On pseudoknot-bordered words and their properties *

Lila Kari and Shinnosuke Seki

Department of Computer Science, University of Western Ontario,
London, Ontario, Canada, N6A 5B7
E-mail: *lila, sseki@csd.uwo.ca*
Tel: *1-519-661-2111*
Fax: *1-519-661-3515*

Abstract

We study a generalization of the classical notions of bordered and unbordered words, motivated by biomolecular computing. DNA strands can be viewed as finite strings over the alphabet {A, G, C, T}, and are used in biomolecular computing to encode information. Due to the fact that A is Watson-Crick complementary to T and G to C, DNA single strands that are Watson-Crick complementary can bind to each other or to themselves forming so-called secondary structures. Most of these secondary structures are undesirable for biomolecular computational purposes since the strands they involve cannot further interact with other strands. This paper studies pseudoknot-bordered words, a mathematical formalization of pseudoknot-like inter- and intra-molecular structures. In this context, pseudoknot-unbordered words model DNA or RNA strands that will be free of such secondary structures. We obtain several properties of pseudoknot-bordered and -unbordered words. We also address following problem: Given a pseudoknot-unbordered word u, does $\{u\}^+$ consist of pseudoknot-unbordered words only? We show that this is not generally true. We find that a sufficient condition for $\{u\}^+$ to consist of pseudoknot-unbordered words only is that u be not primitive. All of our results hold for arbitrary antimorphic involutions, of which the DNA Watson-Crick complementarity function is a particular case.

Key words: DNA computing, bordered words, unbordered words, RNA pseudoknots, pseudoknot-bordered words
1991 MSC: 68R15, 92B05, 03D40, 68Q42

1 Introduction

In this paper we study pseudoknot-bordered and pseudoknot-unbordered words, which are generalizations of the classical notions of bordered and unbordered words, motivated by the need of optimally encoding information as DNA strands for biomolecular computing purposes.

A DNA single strand is a linear chain made up of four different types of nucleotides, each consisting of a sugar-phosphate unit and a base (Adenine, Cytosine, Guanine or Thymine). The sugar-phosphate units are linked together by strong covalent bonds, to form the backbone of the DNA strand. Since nucleotides may differ only by their bases, a DNA single strand can be viewed as a string over the DNA alphabet of bases $\{A, C, G, T\}$. A DNA single strand has an orientation, with one end known as the 3' end, and the other known as the 5' end, based on their chemical properties. By convention, a word over the DNA alphabet represents a DNA single strand in its 5' to 3' orientation. An essential biochemical property of DNA single strands is that of Watson-Crick complementarity, wherein A can bind to T and C can bind to G by weak hydrogen bonds. (In the case of RNA, T is replaced by U, and U is complementary to A, though the binding U-G may also occur.) Two Watson-Crick complementary DNA single strands of opposite orientation can bind to each other to form a DNA double strand. This and other biochemical properties of DNA have all been harnessed in biomolecular computing [1], in which information is encoded as DNA single strands, and processed through bio-operations [4].

One of the problems encountered when encoding information as DNA single strands is that the Watson-Crick complementarity often results in information-encoding DNA single strands either folding onto themselves to form intra-molecular structures, or interacting with each other to form inter-molecular structures. While these so-called secondary structures optimize biochemical determinants such as the Gibbs free-energy [17] and often have a significant role in determining the biochemical functions of real-life nucleic acids (DNA or RNA), in DNA computing they are often seen as a disadvantage. This is because it is very likely that the secondary structure formation of DNA strands will prevent them from interacting with other DNA strands in the expected, pre-programmed ways. Consequently, the property of a set of information-encoding strands to be free of unwanted intra- and inter-molecular structures has been intensively studied from many different points of view. These include design of algorithms based on free energy [2,3,14], algebra [13], and formal language theory [8–11].

In this context, the notion of *antimorphic involution* θ was proposed, as the most natural mathematical formalization of the notion of DNA Watson-Crick complementarity [7,9,12]. Using this notion, Kari and Mahalingam [11] intro-
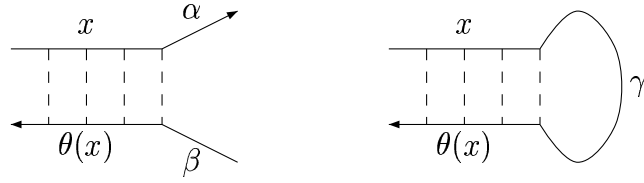
Fig. 1. Inter- and intra-molecular structures which θ-unbordered words avoid.
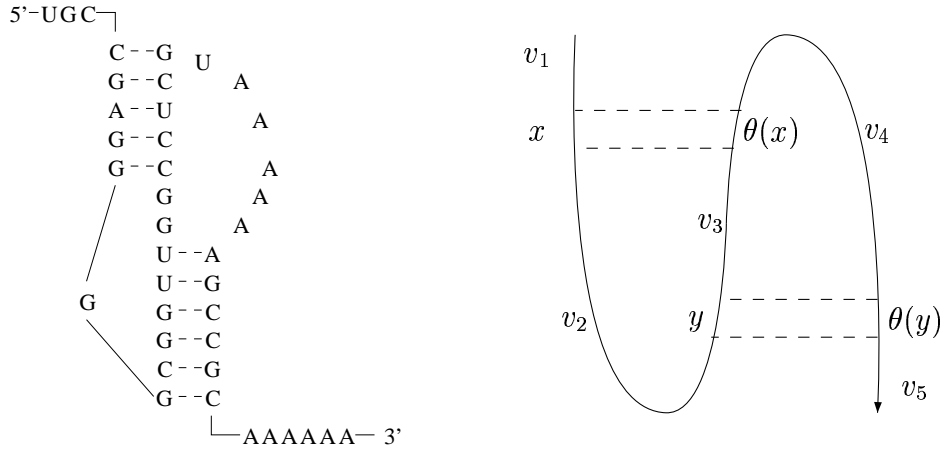

Fig. 2. *Left:* A pseudoknot found in *E. Coli* transfer-messenger-RNA. (From Rfam [6]). *Right:* A depiction of a string modelling the pseudoknot in *Left*, as a word $v_1xv_2yv_3\theta(x)v_4\theta(y)v_5$. Here, $v_1 = $ UGC, $x = $ CGAGG, $v_2 = $ G, $y = $ GCGGUU, $v_3 = $ GG, $v_4 = $ UAAAAA, and $v_5 = $ AAAAAA.

duced and investigated the concept of a θ-unbordered word, as a formalization of DNA strands that avoid some of the most common inter- and intra-molecular structures. A θ-bordered word is a nonempty word which has a nonempty prefix x, and a suffix $\theta(x)$. If the alphabet under consideration is the DNA alphabet, and θ is the Watson-Crick complementarity function, then a θ-unbordered word represents a population of identical DNA single strands that are free from both inter-molecular structures such as the ones shown in Figure 1 (Left), and *hairpins* (words of the form $x\gamma\theta(x)$, shown in Figure 1, (Right)), one of the most common DNA intra-molecular structures. In addition to being relevant for DNA computing, the notions of θ-bordered and θ-unbordered words are generalizations of classical notions in combinatorics of words, namely those of bordered [5] (a.k.a. overlapping [20,23], unipolar [21] words), respectively unbordered words (a.k.a. *d*-primitive, dipolar words).

