Available online at www.sciencedirect.com

ScienceDirect

C. R. Acad. Sci. Paris, Ser. I 346 (2008) 813–818

COMPTES RENDUS MATHEMATIQUE

http://france.elsevier.com/direct/CRASS1/

Combinatorics

On the triplex substitution – combinatorial properties ☆

Bo Tan [a], Zhi-Xiong Wen [a], Yiping Zhang [b]

[a] *Department of Mathematics, Huazhong University of Science and Technology, Wuhan 430074, PR China*
[b] *Department of Mathematics, Wuhan University, Wuhan 430072, PR China*

Received 18 February 2008; accepted after revision 18 June 2008

Available online 24 July 2008

Presented by Jean-Pierre Kahane

Abstract

If a substitution τ over a three-letter alphabet has a positively linear complexity, that is, $P_\tau(n) = C_1 n + C_2$ $(n \geqslant 1)$ with $C_1, C_2 \geqslant 0$, there are only 4 possibilities: $P_\tau(n) = 3$, $n+2$, $2n+1$ or $3n$. The first three cases have been studied by many authors, but the case $3n$ remained unclear. This leads us to consider the triplex substitution $\sigma : a \mapsto ab$, $b \mapsto acb$, $c \mapsto acc$. Studying the factor structure of its fixed point, which is quite different from the other cases, we show that it is of complexity $3n$. We remark that the triplex substitution is also a typical example of invertible substitution over a three-letter alphabet. ***To cite this article: B. Tan et al., C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Résumé

Sur la substitution triplexe – propriétés combinatoires. Si une substitution τ sur un alphabet de trois lettres a une complexité positivement linéaire, c'est-à-dire $P_\tau(n) = C_1 n + C_2$ $(n \geqslant 1)$ où $C_1, C_2 \geqslant 0$, alors il n'y a que quatre possibilités : $P_\tau(n) = 3$, $n+2$, $2n+1$ ou $3n$. Les trois premiers cas ont été étudiés par différents auteurs, mais le cas $3n$ reste non entièrement élucidé. Nous considérons donc la substitution triplexe $\sigma : a \mapsto ab$, $b \mapsto acb$, $c \mapsto acc$. Analysant la structure des facteurs de son point fixe nous montrons que sa complexité est $3n$. La substitution triplexe est un exemple typique de substitution inversible sur un alphabet de trois lettres. ***Pour citer cet article : B. Tan et al., C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Version française abrégée

Nous considérons la substitution $\sigma = (ab, acb, acc)$ (c'est-à-dire, $a \mapsto ab$, $b \mapsto acb$, $c \mapsto acc$), appelée substitution triplexe. Elle a un point fixe unique $\xi = \xi_1\xi_2\xi_3\cdots = abacbabaccacb\cdots$.

Le but est de démontrer que ξ a pour complexité $3n$. Pour cela on analyse ses facteurs spéciaux :

Lemme 1. *Pour $n \geqslant 1$, les facteurs suivants sont spéciaux : $b\sigma^n(b)b^{-1}$, $c\sigma^n(b)b^{-1}$, $b\sigma^n(ab)b^{-1}$, $c\sigma^n(ab)b^{-1}$, $\sigma^{n-1}(accab)b^{-1}$, $\sigma^n(bab)b^{-1}$.*

☆ Research supported by NSFC No. 10501035, 10631040, 10571140 and 10671150.
E-mail addresses: bo_tan@163.com (B. Tan), wen.gong@263.net (Z.-X. Wen), ypzhang@whu.edu.cn (Y. Zhang).

1631-073X/$ – see front matter © 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crma.2008.06.013

On obtient ainsi un arbre de facteurs spéciaux (voir Fig. 1), qui démontre le :

Lemme 2. *Pour $k \geqslant 2$, il existe au moins 3 facteurs spéciaux de longueur k.*

Pour obtenir une égalité, on contrôle le nombre de facteurs spéciaux de longueur $|A_n|$:

Proposition 1. *Le nombre de facteurs spéciaux de longueur $|A_n|$ est au plus $3|A_n|$.*

Par une méthode d'interpolation, la proposition et les lemmes précédents nous permettent de démontrer

Théorème 1. *La complexité de ξ est $P(n) = 3n$ $(n \geqslant 1)$.*

En particulier, l'arbre de la Fig. 1 est complet : il donne tous les facteurs spéciaux de ξ. Par conséquent, le language engendré par ξ ou σ est complètement déterminé.

