



Counting Distinguishable RNA Secondary Structures

MASARU NAKAJIMA¹ and ANDREW D. SMITH²

ABSTRACT

RNA secondary structures are essential abstractions for understanding spacial folding behaviors of those macromolecules. Many secondary structure algorithms involve a common dynamic programming setup to exploit the property that secondary structures can be decomposed into substructures. Dirks et al. noted that this setup cannot directly address an issue of distinguishability among secondary structures, which arises for classes of sequences that admit nontrivial symmetry. Circular sequences are among these. We examine the problem of counting distinguishable secondary structures. Drawing from elementary results in group theory, we identify useful subsets of secondary structures. We then extend an algorithm due to Hofacker et al. for computing the sizes of these subsets. This yields a cubic-time algorithm to count distinguishable structures compatible with a given circular sequence. Furthermore, this general approach may be used to solve similar problems for which the RNA structures of interest involve symmetries.

Keywords: counting algorithms, RNA secondary structure.

1. INTRODUCTION

THE RNA SECONDARY STRUCTURE MODEL is a convenient framework for analyzing and understanding RNA conformations (Mathews, 2006). A secondary structure is a specifically constrained set of paired positions in a sequence. Algorithms to identify optimal structures for given RNA sequences, that is, to “fold” RNA, focused on maximizing the number of paired bases (Nussinov and Jacobson, 1980) and finding minimum free energy secondary structures (Waterman, 1978; Zuker and Stiegler, 1981). Although models of varying intricacies exist, free energy for a secondary structure is typically evaluated as a function of energies associated with bases at paired positions. The algorithm of McCaskill (1990) efficiently computes the partition function for the set of secondary structures, and enables a detailed characterization of equilibrium secondary structural features.

Secondary structures with so-called pseudoknots are biologically relevant, but inclusion of such structures in the analysis dramatically increases its complexity (Akutsu, 2000; Dirks and Pierce, 2003; Rivas and Eddy, 1999).

Most of the algorithmic work on secondary structures has focused on ordinary RNA sequences, that is, strings, but certain variations have also emerged as important. For example, circular RNAs are relevant in

Departments of ¹Physics and Astronomy and ²Quantitative and Computational Biology, University of Southern California, Los Angeles, California, USA.

studies of certain viruses (Gudima et al., 2004) and viroids (Flores et al., 2004). Although circular RNAs fundamentally differ from ordinary RNA sequences, the usual algorithms can often be adapted by slight modifications to solve corresponding problems on circular sequences. Zuker and Sankoff (1984) and Zuker (1989) established a folding algorithm for circular RNAs. Secondary structures for interacting RNA sequences have also been studied, driven, in part, by applications in biotechnology (Isaacs et al., 2006).

Algorithms to fold multiple interacting RNA sequences have been developed by Dimitrov and Zuker (2004) and Andronescu et al. (2005). Dirks et al. (2007) gave a cubic time algorithm to compute the partition function for a multiset of RNA sequences under restrictions that remain useful in practice.

The core of most RNA folding and partition function algorithms is dynamic programming that exploits properties of substructures. The recurrences for such algorithms closely resemble those for context-free grammars (Sakai, 1961) and may be viewed as variants thereof. In this article, we refer to this type of dynamic programming algorithm as the “standard” dynamic programming for convenience. For ordinary RNA sequences, the secondary structures accounted for by such algorithms all represent physically distinguishable conformations. However, this need not hold for circular RNAs or multiple interacting RNAs. For circular RNAs, if the sequence consists of repeated substrings, different secondary structures may be indistinguishable (Hofacker et al., 2012). Dirks et al. (2007) also observed that secondary structures of interacting RNA sequences can be indistinguishable.