The pseudoknot is another intra-molecular structure of biological significance, formed primarily by RNA strands. A pseudoknot found in *E.Coli* transfer-messenger-RNA is shown in Figure 2 (Left). This type of pseudoknot, which is the simplest and hence the most common, can be modelled as a word of the
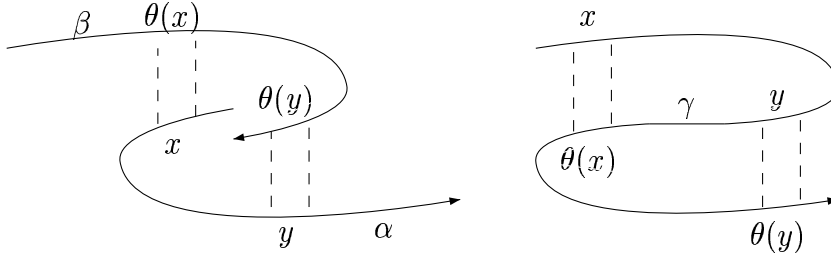
4

Fig. 3. An inter-molecular structure and intra-molecular structure which θ-pseudoknot-unbordered words avoid.

form $v_1 x v_2 y v_3 \theta(x) v_4 \theta(y) v_5$, as shown in Figure 2 (Right).

In this paper, we investigate not only such θ-pseudoknot-bordered words, but θ-pseudoknot-unbordered words, the latter being models of DNA or RNA strands that will not form pseudoknot-like inter- and intra-molecular structures. A nonempty word w is θ-pseudoknot-bordered if $w = xy\alpha = \beta\theta(yx)$ for some words x, y, α, and β. Thus, θ-pseudoknot-unbordered words avoid both inter-molecular bonds between identical strands of the type depicted in Figure 3 (Left), and intra-molecular structures of the form $xy\gamma\theta(x)\theta(y)$ shown in Figure 3 (Right). Note that this a particular case of the general model of pseudoknots, namely the case where $v_1 = v_2 = v_4 = v_5 = \lambda$.

The paper is organized as follows. Using the notations and terminology given in Section 2, we propose the notion of θ-pseudoknot-bordered words in Section 3 and present some of their basic properties. We also show that the notion of θ-pseudoknot-bordered word is a proper generalization of that of θ-bordered word, and thus also properly generalizes the classical notion of bordered word. Since information-encoding DNA single strands often need to be concatenated together in the course of biocomputations, another problem of interest, which we address in Section 4, is whether the property of being pseudoknot-unbordered is preserved by catenation. Here we address the simplest case of this problem: Given a pseudoknot-unbordered word u, are all the words in $\{u^+\}$ still pseudoknot-unbordered? This turns out not to be always the case. However, we find a sufficient condition for a θ-pseudoknot-unbordered word u to satisfy the property that any power of u remains θ-pseudoknot-unbordered: the condition is that u be non-primitive (Corollary 21). Section 5 discusses possible further directions of research.

2   Preliminaries

In this section we introduce the terminology and notations used in the paper. For details, we refer the reader to [19,20,23].

5

Let Σ be a finite alphabet. We denote by Σ^* the set of all words over Σ, and by Σ^+ the set of all nonempty words over Σ. Let λ be the empty word. Then $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$. For a word $w \in \Sigma^*$, $|w|$ denotes the length of w. A word u is said to be a *prefix* (*suffix*) of w if $w = uv$ (resp. $w = vu$) for some $v \in \Sigma^*$; here if $v \neq \lambda$, then the prefix (suffix) u is said to be *proper*. Let $\mathrm{Pref}(w)$ ($\mathrm{Suff}(w)$) be the set of all prefixes (resp. suffixes) of w.

A word $z \in \Sigma^*$ is said to be a *border* of a word $w \in \Sigma^*$ if $w = uz = zv$ for some words u and v in Σ^*. A nonempty word is said to be *bordered* if it admits a nonempty border, and it is said to be *unbordered* otherwise. A word $w \in \Sigma^+$ is called *primitive* if it cannot be written as a power of another word, i.e., $w = u^n$ with $u \in \Sigma^+$ implies $n = 1$. For a word $w \in \Sigma^+$, the shortest $u \in \Sigma^+$ satisfying $w = u^n$ for some $n \geq 1$ is called the *primitive root* of w. It is well known [15] that every nonempty word has a unique primitive root. Moreover, we have the following result due to Lyndon and Schützenberger.

Theorem 1 *For $u, v \in \Sigma^+$, $uv = vu$ implies that u and v have the same primitive root.*

For a word $w \in \Sigma^*$, a word $v \in \Sigma^*$ is called a *cyclic permutation* of w if there exist two words $x, y \in \Sigma^*$ such that $w = xy$ and $v = yx$. We denote the set of all cyclic permutations of w by $\mathrm{Cp}(w)$, that is, $\mathrm{Cp}(w) := \{yx \mid w = xy, x, y \in \Sigma^*\}$. Moreover, for a language $L \subseteq \Sigma^*$, we define $\mathrm{Cp}(L) := \bigcup_{w \in L} \mathrm{Cp}(w)$.

An *involution* $\theta : \Sigma \to \Sigma$ of a set Σ is a function such that θ^2 equals the identity function, i.e., $\theta(\theta(a)) = a$ for all $a \in \Sigma$. A *morphism* (*antimorphism*) θ on Σ^* is a function such that $\theta(xy) = \theta(x)\theta(y)$ (resp. $\theta(xy) = \theta(y)\theta(x)$) for all $x, y \in \Sigma^*$. A *d-morphism* is a generic term that refers to a function that is either a morphism or an antimorphism. An involution θ can be extended to a function $\theta : 2^{\Sigma^*} \to 2^{\Sigma^*}$, for a given language $L \subseteq \Sigma^*$, as follows: $\theta(L) := \{\theta(w) \mid w \in L\}$. In order to prove that the notion of θ-pseudoknot-bordered word is a proper generalization of the notion of a bordered word, in Section 3 we consider both morphic and antimorphic involutions. However, the problem of investigating whether catenations of pseudoknot-unbordered words have the same property is motivated mainly by DNA/RNA computing. Thus, in Section 4 we focus only on the mathematical formalization of the Watson-Crick complementarity, i.e., we consider only the case of antimorphic involutions.

A few words about morphic and antimorphic involutions are in order. Note that, if the alphabet Σ has m letters, and if we regard involutions that are isomorphic to each other as identical, the number of different involutions on Σ^* is $\lfloor m/2 \rfloor + 1$. For example, on a binary alphabet $\Sigma = \{a, b\}$, there exist only two essentially different involutions: θ defined as $\theta(a) := b$ and $\theta(b) := a$, and the identity function. Each of these $\lfloor m/2 \rfloor + 1$ involutions can be extended to a morphic or antimorphic involution. With applications to the Watson-

Crick complementarity in mind, herein we deal only with functions that are not the identity function. Thus, implicitly, we also exclude singleton alphabet sets. Note also that for any d-morphic involution θ that is not the identity, there exist two distinct characters $a, b \in \Sigma$ such that $\theta(a) = b$ and $\theta(b) = a$. We assume that in all the examples of this paper, for a given non-identity d-morphic involution θ, such $a, b \in \Sigma$ are chosen.

3 θ-pseudoknot-bordered words

In this section we propose the notion of θ-*pseudoknot-bordered words* for a morphic or antimorphic involution θ. If we consider the DNA alphabet $\{A, C, G, T\}$, wherein θ is the Watson-Crick complementarity function, then a word that is θ-pseudoknot-unbordered will not form pseudoknot-like secondary structures such as the ones in Figure 3. We show that the notion of θ-pseudoknot-bordered word is a proper generalization of the notion of θ-bordered word proposed in [11], and thus a proper generalization of the notion of bordered word. We also provide several properties of θ-pseudoknot-(un)bordered words.

