International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6S3, September 2019

Computation of Splicing Languages from DNA Splicing System Based on Sequences of Restriction Enzymes

Nurul Izzaty Ismail, Wan Heng Fong, Nor Haniza Sarmin

Abstract: ***In DNA splicing systems, restriction enzymes and ligases cleave and recombine DNA molecules based on the cleavage pattern of the restriction enzymes. The set of molecules resulting from the splicing system depicts a splicing language. In this research, an algorithm for DNA splicing systems is developed using C++ visual programming. The splicing languages which have been characterised through some theorems based on the crossings and sequences of the restriction enzymes, are generated as the output from this computation. In order to generate the splicing languages, the algorithm detects and calculates the number of cutting sites of the restriction enzymes found in the initial molecules, determines whether the sequence of restriction enzyme is a palindrome or not, and if the restriction enzymes have the same or different crossings. The results from this research depict the splicing languages obtained from the manual computations, which contributes to the development of computational software in DNA computing.***

Index Terms: ***C++ visual programming, DNA, palindrome, restriction enzyme, splicing system.***

I. INTRODUCTION

Deoxyribonucleic acid (DNA) splicing system is mathematically developed based on some models in DNA computing. The idea of DNA computing is introduced by Feynman [1] in 1959 involving computation in molecular biology. Adleman's experiment [2] on solving Hamiltonian path problem contributes to the development of DNA computing using two features: massive parallelism of DNA strands and the Watson-Crick complementarity. DNA strand is a nucleotide chain which is made up of four bases: adenine (A), guanine (G), cytosine (C) and thymine (T) [3]. Watson-Crick complementarity is generated from nitrogenous base pairings: adenine pairs with thymine, while cytosine pairs with guanine for the formation of double-stranded DNA (dsDNA) [3].

The research relating between formal language theory and molecular biology in DNA computing leads to the modelling of DNA splicing system. The splicing system is introduced by Head [4] in 1987 which is also known as Head's splicing system. In splicing systems, DNA molecules are cut and recombined when mixed with a ligase and restriction enzymes which are biologically called as endodeoxyribonucleases [5].

The set of molecules resulting from a DNA splicing system is called a splicing language which is simulated using formal language theory. A formal language consists of a set of strings of symbols from an alphabet [6]. Some notations of regular expression in formal language theory, namely λ, +, * and {} or () denote the empty string, union, star-closure and parentheses respectively, are used in this research [3]. For example, the language L for the regular expression $(a+b)\cdot c^*$ where a, b, and c are the set of symbols derived from an alphabet is shown in the following:

$$\begin{aligned} L((a+b)\cdot c^*) &= L((a+b))\,L(c^*) \\ &= L(a+b)\,L(c^*) \\ &= (L(a)\cup L(b))\,(L(c))^* \\ &= (\{a\}\cup\{b\})(\{c\})^* \\ &= \{a,b\}\{\lambda,c,cc,\ldots\} \\ &= \{a,ac,acc,\ldots,b,bc,bcc\ldots\}. \end{aligned}$$

By using the concepts in formal language theory, the splicing system is associated with three sets. The first set is the set of dsDNA symbols from nitrogenous base pairings [3]. The second set consists of initial DNA molecules or strings taken from the sub sequences or pattern in protein or nucleotide chains [7]. Lastly, the third set consists of rules for the cleavage pattern of restriction enzymes. The cleavage pattern of restriction enzymes is made up of three sites namely the crossing, left and right context [8].

Throughout the years, notations in Head's splicing system have been extended and variants of splicing models have been developed namely Paun [9], Pixton [10], Goode-Pixton [11] and Yusof-Goode [12] splicing systems. The variants of splicing systems resulted in many types of splicing languages. The splicing languages from different models of splicing system can be obtained based on the specific sequences of restriction enzymes. This research focuses on palindromic and non-palindromic sequences of restriction enzyme. Palindrome is a sequence of string that reads the same forwards and backwards [13]. Previously, research on generalised splicing languages from DNA splicing systems with palindromic and non-palindromic restriction enzymes has been done in [14]-[18], in which the generalised splicing languages from the respective splicing systems are presented as theorems.

Manuscript published on 30 September 2019.
***** Correspondence Author (s)
Nurul Izzaty Ismail, Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.
Wan Heng Fong, Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.
Nor Haniza Sarmin, Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.

The name, sequence and cleavage for all restriction enzymes used in this research are obtained from the New England Biolabs (NEB) catalogue [19].

Next, some preliminaries related to this research are given. DNA splicing systems with palindromic and non-palindromic restriction enzymes are modelled using Head's splicing system in this research. The definitions of Head's splicing system and splicing language are stated in the following.

Definition 1 [4] 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 $L = L(S)$.

After that, the definition of a palindromic string is stated.

Definition 2 [20] Palindromic String

A string I of a dsDNA is 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.

For example, the enzyme *Bfa*I $\begin{matrix}5'-\text{CTAG}-3'\\3'-\text{GATC}-5'\end{matrix}$ is a palindromic restriction enzyme since the upper single strand of enzyme *Bfa*I matches with the lower single strand when read from backwards; while the enzyme *Bbv*CI $\begin{matrix}5'-\text{CCTCAGC}-3'\\3'-\text{GGAGTCG}-5'\end{matrix}$ is a non-palindromic restriction enzyme since the upper single strand of enzyme *Bbv*CI does not match with the lower single strand when read from backwards.

The aim of this research is to develop an algorithm and a coding for DNA splicing systems with one or two restriction enzymes using C++ visual programming by developing a user friendly graphical user interface that generates the splicing languages from the corresponding splicing systems. The generalised splicing languages resulting from DNA splicing system with palindromic and non-palindromic restriction enzymes are applied in this coding to generate the results. The interface is designed for random input of initial DNA string and restriction enzyme(s) for the purpose of generating the splicing languages in place of the time-consuming manual computations.

In the next section, the generalised splicing languages from DNA splicing system with one and two restriction enzymes are discussed.

II. METHODOLOGY

The splicing languages from the DNA splicing systems are generalised based on the number of cutting sites, the sequences of restriction enzymes and the crossings for one and two different restriction enzymes, where these generalisations have been given as theorems in [14]-[18].

Table I shows the generalised splicing languages from DNA splicing systems with one palindromic restriction enzyme where Theorem 1 involves one cutting site; and Theorem 2 involves two non-overlapping cutting sites of a palindromic restriction enzyme. The symbols $\begin{matrix}N_1\\N_1'\end{matrix}, \begin{matrix}X_1\\X_1'\end{matrix}, \begin{matrix}Y\\Y'\end{matrix}, \begin{matrix}X_2\\X_2'\end{matrix}, \begin{matrix}M\\M'\end{matrix}, \begin{matrix}W_1\\W_1'\end{matrix}, \begin{matrix}Z\\Z'\end{matrix}, \begin{matrix}W_2\\W_2'\end{matrix}$ and $\begin{matrix}N_2\\N_2'\end{matrix}$ denote arbitrary dsDNA symbol(s), where N_1', X_1', Y', X_2', M', W_1', Z', W_2' and N_2' are complementaries for N_1, X_1, Y, X_2, M, W_1, Z, W_2 and N_2 respectively, and $\begin{matrix}Y\\Y'\end{matrix}$ and $\begin{matrix}Z\\Z'\end{matrix}$ are the crossings.

