PAPER • OPEN ACCESS

Differentiating the persistency and permanency of some two stages DNA splicing language via Yusof-Goode (Y-G) approach

To cite this article: M H Mudaber *et al* 2017 *J. Phys.: Conf. Ser.* **890** 012081

View the article online for updates and enhancements.

Differentiating the persistency and permanency of some two stages DNA splicing language via Yusof-Goode (Y-G) approach

M H Mudaber[1], Y Yusof[2] and M S Mohamad[2]

[1]Mathematics Department, Faculty of Natural Sciences, Kabul Education University, Afshar District, Kabul, Afganistan
[2]Applied & Industrial Mathematics (AIMs) Research Cluster, Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Kuantan Pahang Darul Makmur.

E-mail: yuhani@ump.edu.my

Abstract. Predicting the existence of restriction enzymes sequences on the recombinant DNA fragments, after accomplishing the manipulating reaction, via mathematical approach is considered as a convenient way in terms of DNA recombination. In terms of mathematics, for this characteristic of the recombinant DNA strands, which involve the recognition sites of restriction enzymes, is called persistent and permanent. Normally differentiating the persistency and permanency of two stages recombinant DNA strands using wet-lab experiment is expensive and time-consuming due to running the experiment at two stages as well as adding more restriction enzymes on the reaction. Therefore, in this research, by using Yusof-Goode (Y-G) model the difference between persistent and permanent splicing language of some two stages is investigated. Two theorems were provided, which show the persistency and non-permanency of two stages DNA splicing language.

1. Introduction

Deoxyribonucleic acid or DNA is a double-stranded molecule, which its structure contains deoxyribose sugar, phosphate group and nitrogenous bases [1]. The backbone of DNA is formed by alternating of deoxyribose sugar and phosphate group. The four bases, which are adenine (A), guanine (G), cytosine (C) and thymine (T) are attached to the sugar and are located inside the sugar-phosphate (backbone) of dsDNA. Since the backbone of dsDNA screws around each other, by screwing them the double helical form of DNA is obtained. The ends of the two strands in a dsDNA are distinct and have different polarities. The two ends of upper and lower single strands of dsDNA are designed by the numbers $5'$ and $3'$, respectively. The c refers to the $5'$ carbon in the deoxyribose sugar, which a phosphate (PO_4) group is attached to that sugar ring, while the $3'$ refers to the $3'$ carbon in the sugar, which a hydroxyl group (OH) is attached to that sugar ring. Asymmetry of the two ends in dsDNA shows that each single strand has a direction as $5'-3'$ and $3'-5'$. According to Watson-Crick base pairing rules, the hydrogen bonding between the bases occur in such a way that A normally pairs with T by two hydrogen bonds and C with G by three hydrogen bonds. By using these pairing rules, the sequence of a DNA fragment can be written as dsDNA form. Restriction enzymes recognize a particular nucleotides sequence along dsDNA and produce $5'$ overhang or $3'$ overhang or blunt ends

dsDNA cut. The fragments of DNA strands will be recombined with their complementary ends at the existence of an appropriate DNA ligase to form recombinant DNA molecules (hybrid DNA).

A mathematical model named splicing system was firstly introduced by Head [2] to present the cutting and pasting process of DNA recombination via formal language theory. Since the recombinant DNA molecules will often be split by the existence of restriction enzymes, based on this actual biological fact Head [2] introduced the concept of persistent. Later Gatterdam [3] introduced another property of splicing system named permanent. Recently, many researches have been done research in the world of splicing system to study the persistency and permanency of splicing systems as well as languages. Some theorems provided in [4,5] which show the sufficient conditions for a splicing system to be persistent and also permanent. The persistency and permanency characteristics of some non-semi simple splicing system were investigated using Y-G approach, see [6]. Mudaber [7] investigated on two stages splicing languages and developed some mathematical theorems to present the relations between two stages splicing languages. In addition, some sufficient conditions for two stages splicing languages to be persistent and permanent were provided [8]. According to [6] and [8], a splicing system consisting of one initial string and one rule as well as two initial strings and two rules as discussed in the conditions of the theorems are persistent and permanent as well. Since these two properties of splicing system as well as splicing languages are always not the same, therefore, in this research, the differences between persistency and permanency of two stages splicing languages are investigated and presented by providing mathematical theorems focussing on Y-G splicing systems consisting of one initial string and three rules, and two initial strings and four rules, respectively.