1. Introduction

The study of substitutions over a finite alphabet plays an important rôle in finite automata, symbolic dynamics, formal languages, number theory, and fractal geometry, and it has various applications to quasi-crystals, computational complexity, information theory, ... (see [2,3,6,12] and the references therein). In addition, substitutions are fundamental objects in combinatorial group theory [8,9].

Given a sequence $\xi = \xi_1\xi_2\xi_3\cdots$ over some finite alphabet $\mathcal{A}$, with $\xi_i \in \mathcal{A}$. We denote by $\mathcal{L}_n(\xi)$ the set $\{\xi_i \cdots \xi_{i+n-1} \mid i \geqslant 1\}$ of factors of ξ with length n $(n \geqslant 1)$. The set $\mathcal{L}_\xi = \bigcup_{n\geqslant 1} \mathcal{L}_n(\xi)$ is called the language of ξ, and the function $P_\xi(n) := \#\mathcal{L}_n(\xi)$ is called the complexity of ξ.

Let $\mathcal{A}^*$ be the free monoid generated by $\mathcal{A}$ (its identity is the empty word ε). A morphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ is called a substitution. Denote by ξ_σ any one of the fixed points of σ (that is $\sigma(\xi_\sigma) = \xi_\sigma$), if it exists.

The study of complexity and substitutions has a long history. Here are some classical results:

- $P_\xi(n) \leqslant n$ for some n if and only if ξ is ultimately periodic, and in this case the complexity is bounded [10];
- A sequence ξ over a two-letter alphabet with complexity $P_\xi(n) = n + 1$ is called Sturmian. There are many equivalent characterizations of Sturmian sequences (e.g., see [12,14,16]);
- Rote [13] constructed a class of sequences with complexity $2n$ by using graphs;
- Mossé [11] studied the case of q-automata (substitutions of constant length). A method to compute $P(n)$ with linear recurrence formulas was given under some technical conditions;
- Over a three-letter alphabet, a class of Tribonacci type substitutions with complexity $2n + 1$ was introduced by Arnoux and Rauzy [4].

However, the complexity of substitutions over a three-letter alphabet presents much more complex phenomena. Few examples can be explicitly worked out, even for invertible substitutions. This is because the structure of invertible substitutions over a three-letter alphabet is quite different from the case of substitutions over a two-letter alphabet: In [17] we showed that the set of invertible substitutions over a three-letter alphabet is not finitely generated. Nevertheless in [15] we were able to characterize the structure of invertible substitutions.

Now notice that for a primitive substitution, the corresponding complexity $P(n)$ satisfies a linear inequality $P(n) \leqslant C_1n + C_2$ $(n \geqslant 1)$ for some positive constants C_1 and C_2. If we confine ourselves to the case of a three-letter alphabet, and if the above inequality becomes equality for all $n \geqslant 1$, there are clearly only four possibilities: (1) $P(n) = 3$; (2) $P(n) = n + 2$; (3) $P(n) = 2n + 1$; (4) $P(n) = 3n$. The first case is trivial (periodic), the second case (Sturmian-like) was studied in [1]. Arnoux and Rauzy discussed the case (3) in [4]. For the case (4), as far as we know, the existence of such substitution is not present in the literature yet! This is one of the motivations of this Note. So we consider the substitution: $\sigma = (ab, acb, acc)$, that is, $a \mapsto ab$, $b \mapsto acb$, $c \mapsto acc$. We call it the *triplex substitution*.

We remark that the triplex substitution, which can be seen as a representative of undecomposable invertible substitutions (remark that the inverse of σ is $a \mapsto ab^{-1}ab^{-1}c$, $b \mapsto c^{-1}ba^{-1}b$, $c \mapsto c^{-1}ba^{-1}c$), plays an important rôle in the study of invertible substitutions over a three-letter alphabet [17,15]. In this Note, we show that its complexity is $3n$, which fulfils the case (4) above. We also see that its special words and factor structure are quite different from the Tribonacci type substitutions. We shall completely characterize the corresponding Rauzy fractal in the next Note.

