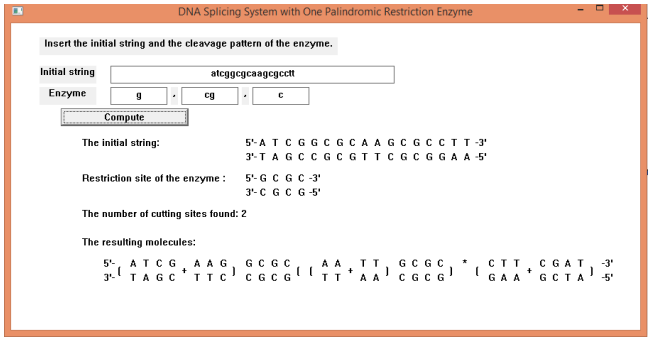


Fig. 5 The output for DNA splicing system involving two non-overlapping cutting sites.

From Fig. 4 and 5, the GUI computes the resulting molecules which is the output of the C++ programming code. The users can also view the initial molecule, the restriction site of the enzyme and the number of cutting sites in the interface.

As an additional feature, the GUI also prompts the user if the inserted enzyme is not palindromic. The output in Fig. 6 shows that the enzyme *Hba*II, $\{g, ccca, a\}$ is non-palindromic.

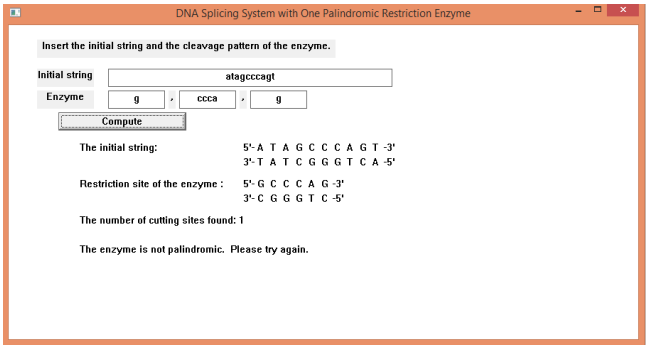


Fig. 6 The output depicting a non-palindromic enzyme.

Besides that, the GUI determines if the cutting sites of the enzyme overlapped. The first and second cutting sites of the enzyme *Sci*NI, $\{g, cg, c\}$, are overlapping in the initial string *atcgcgcgcatt* as shown in Fig. 7.

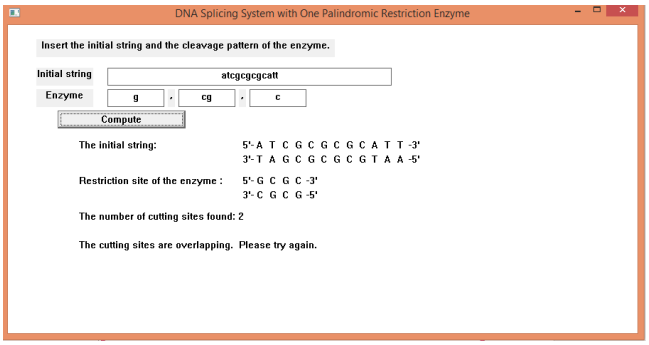


Fig. 7 The output showing overlapped cutting sites.

Furthermore, some messages are also displayed on the interface if the number of cutting sites of the enzyme found is more than two. Fig. 8 shows an example of DNA splicing system with three cutting sites of the enzyme *Mse*I,$\{t, ta, a\}$, which is found in the initial string *aattaagcgttaagttaact*.

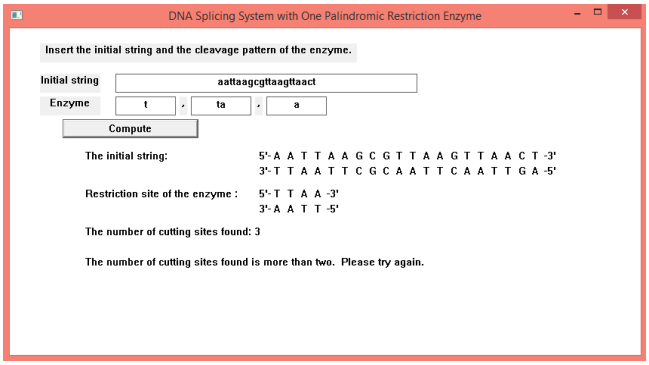


Fig. 8 The output showing three cutting sites.

