Towards the Algorithmic Molecular Self-Assembly of Fractals by Cotranscriptional Folding[*]

Yusei Masuda, Shinnosuke Seki[**], and Yuki Ubukata

Department of Computer and Network Engineering, The University of
Electro-Communications, 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*, and Σ^* be the set of finite sequences of beads. A bead of type $a \in \Sigma$ is called an a-bead. Let $w = b_1 b_2 \cdots b_n \in \Sigma^*$ be a string 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 $1 \leq i \leq j \leq n$, we let $w[i..j]$ refer to the subsequence $b_i b_{i+1} \cdots b_{j-1} b_j$; if $i = j$, then we simplify $w[i..i]$ as $w[i]$. For $k \geq 1$, $w[1..k]$ is called a *prefix* of w.

Oritatami systems fold their transcript, a sequence of beads, over the triangular grid graph $\mathbb{T} = (V, E)$ as suggested in Fig. **??** cotranscriptionally based on hydrogen-bond-based interactions (*h-interactions* for short) which the rule set of the system allow for between adjacent beads of particular types. 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 $1 \leq i < n$. Its i-th point is referred to as $P[i]$. A *rule set* $\mathcal{H} \subseteq \Sigma \times \Sigma$ is a symmetric relation over the set of pairs of bead types, that is, for all bead types $a, b \in \Sigma$, $(a, b) \in \mathcal{H}$ implies $(b, a) \in \mathcal{H}$.

A *conformation C* 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 $H \subseteq \{\{i, j\} \mid 1 \leq i, i + 2 \leq j, \{P[i], P[j]\} \in E\}$. This is to be interpreted as the sequence w being folded in such a manner that its i-th bead $w[i]$ is placed on the i-th point $P[i]$ along the path and there is an h-interaction between the i-th and j-th beads if and only if $\{i, j\} \in H$. The condition $i + 2 \leq j$ represents the topological restriction that two consecutive beads along the path cannot form an h-interaction between them. Let $\mathcal{H}$ be a rule set. An h-interaction $(i, j) \in H$ is *valid with respect to* $\mathcal{H}$,

or simply $\mathcal{H}$-*valid*, if $(w[i], w[j]) \in \mathcal{H}$. This conformation C is $\mathcal{H}$-valid if all of its h-interactions are $\mathcal{H}$-valid. For an integer $\alpha \geq 1$, C is *of arity* α if C contains a bead that forms α h-interactions and no bead of C forms more. By $\mathcal{C}_{\leq \alpha}$, we denote the set of all conformations of arity at most α.

Oritatami systems grow conformations by elongating them under their own rule set. Given a rule set $\mathcal{H}$ and an $\mathcal{H}$-valid finite 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 $C_1 \xrightarrow{\mathcal{H}}_b C_2$, if $C_2 = (Pp, wb, H \cup H')$ for some point p not along the path P and set of h-interactions $H' \subseteq \{\{i, |w| + 1\} \mid 1 \leq i < |w|, \{P[i], p\} \in E, (w[i], b) \in \mathcal{H}\}$, which can be empty. Note that C_2 is also $\mathcal{H}$-valid. This operation is recursively extended to the elongation by a finite sequence of beads as: for any conformation C, $C \xrightarrow{\mathcal{H}}{}^*_\lambda C$; 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 $C_1 \xrightarrow{\mathcal{H}}{}^*_{wb} C_2$, if there is a conformation C' that satisfies $C_1 \xrightarrow{\mathcal{H}}{}^*_w C'$ and $C' \xrightarrow{\mathcal{H}}_b C_2$.

A finite *oritatami system* (OS) is a 5-tuple $\Xi = (\mathcal{H}, \alpha, \delta, \sigma, w)$, where $\mathcal{H}$ is a rule set, α is an arity, $\delta \geq 1$ is a parameter called the *delay*, σ is an initial $\mathcal{H}$-valid conformation of arity α called the *seed*, upon which its finite *transcript* $w \in \Sigma^*$ 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 h-interactions a conformation has, the more stable it gets. The set $\mathcal{F}(\Xi)$ of conformations *foldable* by this system 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

$$C_{i+1} \in \underset{\substack{C \in \mathcal{C}_{\leq \alpha} s.t. \\ C_i \xrightarrow{\mathcal{H}}_{w[i+1]} C}}{\arg \min} \min \left\{ \Delta G(C') \mid C \xrightarrow{\mathcal{H}}{}^*_{w[i+2...i+k]} C', k \leq \delta, C' \in \mathcal{C}_{\leq \alpha} \right\}. \quad (1)$$

We say that the bead $w[i + 1]$ and the h-interactions it forms are *stabilized* according to C_{i+1}. Note that an arity-α OS cannot fold any conformation of arity larger than α. The OS Ξ is *deterministic* if for all $i \geq 0$, there exists at most one C_{i+1} that satisfies (1).