Table- I: Generalised Splicing Languages from DNA Splicing Systems with One Palindromic Restriction Enzyme

Theorem	Theorem 1 [14]	Theorem 2 [14]
Initial String	$\begin{matrix}N_1N_1...N_1\ X_1\ Y\ X_2\ N_2N_2...N_2\\N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ N_2'N_2'...N_2'\end{matrix}$	$\begin{matrix}N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ X_1\ Y\ X_2\ N_2N_2...N_2\\N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ X_1'\ Y'\ X_2'\ N_2'N_2'...N_2'\end{matrix}$
Restriction Enzyme	$\begin{matrix}X_1\ Y\ X_2\\X_1'\ Y'\ X_2'\end{matrix}$	$\begin{matrix}X_1\ Y\ X_2\\X_1'\ Y'\ X_2'\end{matrix}$
Splicing Language	$\begin{pmatrix}N_1N_1...N_1 & N_2'N_2'...N_2'\\N_1'N_1'...N_1' & N_2N_2...N_2\end{pmatrix}\begin{matrix}X_1\ Y\ X_2\\X_1'\ Y'\ X_2'\end{matrix}\begin{pmatrix}N_2N_2...N_2 & N_1'N_1'...N_1'\\N_2'N_2'...N_2' & N_1N_1...N_1\end{pmatrix}$	$\begin{pmatrix}N_1N_1...N_1 & N_2'N_2'...N_2'\\N_1'N_1'...N_1' & N_2N_2...N_2\end{pmatrix}\begin{matrix}X_1\ Y\ X_2\\X_1'\ Y'\ X_2'\end{matrix}\left(\begin{pmatrix}M\ M\ ...M & M'M'....M'\\M'M'....M' & M\ M\ ...M\end{pmatrix}\begin{matrix}X_1\ Y\ X_2\\X_1'\ Y'\ X_2'\end{matrix}\right)^*\begin{pmatrix}N_2N_2...N_2 & N_1'N_1'...N_1'\\N_2'N_2'...N_2' & N_1N_1...N_1\end{pmatrix}$

Table II shows the generalised splicing languages from DNA splicing systems with one non-palindromic restriction enzyme where Theorem 3 involves one cutting site; and Theorem 4 involves two non-overlapping cutting sites of a non-palindromic restriction enzyme.

Table- II: Generalised Splicing Languages from DNA Splicing Systems with One Non-Palindromic Restriction Enzyme

Theorem	Theorem 3 [15]	Theorem 4 [15]
Initial String	$N_1N_1...N_1\ X_1\ Y\ X_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ N_2'N_2'...N_2'$	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ X_1\ Y\ X_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ X_1'\ Y'\ X_2'\ N_2'N_2'...N_2'$
Restriction Enzyme	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$
Splicing Language	$N_1N_1...N_1\ X_1\ Y\ X_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ N_2'N_2'...N_2'$	$N_1N_1...N_1\ X_1\ Y\ X_2\begin{pmatrix} M\ M\ ...M\ X_1\ Y\ X_2 \\ M'M'....M'\ X_1'\ Y'\ X_2' \end{pmatrix}^* N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ N_2'N_2'...N_2'$

Table III shows the generalised splicing languages from DNA splicing systems with two palindromic restriction enzymes where Theorem 5 involves the same crossing of palindromic restriction enzymes; while Theorem 6 involves different crossings of palindromic restriction enzymes.

Table IV shows the generalised splicing languages from DNA splicing systems with two non-palindromic restriction enzymes where Theorem 7 involves the same crossing of non-palindromic restriction enzymes, while Theorem 8 involves different crossings of non-palindromic restriction enzymes.

Table- III: Generalised Splicing Languages from DNA Splicing Systems with Two Palindromic Restriction Enzymes

Theorem	Theorem 5 [16]	Theorem 6 [16]
Initial String	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Y\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Y'\ W_2'\ N_2'N_2'...N_2'$	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Z\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Z'\ W_2'\ N_2'N_2'...N_2'$
First Restriction Enzyme	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$
Second Restriction Enzyme	$W_1\ Y\ W_2$ $W_1'\ Y'\ W_2'$	$W_1\ Z\ W_2$ $W_1'\ Z'\ W_2'$
Splicing Language	$\begin{pmatrix} N_1N_1...N_1\ X_1 \\ N_1'N_1'...N_1'\ X_1' \end{pmatrix} + \begin{pmatrix} N_2'N_2'...N_2'\ W_1 \\ N_2N_2...N_2\ W_1' \end{pmatrix}$ $Y\left\{\begin{pmatrix} X_2\ M\ M\ ...M\ W_1 \\ X_2'\ M'M'....M'\ W_1' \end{pmatrix} + \begin{pmatrix} W_2\ M'M'....M'\ X_1 \\ W_2'\ M\ M\ ...M\ X_1' \end{pmatrix}\right\}^*$ $Y\begin{pmatrix} W_2\ N_2N_2...N_2 \\ W_2'\ N_2'N_2'...N_2' \end{pmatrix} + \begin{pmatrix} X_2\ N_1'N_1'...N_1' \\ X_2'\ N_1N_1...N_1 \end{pmatrix}$	$\left\{\begin{pmatrix} N_1N_1...N_1\ X_1\ Y\ X_2\begin{pmatrix} M\ M\ ...M\ W_1\ Z\ W_2\ M'M'....M'\ X_1\ Y\ X_2 \\ M'M'....M'\ W_1'\ Z'\ W_2'\ M\ M\ ...M\ X_1'\ Y'\ X_2' \end{pmatrix}^* \\ \begin{pmatrix} N_1'N_1'...N_1' \\ N_1N_1...N_1 \end{pmatrix} + \begin{pmatrix} M\ M\ ...M\ W_1\ Z\ W_2\ N_2N_2...N_2 \\ M'M'....M'\ W_1'\ Z'\ W_2'\ N_2'N_2'...N_2' \end{pmatrix} \end{pmatrix}\right\}^* +$ $\left\{\begin{pmatrix} N_2'N_2'...N_2'\ W_1\ Z\ W_2\begin{pmatrix} M'M'....M'\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Z\ W_2 \\ M\ M\ ...M\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Z'\ W_2' \end{pmatrix}^* \\ \begin{pmatrix} N_2N_2...N_2 \\ N_2'N_2'...N_2' \end{pmatrix} + \begin{pmatrix} M'M'....M'\ X_1\ Y\ X_2\ N_1'N_1'...N_1' \\ M\ M\ ...M\ X_1'Y'\ X_2'\ N_1N_1...N_1 \end{pmatrix} \end{pmatrix}\right\}^*$

Table- IV: Generalised Splicing Languages from DNA Splicing Systems with Two Non-Palindromic Restriction Enzymes

Theorem	Theorem 7 [17]	Theorem 8 [17]
Initial String	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Y\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Y'\ W_2'\ N_2'N_2'...N_2'$	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Z\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Z'\ W_2'\ N_2'N_2'...N_2'$
First Restriction Enzyme	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$

Second Restriction Enzyme	$W_1\ Y\ W_2$ $W_1'\ Y'\ W_2'$	$W_1\ Z\ W_2$ $W_1'\ Z'\ W_2'$
Splicing Language	$N_1N_1...N_1\ X_1\left(Y\ X_2\ M\ M\ ...M\ W_1\right)^*\ Y\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\left(Y'\ X_2'\ M'M'....M'\ W_1'\right)\ Y'\ W_2'\ N_2'N_2'...N_2'$	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Z\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Z'\ W_2'\ N_2'N_2'...N_2'$

Table V shows the generalised splicing languages from DNA splicing systems with one palindromic and one non-palindromic restriction enzymes where Theorem 9 involves the same crossing of the restriction enzymes; while Theorem 10 involves different crossings of the restriction enzymes.