Such indistinguishable secondary structures can lead to redundancy when attempting to count or compute partition functions. Moreover, the extent of such redundancy depends on the symmetry (defined formally below) of each structure (Dirks et al., 2007). Since this symmetry is a global property for any individual secondary structure, the standard dynamic programming schemes, which decompose a problem as independent subproblems, cannot be adapted through local adjustments. Dirks et al. (2007) observed that, in computing the partition function, the above fact does not cause an issue, and had the insight that a symmetry correction to the free energy can compensate for any redundancy. However, redundancy cannot be ignored when counting distinguishable secondary structures. To our knowledge, no efficient algorithm to count distinguishable secondary structures for circular RNAs has been given.

We introduce a general approach to count distinguishable secondary structures on circular sequences. We use a result from group theory to identify certain subsets of secondary structures, the sizes of which are related to the number of distinguishable secondary structures. This allows us to build on an algorithm given by Hofacker et al. (2012), extending it to count distinguishable secondary structures for circular sequences in cubic time, in a model of computation that allows constant time multiplication.

2. PRELIMINARIES

A circular RNA sequence w is a cyclically ordered multiset of RNA bases $\{A, C, G, U\}$. Cyclic ordering implies, for example, that circular sequences spelled ACG, CGA, GAC are all identical. The length of a circular sequence is the size of the multiset of bases. To reference the positions of bases in a circular sequence w of length n , we assume a representative spelling has been arbitrarily chosen (e.g., the spelling ACG for the above example). We then index from 0 to $n-1$ in that spelling, such that $w_i \in \{A, C, G, U\}$ for each $i \in [0, n)$. We call these indices *sites*.

A *secondary structure* s for w is a set of unordered pairs from $\{0, \dots, n-1\}$ such that:

S1. For all $\{i, j\}$ in s , $i \neq j$.

S2. For all $\{i, j\}$ and $\{k, l\}$ in s , either $\{i, j\} = \{k, l\}$ or $\{i, j\} \cap \{k, l\} = \emptyset$.

S3. For all $\{i, j\}$ and $\{k, l\}$ in s , if $i < k < j$ then $i < l < j$.

S4. For all $\{i, j\}$ in s , $\{w_i, w_j\} \in B$, where $B = \{\{A, U\}, \{C, G\}, \{G, U\}\}$ are the allowed base pairs.

Condition S1 prevents a base from pairing with itself. In practical applications, steric constraints such as $|j-i| > 3$ are used (Mathews et al., 1999); here we ignore these for simplicity. Condition S2 ensures that no site is involved in more than one pair. Condition S3 prevents two pairs $\{i, j\}$ and $\{k, l\}$ from satisfying $i < k < j < l$, which is called a pseudoknot. Condition S4 ensures that the pairs are valid RNA base pairs. We let Ω_w denote the set of secondary structures for w . From here onward, we simply write structures to mean secondary structures.

The problem of counting structures for a given circular sequence w is closely related to the partition function problem for ordinary sequences, which can be solved in cubic time (Lyngsø et al., 1999; McCaskill, 1990). We note that, while the last positions for the sequence w is $n-1$, the maximum index j in

the quantity $C_{i,j}$ is n , as the quantities $C_{i,n}$ are for the substrings $w[i..n-1]$. For convenience, we define the base cases as $C_{i,i} = 1$ for $0 \leq i \leq n$, and for $0 \leq i < j \leq n$ the recurrence can be expressed as follows:

$$C_{i,j} = C_{i+1,j} + \sum_{i < k \leq j} C_{i,k}^b C_{k,j}, \tag{1}$$

where $C_{i,j}^b$ is the number of structures for $w[i..j-1]$ and containing the pair $\{i, j-1\}$. In Equation (1), the first term corresponds to structures with site i unpaired. The second term accounts for structures containing pair $\{i, k-1\}$, for all $k \in (i, j]$. For $0 \leq i < j \leq n$, $C_{i,j}^b = 0$ if $\{w_i, w_{j-1}\} \notin B$, and otherwise

$$C_{i,j}^b = C_{i+1,j-1}. \tag{2}$$

The quantities $C_{i,j}$ and $C_{i,j}^b$ for all $0 \leq i \leq j \leq n$ can be computed in $\mathcal{O}(n^3)$ time by dynamic programming. This implies that we can compute $|\Omega_w|$ in cubic time. Here, and throughout this work, we assume that integer multiplication takes constant time. We discuss this assumption in Section 6.

