Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Rule set design problems for oritatami system

Makoto Ota, Shinnosuke Seki *,[1]

University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo, 1828585, Japan

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form 1 September 2016
Accepted 17 September 2016
Available online xxxx

Keywords:
Co-transcriptional folding
Oritatami system
Algorithmic self-assembly
Time complexity
Linear-time 2SAT solver

ABSTRACT

A single-stranded RNA is transcribed from its DNA template by an RNA polymerase enzyme. The RNA transcript begins to fold upon itself while it is still being transcribed. This ubiquitous phenomenon is called cotranscriptional folding and was recently used as an engineering tool to self-assemble "RNA origami" tile by Geary, Rothemund, and Andersen (2014) [8]. The oritatami system (OS) is a new mathematical model of algorithmic self-assembly by cotranscriptional folding, proposed by Geary, Meunier, Schabanel, and Seki (2016). A problem of designing OSs is studied in this paper. We provide a sharp boundary between the **NP**-hardness and polynomial-time computability of this problem with respect to the relative speed of transcription to folding and valence of molecules.

1. Introduction

Information in nucleic acids (DNA, RNA) is encoded onto a chain of nucleotides (nts) (A, C, G, T for DNA, and U is in place of T for RNA) in a directional manner. An RNA sequence is synthesized through a process called *transcription*, in which RNA polymerase enzyme binds to a template DNA and produces the complementary RNA sequence (transcript) from the 5' to 3'-end until it is dissociated from the template. Transcripts fold upon itself by forming/dissociating helices through cooperative base pairing/unpairing via hydrogen bonds into an intricate functional 3D-structures such as the ribosome, arguably the most essential biomolecule in all living cells.

The directionality of synthesis and the rate at which nucleotides are added to the growing chain allow the RNA to fold over a preprogrammed pathway into a non-equilibrium structure [1]. Nucleotides are transcribed much more slowly (on millisecond/nt scale) than the RNA transcript undergoes conformational search to form helices (on μs time scale). This phenomenon is called *cotranscriptional (or kinetically-driven) folding*. Conventional folding, in which a fully-synthesized chain is folded for a minimum-free energy (MFE) structure at equilibrium, has been intensively studied (see [2] and references therein). It is computationally-hard in general [3]. Prediction of MFE structures of RNA has been made in various models [4–6,2]. In contrast, little has been done on cotranscriptional folding. Kinefold [7] provides a simulation of RNA cotranscriptional folding but is limited to 400 nts as of now.

Cotranscriptional folding is a novel challenge in the field of algorithmic self-assembly. RNA tiles [8] and cubic scaffolds [9] have been already assembled cotranscriptionally. The architecture for RNA tiles in [8] named *RNA origami* (Fig. 1) is striking in that it provides a method to design a single-stranded RNA (ssRNA) that cotranscriptionally folds into a target structure.

* Corresponding author.
E-mail address: s.seki@uec.ac.jp (S. Seki).

[1] His work is in part supported by Japan Science and Technology Agency Program to Disseminate Tenure Tracking System No. 6F36, JSPS KAKENHI Grant-in-Aid for Research Activity Start-up No. 15H06212 and for Young Scientists (A) No. 16H05854, and UEC Research Support for Newly Appointed Faculty Members No. XZ002.

http://dx.doi.org/10.1016/j.tcs.2016.09.010
0304-3975/

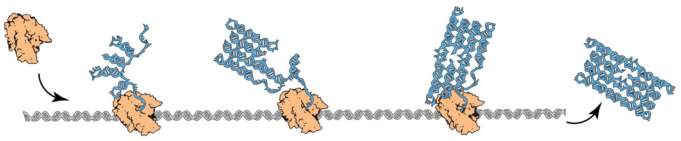

Fig. 1. RNA origami [8], an architecture of a ssRNA that folds into a tile cotranscriptionally. The artwork is by Cody Geary.