Table- V: Generalised Splicing Languages from DNA Splicing Systems with One Palindromic and One Non-Palindromic Restriction Enzymes

Theorem	**Theorem 9 [18]**	**Theorem 10 [18]**
Initial String	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Y\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Y'\ W_2'\ N_2'N_2'...N_2'$	$N_1N_1...N_1\ X_1\ Y\ X_2\ M\ M\ ...M\ W_1\ Z\ W_2\ N_2N_2...N_2$ $N_1'N_1'...N_1'\ X_1'\ Y'\ X_2'\ M'M'....M'\ W_1'\ Z'\ W_2'\ N_2'N_2'...N_2'$
Palindromic Restriction Enzyme	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$	$X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$
Non-Palindromic Restriction Enzyme	$W_1\ Y\ W_2$ $W_1'\ Y'\ W_2'$	$W_1\ Z\ W_2$ $W_1'\ Z'\ W_2'$
Splicing Language	$\begin{pmatrix} N_1N_1...N_1 & N_2'N_2'...N_2'\ W_1'\ Y\ W_2'\ M'M'....M' \\ N_1'N_1'...N_1' & N_2N_2...N_2\ W_1\ Y'\ W_2\ M\ M\ ...M \end{pmatrix}$ $X_1\left(Y\ X_2\ M\ M\ ...M\ W_1\right)^*$ $X_1'\left(Y'\ X_2'\ M'M'....M'\ W_1'\right)$ $Y\begin{pmatrix} W_2\ N_2N_2...N_2 & X_2\ N_1'N_1'...N_1' \\ W_2'\ N_2'N_2'...N_2' & X_2'\ N_1N_1...N_1 \end{pmatrix}$	$\begin{pmatrix} N_1N_1...N_1 & N_2'N_2'...N_2'\ W_1'\ Z'\ W_2'\ M'M'....M' \\ N_1'N_1'...N_1' & N_2N_2...N_2\ W_1\ Z\ W_2\ M\ M\ ...M \end{pmatrix}$ $X_1\ Y\ X_2$ $X_1'\ Y'\ X_2'$ $\begin{pmatrix} M\ M\ ...M\ W_1\ Z\ W_2\ N_2N_2...N_2 & N_1'N_1'...N_1' \\ M'M'....M'\ W_1'\ Z'\ W_2'\ N_2'N_2'...N_2' & N_1N_1...N_1 \end{pmatrix}$

The generalised splicing languages from DNA splicing systems with palindromic and non-palindromic restriction enzymes discussed above are applied in this coding to generate the results. The coding is designed by inserting the initial string and cleavage pattern of the restriction enzyme(s) in the interface. In order to generate the splicing languages, the coding detects and calculates the number of cutting sites of the restriction enzyme(s) found in the initial string, determines whether the sequence of restriction enzyme is a palindrome or not, and determine if the restriction enzymes have the same or different crossings. The procedure and design of developing the algorithm of the C++ coding are illustrated in Fig. 1.

The outputs for the coding to generate splicing languages from DNA splicing systems with one or two restriction enzymes are discussed next.

III. RESULTS AND DISCUSSION

The coding for DNA splicing systems with one or two palindromic and non-palindromic restriction enzymes is designed to develop a graphical user interface. The default interface for the DNA splicing system is illustrated in Fig. 2.

Firstly, user inserts the initial string and cleavage pattern of the restriction enzyme(s) in the interface to generate the resulting molecules which are the results for the splicing languages from the corresponding DNA splicing system. By clicking the button 'Compute' in the interface, the coding is run to generate the results; while the button 'Clear' is used to reset the interface.

In Fig. 3, the output for the DNA splicing system involving one cutting site of a palindromic restriction enzyme is presented in the interface where the initial string is *aggactagtct* consisting the cleavage pattern of the enzyme *Bfa*I which is {*c*, *ta*, *g*}.

In Fig. 4, the output for the DNA splicing system involving two non-overlapping cutting sites of a palindromic restriction enzyme is presented in the interface where the initial string is *taccggaattccggaa* consisting the cleavage pattern of the enzyme *Msp*I which is {*c*, *cg*, *g*}.

In Fig. 5, the output for the DNA splicing system involving one cutting site of a non-palindromic restriction enzyme is presented in the interface where the initial string is *ttcccagcgac* consisting the cleavage pattern of the enzyme *BseY*I which is {*c*, *ccag*, *c*}.