In the next section, the fundamental concepts which related towards the new theorems constructed are discussed.

2. Preliminaries

In this section, the definitions of splicing system, persistent and permanent, which were introduced by Yusof [9], Head [2] and Gatterdam [3], are reviewed. Splicing system was formulated as a mathematical model by connecting the ideas from DNA molecular and formal language theory. In this formulism, the set of alphabet represents the four bases namely adenine, guanine, cytosine and thymine, and the set of initial strings represents the initial double-stranded DNA and the set of splicing rules represents the restriction enzymes acting on DNA molecules. Besides, the splicing languages, which are generated by splicing system, represent the recombinant DNA strands. The definition of Y-G splicing system is first given below.

Definition 1[9]: Yusof-Goode (Y-G) Splicing System
A Y-G splicing system is the form $S = (A, I, R)$, where A is the set of four alphabets a, g, c and t, I is the set of initial strings of double-stranded DNA and R is the set of splicing rules that indicates the set of enzymatic operation. The rule R in Y-G model is presented as $(u; x, v : y; x, z)$ and $(u, x; v : y, x; z)$ which show the left pattern and right pattern, respectively. However, the notation $(u, x, v : y, x, z)$ indicates that both hands patterns of rules were applied on DNA string. If $r \in R$, where $r = (u, x, v : y, x, z)$ and $s_1 = \alpha u x v \beta$ $s_1 = \alpha u x v \beta$ and $s_2 = \gamma y x z \delta$ are elements of I, then splicing s_1 and s_2 using r produce the initial string I together with $\alpha u x z \delta$ and $\gamma y x v \beta$, presented in either order where $\alpha, \beta, \gamma, \delta, u, v, x$ and $y \in A^*$ are free monoid generated by A with the concatenation operation and 1 as the identity element.□

In real situation, persistent and permanent are two properties of the recombinant DNA molecules, which show the existence of restriction enzyme(s) sequence(s) on the hybrid DNA strands. In the other words, the recombinant DNA strands will split by the existence of restriction enzyme(s) if they

are persistent and permanent. There is a main difference between persistent and permanent splicing systems. This difference is stated in the definitions of persistent and permanent. According to definitions, the two words, that make these two concepts difference, are **contain** in persistent and **is** in permanent. Hence, the definitions of persistent and permanent are stated below.

Definition 2[2]: Persistent

Let $S = (A, I, R)$ be a splicing system. Then S is persistent if for each pair of strings $ucxdv$ and $pexfq$, in A^* with (c, x, d) and (e, x, f) patterns of the same hands: if y is a sub segment of ucx (respectively xfq) that is crossing of a site in $ucxdv$ (respectively $pexfq$) then this same sub segment y of $ucxfq$ contains an occurrence of a crossing of a site in $ucxfq$.□

In the next, the definition of permanent is stated.

Definition 3[3]: Permanent

Let $S = (A, I, R)$ be a splicing system. Then S is permanent if for each pair of strings $ucxdv$ and $pexfq$, in A^* with (c, x, d) and (e, x, f) patterns of the same hands: if y is a sub segment of ucx (respectively xfq) that is crossing of a site in $ucxdv$ (respectively $pexfq$) then this same sub segment y of $ucxfq$ is an occurrence of a crossing of a site in $ucxfq$.□

Since this research presents the difference between persistency and permanency of two stages DNA splicing languages, therefore the definition of two stages splicing languages is given below.