Let θ be a d-morphic involution. A word $u \in \Sigma^*$ is said to be a *proper θ-border* of a word $w \in \Sigma^+$ if u is a proper prefix of w and $\theta(u)$ is a proper suffix of w, i.e., $w = u\alpha = \beta\theta(u)$ for some $\alpha, \beta \in \Sigma^+$. $\mathrm{L}_{\mathrm{d}}^{\theta}(w)$ denotes the set of all proper θ-borders of a nonempty word w. Note that $\lambda \in \mathrm{L}_{\mathrm{d}}^{\theta}(w)$ for all $w \in \Sigma^+$. A word $w \in \Sigma^+$ is said to be *θ-bordered* if it has a proper θ-border other than λ, i.e., $|\mathrm{L}_{\mathrm{d}}^{\theta}(w)| \geq 2$; otherwise, it is *$\theta$-unbordered*. Define now $\mathrm{D}_{\theta}(i) := \{w \in \Sigma^+ \mid |\mathrm{L}_{\mathrm{d}}^{\theta}(w)| = i\}$. Then $\mathrm{D}_{\theta}(1)$ is the set of all θ-unbordered words.

We call a word $u \in \Sigma^*$ a *θ-pseudoknot-border* (or θ-pk-border) of a word $w \in \Sigma^*$ if there exists a cyclic permutation v of u such that $w = u\alpha = \beta\theta(v)$ for some $\alpha, \beta \in \Sigma^*$. We also employ the expression "*xy is a θ-pk-border of w*" to mean "v is a θ-pk-border of w such that $v = xy$ and $w = xy\alpha = \beta\theta(yx)$ for some $\alpha, \beta \in \Sigma^*$". Let $\mathrm{L}_{\mathrm{cd}}^{\theta}(w)$ denote the set of all θ-pk-borders of a nonempty word w, and $\mathrm{K}_{\theta}(i) := \{w \in \Sigma^+ \mid |\mathrm{L}_{\mathrm{cd}}^{\theta}(w)| = i\}$. We call a nonempty word *θ-pseudoknot-bordered* (or θ-pk-bordered) if it has a nonempty θ-pk-border; otherwise, it is *θ-pseudoknot-unbordered*. Note that $\lambda \in \mathrm{L}_{\mathrm{cd}}^{\theta}(w)$ for all $w \in \Sigma^+$. Note also that no word in $\mathrm{K}_{\theta}(1)$ has θ-pk-borders other than λ, and hence $\mathrm{K}_{\theta}(1)$ is the set of all θ-pk-unbordered words.

Example 2 *Let θ be an antimorphic involution on Σ^* and $w = aababbb$. As mentioned in Section 2, $a, b \in \Sigma$ are chosen so as to satisfy $\theta(a) = b$ and $\theta(b) = a$. Then $\mathrm{L}_{\mathrm{cd}}^{\theta}(w) = \{\lambda, a, aa, aaba\}$. In particular, setting $x = aab$ and $y = a$ shows that $w = xybbb = aab\theta(yx)$ and hence $aaba \in \mathrm{L}_{\mathrm{cd}}^{\theta}(w)$. Note that $w \in \mathrm{K}_{\theta}(4)$.*

7

A word may have itself as its θ-pk-border, in both cases of θ being morphic and being antimorphic, as shown by the following examples.

Example 3 *Let θ be a morphic involution on Σ^* and $w = abbaabba$. Then w can be written as $w = xy = \theta(yx)$ by letting $x = abbaab$ and $y = ba$.*

Example 4 *Let θ be an antimorphic involution on Σ^* and $w = ababbbaa$. Then $w = xy = \theta(yx)$ by letting $x = abab$ and $y = bbaa$.*

Observe that the definitions of $L_d^\theta(w)$ and $L_{cd}^\theta(w)$ are different in that the former does not contain w while the latter can, if w is a θ-pk-border of itself. This scenario is different also from the classical case of bordered words, a particular case of θ-pk-bordered words where θ, as well as the permutation involved, are the identity. In the classical case, $L_d(w)$ denotes the set of proper borders of w, i.e., it does not contain w, since w is trivially always a border of itself. This definition was followed closely when defining $L_d^\theta(w)$, the set of proper θ-borders of a word w. However, in the case of θ-pseudoknot-bordered words we strayed from this model in defining $L_{cd}^\theta(w)$. This was because a word may, or may not, be a θ-pk-border of itself, and thus it is meaningful to observe for a word w, whether or not w belongs to $L_{cd}^\theta(w)$. This choice implies that, while all other notions proposed here are strict generalizations of the corresponding notions related to θ-bordered and bordered words, $L_{cd}^\theta(w)$ does not strictly generalize $L_d^\theta(w)$ and $L_d(w)$. Observe, however, that all the results obtained in this paper hold for the other definition choice for $L_{cd}^\theta(w)$ as well, either unchanged or augmented by a weak additional condition.

Since a word is a cyclic permutation of itself, if a word has a θ-border, then the θ-border also becomes a θ-pk-border of the word. Hence, the following lemma and its corollary hold.

Lemma 5 *Let θ be a d-morphic involution on Σ^* and $w \in \Sigma^+$. Then $L_d^\theta(w) \subseteq L_{cd}^\theta(w)$ holds.*

Corollary 6 *Let θ be a d-morphic involution on Σ^*. Then $K_\theta(1) \subseteq D_\theta(1)$.*

As shown in the following example, there exist a word w and a d-morphic involution θ for which $L_d^\theta(w)$ is strictly included in $L_{cd}^\theta(w)$.

Example 7 *Let θ be a d-morphic involution on Σ^* and $w = aaababbb$. For both cases of θ being morphic or antimorphic, $L_d^\theta(w) = \{\lambda, a, aa\}$ but $L_{cd}^\theta(w) = \{\lambda, a, aa, aaba\}$.*

In the preceding example, $L_{cd}^\theta(w)$ happens to be the same whether the involution defined as $\theta(a) = b$ and vice versa is extended to a morphism, or to an antimorphism of Σ^*. This is not always the case, as indicated in the following two examples.

Example 8 *Let θ be a d-morphic involution on Σ^* and $w = aabbabaababb$. When θ is morphic, $w = xyaababb = aabbab\theta(yx)$ for $x = aa$ and $y = bbab$, and hence $aabbab \in \mathrm{L}^\theta_{\mathrm{cd}}(w)$. On the other hand, $aabbab \notin \mathrm{L}^\theta_{\mathrm{cd}}(w)$ if θ is anti-morphic.*

Example 9 *Let θ be a d-morphic involution on Σ^* and $w' = aabbabbbabaa$. When θ is antimorphic, $w = xybbabaa = aabbab\theta(yx)$ for $x = aa$ and $y = bbab$, and hence $aabbab \in \mathrm{L}^\theta_{\mathrm{cd}}(w')$. On the other hand, $aabbab \notin \mathrm{L}^\theta_{\mathrm{cd}}(w')$ for θ being morphic.*

There exist alphabets Σ, and d-morphic involutions θ on Σ^*, for which the inclusion relation of Corollary 6 is proper. Indeed, let us consider a morphic involution θ, and a word $w \in \mathrm{D}_\theta(1)$ such that $w \notin \mathrm{K}_\theta(1)$. This implies that $w = xy\alpha = \beta\theta(y)\theta(x)$ for some $x, y, \alpha, \beta \in \Sigma^*$. If x were a proper prefix of w, then w would be θ-bordered, and hence $w = x = \theta(x)$. Hence, if there exists $c \in \Sigma$ such that $\theta(c) = c$, a word $w \in \mathrm{D}_\theta(1) \setminus \mathrm{K}_\theta(1)$ exists, and the inclusion relation is proper as seen by choosing $w = c$; otherwise $\mathrm{D}_\theta(1) = \mathrm{K}_\theta(1)$. For an antimorphic involution θ, we have the following example.