3. DISTINGUISHABILITY AMONG STRUCTURES

Here we lay out the concepts of symmetry for structures over circular sequences, and distinguishability of structures. Our explanation here follows the analogous framework given by Dirks et al. (2007) for particular interacting sequences (which we discuss in Section 5).

We say that circular sequence w has *symmetry* p if p is the largest integer such that w may be formed by concatenating one string p times. For example, the circular sequence $w = \text{ACACACACAC}$ has symmetry $p = 5$. Evidently, the symmetry of a circular sequence divides its length (i.e., the number of bases). If w has symmetry p , then the bases in w satisfy $w_i = w_{(i+t) \bmod n}$ for all $i \in [0, n)$, where $t = n/p$. We say that a circular sequence has nontrivial symmetry if $p > 1$.

Nontrivial symmetry in a circular sequence renders some structures indistinguishable. Consider a structure s for circular sequence w of length n and symmetry $p > 1$. Dirks et al. (2007) recognized that, when there is a rotational symmetry in the sequence, a structure s is indistinguishable from another structure s' if s' can be obtained by cyclically shifting every pair in s by ht , for any $h \in [0, p)$. Formally, we define the rotation of a pair $\{i, j\}$ as a mapping

$$Y_c(\{i, j\}) = \{(i+c) \bmod n, (j+c) \bmod n\},$$

for any $c \in [0, n)$. We extend this operation to structures:

$$Y_c(s) = \{Y_c(\{i, j\}) : \{i, j\} \in s\}.$$

Let $\mathbb{Z}/p\mathbb{Z}$ denote the cyclic group of order p under addition mod p . Given a circular sequence w of length n and symmetry p , the action of $h \in \mathbb{Z}/p\mathbb{Z}$ is defined as the rotation Y_{ht} , where again $t = n/p$. The orbit of structure $s \in \Omega_w$ under $\mathbb{Z}/p\mathbb{Z}$ is then

$$\text{Orb}_p(s) = \{Y_{ht}(s) : h \in \mathbb{Z}/p\mathbb{Z}\}.$$

Therefore, the orbits induced by $\mathbb{Z}/p\mathbb{Z}$ partition Ω_w into sets of structures. Two structures s and s' are indistinguishable if they belong to the same orbit, and distinguishable otherwise. Such structures have equivalent loop decomposition and therefore identical free energy. The Orbit-Stabilizer theorem states

$$|\text{Orb}_p(s)| |\text{Stab}_p(s)| = p,$$

where the stabilizer subgroup of s under the action of $\mathbb{Z}/p\mathbb{Z}$ is

$$\text{Stab}_p(s) = \{h : h \in \mathbb{Z}/p\mathbb{Z} \text{ and } Y_{ht}(s) = s\}.$$

We define the *symmetry of a structure* s as the size of its stabilizer. Note that the order of the cyclic group $\mathbb{Z}/p\mathbb{Z}$ is p , the symmetry of the given sequence w . The symmetry of the structure s must then divide the symmetry p of the circular sequence w . The number of indistinguishable structures for a particular structure

