# Some Analysis on Certain Types of Splicing Systems

Fong Wan Heng
Ibnu Sina Institute for Fundamental Science Studies,
Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Johor, Malaysia.
fwh@utm.my

Nor Haniza Sarmin, Yuhani Yusof
Department of Mathematics,
Faculty of Science, Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Johor, Malaysia.
nhs@utm.my, yuhani@ump.edu.my

***Abstract*— A mathematical model of a splicing system was firstly developed by Head in 1987. This model is abstractly analyzed in the framework of Formal Language Theory, which is a branch of Applied Discrete Mathematics and Theoretical Computer Science. This model consists of a finite set of initial strings over an alphabet that acts upon a finite set of rules. In this paper, some analysis on certain types of splicing systems namely, null-context, uniform, simple and $S_kH$ system are presented as propositions and corollaries. Besides, some counterexamples are given to illustrate these relations.**

***Keywords-mathematical model; splicing systems; Formal Language Theory***

## I. INTRODUCTION

Every living organism has a unique deoxyribonucleic acid (DNA). Strands of DNA are different from each other by the sequence of their bases namely: *Adenine* (*A*), *Guanine* (*G*), *Cytosine* (*C*) and *Thymine* (*T*). These bases are tied together by hydrogen bonds using base-pairing rules, where *A* pairs with *T*, *G* pairs with *C* and vice-versa [1]. These rules of pairing can be denoted as [*A*/*T*], [*G*/*C*], [*C*/*G*] and [*T*/*A*] or simply be written as *a*, *g*, *c* and *t*, respectively.

In New England Biolabs 2007-08 Catalogue & Technical Reference [2], there exist more than 200 types of restriction enzymes. These restriction enzymes are found in bacteria which can splice the DNA molecules at specific places, resulting in molecules with sticky or blunt ends based on their restriction sites. For example, the restriction enzyme *Afe*I produces blunt end, whereas the restriction enzyme *Acl*I produces sticky end during splicing. The resulting DNA molecules will recombine with the existence of a ligase.

The first mathematical modelling of splicing system was formally illustrated in the framework of Formal Language Theory by Head in 1987 [3]. This mathematical model represents the initial string of DNA as a set of finite initial string *I*, that acts upon a finite set of alphabet *A*, and with a finite set of rules *R* that will be defined in Section II later.

There are various types of splicing systems, including null-context, uniform, simple and $S_kH$ system. The null-context and uniform splicing systems have been introduced by Head [3]. This paper shows that each null-context splicing system is persistent and if a language *L* is a persistent splicing language, *L* is also uniform. Mateescu et al. [4] introduced the notion of simple splicing systems in 1998. Later, in 2008, some concepts involving simple splicing system using Formal Language Theory was done by Fong [5].

The sequence of language families $S_kH$ was introduced in 1998 [6]. A decade after, Fong et al. [7] reduced the $S_kH$ system to the simple splicing system using solid codes. Meanwhile in [8], Fong et al. introduced the concepts and examples of firmness and maximal firm subword (MFS) with their regular expressions and *SH*-automata applying on the reduction of splicing system.

This paper is organized into four sections. The first section is the introduction, followed by Section II which includes some definitions used in this research. In Section III, some analysis on certain types of splicing systems are done and presented as propositions, corollaries and counterexamples. Finally, the conclusion is given in the last section.

## II. PRELIMINARIES

Some formal definitions related to this research will be presented in this section. The main definition which is splicing system will first be defined.

Let *A* be defined as a fixed finite set to be used as an alphabet and $A^*$ as a free monoid that consists of all strings of symbols in *A*, including the null string.

**Definition 1:** [3] **(Splicing System)**
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. □

Next, the definitions of four types of splicing system that will be analyzed in this paper are presented.