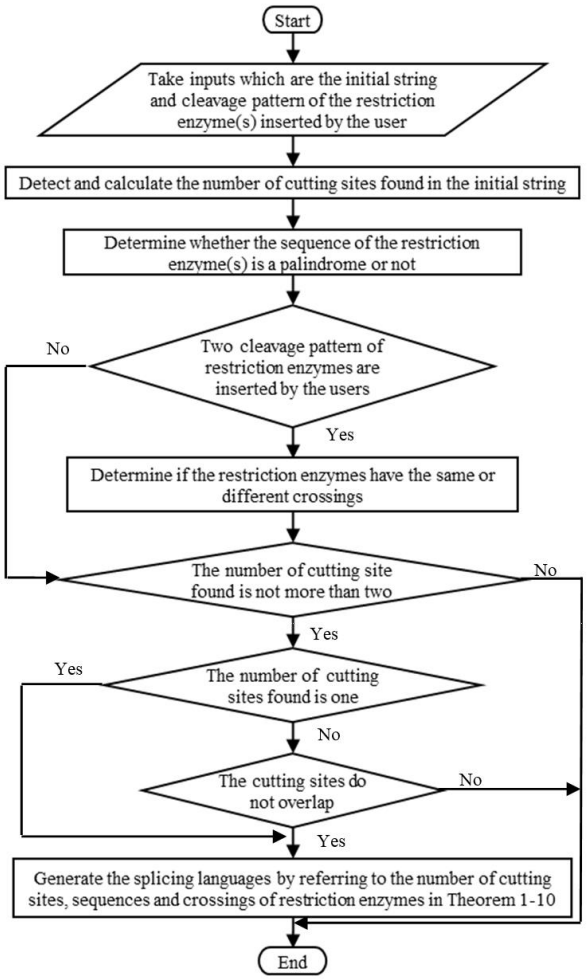

Fig. 1. Procedure and design for the algorithm

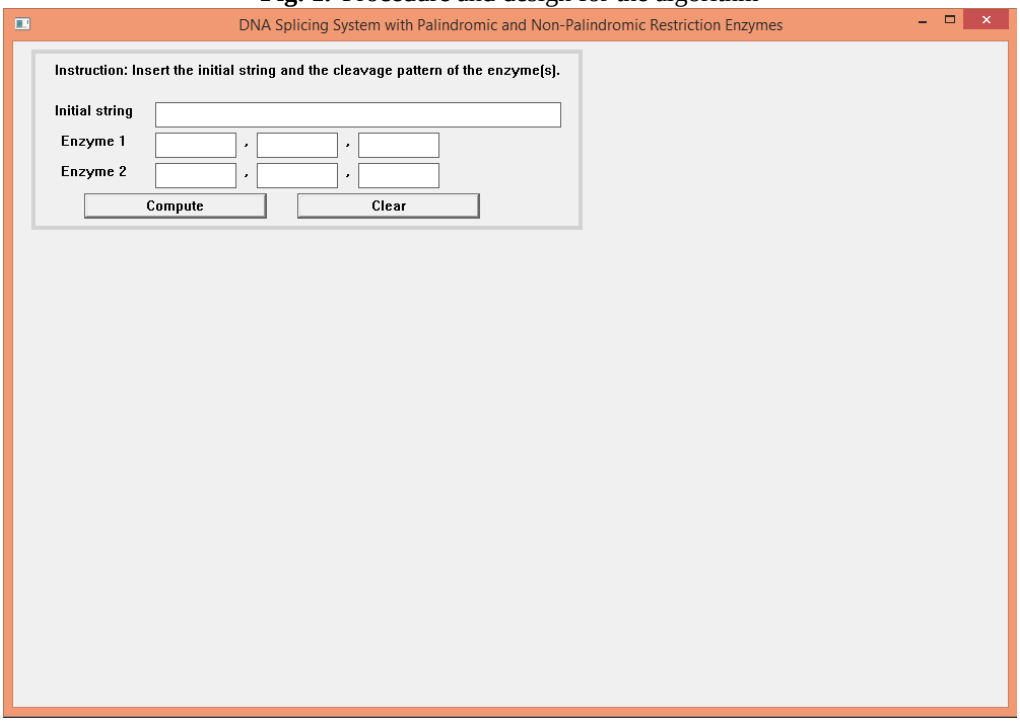

Fig. 2. The default graphical user interface

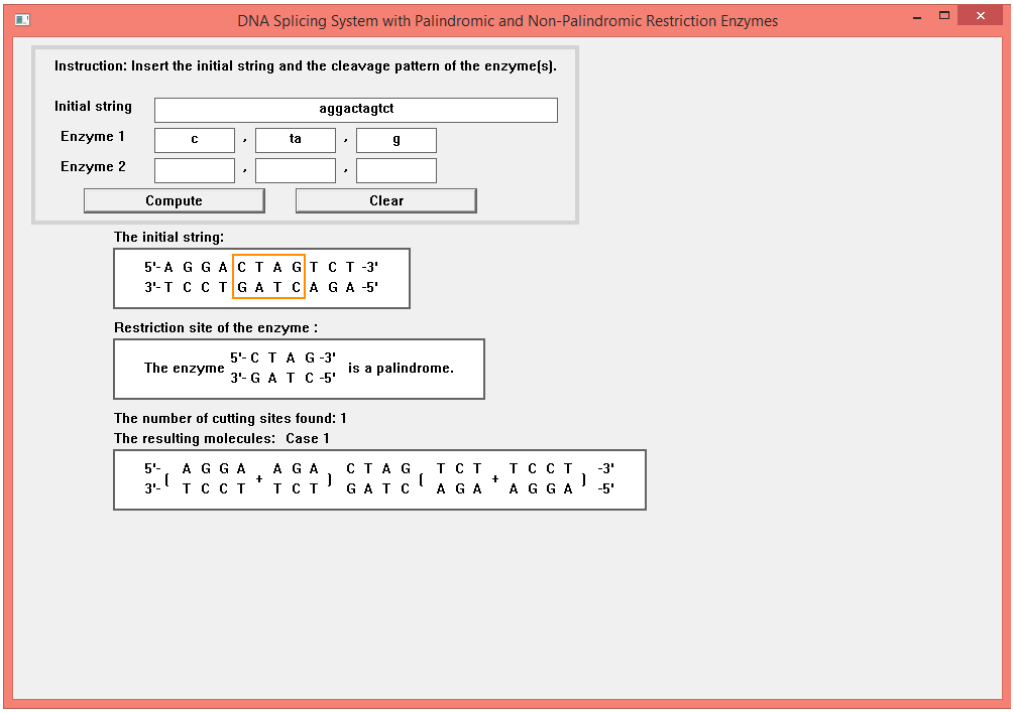

Fig. 3. The output for the DNA splicing system involving one cutting site of a palindromic restriction enzyme

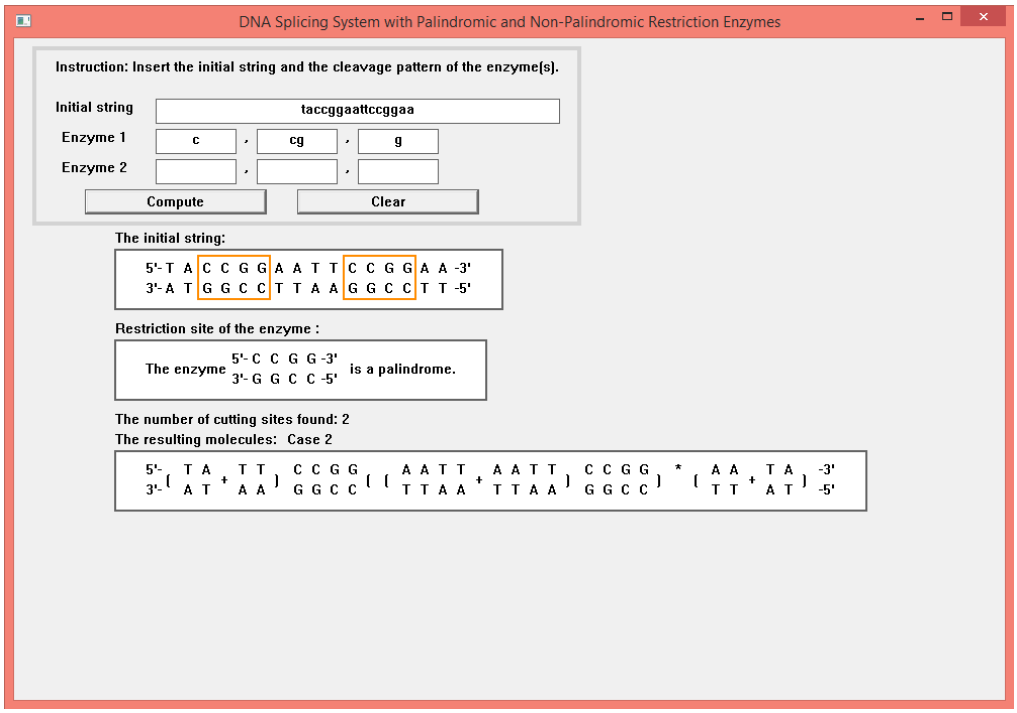

Fig 4. The output for the DNA splicing system involving two non-overlapping cutting sites of a palindromic restriction enzyme

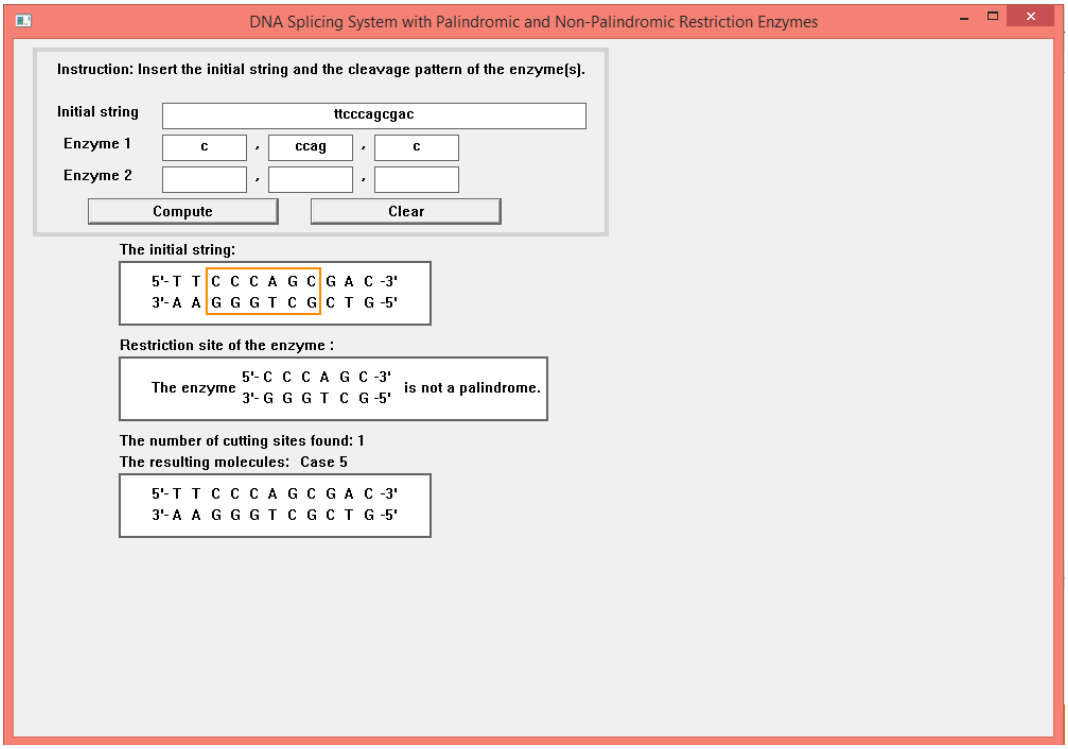

Fig. 5. The output for the DNA splicing system involving one cutting site of a non-palindromic restriction enzyme

In Fig. 6, the output for the DNA splicing system involving two non-overlapping cutting sites of a non-palindromic restriction enzyme is presented in the interface where the initial string is *ctcacgagttcacgagga* consisting the cleavage pattern of the enzyme *BssS*I which is {*c*, *acga*, *g*}.