Definition 4[7]: Two Stages Splicing Languages

Let $S = (A, I, R)$ is a splicing system. Furthermore, let $L = L(S)$ is the set of stage one splicing languages produced by splicing system S and $L' = L'(S)$ is the set of stage two splicing languages produced by S that consists of $L = L(S)$ and all splicing languages that can be resulted by splicing L. Then, the union of stage one and stage two splicing languages are called two stages splicing languages.□

In the next section, the results and discussion of this paper is presented.

3. Results and Discussion

In this section, the main difference between the two concepts persistent and permanent will be investigated. Since permanent is a sub set of persistent, thus every permanent splicing system is persistent. However, the reverse of statement is not always true. Therefore, to show there are persistent splicing languages, which are not permanent, two theorems are provided. Theorem 1 presents the persistency and non-permanency of two stages splicing languages, which are generated by a Y-G splicing system, involving one initial string and three rules. However, the second theorem presents the persistency and permanency of two stages DNA splicing languages, which are generated by a Y-G splicing system, involving two initial strings and four splicing rules.

Theorem 1: Let $S = (A, I, R)$ is a Y-G splicing system involving one initial string and three splicing rules such that the crossing sites of the rules are identical, and the left and right contexts of the first rule is same and the whole sequences of second and third rules are palindromic, then the set of two stages splicing language, which is produced by S, is persistent, but not permanent.□

Proof: Assume the set R consisting of three splicing rules having left pattern $R = \{r_1, r_2, r_3\}$ where $r_1 = (u_1; u_1u_2, u_1 : u_1; u_1u_2, u_1)$, $r_2 = (u_1; u_1u_2, u_2 : u_1; u_1u_2, u_2)$ and $r_3 = (u_2; u_1u_2, u_1 : u_2; u_1u_2, u_1)$. Suppose $I = \{pu_1u_1u_2u_1q\}$ is the set of initial string, which is only cut by the rule r_1, such that $[p/p']$ and $[q/q']$ and $u_1, u_2, p, q \in A^*$. To prove the set of two stages DNA splicing language is persistent and not permanent, the two stages splicing language need to be obtained. By applying the rule r_1 on the initial string $pu_1u_1u_2u_1q$ in the set, I two distinct DNA splicing languages can be yielded at stage one presented as $L(S)$.

$$L(S) = I \cup \{pu_1u_1u_2u_2 p', q'u_2u_1u_2u_1q\}$$

To achieve the splicing operation at stage two, both of splicing rules are simultaneously applied on the produced DNA splicing languages $L(S)$. When splicing takes place, no distinct DNA splicing languages will be produced at stage two. Hence, the set of stage one and stage two splicing languages are the same as presented in the following set.

$$L'(S) = L(S) = I \cup \{pu_1u_1u_2u_2 p', q'u_2u_1u_2u_1q\}$$

The above resulted two stages DNA splicing languages are persistent, but not permanent due to the following statements:

First of all, the two stages splicing language is proven to be persistent. By taking u_1u_2 as sub fragment of $pu_1u_1u_2$, that is crossing of the site in $pu_1u_1u_2u_1q$. This same sub segment u_1u_2 also contains an occurrence of the crossing of a site in the yielded strings $pu_1u_1u_2u_2 p'$ and $q'u_2u_1u_2u_1q$. Thus, according to definition of persistent, the yielded two stages splicing languages is persistent.

Second, the non-permanency of two stages DNA splicing languages is proven. According to definition of permanent, if u_1u_2 be a sub segment of $pu_1u_1u_2$, that is crossing of $pu_1u_1u_2u_1q$. This same sub segment u_1u_2 is not a crossing of a site in the yielded strings $pu_1u_1u_2u_2 p'$ and $q'u_2u_1u_2u_1q$. Hence, the two stages DNA splicing language is not permanent. ∎

In the next theorem, the difference between persistent and permanent perspectives of two stages splicing languages is investigated. To distinguish these two concepts, a Y-G splicing system associated with two initial strings and four splicing rules is considered.