Example 10 *Let θ be an antimorphic involution on Σ^* and $w = aba$. Then $w \in \mathrm{D}_\theta(1)$, but $w \notin \mathrm{K}_\theta(1)$ because $w = xya = a\theta(yx)$ for $x = a$ and $y = b$.*

For a given d-morphic involution θ on Σ^*, a few remarks are in order regarding the set of all θ-pseudoknot-bordered words over Σ, i.e., the complement of $\mathrm{K}_\theta(1)$. For an antimorphic involution, which is of the most interest because of the biological motivation of this study, the cross-dependency existing in any θ-pk-bordered word indicates that the set of all θ-pk-bordered words is not context-free. This can indeed be proved by using the Pumping Lemma for context-free languages by choosing, e.g., an alphabet Σ, an antimorphic involution θ that maps a to b and vice versa, and the θ-pk-bordered word $a^n b^n a^n$, where n is the constant given by the Pumping Lemma. The fact that several (mild-)context-sensitive grammars or their stochastic variants were proposed to model pseudoknot structures [16,18,22] suggests that, for an antimorphic involution θ, the set of all θ-pk-bordered words over Σ is context-sensitive. This is indeed true, but we omit here the lengthy but straightforward construction of such a context-sensitive grammar, and the proof.

We conclude this section with some basic properties of θ-pk-borders, which will be used mainly in the proofs of the next section.

Lemma 11 *Let θ be a d-morphic involution on Σ^*. The following hold:*

(1) If a word $w \in \Sigma^+$ has a θ-pk-border of length n, then, for every $a \in \Sigma$, the number of occurrences of the letter a in the prefix of length n of w is equal to the number of occurrences of the letter $\theta(a)$ in the suffix of length n of w.

(2) *For all $a \in \Sigma$ such that $a \neq \theta(a)$, a^k is θ-pk-unbordered for all $k \geq 1$.*

(3) *For words $v, w \in \Sigma^+$ and $n \geq 1$, if $v \in \mathrm{L}_{\mathrm{cd}}^{\theta}(w^n)$ and $|w^{m-1}| < |v| \leq |w^m|$*
    for some $m \geq 1$, then $v \in \mathrm{L}_{\mathrm{cd}}^{\theta}(w^k)$ for all k with $m \leq k \leq n$.

4 Primitive and θ-pseudoknot-unbordered words

One of the processes that are essential and often unavoidable in biocomputing
algorithms is the concatenation of information-encoding DNA single strands.
Thus, a question that is often asked is: Given some DNA strands having a cer-
tain "good" encoding property, will the catenation of these strands preserve
this property? In this section we make steps towards answering this question
in the case of the property of a word being θ-pseudoknot-unbordered. That is,
for an antimorphic involution θ, we first address the following question: "Given
a θ-pk-unbordered word u, is every word in $\{u\}^+$ also θ-pk-unbordered?" This
question was answered positively for θ-unbordered words in [11]: A power of
a θ-unbordered word is always θ-unbordered. We show that, in contrast, the
question is answered negatively for θ-pk-unbordered words. Moreover, we pro-
vide a sufficient condition for a θ-pk-unbordered word to satisfy the condition
that all of its powers are θ-pk-unbordered (Corollary 21).

We begin by providing a necessary and sufficient condition for a word to be
θ-pk-unbordered, which follows directly from the definition of a θ-pk-bordered
word.

Lemma 12 *Let θ be an antimorphic involution on Σ^*. Then a word $u \in \Sigma^+$
is θ-pk-unbordered if and only if $\theta(\mathrm{Cp}(\mathrm{Pref}(u))) \cap \mathrm{Suff}(u) = \emptyset$.*

For a d-morphic involution θ on Σ^*, a word $w \in \Sigma^*$ is called θ-*palindrome* if
$w = \theta(w)$. Let P_{θ} denote the set of all θ-palindromes over Σ.

Lemma 13 *Let θ be an antimorphic involution on Σ^*, and x, y be θ-palindromes
such that $xy \neq \lambda$. If a word $u \in \Sigma^+$ has xy as both its prefix and suffix, then
u is θ-pk-bordered.*

PROOF. Let $u = xy\alpha = \beta xy$ for some $\alpha, \beta \in \Sigma^*$. The fact that $x, y \in \mathrm{P}_{\theta}$
implies that $u = \beta\theta(x)\theta(y) = \beta\theta(yx)$, Therefore, u is θ-pk-bordered. $\square$

Recall the following result from [11].

Lemma 14 *Let θ be an antimorphism on Σ^* and let $u \in \Sigma^+$. Then $u \in \mathrm{D}_{\theta}(1)$
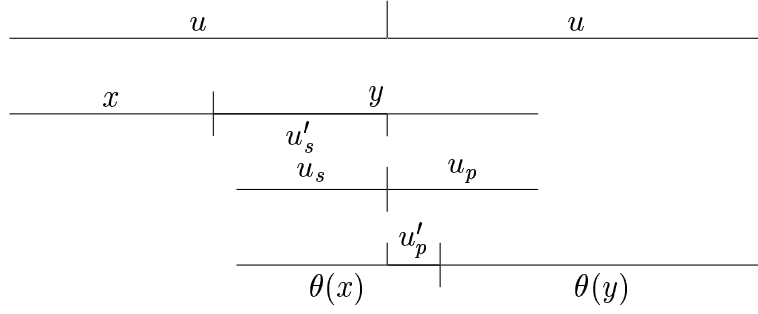if and only if $u^+ \subseteq \mathrm{D}_{\theta}(1)$.*

10

Fig. 4. A pictorial representation of Case 2 of the proof of Proposition 16.

In contrast, the following example shows that there exist θ-pk-unbordered words u such that at u^k is θ-pk-bordered for some $k \geq 2$.

Example 15 *Let θ be an antimorphic involution on Σ^* and $u = aabbbbaba$. Although u is θ-pk-unbordered, u^2 is θ-pk-bordered. In fact, $u^2 = xyabbbbaba = aabbbbab\theta(x)\theta(y)$ for $x = aabbb$ and $y = babaa$.*

In the following, we give a characterization of θ-pk-unbordered words u with the property that u^k is θ-pk-bordered for some $k \geq 2$, that takes into account the relative length of the θ-pk-borders of u^2.

Proposition 16 *Let θ be an antimorphic involution on Σ^*. Then for a θ-pk-unbordered word u, if there exists $k \geq 2$ such that u^k has a nonempty θ-pk-border w, then $|u| < |w| < \frac{4}{3}|u|$ holds.*

PROOF. Suppose for some $k \geq 2$, there were a $w \in \mathrm{L}^\theta_{\mathrm{cd}}(u^k)$ such that either $|w| \leq |u|$ or $\frac{4}{3}|u| \leq |w|$ hold. If $|w| \leq |u|$, then this w leads us to a contradiction immediately. Next we consider the case $\frac{4}{3}|u| \leq |w| < 2|u|$. Then $w \in \mathrm{L}^\theta_{\mathrm{cd}}(u^k)$ implies $w \in \mathrm{L}^\theta_{\mathrm{cd}}(u^2)$. In other words, there exists a decomposition $w = xy$ such that $uu = xy\alpha = \beta\theta(x)\theta(y)$ for some $\alpha, \beta \in \Sigma^+$. Since $|w| \geq \frac{4}{3}|u|$, we have $xy = uu_p$ and $\theta(x)\theta(y) = u_s u$, where $u_p \in \mathrm{Pref}(u)$, and $u_s \in \mathrm{Suff}(u)$. Now we have the following two cases:

(1) $|x| \geq |u|$ or $|y| \geq |u|$ holds,
(2) $|x| < |u|$ and $|y| < |u|$ hold.