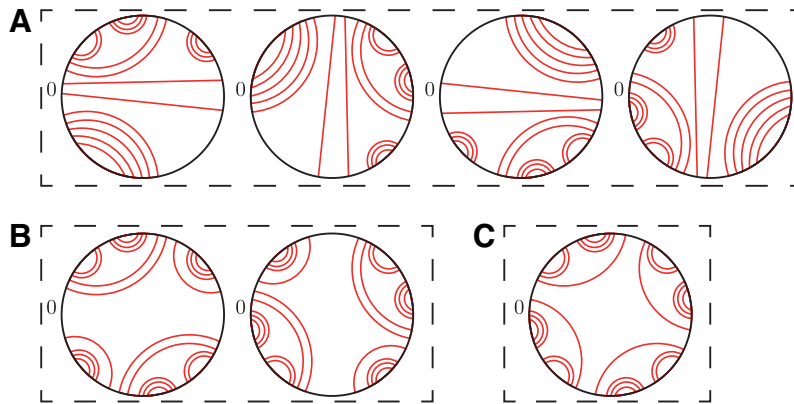


FIG. 1. Orbits of different sizes for a circular sequence of symmetry 4 (0 indicates the location of the first site). (A) Orbit size of 4 and symmetry 1. (B) Orbit size of 2 and symmetry 2. (C) Orbit size of 1 and symmetry 4.

is inversely proportional to the symmetry of that structure. Figure 1 shows example orbits of varying sizes and symmetries for a circular sequence of symmetry 4. For circular sequence w of symmetry p , the set of distinguishable structures is defined as the quotient set

$$\Lambda_w = \{\text{Orb}_p(s) : s \in \Omega_w\}.$$

That is, we treat each orbit as a distinguishable structure.

4. COUNTING DISTINGUISHABLE STRUCTURES

Our computational problem is the counting problem defined as follows:

Problem: Distinguishable Count

Input: A circular sequence w of length n and symmetry p .

Question: What is the size of Λ_w ?

To our knowledge, there is no efficient algorithm for Distinguishable Count problem in the literature. We provide an efficient algorithm by defining subsets consisting of structures having certain periodic characteristics. We first establish some properties of these subsets and then exploit those properties to design an efficient algorithm for counting distinguishable structures compatible with a circular sequence.

Consider a circular sequence w of length n and symmetry p , and let $t = n/p$. In the previous section, we showed that counting distinguishable structures amounts to counting the orbits under the action of cyclic group $\mathbb{Z}/p\mathbb{Z}$. Burnside's lemma relates the number of orbits to particular subsets of Ω_w :

$$|\Lambda_w| = \frac{1}{p} \sum_{h \in \mathbb{Z}/p\mathbb{Z}} |\text{Fix}(h)|, \quad (3)$$

where

$$\text{Fix}(h) = \{s \in \Omega_w : Y_{ht}(s) = s\} \quad (4)$$

is the subset of Ω_w fixed under the group action of $h \in \mathbb{Z}/p\mathbb{Z}$. We call $\text{Fix}(h)$ the h -periodic subset of Ω_w . For any $h \in \mathbb{Z}/p\mathbb{Z}$, if $s \in \text{Fix}(h)$, then $s \in \text{Fix}(h')$ for all $h' \in \langle h \rangle$, where $\langle h \rangle$ is the subgroup of $\mathbb{Z}/p\mathbb{Z}$ generated by h . Since $\langle h \rangle = \langle \gcd(h, p) \rangle$ for any $h \in \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$, we conclude that

$$\text{Fix}(h) = \text{Fix}(\gcd(h, p)),$$

for any $h \in \mathbb{Z}/p\mathbb{Z} \setminus \{0\}$. Furthermore, $\text{Fix}(0) = \Omega_w$, whose size is given by $C_{0,n}$. To compute $|\Lambda_w|$, we therefore only need to compute the size of $\text{Fix}(g)$ for every proper divisor g of p .

Hofacker et al. (2012) showed that in a structure with nontrivial symmetry, the pairs can be said to be repeated. More precisely, assume w is a circular sequence of length n and symmetry $p > 1$. If s is a structure for w and s has symmetry p , then $\{i, j\} \in s$ implies $Y_t(\{i, j\}) \in s$, where $t = n/p$. This property