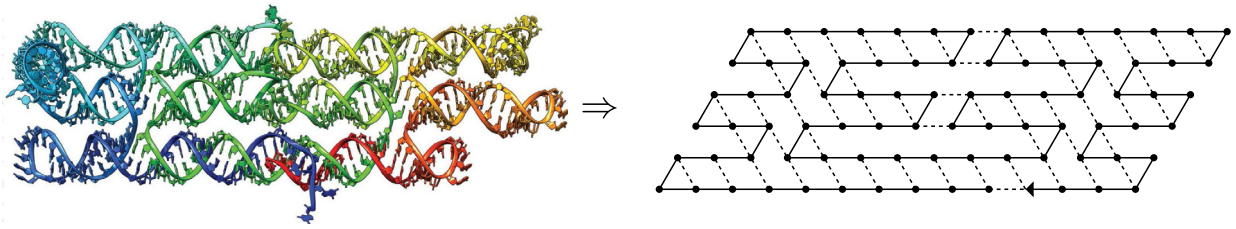

Fig. 2. Abstraction of a design of RNA tile (Left) as a directed path over the hexagonal grid with pairings (Right). The idea and artwork were provided by Cody Geary [12].

The design of RNA origami proceeds in two stages; first, the kinetic folding pathway is designed so as to maximize the locality of interactions, and then its actual RNA sequence is quested for through a stochastic search on subsequences, with the aid of software such as ViennaRNA package [10] or Nupack [11], to evaluate the stability of short regions. [12]. Efficient folding pathways result from intricate networks of events to form/dissociate strong double-helical segments, loops, and comparatively weaker paranemic helices for long-distance interactions to induce tertiary structures [13,1]. The intricacy arises from kinetic and topological restrictions on the order of these events.

(2-dimensional) *oritatami system* (OS) is a first mathematical model to study and engineer algorithmic self-assembly system by cotranscriptional folding at arbitrary levels of abstraction, proposed in [14,15]. The OS models a ssRNA as a directed ($5' \to 3'$) chain of beads each of which may represent one nucleotide or domain of nucleotides. An RNA structure is modeled as a path over some grid labeled with such a bead chain along with a set of pairings between beads at unit distance. See Fig. 2, where a design of RNA tile is thus abstracted such that one bead represents 4 nts, that is, one pairing corresponds to 4/11 turn of a helix. In the OS, structures are called *conformations*. A *rule set*, a parameter of OS, specifies what types of beads can be paired. Pairings are based on hydrogen bonds, but depending on the level of abstraction, one pairing may be a hydrogen bond, a paranemic helix, a full helix, etc. Thus, we collectively call them an *h-interaction*. In OSs, at the 3'-end of an intermediate conformation is a fragment of δ nascent beads, where $\delta \geq 1$ is a parameter called *delay*, which represents the relative speed of transcription to folding. These nascent beads can form and dissociate h-interactions with others or between them as long as topologically possible. The OS thus goes through conformational search, and according to the most stable conformation(s), it perpetuates the eldest (far-5'-end) nascent bead pointwise and h-interaction-wise. The next bead, if any, is then transcribed. Starting from an initial conformation (*seed*), the OS assembles conformation(s) by repeating the perpetuation and transcription. RNA origami [8] requires no seed to fold upon. On the other hand, having an RNA transcript first bind to a pre-built structure and fold upon it seems to narrow possible conformations, simplifying the energy landscape somewhat [12]. The concept of seed in the OS is thus motivated (cf. [16] for the seedless OS). The OS is efficiently Turing universal [14].

One practical goal of OS is to provide an efficient kinetic folding pathway over a chain of domains, based on which an RNA origami is manufactured through sequence design for the domains such that domains interact as designed. The logical design problem is formulated as a *rule set design problem for unique conformation* (RSD-UniqCFM), in which a chain w of domains and a target conformation C are given, and one is asked to find a rule set $\mathcal{H}$ with which an OS cotranscriptionally folds w into C. Recall that cotranscriptional folding pathways may involve dissociations of interactions for efficiency. This means that $\mathcal{H}$ may allow for some interactions that never get permanent, that is, do not appear in C but are indispensable to cotranscriptional folding of w into C. In [14,15], a similar problem was studied. The only difference is that how beads should be paired is not specified as input, that is, just a directed path P is given instead of being along with a set of pairings (conformation). The problem asks for a rule set to fold a given primary structure into a conformation whose path is P. Our problem focuses on interactions to be dissociated and may provide insight into their role in efficient cotranscriptional folding pathways.

We study the time complexity of RSD-UniqCFM. As shown in Fig. 3, we provide a sharp boundary between the **NP**-hardness (Theorems 2 and 3) and tractability (Theorems 1 and 4) w.r.t. delay δ and another OS parameter α called *arity*, that bounds the number of h-interactions a bead can form.

2. Preliminaries

In the introduction to oritatami systems below, we borrow terminologies from graph theory and formal language theory [17,18].

Please cite this article in press as: M. Ota, S. Seki, Rule set design problems for oritatami system, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.09.010