In the first case, for reasons of symmetry, we only have to consider the case $|x| \geq |u|$. Since $\theta(x)\theta(y) = u_s u$, we can write $\theta(x) = u_s u'_p$ for some $u'_p \in \mathrm{Pref}(u)$. Let $u = u'_p u'_s$, and we can easily check that $u'_s \in \mathrm{Suff}(u_s)$. Therefore, $u'_s u'_p \in \mathrm{Suff}(\theta(x))$, which equals $\theta(u'_p)\theta(u'_s) \in \mathrm{Pref}(x)$. This means that $\theta(u'_p)\theta(u'_s) = u$ because u and $\theta(u'_p)\theta(u'_s)$ are prefixes of x and they have equal lengths. Since $u = u'_p u'_s$, we conclude that both u'_p and u'_s are θ-palindromes. The application of Lemmata 12 and 13 leads now to a contradiction.

Next we consider the second case (see Figure 4). This figure shows $xy = uu_p$ and $\theta(x)\theta(y) = u_s u$. Since both x and y are shorter than u, these equations imply that $u = xu'_s = u'_p\theta(y)$, where $u'_p \in \mathrm{Pref}(u)$ and $u'_s \in \mathrm{Suff}(u)$. Comparing this equation with $xy = uu_p$ we derive $y = u'_s u_p$, and hence $u = u'_p\theta(u_p)\theta(u'_s)$. This result, together with $u = xu'_s$, implies that u'_s is a θ-palindrome and $x = u'_p\theta(u_p)$. Substituting this x and $u = u'_p\theta(y)$ into $\theta(x)\theta(y) = u_s u$ gives $u_p\theta(u'_p)\theta(y) = u_s u'_p\theta(y)$, which means that $u_p = u_s$ and u'_p is a θ-palindrome.

Let us bring now into the picture the original condition $\frac{4}{3}|u| \le |w| < 2|u|$. Since $|w| = |u| + |u_p|$, $\frac{4}{3}|u| \le |w|$ means $\frac{1}{3}|u| \le |u_p|$. Hence, $|xy| = |uu_p| \le 4|u_p|$. This implies that either $|x| \le 2|u_p|$ or $|y| \le 2|u_p|$ holds. We assume the former case holds. Then $\theta(x) = u_s u'_p$ implies $|u'_p| \le |u_s|$ because $|\theta(x)| = |x| \le 2|u_p| = 2|u_s|$. Let $u_s = u_1 u_2$ such that $|u_1| = |u'_p|$. Note that $u_s \in \mathrm{Pref}(x)$ because $u_p, x \in \mathrm{Pref}(u)$, $|u_s| < |x|$, and $u_p = u_s$. Comparing $u_s = u_1 u_2$ with $x = \theta(u_1 u_2 u'_p)$ based on $u_s \in \mathrm{Pref}(x)$ results in $u_2 = \theta(u_2)$ and $u_1 = \theta(u'_p)$, which in turn implies $u_1 = \theta(u_1)$ because $u'_p = \theta(u'_p)$. Now Lemmata 12 and 13 lead to a contradiction because u contains the concatenation of two θ-palindromes u_1 and u_2 as its prefix u_p and suffix u_s.

What remains to consider is the case where $|w| \ge 2|u|$. Let $w = xy = u^n u_p$ and $\theta(x)\theta(y) = u_s u^n$ for some $n \ge 2$, $u_p \in \mathrm{Pref}(u)$, and $u_s \in \mathrm{Suff}(u)$. Then either $|x| \ge |u|$ or $|y| \ge |u|$ holds. Let us assume that $|x| \ge |u|$ holds. Then $\theta(x) = u_s u^m u'_p$ and $\theta(y) = u'_s u^{n-m-1}$ for some $m < n$, where $u'_p u'_s = u$. If $u'_s \in \mathrm{Suff}(u_s)$, then $u'_s u^m u'_p \in \mathrm{Suff}(\theta(x))$, which implies $\theta(u'_p)\theta(u)^m\theta(u'_s) \in \mathrm{Pref}(x)$. Since x is not shorter than u, $u = \theta(u'_p)\theta(u'_s)$. Thus, u can be factorized into two θ-palindromes, a contradiction. If $u_s \in \mathrm{Suff}(u'_s)$, then m must be at least 1; otherwise, $|\theta(x)| = |u_s u'_p| < |u'_s u'_p| = |u|$. Therefore, $u'_s u'_p \in \mathrm{Suff}(\theta(x))$, i.e., $\theta(u'_p)\theta(u'_s) \in \mathrm{Pref}(x)$. Now we have $u = \theta(u'_p)\theta(u'_s)$, and this leads to the same contradiction as above. □

Note that, in Example 15, the θ-pk-border xy of u^2 satisfies $|u| < |xy| < \frac{4}{3}|u|$.

Corollary 17 *Let θ be an antimorphic involution on Σ^*. For a word $u \in \mathrm{K}_\theta(1)$, $u^+ \not\subseteq \mathrm{K}_\theta(1)$ if and only if $u^2 \notin \mathrm{K}_\theta(1)$.*

The next lemma is a consequence of the proof of Proposition 16, and will be a useful tool in obtaining several additional properties of θ-pk-unbordered words whose square is θ-pk-bordered.

Lemma 18 *Let θ be an antimorphic involution on Σ^*, and u be a θ-pk-unbordered word. If $xy \in \mathrm{L}^\theta_{\mathrm{cd}}(u^2)$ such that $xy = uu_p$ for some $u_p \in \mathrm{Pref}(u)$, then $2|u_p| < |x| < |u|$ and $2|u_p| < |y| < |u|$ hold.*

In what follows, we give a characterization of θ-pk-unbordered words whose square is θ-pk-bordered.

Lemma 19 *Let θ be an antimorphic involution on Σ^*, and u be a θ-pk-unbordered word. Then u^2 is θ-pk-bordered if and only if $u = u_p \alpha \theta(u_p) \beta u_p$ for some $u_p, \alpha, \beta \in \Sigma^+$ such that $u_p \alpha, \beta u_p$ are θ-palindromes.*

PROOF. (Only if) Let $u^2 = xy\gamma_1 = \gamma_2 \theta(x)\theta(y)$ for xy such that $|u| < |xy| < \frac{4}{3}|u|$. Then $xy = uu_p$ and $\theta(x)\theta(y) = u_s u$ for some $u_p \in \operatorname{Pref}(u)$ and $u_s \in \operatorname{Suff}(u)$, which satisfy $|u_p| = |u_s| < \frac{1}{3}|u|$. In addition, Lemma 18 enables us to assume that $2|u_p| < |x| < |u|$ and $2|u_p| < |y| < |u|$. Now we have $\theta(x) = u_s u_p \alpha$ and $y = \beta u_s u_p$ for some $\alpha, \beta \in \Sigma^+$. Then $x = \theta(u_p \alpha)\theta(u_s)$ and $\theta(y) = \theta(u_p)\theta(\beta u_s)$. Substituting these into $xy = uu_p$ and $\theta(x)\theta(y) = u_s u$ gives that $u = \theta(u_p \alpha)\theta(u_s)\beta u_s = u_p \alpha \theta(u_p)\theta(\beta u_s)$. This means that both $u_p \alpha$ and βu_s are θ-palindromes and $\theta(u_p) = \theta(u_s)$. Thus, $u_p = u_s$ and then $u = u_p \alpha \theta(u_p)\beta u_p$.
(If) Let $x = u_p \alpha \theta(u_p)$ and $y = \beta u_p u_p$. Then $\theta(x)\theta(y) = u_p \theta(u_p \alpha)\theta(u_p)\theta(\beta u_p)$. We can rewrite the right-hand side as $u_p u_p \alpha \theta(u_p)\beta u_p$ because $u_p \alpha, \beta u_p \in \operatorname{P}_\theta$. This means $\theta(x)\theta(y) = u_p u \in \operatorname{Suff}(uu)$. $\square$