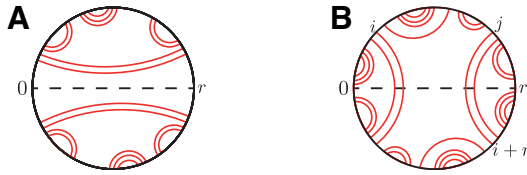


FIG. 2. Example secondary structures for circular sequence of symmetry 4. **(A)** Structure in $\text{Fix}(2)$ with only internal pairs. **(B)** Structure in $\text{Fix}(2)$ containing external pairs.

allows any pair in s to be either of form $Y_{hr}(\{i, j\})$ (*internal pair*) or of form $Y_{hr}(\{i+t, j\})$ (*external pair*), for some $h \in [0, p)$ and $0 \leq i < j < t$. We extend this property to the periodic subsets as follows. Again, assume w is a circular sequence of length n and symmetry $p > 1$. Let g be a proper divisor of p , and let $q = p/g$. If s is a structure for w and $s \in \text{Fix}(g)$, then $\{i, j\} \in s$ implies $Y_r(\{i, j\}) \in s$, where $r = gt = n/q$. We can therefore write any pair in s as an internal pair $Y_{hr}(\{i, j\})$ or an external pair $Y_{hr}(\{i+r, j\})$ for some $h \in [0, q)$ and $0 \leq i < j < r$.

Let n, p, g, q , and r be as defined above. When considering structures in $\text{Fix}(g)$, we must consider two subsets: (i) structures containing no external pairs, and (ii) those with at least one external pair. In case (i), similar to the one shown in Figure 2A, it can be expressed as follows:

$$s = \bigcup_{h \in [0, q)} Y_{hr}(s_{[0, r)}),$$

where substructure $s_{[0, r)}$ is the subset of the pairs in s involving only sites in the indicated range. That is, s is determined by $s_{[0, r)}$. This implies that the number of structures in the first case is $C_{0, r}$.

In case (ii), as depicted in Figure 2B, there is a unique external pair $\{i+r, j\}$ ($0 \leq i < j < r$) such that any other external pair $\{k+r, l\}$ ($0 \leq k < l < r$) in s satisfies $k < i < j < l$. We call $\{i+r, j\}$ the *central pair* of s . The central pair $\{i+r, j\}$ implies that any pair $\{k, l\}$ with $k \in (i, j)$ satisfies $l \in (i, j)$, leading to the existence of the substructure $s_{(i, j)}$. Moreover, the external pair $\{i+r, j\}$ defines another substructure $s_{[j, i+r]}$. The union of these two substructures is another substructure $s_{(i, i+r]}$. Since $Y_r(s) = s$, we can express s as follows:

$$s = \bigcup_{h \in [0, q)} Y_{hr}(s_{(i, i+r]}).$$

In other words, the structures in $\text{Fix}(g)$ with the central pair $\{i+r, j\}$ are determined by $s_{(i, i+r]}$. The number of such structures is given by the product $C_{i+1, j} C_{j, i+r+1}^b$, where the two factors correspond to the numbers of the substructures $s_{(i, j)}$ and $s_{[j, i+r]}$, respectively. Since the central pair is unique, the subsets of $\text{Fix}(g)$ containing different central pairs do not intersect. We can therefore determine the number of structures in $\text{Fix}(g)$ containing at least one external pair to be $\sum_{0 \leq i < j < r} C_{i+1, j} C_{j, i+r+1}^b$. Combining the two subsets of $\text{Fix}(g)$, we can express its size as

$$|\text{Fix}(g)| = C_{0, r} + \sum_{0 \leq i < j < r} C_{i+1, j} C_{j, i+r+1}^b, \text{ where } r = gn/p. \tag{5}$$

Using Equation (5), we develop an algorithm to solve the Distinguishable Count problem. The pseudocode is shown in Algorithm 1. We state the main result of this work in the following theorem.

Theorem 1. *With a model of computation in which addition and multiplication take constant time, the Distinguishable Count problem can be solved in $\mathcal{O}(n^3)$ time and $\mathcal{O}(n^2)$ space.*

Algorithm 1: Counting distinguishable structures for a given circular sequence.