CONCLUSION

In this research, the computation of DNA splicing systems with one palindromic restriction enzyme is done using a tool developed using C++ visual programming. The splicing strings (or the language of) of DNA splicing system with one palindromic restriction enzyme for one and two (non-overlapping) cutting sites are displayed in the GUI. The interface also displays the initial molecule and the restriction site of the enzyme, calculates the number of cutting sites and determines if the enzyme is not palindromic, if the cutting sites overlap and if the number of cutting sites is more than two. The results from this research simplify the manual computation of splicing languages of DNA splicing systems with one palindromic restriction enzyme using formal language theory, which contributes to the advancement of splicing systems in DNA computing. For future research, the splicing languages generated from DNA splicing systems involving more than one palindromic restriction enzymes or non-palindromic restriction enzymes can also be generalized as an extension to this research.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Higher Education (MOHE) and Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the financial funding through Research University Grant Vote No. 13H18.

REFERENCES

[1] G. Paun., G. Rozenberg, A. Salomaa, *DNA Computing: New Computing Paradigms*, Springer -Verlag Berlin Heidelberg, Germany, 1998, p. 1-41.
[2] T. Head, *B. Math. Biol*, 49 (1987) 737-759.
[3] G. Paun, *Discrete. Appl. Math*, 70 (1996) 57-79.
[4] D. Pixton, *Discrete. App. Math*, 69 (1996) 101-124.
[5] E. G. Laun, *Constant and Splicing Systems*, Ph.D. Thesis, State University of New York at Binghamton, 1999.
[6] Y. Yusof, N. H. Sarmin, W. H. Fong, T. E. Goode, M. A. Ahmad, An Analysis of Four Variants of Splicing System, *AIP. Conf. Proc*, Putrajaya, Malaysia, 18-20 December 2012, Melville, NY, 2013, p. 888-895.
[7] National Center for Biotechnology Information, *Expressed Sequence Tag*. Available from: <https://www.ncbi.nlm.nih.gov/nucest>. [11 June 2017].
[8] P. Linz, *An Introduction of Formal Language and Automata*, John and Barlett Publisher, USA, 2006, p. 1-36.
[9] S. M. Kim, *SIAM. J. Comput*. 26 (1997) 1284-1309.
[10] I. Tomohiro, S. Inenaga, M. Takeda, *Theor. Comput. Sci*, 483 (2013) 162-170.
[11] Y. Yusof, *DNA Splicing System Inspired by Bio Molecular Operation*, Ph.D. Thesis, Universiti Teknologi Malaysia, 2012.
[12] H. M. Eun, *Enzymology Primer for Recombinant DNA Technology*, Academic Press, USA, 1996, p. 1-108
[13] W. H. Fong, *Modelling of Splicing Systems using Formal Language Theory*, Ph.D. Thesis, Universiti Teknologi Malaysia, 2008.
[14] T. Head, *Discrete. Appl. Math*,. 87 (1998) 139-147.
[15] M. H. Mudaber, Y. Yusof, M. S. Mohamad, Some Sufficient Conditions for Persistency and Permanency of Two Stages DNA Splicing Languages via Yusof-Goode Approach, *AIP. Conf. Proc*, Penang, Malaysia, p. 6-8 November 2013, Melville, NY , 2014, 591-595.
[16] M. A. Ahmad, N. H. Sarmin, Y. Yusof, W. H. Fong, Some Restrictions on the Existence of Second Order Limit Language, *AIP. Conf. Proc*, Selangor, Malaysia, 24-26 November 2014, Melville, NY, 2015, p. 020048.
[17] W. L. Lim, Y. Yusof, M. H. Mudaber, Modeling of DNA Single Stage Splicing Language via Yusof-Goode Approach: One String with Two Rules, *AIP. Conf. Proc*, Kuantan, Malaysia, 12-14 August 2014, Melville, NY, 2015, p. 695-699.
[18] New England Biolabs Inc, *NEB 2017-18 Catalog & Technical Reference, Catalogue*. 2017.
[19] W. H. Fong, N. I. Ismail, Generalisations of DNA Splicing Systems with One Palindromic Restriction Enzyme, *Malaysian Journal of Industrial and Applied Mathematics*, 2017, Accepted.