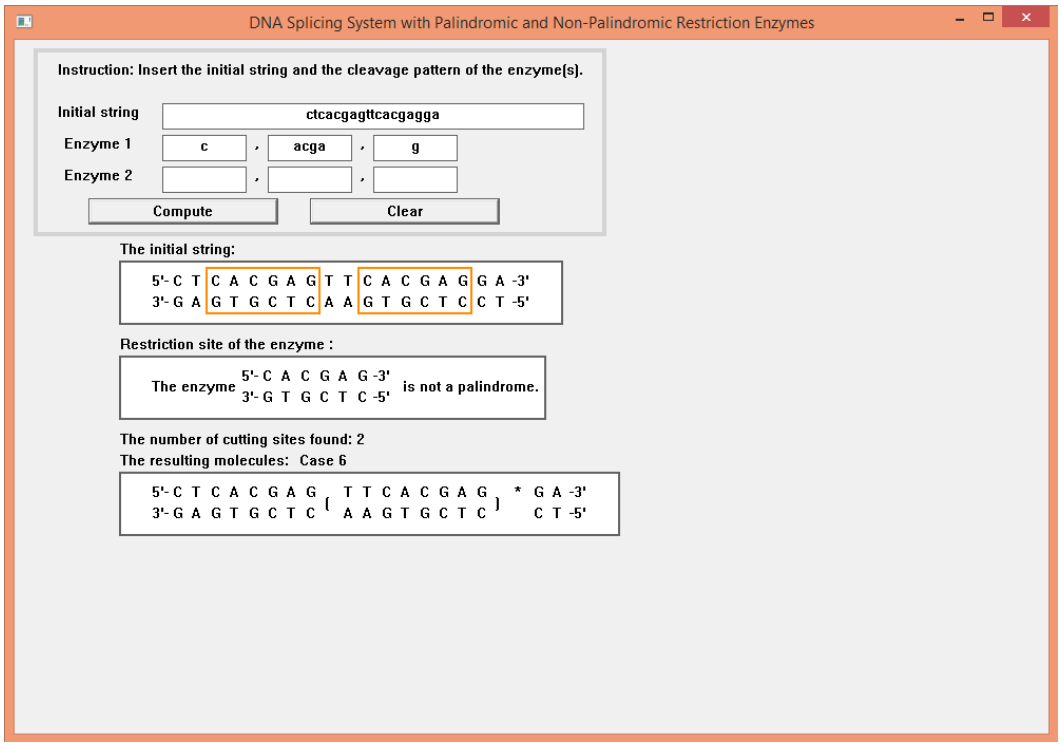

Fig. 6. The output for the DNA splicing system involving two non-overlapping cutting sites of a non-palindromic restriction enzyme

In Fig. 7, the output for the DNA splicing system involving one cutting site each of two palindromic restriction enzymes with the same crossing is presented in the interface where the initial string is *atttaactgtacaga* consisting the cleavage patterns of the enzymes *Mse*I and *CviQ*I which are {*t*, *ta*, *a*} and {*g*, *ta*, *c*} respectively.

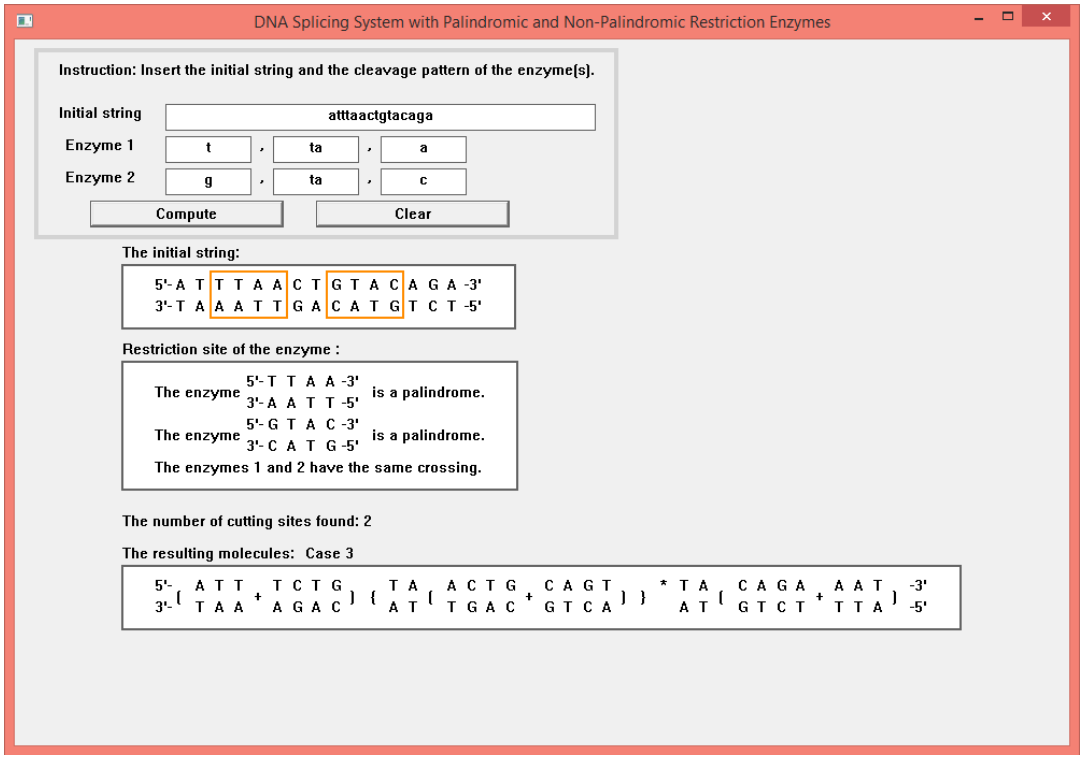

Fig. 7. The output for the DNA splicing system involving two palindromic restriction enzymes with the same crossing

In Fig. 8, the output for the DNA splicing system involving one cutting site each of two palindromic restriction enzymes with different crossings is presented in the interface where the initial string is *ttcgaattcagcgcag* consisting the cleavage patterns of the enzymes *Eco*RI and *Hin*P1I which are {*g*, *aatt*, *c*} and {*g*, *cg*, *c*} respectively.

The example of output for the DNA splicing system involving one cutting site each of two non-palindromic restriction enzymes with the same crossing is not given since there is no such enzyme available in the New England Biolabs Catalogue [19].

In Fig. 9, the output for the DNA splicing system involving one cutting site each of two non-palindromic restriction enzymes with different crossings is presented in the interface where the initial string is *attagcccaaggcacgagttc* consisting the cleavage patterns of the enzymes *Hba*I and *BssS*I which are {*g*, *ccca*, *a*} and {*c*, *acga*, *g*} respectively.