2. Preliminary

Let $\mathcal{A} = \{a, b, c\}$ be a three-letter alphabet. Let $\mathcal{A}^*$ stand for the free monoid generated by $\mathcal{A}$.

If $w \in \mathcal{A}^*$, we denote by $|w|$ its length and by $|w|_a$ (resp. $|w|_b$, $|w|_c$) the number of occurrences of the letter a (resp. b, c) in w. The Parikh vector of w is $P(w) = (|w|_a, |w|_b, |w|_c)^t \in \mathbb{N}^3$.

A word v is a factor of a word w, and then we write $v \prec w$, if there exist $u, u' \in \mathcal{A}^*$, such that $w = uvu'$. We say that v is a prefix (resp. a suffix) of a word w, and then we write $v \lhd w$ (resp. $v \rhd w$), if there exists $u \in \mathcal{A}^*$ such that $w = vu$ (resp. $w = uv$). The notions of factor and prefix extend to infinite words in a natural way.

Let τ be a substitution over $\mathcal{A}$. The matrix $M_\tau = (P(\tau(a)), P(\tau(b)), P(\tau(c)))$ is called the substitution matrix of τ. A substitution is said to be primitive if its matrix is.

If $w = uv$, then $wv^{-1} := u$ and $u^{-1}w := v$ by convention.

For $0 \leqslant k < |w|$, we define the kth conjugate of w by $C_k(w) := x_k \cdots x_{n-1}x_0x_1 \cdots x_{k-1}$. We set $C(w) := \{C_k(w) \mid 0 \leqslant k < |w|\}$. A word $w \in \mathcal{A}^*$ is said to be primitive if the equation $w = u^p$ ($p \in \mathbb{N}$) implies $p = 1$, in other words, if all the conjugates of w are distinct (see [7]).

In the sequel, we consider the triplex substitution $\sigma = (ab, acb, acc)$. Its unique fixed point is $\xi = \xi_1\xi_2\xi_3 \cdots = abacbabaccacb \cdots$. For $n \geqslant 0$, we set $A_n = \sigma^n(a)$, $B_n = \sigma^n(b)$ and $C_n = \sigma^n(c)$.

The lengths of words A_n, B_n and C_n are related to the Fibonacci numbers defined by the recurrence formula $f(n) = f(n-1) + f(n-2)$, with the initial conditions $f(-2) = -1$, $f(-1) = 1$.

Proposition 1. *Let M be the substitution matrix of σ. We have*:

(i) $$M^n = \begin{bmatrix} f(2n-1) & f(2n) & f(2n) \\ f(2n-2)+1 & f(2n-1) & f(2n-1)-1 \\ f(2n-1)-1 & f(2n) & f(2n)+1 \end{bmatrix} \quad \text{for } n \geqslant 0;$$

(ii) *A_n and B_n begin with the letter a and end with b, while C_n begins with a and ends with c ($n \geqslant 1$);*

(iii) *For $n \geqslant 0$, the word B_n differs from C_n by exactly their last letter, i.e., $C_n = B_nb^{-1}c$;*

(iv) *$|A_n| = f(2n+1)$, $|B_n| = |C_n| = f(2n+2)$. The words A_n, B_n and C_n all are primitive.*

Proof. (i) holds by induction. (ii) and (iii) can be verified directly. (i) gives the first statement in (iv); the fact $\gcd(|A_n|_a, |A_n|_b, |A_n|_c) = 1$ yields that A_n is primitive. Likewise, B_n and C_n are primitive. □

3. Factor structure and complexity

In this section, we consider the factors of length $|A_n| = f(2n+1)$ and the structure of special words. From this, we determine the complexity of σ.

3.1. The factors of length $|A_n|$

Recall that ξ denotes the fixed point of σ, hence, for any $n \geqslant 1$,

$$\xi = \sigma^n(\xi) = \sigma^n(abacb\cdots) = A_nB_nA_nC_nB_n\cdots.$$

In particular, if w is a factor of ξ, so is $\sigma^n(w)$.

Since $|A_n| < |B_n| = |C_n|$, the factors of length $|A_n|$ can be divided into two classes:

Class I factors of either A_n, B_n or C_n.