Input: Circular sequence w of length n and symmetry p .

Output: The size of Λ_w .

- 1: Compute $C_{i, j}$ and $C_{i, j}^b$, for $0 \leq i \leq j \leq n$.
- 2: $f(h) \leftarrow 0$ for all $h \in [0, p)$
- 3: $f(0) \leftarrow C_{0, n}$
- 4: **for** h from 1 to $p-1$ **do**
- 5: $g \leftarrow \text{gcd}(h, p)$

Algorithm 1: Counting distinguishable structures for a given circular sequence. (Continued)

```

6:   if  $g \neq h$  then
7:      $f(h) \leftarrow F(g)$ 
8:   else
9:      $r \leftarrow ng/p$ 
10:     $f(g) \leftarrow C_{0,r}$ 
11:    for  $i \leftarrow 0$  to  $r$  do
12:      for  $j \leftarrow i+1$  to  $r$  do
13:         $f(g) \leftarrow f(g) + C_{i+1,j} C_{j,i+r+1}^b$ 
14:   $\Lambda \leftarrow 0$ 
15:  for  $h$  from 0 to  $p-1$  do
16:     $\Lambda \leftarrow \Lambda + f(h)$ 
17:  return  $\Lambda/p$ 

```

Proof. We refer to Algorithm 1. A standard dynamic programming scheme to compute $C_{i,j}$ and $C_{i,j}^b$ takes $\mathcal{O}(n^3)$ time and $\mathcal{O}(n^2)$ space. Note that the remaining steps require only $\mathcal{O}(pn^2)$ time. Since $p \leq n$, the theorem holds. \square

5. APPLICATIONS OF THE ALGORITHM

Analyzing unlabeled sequences is relevant in studying the combinatoric characteristics of structures. Previous studies (Cuesta and Manrubia, 2017; Hofacker et al., 2012) involved enumeration of secondary structures for circular sequences. However, no study has been done on enumerating distinguishable structures.

Waterman (1978) derived a recursion relation for the number of structures compatible with an unlabeled sequence of length n . This removes the burden of ensuring that every pair forms a valid base pair, since this would hold for some sequence. Let $C(n)$ be the number of structures for a sequence of length n , that is, those satisfying conditions S1–S3 (excluding Condition S4). It can be recursively expressed as follows:

$$C(n) = C(n-1) + \sum_{k=0}^{n-2} C(k)C(n-k-2), \quad (6)$$

for $n > 0$ and with base case $C(0) = 1$. As Equation (6) suggests, $C(n)$ can be computed in $\mathcal{O}(n^2)$ time by computing $C(k)$ for $k \in [1, n]$ in increasing order.

We carry out the same analysis on distinguishable structures for unlabeled circular sequences. We define $\Lambda(n, p)$, for a divisor p of n , as the number of distinguishable structures for the unlabeled sequence of length n and symmetry p . Based on Equation (3), we have

$$\Lambda(n, p) = \frac{1}{p} \left(\sum_{h \in (0, p)} F(\gcd(h, p)n/p) + C(n) \right), \quad (7)$$

where $F(r)$, for $r > 0$, is the number of g -periodic structures for a circular sequence of length n and symmetry p such that $r = gn/p$. From Equation (5),

$$F(r) = C(r) + \sum_{k=0}^{r-2} (r-k-1)C(k)C(r-k-2). \quad (8)$$

We note that the case of $p=2$ is special as it allow the pair $\{i, i+r\}$, which cannot be a valid base pair [see Hofacker et al. (2012) for details]. We omitted the contribution from this case in our analysis, but it can be handled separately. We can compute $F(k)$, for $k \in [1, r]$, in $\mathcal{O}(r^2)$ time by dynamic programming. Therefore, $\Lambda(n, p)$ can be computed in $\mathcal{O}(n^2)$ time.