**Definition 2:** [3] **(Null-Context Splicing System)**
A **null-context splicing system** is a splicing system $S = (A, I, B, C)$ for which each cleavage pattern in *B* and *C* has the form $(1, x, 1)$. □

**Definition 3:** [3] **(Uniform Splicing System)**
A **uniform splicing system** is a null-context splicing system $S = (A, I, X, X)$ for which there is a positive integer *P* such that $X = A^P$. □

**Definition 4:** [9] **(Simple Splicing System)**
Let $S = (A, I, R)$ be a splicing system in which all rules in *R* have the form $(a, 1; a, 1)$ where $a \in A$. Then *S* is called a **simple splicing system**. □

**978-1-4244-6439-5/10/$26.00 ©2010 IEEE**

**Definition 5:** [6] **($S_k$ Splicing System, $S_k$ Splicing Language)** Let $k$ be an integer $\geq -1$ . An **$S_k$ splicing system** ($S_kH$ system) is a null-context splicing system $G$ = ($A$, $I$, $R$) for which, for each string $r$ in $R$, length $r \leq k$. □

Note that when $k$ = 1, the $S_1H$ system is just the simple splicing system, denoted by $SH$.

## III. Some Analysis On Null-Context, Uniform, Simple and $S_kH$ Splicing System

In this section, some relations on different types of splicing system are analyzed and presented as propositions and corollaries. Besides, some counterexamples are given to illustrate them. Since Head's model is being used, the rules $R$ will be presented in notation of triples. Hence, the rule $R$ of simple splicing system in Definition 4 can be rewritten as $R = \{(1,a,1;1,a,1)\}$, where $a \in A$.

In the first proposition, a relation between simple and uniform splicing systems is presented.

**Proposition 1**
Every simple splicing system $S$ is uniform splicing system where $S$ = ($A$, $I$, $X$, $X$). □

**Proof**
Assume that $q$ is not an element of a uniform splicing system for which each crossing site in $X$ has the form of $(1,x,1)$ where $X = A^P$. Hence, $q$ is not an element of a simple splicing system since the crossing site of this splicing system is not in the form of $(1,x,1)$ and $A$ is a subset of $A^P$. ■

However, there exists a uniform splicing system that is not simple as presented in Example 1 below.

**Example 1**
Let $S = (\{a,g,c,t\}, I(\text{unspecified}), \{1, gatc, 1; 1, catg, 1\}, \varnothing)$ .
The left pattern consists of two restriction enzymes namely, *Dpn*II and *Fat*I, with the crossing sites as follows:
Crossing site for the enzyme *Dpn*II:

5'…▼*GATC* …3'
3'… *CTAG*▲…5' .

Crossing site for the enzyme *Fat*I:

5'…▼*CATG* …3'
3'… *GTAC*▲…5' .

Thus, $S$ is a uniform splicing system since both restriction enzymes, *Dpn*II and *Fat*I, have the same length of crossing with $P$ = 4. However, $S$ is not a simple splicing system since crossings of *Dpn*II and *Fat*I are *gatc* and *catg* respectively, which are two different elements that are not in $A$. ■

In the next proposition, the relation between uniform splicing system and $S_kH$ system is given.

**Proposition 2**
Every uniform splicing system is an $S_kH$ system $G$ where $G$ = ($A$, $I$, $R$). □

**Proof**
Suppose that $q$ is not an element of an $S_kH$ system for which each crossing site in $R$ has the form of $(1,x,1)$ and $r$ in $R$, length $r \leq k$ for $k \geq -1$ . Hence, $q$ is not an element of a uniform splicing system since $q$ itself is not an element of a null-context splicing system. ■

However, there exists an $S_kH$ system that is not uniform as illustrated in the following counterexample.

**Example 2**
Let $S = (\{a,g,c,t\}, I(\text{unspecified}), \varnothing, \{1, cgwcg, 1; 1, catg, 1\})$ be a splicing system where $w$ = $a$ or $t$. The right pattern consists of two restriction enzymes, namely *Hpy*99I and *Nla*III, with the crossing sites as follows:
Crossing site for the enzyme *Hpy*99I:

5'… *CGWCG*▼…3'
3'…▲*GCWGC* …5' .

Crossing site for the enzyme *Nla*III:

5'… *CATG*▼…3'
3'…▲*GTAC* …5' .

Thus, $S$ is an $S_kH$ splicing system with $k$ = 5 since five is the longest crossing for this splicing system. However, $S$ is not a uniform splicing system since restriction enzymes *Hpy*99I and *Nla*III have two different lengths of crossings $r$, which is five and four, respectively. ■

Propositions 1 and 2 lead to Corollary 1.