3   Folding the n-bit Heighway dragon

We propose a generic design of deterministic cyclic OS that allows us to fold an arbitrary finite portion $P[j_1..j_2]$ of the slanted Heighway dragon. Independently of j_1, j_2, the design sets both delay and arity to 3 and employs 567 bead types with a fixed rule set $\mathcal{H}$ (some of the bead types might be saved but not easily due to the NP-hardness of minimizing the number of bead types [2]).

One period is semantically divided into six parts (*modules*) as $CF_v TCF_h T$, where

- C is called a *counter module*, which increments i by 1 and propagates it;
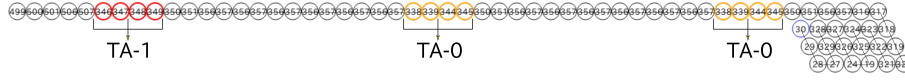
Fig. 1. The seed conformation for the 3-bit Heighway dragon encoding the initial count 100.

- F_v, F_h are called a *DFAO module*, which computes $P[i]$ and interprets it properly as A(cute) or O(btuse);
- T is called a *turning module*, which makes a turn according to A/O.

The first C and F_v modules fold into a vertical line segment, while the second C and F_h fold into the next line segment, which is guaranteed to be horizontal. The DFAO modules F_v, F_h differ only in their way to interpret their intermediate outcome $P[i]$. The slanted dragon involves two types of left turn as well as two types of right turn: acute and obtuse. Observe that after (slanted) vertical line segments, the dragon turns left obtusely and right acutely, whereas after horizontal ones, it turns left acutely and right obtusely. Moreover, vertical and horizontal segments alternate on the dragon. Therefore, it suffices for F_v and F_h to compute $P[i] \in \{L, R\}$ in the same way and interpret it as A(cute)/O(btuse) in an opposite fashion as F_v converts L into O and R into A, while F_h converts L into A and R into O.

Seed conformation encodes the initial count j_1 in its binary representation $b_n b_{n-1} \cdots b_1$ as

$$499{\to}500{\to}501{\to}506{\to}507{\to}\bigodot_{k=n}^{1}\left(\mathtt{w_{t,b_k}}{\to}350{\to}351{\to}(356{\to}357{\to})^6\right) \quad (2)$$

where $\mathtt{w_{t,0}} = 338{\to}339{\to}344{\to}345$ and $\mathtt{w_{t,1}} = 346{\to}347{\to}348{\to}349$.

Counter module is implemented essentially in the same manner as in [1]. It takes the current count i formatted as (2), which is fed by the seed or by the previous turner module, and increments the count by 1 unless it is preceded by the seed, and outputs the resulting count in its binary representation $a_n a_{n-1} \cdots a_1$ in the following format:

$$44{\to}45{\to}46{\to}75{\to}76{\to}\bigodot_{k=n}^{1}\left(\mathtt{w_{c,a_k}}{\to}(69{\to}70{\to})^5 75{\to}76{\to}\right) \quad (3)$$

where $\mathtt{w_{c,0}} = 51{\to}52{\to}57{\to}58{\to}63{\to}64$ and $\mathtt{w_{c,1}} = 59{\to}60{\to}61{\to}62{\to}63{\to}64$.
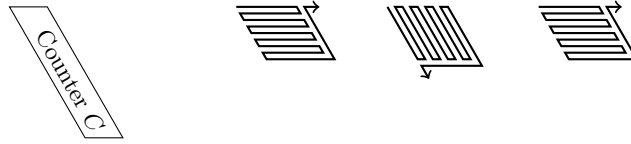
DFAO module

Turner module

References

1. C. Geary, P.-É. Meunier, N. Schabanel, and S. Seki. Programming biomolecules that fold greedily during transcription. In *Proc. MFCS2016*, LIPIcs 58, pages 43:1–43:14, 2016.

2. Y-S. Han and H. Kim. Ruleset optimization on isomorphic oritatami systems. In *Proc. DNA23*, LNCS 10467, pages 33–45. Springer, 2017.

A   Module automaton

Fig. 2. Module automaton

B   Transcript