Dirks et al. (2007) defined a class of pseudoknot-free secondary structures for interacting sequences. In their model, m sequences are concatenated in a given fixed order and treated as a linear sequence along with

the locations of the concatenation points. Structures of interest must satisfy S1–S4 as well as a condition involving the concatenation points to ensure the individual sequences are “connected” [see Dirks et al. (2007) for details]. If the concatenated sequence has length n , we define $H_{i,j}$ and $H_{i,j}^b$ for that sequence analogously to the $C_{i,j}$ and $C_{i,j}^b$ we defined for a given circular sequence. A dynamic programming technique, similar to the one used in Dirks et al. (2007) for the partition function problem, can be used to compute $H_{i,j}$ and $H_{i,j}^b$ in $\mathcal{O}(n^3)$ time. The entry $H_{0,n}$ is then the number of structures for these concatenated “interacting” sequences.

A key observation by Dirks et al. (2007) is that, if the m sequences involve some that are identical, certain orderings of the sequences may have symmetry. Specifically, the symmetry of the ordering is the greatest divisor p of m such that the concatenated sequence does not change when cyclically shifting the ordering by m/p sequences. For example, given two sequences x and y , with $x \neq y$, the concatenated sequence $xyyxyxy$ has symmetry 3. With symmetry p , the structures for the concatenated sequence of length n are partitioned into orbits of indistinguishable structures under the action of $\mathbb{Z}/p\mathbb{Z}$ through rotation Y_{ht} where $t=n/p$ and $h \in \mathbb{Z}/p\mathbb{Z}$. This property is identical to that of a circular sequence of symmetry p . Therefore, Equation (3) applies to count the distinguishable structures (i.e., the orbits) for interacting sequences within this model. The only difference is in how the sizes of g -periodic subsets are computed:

$$|\text{Fix}(g)| = \sum_{0 \leq i < j \leq r} H_{i+1,j} H_{j,i+r+1}^b, \quad (9)$$

for proper divisor g of p , where $r=gn/p$. Equation (9) is obtained from Equation (5) by replacing $C_{i,j}$ ($C_{i,j}^b$) with $H_{i,j}$ ($H_{i,j}^b$). We omit the first term because structures with no external pairs would not be connected. Thus, in this model of interacting sequences, an adaptation of Algorithm 1 can compute the number of distinguishable structures in $\mathcal{O}(n^3)$ time.

6. CONCLUSIONS

In this work, we addressed the problem of counting distinguishable structures for circular RNA sequences, a problem that, to our knowledge, has not yet seen an efficient algorithm. We applied a result from group theory to identify useful subsets, the periodic subsets, of the structures. We then extended the algorithm of Hofacker et al. (2012), applying it to these periodic subsets, allowing us to compute their sizes. This leads to a cubic-time algorithm. We also showed that the developed approach could be applied to count distinguishable structures for unlabeled circular sequences as well as for multiple interacting sequences. While we focused on secondary structures without pseudoknots, the approach laid out here can also be applied to structures with pseudoknots.

Our claim of cubic time assumes that multiplication takes constant time, which is intertwined with restrictions on the magnitudes of numbers and the sizes of their representations. It is well known that for ordinary sequence, the number of secondary structures grows exponentially with sequence length (Waterman, 1978). Since the fastest current algorithm to multiply c -bit integers takes $\Theta(c \log c)$ time (Harvey and Van Der Hoeven, 2021), in a random access machine model, the time complexity would become $\mathcal{O}(n^4 \log n)$. The performance can be improved by acceleration techniques that apply to many secondary structure algorithms (Zakov et al., 2010). This detail regarding the assumed model of computation also emerges in the partition function problem, of which the counting problem is a variant. Algorithms for computing partition functions for RNA structures (Dirks et al., 2007; McCaskill, 1990) have cubic-time performance, and this requires a floating point arithmetic, which can lead to arbitrary numerical error.