Class II factors appearing at the concatenation of these A_n, B_n and C_n's. More precisely, the factors are of the form $w = sp \prec \sigma^n(\alpha)\sigma^n(\beta) = \sigma^n(\alpha\beta)$, where s is a proper suffix of $\sigma^n(\alpha)$, p is a proper prefix of $\sigma^n(\beta)$, and $\alpha\beta \in \mathcal{L}_2$ (recall that $\mathcal{L}_2$ stands for the set of factors of ξ of length 2).

For the factors in **Class I**, we have:

(i) $w = A_n$ is the only factor of A_n of length $|A_n|$;
(ii) B_n contains at most $|B_n| - |A_n| + 1$ factors of length $|A_n|$;
(iii) since $C_n = B_n b^{-1} c$, all the factors appearing in C_n are the same as those appearing in B_n except the last one (the suffix of C_n of length $|A_n|$).

Now we consider **Class II**: let $sp \in \sigma^n(\alpha\beta)$ be a word with $s \rhd \sigma^n(\alpha)$, $p \lhd \sigma^n(\beta)$, $s, p \neq \varepsilon$ and $\alpha\beta \in \mathcal{L}_2$.

It is readily checked that $\mathcal{L}_2 = \{ab, ac, ba, ca, cb, cc\}$. Hence we have to consider the following corresponding sub-cases:

- $A_n B_n$, more precisely, $sp \in A_n B_n$, $s \rhd A_n$, $p \lhd B_n$, $s, p \neq \varepsilon$:
  First notice that $\sigma^k(b^{-1}c) = b^{-1}c$ $(k \geqslant 0)$. We verify easily that $B_n = A_n b^{-1} c B_{n-1}$. This together with the assumption $p \lhd B_n$ (with $|p| < |A_n| < |B_n|$) implies that p is in fact a prefix of A_n. Thus we have $sp \prec A_n A_n$ (recall that $|sp| = |A_n|$), which shows that sp is a conjugate of A_n;
- $A_n C_n$: just the same as the above case, sp is a conjugate of A_n;
- $B_n A_n$: Consider the position where sp appears in
  $$B_n A_n = (A_{n-1} C_{n-1} B_{n-1})(A_{n-1} B_{n-1}).$$
  If s $(s \rhd A_{n-1} C_{n-1} B_{n-1})$ appears as a suffix of B_{n-1} in the above representation, B_{n-1} being a suffix of A_n (and p is a prefix of A_n), we know that sp is a conjugate of A_n. Now we consider the other positions of the factor sp. Since s is a proper suffix of B_n and p is a proper prefix of A_{n-1}. The number of this kind of words is at most $|A_{n-1}| - 1$;
- $C_n A_n$: the factor sp satisfies that s is a proper suffix of C_n and p is a proper prefix of A_n. The number of this kind of words is at most $|A_n| - 1$;
- $C_n B_n$ and $C_n C_n$: whence p is also a proper prefix of A_n, and there are no new factors sp other than those of $C_n A_n$.

Counting out all possibilities discussed above, the number of factors of length $|A_n|$ is at most

$$|A_n| + \big(|B_n| - |A_n| + 1\big) + 1 + \big(|A_{n-1}| - 1\big) + \big(|A_n| - 1\big) = 3|A_n|.$$

Up to now, we have shown the following:

Lemma 2. *For $n \geqslant 1$, the factors on length $|A_n|$ appearing in ξ are either*:

- *a conjugate word of A_n*;
- *a factor of B_n or C_n*;
- *a word of form sp with s a proper suffix of B_n and p a proper prefix of A_{n-1}*;
- *a word of form sp with s a proper suffix of C_n and p a proper prefix of A_n.*

In total, the number of factors of length $|A_n|$ is at most $3|A_n|$, i.e., $P_\xi(|A_n|) \leqslant 3|A_n|$.

3.2. Special words

The notion of "special words" is very useful for computing the complexity.

Let w be a factor of ξ and $\alpha \in \mathcal{A}$. If $w\alpha$ is also a factor of ξ, then we say that $w\alpha$ is a right extension (extension for short) of w. A word is called a right special word (special word for short) of ξ if it has more than one extension. Similarly we define "left extension" and "left special word". Notice that a suffix of a special word is also a special word. See [7,5] for more information.

