On finiteness of structures produced deterministically by oritatami co-transcriptional folding *

Szilárd Zsolt Fazekas[1] and Shinnosuke Seki[2]**

[1] Akita University, Graduate School of Engineering Science, 1-1 Tegate Gakuen-machi, Akita, 0108502, Japan
szilard.fazekas@ie.akita-u.ac.jp

[2] The University of Electro-Communications, Graduate School of Informatics and Engineering, 1-5-1 Chofugaoka, Chofu, Tokyo, 1828585, Japan
s.seki@uec.ac.jp

Abstract.

1 Introduction

2 Preliminaries

Let Σ be a set of types of abstract molecules, or *beads*. A bead of type $a \in \Sigma$ is called an a-bead. By Σ^* and Σ^ω, we denote the set of finite sequences of beads and that of one-way infinite sequences of beads, respectively. The empty sequence is denoted by λ. Let $w = b_1 b_2 \cdots b_n \in \Sigma^*$ be a sequence of length n for some integer n and bead types $b_1, \ldots, b_n \in \Sigma$. The *length* of w is denoted by $|w|$, that is, $|w| = n$. For two indices i, j with $1 \le i \le j \le n$, we let $w[i..j]$ refer to the subsequence $b_i b_{i+1} \cdots b_{j-1} b_j$; if $i = j$, then $w[i..i]$ is simplified as $w[i]$. For $k \ge 1$, $w[1..k]$ is called a *prefix* of w.

Oritatami systems fold their transcript, which is a sequence of beads, over the triangular grid graph $\mathbb{T} = (V, E)$ cotranscriptionally. We designate one point in V as the origin O of $\mathbb{T}$. For a point $p \in V$, let $\hexagon_p^d$ denote the set of points which lie in the regular hexagon of radius d centered at the point p. A directed path $P = p_1 p_2 \cdots p_n$ in $\mathbb{T}$ is a sequence of *pairwise-distinct* points $p_1, p_2, \ldots, p_n \in V$ such that $\{p_i, p_{i+1}\} \in E$ for all $1 \le i < n$. Its i-th point is referred to as $P[i]$. Now we are ready to abstract RNA single-stranded structures in the name of conformation. A *conformation* C (over Σ) is a triple (P, w, H) of a directed path P in $\mathbb{T}$, $w \in \Sigma^*$ of the same length as P, and a set of h-interactions $H \subseteq \big\{\{i,j\} \bigm| 1 \le i, i+2 \le j, \{P[i], P[j]\} \in E\big\}$. This is to be interpreted as the sequence w being folded along the path P in such a manner that its i-th

* This work is supported in part by KAKENHI Grant-in-Aid for Challenging Research (Exploratory) No. 18K19779 and JST Program to Disseminate Tenure Tracking System No. 6F36 granted to S. S.

** Corresponding author

bead $w[i]$ is placed at the i-th point $P[i]$ and the i-th and j-th beads are bound (by a hydrogen-bond-based interaction) if and only if $\{i,j\} \in H$. The condition $i+2 \le j$ represents the topological restriction that two consecutive beads along the path cannot be bound. A *rule set* $R \subseteq \Sigma \times \Sigma$ is a symmetric relation over Σ, that is, for all bead types $a, b \in \Sigma$, $(a,b) \in R$ implies $(b,a) \in R$. A bond $\{i,j\} \in H$ is *valid with respect to* R, or simply R-valid, if $(w[i], w[j]) \in R$. This conformation C is *R-valid* if all of its bonds are R-valid. For an integer $\alpha \ge 1$, C is *of arity* α if it contains a bead that forms α bonds but none of its bead forms more. By $\mathcal{C}_{\le\alpha}(\Sigma)$, we denote the set of all conformations over Σ whose arity is at most α; its argument Σ is omitted whenever Σ is clear from the context.

The oritatami system grows conformations by an operation called elongation. Given a rule set R and an R-valid conformation $C_1 = (P, w, H)$, we say that another conformation C_2 is an elongation of C_1 by a bead $b \in \Sigma$, written as $C_1 \xrightarrow{R}_b C_2$, if $C_2 = (Pp, wb, H \cup H')$ for some point $p \in V$ not along the path P and set $H' \subseteq \big\{\{i, |w|+1\} \bigm| 1 \le i < |w|, \{P[i], p\} \in E, (w[i], b) \in R\big\}$ of bonds formed by the b-bead; this set H' can be empty. Note that C_2 is also R-valid. This operation is recursively extended to the elongation by a finite sequence of beads as: for any conformation C, $C \xrightarrow{R}{}^*_\lambda C$; and for a finite sequence of beads $w \in \Sigma^*$ and a bead $b \in \Sigma$, a conformation C_1 is elongated to a conformation C_2 by wb, written as $C_1 \xrightarrow{R}{}^*_{wb} C_2$, if there is a conformation C' that satisfies $C_1 \xrightarrow{R}{}^*_w C'$ and $C' \xrightarrow{R}_b C_2$.