Lemma 20 *Let θ be an antimorphic involution on Σ^*, and u be a θ-pk-unbordered word. If u^2 is θ-pk-bordered, then u is primitive.*

PROOF. Since u^2 has a θ-pk-border uu_p for some $u_p \in \operatorname{Pref}(u)$, Lemma 19 implies that u can be written as $u_p \alpha \theta(u_p)\beta u_p$ for some $\alpha, \beta \in \Sigma^*$ such that $u_p \alpha, \beta u_p \in \operatorname{P}_\theta$. Suppose u were not primitive, i.e., $u = w^r$ for some $w \in \Sigma^+$ and $r \geq 2$. To begin with, we consider the case $|u_p| \leq |w|$. This case has the two subcases depending on whether there exists an integer n such that $|u_p \alpha| < |w^n| < |u_p \alpha \theta(u_p)|$, where $1 \leq n \leq r-1$, or not. If such n exists, the infix $\theta(u_p)$ overlaps with the n-th occurrence of w and with the $(n+1)$-th occurrence of w, counted from the left. Let $\theta(u_p) = \theta(u_2)\theta(u_1)$ such that $\theta(u_2) \in \operatorname{Suff}(w)$ and $\theta(u_1) \in \operatorname{Pref}(w)$. Then we have $u_p = u_1 u_2$. Both u_p and w are prefixes of u and $|u_p| \leq |w|$ so that $u_1 \in \operatorname{Pref}(w)$, and hence $u_1 = \theta(u_1)$. In the same way, $u_2 = \theta(u_2)$. Then Lemma 13 leads to a contradiction with the fact that $u \in \operatorname{K}_\theta(1)$.

Next we consider the other subcase. We can rewrite this subcase as follows: There exists an integer n such that $|w^n| \leq |u_p \alpha|$ and $|u_p \alpha \theta(u_p)| \leq |w^{n+1}|$, where $0 \leq n \leq r-1$. Depending on the value of n, there exist 3 possibilities to be taken into account: (a) $n = 0$, (b) $n = r-1$, and (c) otherwise. In case (a), we can write $w = u_p \alpha \theta(u_p)\beta_p$, $\beta_i = w^{r-2}$, and $w = \beta_s u_p$ for some $\beta_p, \beta_i, \beta_s \in \Sigma^*$ such that $\beta_p \beta_i \beta_s = \beta$. Then $\beta u_p = \beta_p w^{r-1}$. Replacing one occurrence of w in the right-hand of this equation with $u_p \alpha \theta(u_p)\beta_p$ gives $\beta u_p = \beta_p u_p \alpha \theta(u_p)\beta_p w^{r-2} = \beta_p w^{r-2} u_p \alpha \theta(u_p)\beta_p$. This means that both β_p and $u_p \alpha \theta(u_p)$ are θ-palindromes because $\beta u_p \in \operatorname{P}_\theta$. Therefore, w is the concatenation of two θ-palindromes. Since u has w as its prefix and suffix, Lemma 13 leads to a contradiction. Case (b) is similar. In case (c), let $w = u_p \alpha_p$, $w^{n-1} = \alpha_i$,

13

$w = \alpha_s \theta(u_p) \beta_p$, $w^{r-n-2} = \beta_i$, and $w = \beta_s u_p$ for $\alpha_p, \alpha_i, \alpha_s, \beta_p, \beta_i, \beta_s \in \Sigma^*$ such that $\alpha_p \alpha_i \alpha_s = \alpha$ and $\beta_p \beta_i \beta_s = \beta$. Then one has $u_p \alpha = w^n \alpha_s$. Substituting $w = \alpha_s \theta(u_p) \beta_p$ into one occurrence of w in the right-hand side of this equation gives $u_p \alpha = \alpha_s \theta(u_p) \beta_p w^{n-1} \alpha_s = w^{n-1} \alpha_s \theta(u_p) \beta_p \alpha_s$. Since $u_p \alpha \in \mathrm{P}_\theta$, both α_s and $\theta(u_p) \beta_p$ are θ-palindromes. Then Lemma 13 leads to a contradiction as above.

Finally, we consider the case where w is shorter than u_p. Then there exist two integers n and h satisfying $|w^n| \leq |u_p \alpha| < |w^{n+1}|$ and $|w^{n+1+h}| < |u_p \alpha \theta(u_p)| \leq |w^{n+1+h+1}|$, where $h \geq 0$. Hence, we can write $\alpha_s \theta(u_p) \beta_p = w^{h+2}$ for some $\alpha_s \in \mathrm{Suff}(\alpha)$ and $\beta_p \in \mathrm{Pref}(\beta)$ such that $|\alpha_s|, |\beta_p| < |w|$. Thus, $w = \gamma \beta_p$ for some $\gamma \in \mathrm{Suff}(\theta(u_p))$, and hence $\alpha_s \theta(u_p) \beta_p = (\gamma \beta_p)^h \gamma \beta_p \gamma \beta_p$. This equation means $\beta_p \gamma \in \mathrm{Suff}(\theta(u_p))$ because $|w| < |\theta(u_p)|$. Therefore, $\theta(\beta_p \gamma) = \theta(\gamma) \theta(\beta_p) \in \mathrm{Pref}(u_p) \subseteq \mathrm{Pref}(u)$. In addition, $w = \gamma \beta_p \in \mathrm{Suff}(u)$. Thus, $u = \theta(\gamma) \theta(\beta_p) v \gamma \beta_p$ for some $v \in \Sigma^*$, which conflicts with $u \in \mathrm{K}_\theta(1)$. $\square$

Corollary 21 *Let θ be an antimorphic involution on Σ^*. If u is a non-primitive θ-pk-unbordered word, then u^2 is θ-pk-unbordered. This further implies that any power of u is θ-pk-unbordered.*

Example 22 *Let θ be an antimorphic involution on Σ^* and $u = abaaabaa$, which is clearly not primitive. It is easy to see that neither u nor u^2 is θ-pk-bordered. Hence, u^k is θ-pk-unbordered for any $k \geq 1$.*

For a θ-pk-unbordered word u whose square is θ-pk-bordered, we now investigate the primitivity of θ-pk-borders of u^2.

Theorem 23 *Let θ be an antimorphic involution on Σ^*, and u be a θ-pk-unbordered word whose square is θ-pk-bordered. Then any θ-pk-border of u^2 is primitive.*

PROOF. Let uu_p be a θ-pk-border of u^2. Lemma 19 says $u = u_p \alpha \theta(u_p) \beta u_p$ for $u_p, \alpha, \beta \in \Sigma^+$ such that $u_p \alpha, \beta u_p \in \mathrm{P}_\theta$. Suppose uu_p is not primitive, that is, $uu_p = w^r$ for $w \in \Sigma^+$ and $r \geq 2$.

To begin with, we assume $|w| = |u_p|$. Then we have $u = w^{r-1}$. The condition $|u_p| < \frac{1}{3}|u|$, which is necessary for $u^+ \not\subseteq K_\theta(1)$, implies $r > 4$. Thus, u would be not primitive, a contradiction.