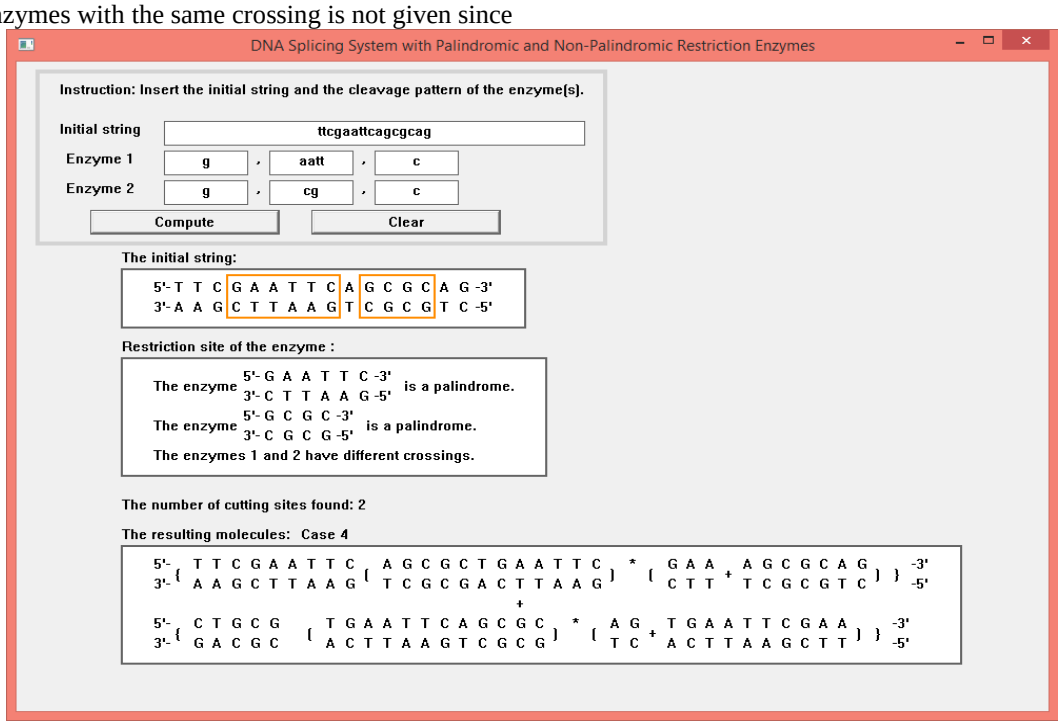

Fig. 8. The output for the DNA splicing system involving two palindromic restriction enzymes with different crossings

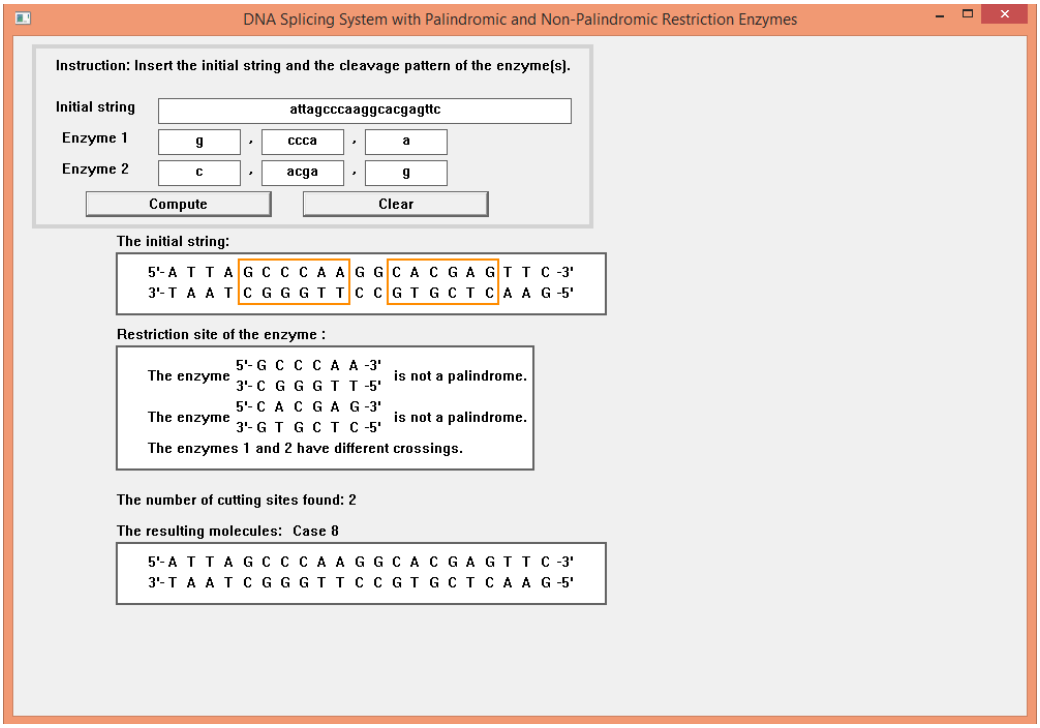

Fig. 9. The output for the DNA splicing system involving two non-palindromic restriction enzymes with different crossings

In Fig. 10, the output for the DNA splicing system involving one cutting site each of one palindromic and one non-palindromic restriction enzymes with the same crossing is presented in the interface where the initial string is *ttcgaacaccgcgc* consisting the cleavage patterns of the enzymes *Taq*I and *Aci*I which are {*t*, *cg*, *a*} and {*c*, *cg*, *c*} respectively.

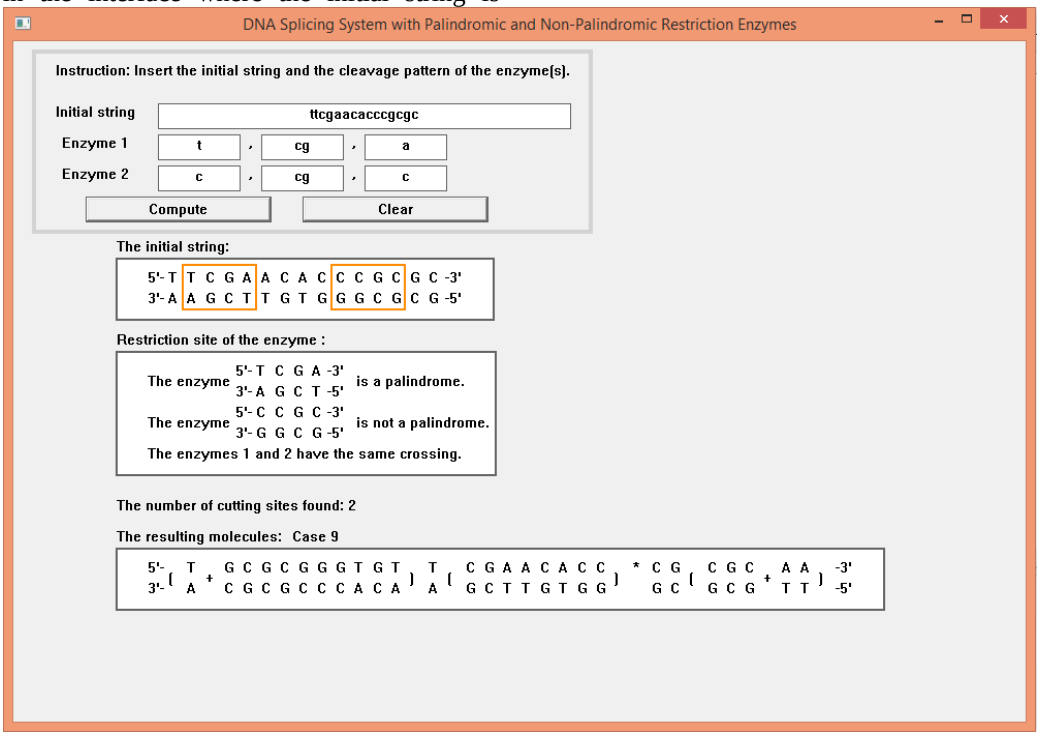

Fig. 10. The output for the DNA splicing system involving one palindromic and one non-palindromic restriction enzymes with the same crossing

In Fig. 11, the output for the DNA splicing system involving one cutting site each of one palindromic and one non-palindromic restriction enzymes with different crossings is presented in the interface where the initial string is *tgtacggaccgcgc* consisting the cleavage patterns of the enzymes *Cvi*QI and *Aci*I which are {*g*, *ta*, *c*} and {*c*, *cg*, *c*} respectively.