In this subsection, we study the special words of ξ, the fixed point of the triplex substitution.

In $\mathcal{L}_1 = \{a, b, c\}$, the words a and c are special, and a has two extensions ab and ac, while c has three extensions ca, cb and cc. In $\mathcal{L}_2$, there are 3 special words ac, ca and ba.

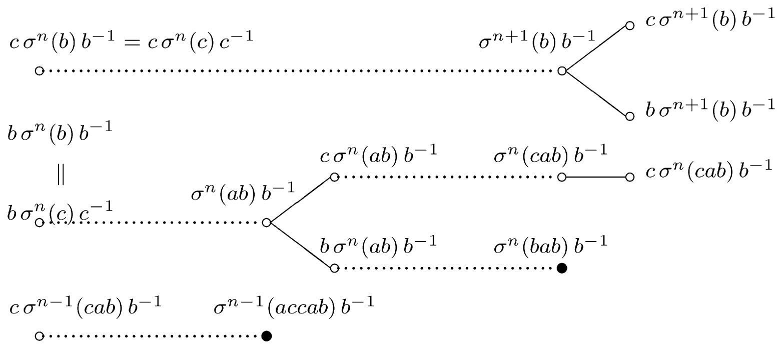

Fig. 1. Graph of special words (the part with length between $f(2n+2)$ to $f(2n+4)$).

Lemma 3. *A factor w with $|w| \geqslant 2$ is special if and only if it has exactly two (right) extensions, namely wb and wc.*

Proof. Since a suffix of a special factor is also a special factor, we need only check the lemma for the words of length 2. And this is just a routine. □

Lemma 4. *For $n \geqslant 1$, we have:*

(i) *both $b\sigma^n(b)b^{-1}$ and $c\sigma^n(b)b^{-1}$ are special; both $b\sigma^n(ab)b^{-1}$ and $c\sigma^n(ab)b^{-1}$ are special;*
(ii) *the words $\sigma^{n-1}(accab)b^{-1}$ and $\sigma^n(bab)b^{-1}$ are special.*

Proof. (1) Since ab and ac are factors of ξ, A_nB_n and A_nC_n are factors also. This, together with the facts that $C_n = B_nb^{-1}c$ and that b is the last letter of A_n, gives that $u = b\sigma^n(b)b^{-1}$ has two right extensions ub ($\prec A_nB_n$) and uc ($\prec A_nC_n$), and thus it is special. Likewise, by considering the factors cb and cc, we can show the word $c\sigma^n(b)b^{-1}$ is special.

Then considering the factor pairs $\{cab, cac\}$ and $\{bab, bac\}$, we can show similarly that both $c\sigma^n(ab)b^{-1}$ and $b\sigma^n(ab)b^{-1}$ are special.

(2) Since both $accab \prec \sigma^3(c)$ and $accac \prec \sigma^2(c)$ are factors, $\sigma^{n-1}(accab)b^{-1}$ is special; since both bab and bac are factors, $\sigma^n(bab)b^{-1}$ is special. □

For more characteristics of the special words, see Remark 7 and Proposition 8.

Since any suffix of a special word is also special, the set of special words can be viewed as an infinite tree, with an edge from w to its left extension αw if αw is special. From Lemma 4, we can get the "partial tree" shown in Fig. 1 of special factors with length between $f(2n+2)$ to $f(2n+4)$ (in the tree, many evident nodes on the path indicated by broken lines are omitted).

Let us remark that the special words are aligned vertically according to their lengths in Fig. 1. The following lemma is then clear:

Lemma 5. *For $k \geqslant 2$, there are at least 3 special factors of length k.*

3.3. Complexity

Theorem 6. *The complexity of ξ is $P(n) = 3n$ for $n \geqslant 1$.*

Proof. It is clear that $P(1) = 3$ and $P(2) = 6$. It suffices to show that $P(n+1) - P(n) = 3$ ($n \geqslant 2$).