Since Dirks et al. (2007) first raised the issue of distinguishability, it has not been obvious whether it is possible to efficiency account for the redundancy in standard dynamic programming algorithms for RNA secondary structures. In this study, we answered the above question in the affirmative. We note that another issue, raised by Dirks et al. (2007) regarding the minimum free energy problem in the presence of entropic energy correction, still remains. The challenge stems from accounting for such a global property as symmetry while using a dynamic programming scheme that operates on local problems. We believe that our strategy, building on the works by Dirks et al. (2007) and Hofacker et al. (2012), forms a framework in which one can address the issue regarding the symmetry in secondary structures.

ACKNOWLEDGMENTS

We thank Dr. Guilherme de Sena Brandine and Dr. Amal Thomas for constructive discussions.

AUTHORS' CONTRIBUTION

Both authors wrote, reviewed, and approved the final article.

AUTHOR DISCLOSURE STATEMENT

The authors declare they have no conflicting financial interests.

FUNDING INFORMATION

No funding was received for this article.

REFERENCES

- Akutsu T. Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *J Appl Math* 2000;104(1–3):45–62.
- Andronescu M, Zhang ZC, Condon A. Secondary structure prediction of interacting RNA molecules. *J Mol Biol* 2005;345(5):987–1001.
- Cuesta JA, Manrubia S. Enumerating secondary structures and structural moieties for circular RNAs. *J Theor Biol* 2017;419:375–382.
- Dimitrov RA, Zuker M. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys J* 2004;87(1):215–226.
- Dirks RM, Bois JS, Schaeffer JM, et al. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev* 2007;49(1):65–88.
- Dirks RM, Pierce NA. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comput Chem* 2003;24(13):1664–1677.
- Flores R, Delgado S, Gas M-E, et al. Viroids: The minimal non-coding RNAs with autonomous replication. *FEBS Lett* 2004;567(1):42–48.
- Gudima SO, Chang J, Taylor JM. Features affecting the ability of hepatitis delta virus RNAs to initiate RNA-directed RNA synthesis. *J Virol* 2004;78(11):5737–5744.
- Harvey D, Van Der Hoeven J. Integer multiplication in time $O(n \log n)$. *Ann Math* 2021;193(2):563–617.
- Hofacker IL, Reidys CM, Stadler PF. Symmetric circular matchings and RNA folding. *Discrete Math* 2012;312(1):100–112.
- Isaacs FJ, Dwyer DJ, Collins JJ. RNA synthetic biology. *Nat Biotechnol* 2006;24(5):545–554.
- Lyngsø RB, Zuker M, Pedersen C. Fast evaluation of internal loops in RNA secondary structure prediction. *Bioinformatics* 1999;15(6):440–445.
- Mathews DH. Revolutions in RNA secondary structure prediction. *J Mol Biol* 2006;359(3):526–532.
- Mathews DH, Sabina J, Zuker M, et al. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 1999;288(5):911–940.
- McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* 1999;29(6–7):1105–1119.
- Nussinov R, Jacobson AB. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci U S A* 1980;77(11):6309–6313.
- Rivas E, Eddy SR. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol* 1999;285(5):2053–2068.
- Sakai I. Syntax in universal translation. In: *Proceedings of International Conference on Machine Translation of Languages and Applied Language Analysis*. 1961.
- Waterman MS. Secondary structure of single-stranded nucleic acids. *Adv Math Suppl Studies* 1978;1:167–212.
- Zakov S, Tsur D, Ziv-Ukelson M. Reducing the worst case running times of a family of RNA and CFG problems, using Valiants approach. *Algo Bioinfo* 2011;6(1):1–22.

- Zuker M. On finding all suboptimal foldings of an RNA molecule. *Science* 1989;244(4900):48–52.
- Zuker M, Sankoff D. RNA secondary structures and their prediction. *Bull Math Biol* 1984;46(4):591–621.
- Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 1981;9(1):133–148.

Address correspondence to:

Dr. Andrew D. Smith

Department of Quantitative and Computational Biology

University of Southern California

Childs Way

Los Angeles, CA 90007

USA

E-mail: andrewds@usc.edu