An *oritatami system* (OS) $\Xi = (\Sigma, R, \delta, \alpha, \sigma, w)$ is composed of

- a set Σ of bead types,
- a rule set $R \subseteq \Sigma \times \Sigma$,
- a positive integer δ called the *delay*,
- a positive integer α called the *arity*,
- an initial R-valid conformation $\sigma \in C_{\le\alpha}(\Sigma)$ called the *seed*, upon which
- its (possibly infinite) *transcript* $w \in \Sigma^* \cup \Sigma^\omega$ is to be folded by stabilizing beads of w one at a time so as to minimize energy collaboratively with the succeeding $\delta-1$ nascent beads.

The energy of a conformation $C = (P, w, H)$, denoted by $\Delta G(C)$, is defined to be $-|H|$; the more bonds a conformation has, the more stable it gets. The set $\mathcal{F}(\Xi)$ of conformations *foldable* by the system Ξ is recursively defined as: the seed σ is in $\mathcal{F}(\Xi)$; and provided that an elongation C_i of σ by the prefix $w[1..i]$ be foldable (i.e., $C_0 = \sigma$), its further elongation C_{i+1} by the next bead $w[i+1]$ is foldable if

$$C_{i+1} \in \mathop{\arg\min}_{\substack{C \in \mathcal{C}_{\le\alpha} s.t. \\ C_i \xrightarrow{R}_{w[i+1]} C}} \min\Big\{\Delta G(C') \Bigm| C \xrightarrow{R}{}^*_{w[i+2\ldots i+k]} C', k \le \delta, C' \in \mathcal{C}_{\le\alpha}\Big\}. \quad (1)$$

Then we say that the bead $w[i+1]$ and the bonds it forms are *stabilized* according to C_{i+1}. Note that an arity-α oritatami system cannot fold any conformation of arity larger than α. A conformation foldable by Ξ is *terminal* if none of its elongations is foldable by Ξ. The oritatami system Ξ is *deterministic* if for all $i \ge 0$, there exists at most one C_{i+1} that satisfies (1). A deterministic oritatami system folds into a unique terminal conformation. An oritatami system with the empty rule set just folds into an arbitrary elongation of its seed nondeterministically. Thus, the rule set is always assumed to be non-empty.

In the second half of this paper, we consider the unary oritatami system. An oritatami system is *unary* if its bead type set Σ is of size 1. Its sole bead type is denoted by a, that is, $\Sigma = \{a\}$. Its only possible rule is (a, a) so that the non-empty rule set assumption implies that its rule set is $R = \{(a, a)\}$. Its transcript is a sequence of a-beads. That is to say, the behavior of a unary oritatami system is fully determined by the delay, arity, and seed.

Proposition 1. *For any rule set R, arity α and conformation $C = (P, w, H)$ it is possible to check whether C is R-valid and whether $C \in \mathcal{C}_{\leq\alpha}$ in time $\mathcal{O}(|H| \cdot |w| \cdot |R|)$.*

Proof. To check whether C is R-valid:

1. FOR each $(i, j) \in H$:
2. IF $(w[i], w[j]) \notin R$ THEN answer NO and HALT
3. answer YES and HALT

Checking the condition in 2. can be done in $\mathcal{O}(|w| \cdot |R|)$ time for any reasonable representation of w and R, hence the whole process takes $\mathcal{O}(|H| \cdot |w| \cdot |R|)$ time. To check the arity constraint $C \in \mathcal{C}_{\leq\alpha}$:

1. FOR each $i \in \{1, \ldots, |w|\}$:
2. IF degree$(i) = |\{j | (i, j) \in H\}| > \alpha$ THEN answer NO and HALT
3. answer YES and HALT

Checking the condition in 2. can be done in $\mathcal{O}(|H|)$ time for any reasonable representation of H, hence the whole process takes $\mathcal{O}(|w| \cdot |H|)$ time.

Theorem 1. *There is an algorithm that simulates any oritatami system $\Xi = (\Sigma, R, \delta, \alpha, \sigma, w)$ in time $2^{\mathcal{O}(\delta)} \cdot |R| \cdot |w|$.*

Proof. Take any step in the computation, up to which some $i \geq 0$ first beads of w have been stabilized, with the last bead at a point p. The number of all possible elongations of the current conformation by the next δ-beads is $(6 \times 5^{\delta-1}) \times ((2^4)^{\delta-1} \times 2^5) \in 2^{O(\delta)}$. By Proposition 1, we can check for each of these elongations whether its arity is at most α or not and whether it is R-valid or not in time $\mathcal{O}((2^4)^{\delta-1} \cdot 2^5 \cdot \delta \cdot |R|) = 2^{\mathcal{O}(\delta)} \cdot |R|$. Therefore, the total running time is $2^{\mathcal{O}(\delta)} \cdot |R| \cdot |w|$.

Corollary 1. *For fixed δ and α, the class of problems solvable by oritatami systems $(\Sigma, R, \delta, \alpha, \sigma, w)$ is included in* DTIME(n^3).

Proof. The claim follows from Theorem 1 and the fact that $|R|$ is implicitly bounded by $|w|^2$.

Because of the time hierarchy theorems, we know that P $\subsetneq$ EXP, so we can conclude that OS which cannot deterministically fold transcripts of length exponential in the length of the seed are not computationally universal.

3 Case of $\delta = 1$

4 Case of $\delta \geq 2$