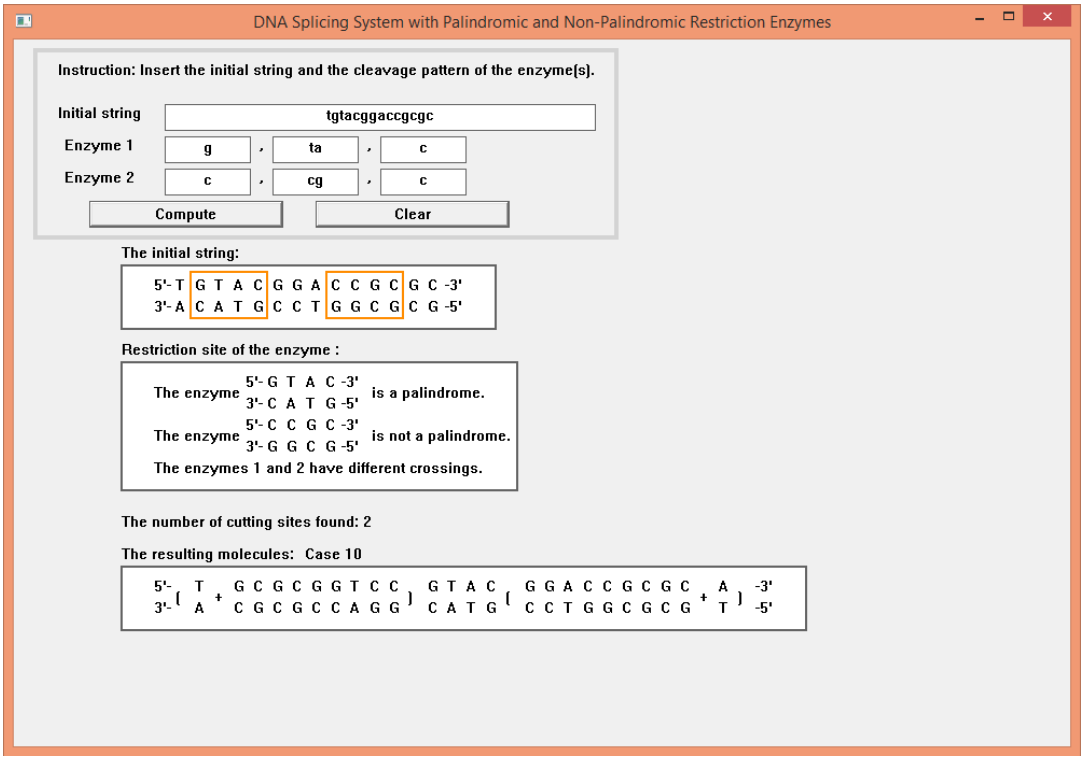

Fig 11. The output for the DNA splicing system involving one palindromic and one non-palindromic restriction enzymes with different crossings

As an additional feature, the interface also prompts the user if the number of cutting sites found exceeds two. The output in Fig. 12 shows the initial string *tgtacggacgcgcgcttgcgcgc* consisting three cutting sites of the enzymes *CviQI* and *BshII* which are {*g*, *ta*, *c*} and {*g*, *cgcg*, *c*} respectively.

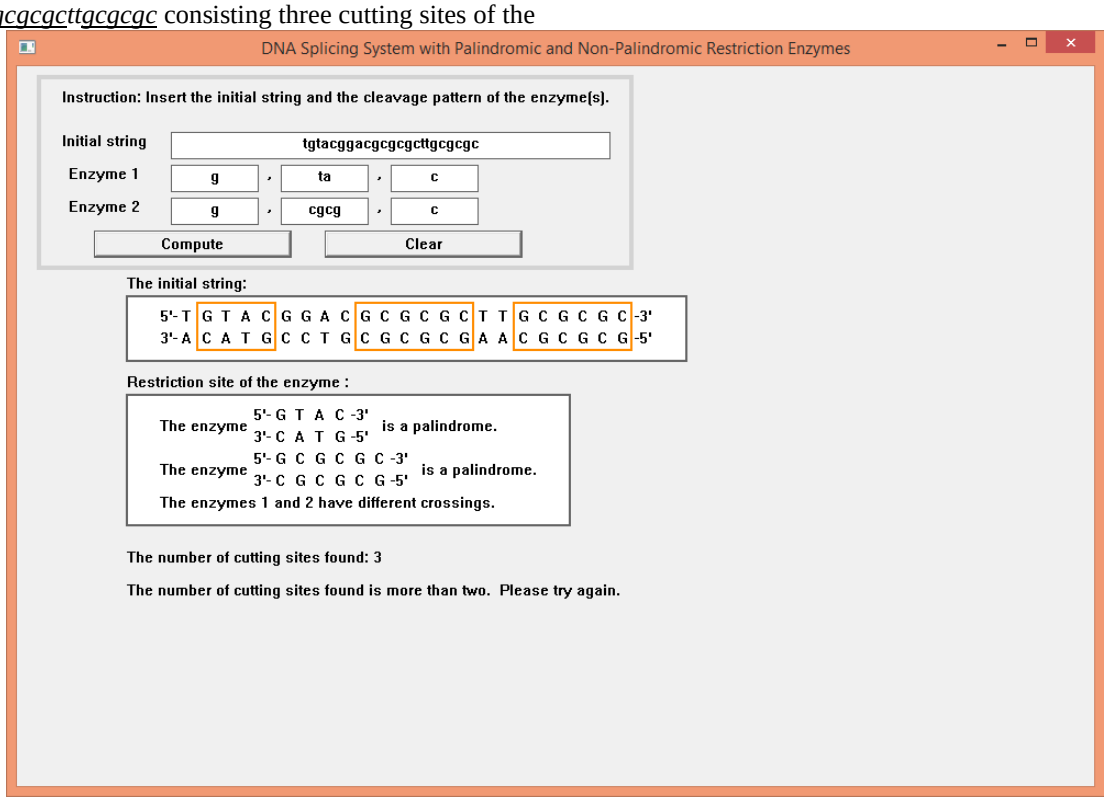

Fig. 12. The output showing three cutting sites

Besides that, the interface determines if the cutting sites of the enzyme overlap. For example, the first and second cutting sites of the enzyme *HinP*1I, {*g*, *cg*, *c*}, are overlapping in the initial string *acggagcgcgcg* as shown in Fig. 13.

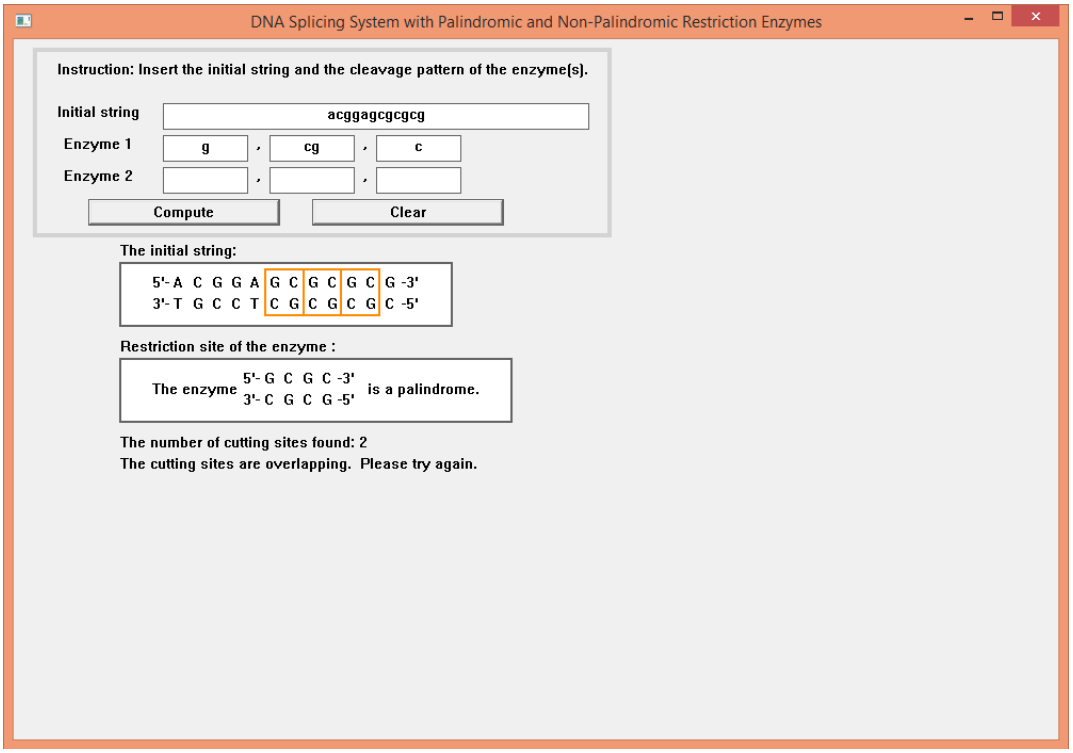

Fig. 13. The output showing overlapped cutting sites

Hence, the interface generates splicing languages from DNA splicing systems with at most two non-overlapping cutting sites of one and two restriction enzyme(s) using the generalised splicing languages from the corresponding splicing systems in Theorems 1-10 based on the crossings and sequences of the restriction enzymes.