By Lemma 5, $P(k+1) - P(k) \geqslant 3$ for all $k \geqslant 2$. If $P(k_0+1) - P(k_0) > 3$ for some k_0, then $P(n) > 3n$ whenever $n > k_0$. In particular $P(|A_n|) > 3|A_n|$ when $|A_n| > k_0$, which contradicts Lemma 2. □

Remark 7. The above theorem implies that there are exact three special words of length $n \geqslant 2$. Hence Fig. 1 presents *all* the special words of length between $f(2n+2)$ and $f(2n+4)$ (of course, do not ignore the omitted nodes).

The following concept is useful to determine the complexity of substitutive sequences:

Definition 1. Let w be a special word of ξ. If w has exactly k left extensions which are special words, we say that w is a k-special word.

By the above remark (and from Fig. 1) we have the following:

Proposition 8. *The special words $\sigma^n(b)b^{-1}$ and $\sigma^n(ab)b^{-1}$ are* 2*-special*; $\sigma^{n-1}(accab)b^{-1}$ *and* $\sigma^n(bab)b^{-1}$ *are* 0*-special*; *all other special words in the tree* (*Fig.* 1) *are* 1*-special.*

We end this note by giving a classes of substitutions with complexity $3n$ (moreover they share the same language with the triplex substitution), of which the proof is easy and omitted:

Proposition 9. *Let τ be the conjugate of the triplex substitution σ, i.e., $\tau = (ba, cba, cca)$, and $\sigma^{(m)} = \sigma_m \circ \cdots \circ \sigma_1$ with $\sigma_i \in \{\sigma, \tau\}$, $1 \leqslant i \leqslant m$. Suppose that ξ' is any one of the accumulation points of $\{\sigma^{(m)}(\alpha)\}_{m \geqslant 1}$, $\alpha \in \mathcal{A}$, then the complexity of the sequence ξ' is $P_{\xi'}(n) = 3n$ $(n \geqslant 1)$.*

References

[1] P. Alessandri, Classification et représentation des suites de complexité $n + 2$, Manuscrit non publié, 1995.
[2] J.-P. Allouche, Automates finis en théorie des nombres, Exposition. Math. 5 (1987) 239–266.
[3] P. Arnoux, S. Ito, Pisot substitutions and Rauzy fractals, in: Journées Montoises d'Informatique Théorique, Marne-la-Vallée, 2000, Bull. Belg. Math. Soc. Simon Stevin 8 (2001) 181–207.
[4] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n + 1$, Bull. Soc. Math. France 119 (1991) 199–215.
[5] J. Cassaigne, Special factors of sequences with linear subword complexity, in: Developments in Language Theory II (DLT'95), Magdeburg, Allemagne, World Sci., 1996, pp. 25–34.
[6] A. Cobham, Uniform tag sequences, Math. System Theory 6 (1972) 164–192.
[7] M. Lothaire, Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley, 1983.
[8] R.C. Lyndon, P.E. Schupp, Combinatorial Group Theory, Spring-Verlag, 1977.
[9] W. Magnus, A. Karrass, D. Solitar, Combinatorial Group Theory: Presentations of Groups in Terms of Generators and Relations, second revised edition, Dover Publications Inc., 1976.
[10] M. Morse, G.A. Hedlund, Symbolic dynamics II: Sturmian trajectories, Amer. J. Math. 62 (1940) 1–42.
[11] B. Mossé, Reconnaissabilité des substitutions et complexité des suites automatiques, Bull. Soc. Math. France 124 (1996) 329–346.
[12] N. Pytheas Fogg, Substitutions in dynamics, arithmetics and combinatorics, in: V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel (Eds.), in: Lecture Notes in Mathematics, vol. 1794, Springer, 2002.
[13] G. Rote, Sequences with subword complexity $2n$, J. Number Theory 46 (1994) 196–213.
[14] P. Séébold, Fibonacci morphisms and Sturmian words, Theoret. Comput. Sci. 88 (1991) 365–384.
[15] B. Tan, Z.-X. Wen, Y. Zhang, The structure of invertible substitutions on a three-letter alphabet, Adv. Appl. Math. 32 (2004) 736–753.
[16] Z.-X. Wen, Z.-Y. Wen, Local isomorphisms of invertible substitutions, C. R. Acad. Sci. Paris Sér. I 318 (1994) 299–304.
[17] Z.-X. Wen, Y. Zhang, Some remarks on invertible substitutions on three letter alphabet, Chinese Sci. Bull. 44 (1999) 1755–1760.