**Corollary 1**
Every simple splicing system is an $S_kH$ system. ■

In the next proposition, the relation between $S_kH$ system and null-context splicing system is presented.

**Proposition 3**
Every $S_kH$ system is null-context splicing system $S$ = ($A$, $I$, $B$, $C$). □

**Proof**
Suppose that $q$ is not an element of a null-context splicing system for which each crossing site in $B$ and $C$ has the form of $(1,x,1)$. Hence, by definition, $q$ is not an element of an $S_kH$ system. ■

However, there exists a null-context splicing system that is not an $S_kH$ system as given in the following counterexample.

**Example 3**
Let $S = (\{a,g,c,t\}, I(\text{unspecified}), \{1, ccwgg, 1; 1, gatc, 1\}, \varnothing)$ be a splicing system where $w$ = $a$ or $t$. The left pattern consists of two restriction enzymes, namely *PspG*I and *Sau*3AI, with the crossing sites as follows:
Crossing site for the enzyme *PspG*I:

5'…▼*CCWGG* …3'
3'… *GGWCC*▲…5' .

Crossing site for the enzyme *Sau*3AI:

5'…▼*GATC* …3'
3'… *CTAG*▲…5' .

Thus, $S$ is a null-context splicing system since both crossing for restriction enzymes *PspG*I and *Sau*3AI are in the form of $(1,x,1)$. However, $S$ is not an $S_kH$ system for $k$ = −1, 0, $n$ where $n$ are elements of $\{\mathbb{Z}^+ \setminus 5\}$ since the longest length of crossing $k$ for this splicing system is five. ■

Corollary 1 and Proposition 3 lead to the following corollary.

**Corollary 2**
Every simple splicing system is null-context splicing system.■

## IV. CONCLUSION

In this paper, some analysis on certain types of splicing systems, namely null-context, uniform, simple and $S_kH$ splicing systems are done and presented as Propositions 1, 2, 3, Corollaries 1, 2 and Examples 1, 2, 3. These relations can be simplified as follows:

simple splicing system $\subseteq$ uniform splicing system $\subseteq$ $S_kH$ splicing system $\subseteq$ null-context splicing system.


## ACKNOWLEDGMENT

We would like to thank the Ministry of Higher Education (MOHE) for the FRGS Vote No. 78482. The third author would also like to acknowledge Universiti Malaysia PAHANG (UMP) for her study leave and to MOHE for her SLAI scholarship.


## REFERENCES


[1] R. H. Tamarin, Principle of Genetics. USA: The McGraw-Hill Companies, 2001.

[2] Research Biolabs Sdn. Bhd., New England Biolabs 2007-08 Catalog & Technical Reference, USA, 2007.

[3] T. Head, "Formal Language Theory and DNA: An Analysis of the Generative Capacity Specific Recombinant Behaviors," Bull. Math. Biology, vol. 49, 1987, pp. 737-759.

[4] A. Mateescu, Gh. Paun, G. Rozenberg and A. Salomaa, "Simple Splicing Systems," Discrete Applied Mathematics, vol. 84, 1998, pp. 145-163.

[5] W. H. Fong, "Modelling of Splicing Systems Using Formal Language Theory," PhD Thesis, Universiti Teknologi Malaysia, 2008.

[6] T. Head, "Splicing Representations of Strictly Locally Testable Languages," Discrete Applied Mathematics, vol. 87, 1998, pp. 139-147.

[7] W. H. Fong, N. H. Sarmin and Z. Ibrahim, "Reduction of Splicing Systems using Solid Codes," Proceeding of the 16th Mathematical Sciences National Symposium, 2008, pp. 37-41.

[8] W. H. Fong, N. H. Sarmin and Z. Ibrahim, "Recognition of Simple Splicing Systems using *SH*-Automaton," Journal of Fundamental Sciences, vol. 4(2), 2008, pp. 337- 342.

[9] E. G. Laun, "Constants and Splicing Systems," Ph.D Thesis, State University of New York at Binghamton, 1999.