Theorem 2: Let $S = (A, I, R)$ is a Y-G splicing system involving two initial strings and four rules such that the crossing sites of the rules are identical and self-closed, then the set of two stages splicing language, which is produced by S, is persistent, but not permanent.□

Proof: Assume the set of splicing rules is left pattern, $R = \{r_1, r_2, r_3, r_4\}$ where $r_1 = (u_1; u_1u_2, u_2 : u_1; u_1u_2, u_2)$, $r_2 = (u_2; u_1u_2, u_1 : u_2; u_1u_2, u_1)$, $r_3 = (u_1; u_1u_2, u_1 : u_1; u_1u_2, u_1)$ and $r_4 = (u_2; u_1u_2, u_2 : u_2; u_1u_2, u_2)$, and $u_1, u_2 \in A^*$. Suppose the set of initial strings, I consists of two initial strings $s_1 = pu_1u_1u_2u_2q$ and $s_2 = ru_2u_1u_2u_1s$ such that $p, q, r, s \in A^*$ and

$[p/p']$, $[q/q']$, $[r/r']$ and $[s/s']$. To show the two stages DNA splicing languages are persistent and not permanent, according to crossing site properties of rules two cases need to be considered.

Case 1: Suppose the crossing site of the rule are palindromic, u_1 is complementary with u_2 and vice-versa. According to the rules assumption, the crossing sites of rules are palindromic and identical. Therefore, when the first rule r_1 is applied on s_1 and the second rule r_2 is applied on s_2, the following eight DNA splicing language are generated at stage one besides initial string I namely,

$$pu_1u_1u_2u_2p', \ pu_1u_1u_2u_1s, \ pu_1u_1u_2u_1r', \ q'u_1u_1u_2u_2q,$$
$$q'u_1u_1u_2u_1s, \ ru_2u_1u_2u_2q, ru_2u_1u_2u_1r', \ s'u_2u_1u_2u_1s$$

To obtain the stage two splicing language, the above four rules are applied on the resulted splicing language of stage one. Thus, when splicing takes place no distinct splicing language are produced.

First, the persistency of two stages DNA splicing language, which is obtained by this splicing system needs to be proven. By taking u_1u_2 as a sub segment of $pu_1u_1u_2$ (respectively $u_1u_2u_1s$), that is crossing of $pu_1u_1u_2u_2q$ (respectively $ru_2u_1u_2u_1s$). Then this same sub segment u_1u_2 contains an occurrence of the crossing of a site in $pu_1u_1u_2u_1s$ and all splicing languages that produce by splicing system. Hence, according to definition of persistent the two stages DNA splicing languages are persistent.

Second, the non-permanency of two stages DNA splicing languages is proven. According to definition of permanent if u_1u_2 be a sub segment of $pu_1u_1u_2$ (respectively $u_1u_2u_1s$), that is crossing of $\alpha a_1a_1a_2a_2\beta$ (respectively $\gamma a_2a_1a_2a_1\delta$). Then, this same sub segment u_1u_2 is not a crossing of a site in the obtained splicing language $pu_1u_1u_2u_1s$. Consequently, the two stages splicing language is not permanent.

Case 2: Assume the crossing sites of the rules are non-palindromic, u_1 is not complementary with u_2 and vice-versa. According to the rules assumption, the crossing site is identical, thus there is a possibility to recombine the fragments of initial strings after the strings are cut by the existence of rules. To generate the splicing languages at stage one, the rule r_1 is applied on s_1 and the rule r_2 is applied on s_2. When splicing operation occurs 2 new splicing languages will be generated at stage one besides the set of initial strings, I as listed below.

$$pu_1u_1u_2u_1s, ru_2u_1u_2u_2q$$

However, when splicing takes place among the splicing languages of stages one at the existence of the above four splicing rules, no distinct splicing languages will be produced. Since these two splicing language are exist in the set of two stages DNA splicing languages that has discussed in **Case 1**, therefore the proof for persistency as well as non-persistency of two stages DNA splicing language follows **Case 1**. Hence, the theorem proved. ∎

If the left and right contexts of the splicing rules in Theorem 1 as well as Theorem 2 be null string, then the generated two stages splicing languages is persistent as well as permanent. This result is presented in the following corollary.