Next we assume $|w| > |u_p|$. As in the proof of Lemma 20, we have to consider the following two cases:

(1) There exists an integer n such that $|u_p \alpha| < |w^n| < |u_p \alpha \theta(u_p)|$, where $1 \leq n \leq r-1$,

14

(2) There exists an integer n such that $|w^n| \le |u_p\alpha|$ and $|u_p\alpha\theta(u_p)| \le |w^{n+1}|$, where $0 \le n \le r-1$.

In case 1, let $\theta(u_p) = \theta(u_2)\theta(u_1)$ such that $\theta(u_2) \in \mathrm{Suff}(w)$ and $\theta(u_1) \in \mathrm{Pref}(w)$. Note that w has u_p as both its prefix and suffix because $w^r\ (= uu_p)$ has u_p both as its prefix and as its suffix, and $|u_p| < |w|$. Hence, we have $\theta(u_2) \in \mathrm{Suff}(u_p)$ and $\theta(u_1) \in \mathrm{Pref}(u_p)$. Since $u_p = u_1u_2$, we also have $u_1 \in \mathrm{Pref}(u_p)$ and $u_2 \in \mathrm{Suff}(u_p)$, which means that both u_1 and u_2 are θ-palindromes. Then Lemma 13 leads to a contradiction to the fact that $u \in \mathrm{K}_\theta(1)$ because u has the product of two θ-palindromes u_1u_2 both as its prefix and as its suffix.

In case 2, there are three possibilities depending on the value of n: (a) $n = 0$, (b) $n = r-1$, and (c) otherwise.

In subcase 2(a), we can write $w = u_p\alpha\theta(u_p)\beta_p$, $w^{r-3} = \beta_i$, $\beta_su_p = ww_p$, and $w = w_pu_p$ for some $w_p \in \mathrm{Pref}(w)$ and $\beta_p, \beta_i, \beta_s \in \Sigma^*$ such that $\beta = \beta_p\beta_i\beta_s$. Now we can rewrite $\theta(u_p)\beta u_pu_p = \theta(u_p)\beta_p\beta_i\beta_su_pu_p = \theta(u_p)\beta_p\beta_iw^2 = \theta(u_p)\beta_p\beta_iwu_p\alpha\theta(u_p)\beta_p$. Then we can say that $\theta(u_p)\beta_p$ is a θ-palindrome because $\theta(u_p)\beta u_pu_p \in \mathrm{P}_\theta$. Therefore, $w = u_p\alpha\theta(u_p)\beta_p = u_p\alpha\theta(\beta_p)u_p$. Compared to $w = w_pu_p$, we have $w_p = u_p\alpha\theta(\beta_p)$. Then,

$$w = u_p\alpha\theta(u_p)\beta_p \in \mathrm{Pref}(u) \Rightarrow \beta_pu_p\alpha\theta(u_p) \in \mathrm{Cp}(\mathrm{Pref}(u)),$$
$$\Leftrightarrow \beta_p\theta(\alpha)\theta(u_p)\theta(u_p) \in \mathrm{Cp}(\mathrm{Pref}(u)),$$
$$\Leftrightarrow u_pu_p\alpha\theta(\beta_p) \in \theta(\mathrm{Cp}(\mathrm{Pref}(u))),$$
$$\Leftrightarrow u_pw_p \in \theta(\mathrm{Cp}(\mathrm{Pref}(u))).$$

The third implication is due to the fact that $u_p\alpha \in \mathrm{P}_\theta$. Since $w^2 = w_pu_pw_pu_p$ is the suffix of uu_p, $u_pw_p \in \mathrm{Suff}(u)$, and hence $\theta(\mathrm{Cp}(\mathrm{Pref}(u))) \cap \mathrm{Suff}(u) \ne \emptyset$, which is a contradiction.

In subcase 2(b), we can write $w = u_p\alpha_p$, $w^{r-2} = \alpha_i$, $w_p = \alpha_s\theta(u_p)\beta u_p$, and $w_pu_p = w$ for some $w_p \in \mathrm{Pref}(w)$ and $\alpha_p, \alpha_i, \alpha_s \in \Sigma^*$ such that $\alpha = \alpha_p\alpha_i\alpha_s$. As in the cases above, since $u_p\alpha = \theta(u_p\alpha)$, α_s is also a θ-palindrome. Starting from $w \in \mathrm{Pref}(u)$, now we can show $u_pw_p \in \theta(\mathrm{Cp}(\mathrm{Pref}(u))) \cap \mathrm{Suff}(u)$.

In subcase 2(c), let $w = u_p\alpha_1$, $w^{n-1} = \alpha_2$, $w = \alpha_3\theta(u_p)\beta_1$, $w^{r-n-3} = \beta_2$, $\beta_3u_p = ww_p$, and $w = w_pu_p$ for some $w_p \in \mathrm{Pref}(w)$ and $\alpha_p, \alpha_i, \alpha_s, \beta_p, \beta_i, \beta_s \in \Sigma^*$ such that $\alpha = \alpha_p\alpha_i\alpha_s$ and $\beta = \beta_p\beta_i\beta_s$. Using these notations, we have $u_p\alpha = w^n\alpha_s = w^{n-1}\alpha_s\theta(u_p)\beta_p\alpha_s = \alpha_s\theta(u_p)\beta_pw^{n-1}\alpha_s$. Since $u_p\alpha \in \mathrm{P}_\theta$, we can observe that both α_s and $\theta(u_p)\beta_p$ are also θ-palindromes. Thus, $w = \alpha_s\theta(u_p)\beta_p = \alpha_s\theta(\beta_p)u_p$. Compared to $w = w_pu_p$, we can say $w_p = \alpha_s\theta(\beta_p)$. Now we can obtain a contradiction to $u \in \mathrm{K}_\theta(1)$ as above. This completes the discussion of the case $|w| > |u_p|$ with its two possibilities (1) and (2).

15

Finally, we consider the case where $|w| < |u_p|$. This means that there exist two integers n and h satisfying $|w^n| \leq |u_p\alpha| < |w^{n+1}|$ and $|w^{n+1+h}| < |u_p\alpha\theta(u_p)| \leq |w^{n+1+h+1}|$, where $h \geq 0$. Hence, we have $\alpha_s\theta(u_p)\beta_p = w^{h+2}$ for some $\alpha_s \in \mathrm{Suff}(\alpha)$ and $\beta_p \in \mathrm{Pref}(\beta)$ such that $|\alpha_s|, |\beta_p| < |w|$. Therefore, $w = \gamma\beta_p$ for some $\gamma \in \mathrm{Suff}(\theta(u_p))$, and hence $\alpha_s\theta(u_p)\beta_p = (\gamma\beta_p)^h\gamma\beta_p\gamma\beta_p$. This implies $\beta_p\gamma \in \mathrm{Suff}(\theta(u_p))$ because $|w| < |\theta(u_p)|$. This means $\theta(\gamma)\theta(\beta_1) \in \mathrm{Pref}(u_p)$. Moreover $w = \gamma\beta_p \in \mathrm{Suff}(u_p)$ because the rightmost occurrence of u_p in $uu_p = w^r$ has w as its suffix. Thus, $u = \theta(\gamma)\theta(\beta_1)v\gamma\beta_1$ for some $v \in \Sigma^*$ because u has u_p both as its prefix and as its suffix. This conclusion contradicts $u \in \mathrm{K}_\theta(1)$. $\square$

We conclude this section by showing that, for a word $u \in \mathrm{K}_\theta(1)$, if $xy, x'y' \in \mathrm{L}_{\mathrm{cd}}^\theta(u^2)$ with $|xy| = |x'y'|$, then $x = x'$ and $y = y'$.