IV. CONCLUSIONS

In this research, the algorithm is developed using C++ visual programming to design a graphical user interface for DNA splicing systems with one and two restriction enzymes. The purpose of this interface is to generate splicing languages, which are the output of this computation by inserting a random initial string and cleavage pattern of the restriction enzyme(s). In order to compute the output, the generalised splicing languages from the theorems are applied in the algorithm. From the theorems, the splicing languages are determined through some cases based on the number of cutting sites, sequences and crossings for restriction enzymes found in the initial string.

ACKNOWLEDGMENT

The first author expresses her gratitude to UTM Zamalah Scholarship for financially supporting her study. The second author and third author would like to thank the Ministry of Education (MOE) and Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the financial funding through Fundamental Research Grant Scheme Vote No. 5F022.

REFERENCES

1. R. P. Feynman, "There's Plenty of Room at the Bottom, " in *Engineering and Science*, H. D. Gilbert, Ed. New York: Reinhold Publishing Company, pp. 22-36, 1961.
2. L. M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems," *Science*, vol. 266, no. 5187, 1994, pp. 1021-1024.
3. G. Paun, G. Rozenberg and A. Salomaa, *DNA Computing: New Computing Paradigms*, Germany: Springer-Verlag Berlin Heidelberg, 1998, pp. 1-41.
4. T. Head, "Formal Language Theory and DNA: An Analysis of the Generative Capacity of Specific Recombinant Behaviors," *B. Math. Biol.*, vol. 49, no. 6, 1987, pp. 737-759.
5. S. M. Kim, "Computational Modeling for Genetic Splicing Systems," *SIAM J. Comput.*, vol. 26, no. 5, 1997, pp. 1284-1309.
6. P. Linz, *An Introduction to Formal Languages and Automata*, 4th ed. USA: Jones and Bartlett Publisher, 2006, pp. 1-36.
7. B. Brejová, C. DiMarco, T. Vinar, S. R. Hidalgo, G. Holguin and C. Patten, "Finding Patterns in Biological Sequences," Unpublished project report for CS798G, University of Waterloo, 2000.
8. T. Head, "Splicing Representations of Strictly Locally Testable Languages," *Discrete. Appl. Math.*, vol. 87, no. 1, 1998, pp. 139-147.
9. G. Păun, "On the Splicing Operation," *Discrete. Appl. Math.*, vol. 70, no. 1, 1996, pp. 57-79.
10. D. Pixton, "Regularity of Splicing Languages," *Discrete. Appl. Math.*, vol. 69, no. 1-2, 1996, pp. 101-124.
11. E. Goode, and D. Pixton, "Splicing to the Limit," in *Aspects of Molecular Computing, Lecture Notes in Computer Science*, N. Jonoska, G. Păun, and G. Rozenberg, Ed. Germany: Springer-Verlag, 2004, pp. 189-201.
12. Y. Yusof, N. H. Sarmin, W. H. Fong, T. E. Goode, and M. A. Ahmad. "An Analysis of Four Variants of Splicing System," *Proceeding of the 20th National Symposium on Mathematical Sciences - Research in Mathematical Sciences: A Catalyst for Creativity and Innovation (SKSM 2012)*. December 18-20, 2012. AIP Conference Proceedings. 1522(1): pp. 888-895.
13. I. Tomohiro, S. Inenaga and M. Takeda, "Palindrome Pattern Matching. *Theor. Comput. Sci.*, vol. 483, no. 2013, 2013, pp. 162-170.
14. W. H. Fong, and N. I. Ismail, "Generalisations of DNA Splicing Systems with One Palindromic Restriction Enzyme," *MATEMATIKA*, vol. 34, no. 1, 2018, pp. 59-71.
15. N. I. Ismail, W. H. Fong, and N. H. Sarmin, "DNA Splicing Systems with at Most Two Cutting Sites of a Non-Palindromic Restriction Enzyme," *Malaysian Journal of Industrial and Applied Mathematics*, to be published.
16. N. I. Ismail, W. H. Fong, and N. H. Sarmin, "Generalisations of Splicing Languages in DNA Splicing Systems Involving Two Palindromic Restriction Enzymes," *Malaysian Journal of Fundamental and Applied Sciences*, to be published.

17. W. H. Fong, N. I. Ismail, and N. H. Sarmin, "The Modelling of DNASplicing Systems with Two Non-Palindromic Restriction Enzymes," *Indian Journal of Public Health Research & Development*, to be published.
18. N. I. Ismail, W. H. Fong, and N. H. Sarmin. "The Mathematical Modelling of DNA Splicing System with Palindromic and Non-Palindromic Restriction Enzymes," *Proceeding of the International Conference on Applied Analysis and Mathematical Modelling*. 20-24 June 2018. ICAAMM 2018, Istanbul Gelism University, pp. 127-138.
19. New England Biolabs Inc, *NEB 2017-18 Catalog & Technical Reference*. Ipswich, United States. Catalogue, 2017.
20. Y. Yusof, *DNA Splicing System Inspired by Bio Molecular Operations*. Ph.D. Thesis. Universiti Teknologi Malaysia, 2012.

AUTHORS PROFILE

Nurul Izzaty Ismail is currently pursuing PhD in Mathematics and the title of her research is Generalisations of DNA Splicing Languages with Palindromic and Non-Palindromic Restriction Enzymes. She received her bachelor degree from Universiti Teknologi Malaysia (UTM) namely BSc (Hons) (Mathematics) in 2017. Then, she was offered from UTM to further her PhD study through the fast track program. To date, she has written and presented more than eight research papers in national and international conferences.

Wan Heng Fong is a senior lecturer at the Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia (UTM). She joined UTM as a lecturer in March 2008. She received her BSc (Hons) (Industrial Mathematics), MSc (Mathematics), and PhD (Mathematics) from UTM with her doctoral thesis titled Mathematical Modelling of Splicing Systems. Her major scientific contribution is in the area of splicing and sticker systems, particularly on the mathematical modelling and characterisations. Apart from that, she also does research on the representation of automata over groups and restrictions in variants of splicing and sticker systems in order to increase the generative capacity of splicing and sticker systems in the Chomsky hierarchy. She has written more than 140 research papers in national and international journals and proceedings, mainly in the area of splicing and sticker systems. She has also supervised 8 PhD students and 7 MSc students under her supervision. She has secured 9 research grants as the leader and 10 as a collaborator.

Nor Haniza Sarmin is a Professor in the Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia (UTM) Johor Bahru, Johor. She joined UTM as a lecturer in May 1991. She received her BSc (Hons) in 1989, MA in 1990 and PhD (1994-1998) in Mathematics from Binghamton University, Binghamton, New York, USA. Her specialization of research is in Group Theory, Formal Language Theory, Splicing Systems and Their Applications. She has written more than 500 research papers in national and international journals and proceedings, mainly in the area of group theory and splicing systems. She has also supervised and graduated 28 PhD students under her supervision. This includes the Youngest PhD Holder in Malaysia Book of Record 2014, Dr Hazzirah Mat Hassim. Prof. Dr. Nor Haniza Sarmin has secured more than 20 research grants as the Leader and more than 25 as a collaborator. She was holding the Deputy Dean's position in the School of Graduate Studies UTM from 2010-2014 and is now the Director (Global Education and Student Experience) in UTM International since April 2014.