Corollary 1: Suppose $S = (A, I, R)$ is a Y-G splicing system involving one two initial string and three rules or two initial strings and four rules such that the crossing sites of the rules are identical and the left and right contexts of the splicing rules are null string, then the set of two stages splicing language, which is produced by S, is persistent and permanent. ∎

In the last section, the conclusion of this paper is presented.

4. Conclusions

In this investigation, the two important properties of some two stages splicing language is studied and differentiated by providing two theorems. The first theorem consists of one initial string and three splicing rules, while the second theorem consists of two initial strings and four splicing rules. Besides, a corollary is provided, which show if the contexts of the rules be null string, the persistent and permanent are same. The result of this research is simplified as below:

Permanent is a subset of persistent if

$$S = \left(\{a, g, c, t\}, \{ pu_1 u_1 u_2 u_1 q \}, \{ (u_1; u_1 u_2, u_1 : u_1; u_1 u_2, u_1), (u_1; u_1 u_2, u_2 : u_1; u_1 u_2, u_2), (u_2; u_1 u_2, u_1 : u_2; u_1 u_2, u_1) \} \right)$$

or

$$S = \left(\{a, g, c, t\}, \{ pu_1 u_1 u_2 u_2 q, ru_2 u_1 u_2 u_1 s \}, \begin{cases} (u_1; u_1 u_2, u_2 : u_1; u_1 u_2, u_2), (u_2; u_1 u_2, u_1 : u_2; u_1 u_2, u_1), \\ (u_1; u_1 u_2, u_1 : u_1; u_1 u_2, u_1), (u_2; u_1 u_2, u_2 : u_2; u_1 u_2, u_2) \end{cases} \right)$$

Permanent is the same as persistent if the left and right contexts of the splicing rules are null string,

$$S = \left(\{a, g, c, t\}, \{ p1u_1 u_2 1q \}, \{ (1, u_1 u_2, 1 : 1, u_1 u_2, 1), (1, u_1 u_2, 1 : 1, u_1 u_2, 1), (1, u_1 u_2, 1 : 1, u_1 u_2, 1) \} \right)$$

or

$$S = \left(\{a, g, c, t\}, \{ p1u_1 u_2 1p, r1a_1 a_2 1s \}, \begin{cases} (1, u_1 u_2, 1 : 1, u_1 u_2, 1), (1, u_1 u_2, 1 : 1, u_1 u_2, 1), \\ (1, u_1 u_2, 1 : 1, u_1 u_2, 1), (1, u_1 u_2, 1 : 1, u_1 u_2, 1) \end{cases} \right).$$

References
[1]   Allison L A 2007 *Fundamental Molecular Biology* Blackwell Pub Publishing
[2]   Head T 1987 *Bulletin of Mathematical Biology* **49** 737-759
[3]   Gatterdam R W 1989 *International Journal of Computer Math.* **31** 63-67
[4]   Karimi F, Sarmin N H and Fong W H 2011 *Australian Journal of Basic and Applied Sciences* **5** 20-24
[5]   Fong W H, Sarmin N H and Karimi F 2012 *Proceeding of International Conference on Enabling Science and Nanotechnology* (ESciNano) p 1-2.
[6]   Yusof Y, Sarmin N H and Fong W H 2014 *Proceeding of SKSM21* p 586-590
[7]   Mudaber M H, Yusof Y and Mohamad M S 2014 *AIP Conference Proceeding* 1602 (NY: Melville) p 254-259
[8]   Mudaber M H, Yusof Y and Mohamad M S 2014 *AIP Conference Proceeding* 1605 (NY: Melville) p 591-595
[9]   Mudaber M H, Yusof Y, Mohamad M S and Lim WL 2015 *AIP Conference Proceeding* 1643 (NY: Melville) p 591-595
[10]   Mudaber M H, Yusof Y, Mohamad M S and Lim W L 2015 *ISCA-Research Journal* of *Mathematical and Statistical Sciences* **3**(10) 15-25