Lemma 24 *Let θ be an antimorphic involution on Σ^*, and $w \in \Sigma^+$. A θ-pk-border v of w is not primitive if and only if there exist words x, y, x', y' satisfying $v = xy = x'y'$, $yx = y'x'$, and $x \neq x'$.*

PROOF. (Only if) Under the assumption, $v = xy = z^n$ for some primitive word $z \in \Sigma^+$ and $n \geq 2$. Then $x = z^i z_p$ and $y = z_s z^{n-i-1}$ for some $i \geq 0$, $z_p, z_s \in \Sigma^*$ such that $z = z_p z_s$. Let $x' = z^j z_p$ and $y' = z_s z^{n-j-1}$ for some $j \neq i$. Since $n \geq 2$, such j exists. Clearly $xy = x'y'$ and we can easily check $y'x' = yx$. **(If)** We can represent w as both $w = xy\alpha = \beta\theta(yx)$ and $w = x'y'\alpha = \beta\theta(y'x')$. Without loss of generality, we can assume $|x'| < |x|$, and then this implies that $\theta(x) = \theta(x')q$ and $\theta(y') = q\theta(y)$ for some $q \in \Sigma^+$. Therefore, $x = \theta(q)x'$ and $y' = y\theta(q)$. Substituting these into $xy = x'y'$, we obtain $xy = \theta(q)x'y = x'y\theta(q)$. Then Theorem 1 implies that v is not primitive. $\square$

The next proposition now follows from Theorem 23 and Lemma 24.

Proposition 25 *Let θ be an antimorphic involution on Σ^* and u be a θ-pk-unbordered word. If w is a nonempty θ-pk-border of u^2, then the factorization of w into x and y such that $u^2 = xy\alpha = \beta\theta(yx)$ for some $\alpha, \beta \in \Sigma^*$ is unique.*

5   Discussion

In this paper, we proposed the notion of a θ-pseudoknot-unbordered word, where θ is a morphic or antimorphic involution. This concept models DNA (or RNA) single strands that do not form some pseudoknot-like secondary

structures. This formulation is general enough to handle intermolecular structures similar to pseudoknots. In addition, this notion is a proper generalization of the notion of θ-unbordered word, and thus of the classical notion of unbordered word. After obtaining some basic properties of θ-bordered and θ-unbordered words, we investigated the question of whether or not all powers of a θ-unbordered words remain θ-unbordered. The question was answered in the negative by providing counterexamples. We also showed that, for a θ-unbordered word u, the fact that u is not primitive is a sufficient condition for u^k to be θ-pseudoknot-unbordered for all $k \geq 1$. This is the first step towards obtaining a condition that a language L of θ-pseudoknot-unbordered words would have to satisfy in order for L^+ to have the same property. Another direction of research is to consider more realistic pseudoknot structures, i.e., to remove the restriction $v_1 = v_2 = v_4 = v_5 = \lambda$ in the general definition of the pseudoknot as a word of the form $v_1 x v_2 y v_3 \theta(x) v_4 \theta(y) v_5$. In particular, the conditions $v_2 = \lambda$ and $v_4 = \lambda$ should be weakened, because pseudoknots occurring in real RNAs rarely satisfy these conditions due to steric effects.

References

[1] L. Adleman, *Molecular computation of solutions to combinatorial problems*, Science 266 (1994) 1021-1024.

[2] M. Andronescu, D. Dees, L. Slaybaugh, Y. Zhao, A. Condon, B. Cohen, S. Skiena, *Algorithms for testing that sets of DNA words concatenate without secondary structure*, In: M. Hagiya, A. Ohuchi (Eds.), Proc. DNA Based Computers 8, LNCS 2568, Springer (2003) 182-195.

[3] A. E. Condon, *Problems on RNA secondary structure prediction and design*, Proc. ICALP'2003, LNCS 2719 (2003) 22-32.

[4] M. Daley, L. Kari, *DNA computing: Models and Implementations*, Comments Theor. Biol. 7(3) (2002) 177-198.

[5] A. Ehrenfeucht, D. Silberger, *Periodicity and unbordered segments of words*, Discrete Math. 26(2) (1979) 101-109.

[6] S. G.-Jones, S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, A. Bateman, *Rfam: annotating non-coding RNAs in complete genomes*, Nucleic Acid Res. 33 (2005) D121-D124.

[7] N. Jonoska, K. Mahalingam, J. Chen, *Involution codes: with application to DNA coded languages*, Natural Computing 4(2) (2005) 141-162.

[8] L. Kari, R. Kitto, G. Thierrin, *Codes, involutions and DNA encodings*, in: W. Brauer, H. Ehrig, J. Karhumäki, A. Salomaa (Eds.), Formal and Natural Computing, LNCS 2300, Springer-Verlag (2002) 376-393.

[9] L. Kari, S. Konstantinidis, E. Losseva, G. Wozniak, *Sticky-free and overhang-free DNA languages*, Acta Inform. 40 (2003) 119-157.

[10] L. Kari, S. Konstantinidis, E. Losseva, P. Sosík, G. Thierrin, *Hairpin structures in DNA words*, in: A. Carbone, N. A. Pierce (Eds.), Proc. DNA11, LNCS 3892, Springer-Verlag (2006) 158-170.

[11] L. Kari, K. Mahalingam, *Involutively bordered words*, Int. J. Found. Comput. S. (2007) 1089-1106.

[12] L. Kari, K. Mahalingam, *Watson-Crick conjugate and commutative words*, in: M. Garzon, H. Yan (Eds.), Proc. DNA 13, LNCS 4848, Springer-Verlag (2008) 273-283.

[13] L. Kari, K. Mahalingam, G. Thierrin, *The syntactic monoid of hairpin-free languages*, Acta Inform. 44 (2007) 153-166.

[14] S. Kobayashi, *Testing structure freeness of regular sets of biomolecular sequences (extended abstract)*. in: C. Ferretti, G. Mauri, C. Zandron (Eds.), Proc. DNA 10, LNCS 3384, Springer-Verlag (2005) 192-201.

[15] A. Lentin, M. P. Schützenberger, *A combinatorial problem in the theory of free monoids*, Proc. Combinatorial Mathematics and its Applications (1967) 128-144,

[16] H. Matsui, K. Sato, Y. Sakakibara, *Pair stochastic tree adjoining grammars for aligning and predicting pseudoknot RNA structures*, Proc. 2004 IEEE Computational Systems Bioinformatics Conference (2004) 1-11.

[17] J. S. McCaskill, *The equilibrium partition function and base pair binding probability for RNA secondary structure*, Biopolymers 29 (1990) 1105-1119.

[18] E. Rivas, S. R. Eddy, *The language of RNA: a formal grammar that includes pseudoknots*, Bioinformatics 16(4) (2000) 334-340.

[19] G. Rozenberg, A. Salomaa (Eds.), *Handbook of Formal Languages*, Springer-Verlag, Berlin, 1997.

[20] H. J. Shyr, *Free Monoids and Languages*, Hon Min Book Company, 2001.

[21] H. J. Shyr, G. Thierrin, S.S. Yu, *Monogenic e-closed languages and dipolar words*, Discrete Math. 126 (1994) 339-348.

[22] Y. Uemura, A. Hasegawa, S. Kobayashi, T. Yokomori, *Tree adjoining grammars for RNA structure prediction*, Theor. Comput. Sci. 210 (1999) 277-303.

[23] S. S. Yu, *Languages and Codes*, Lecture Notes, Department of Computer Science, National Chung-Hsing University, Taichung, Taiwan 402, 2005.