COMPUTATIONAL MATHEMATICS

MSC 57M99 **DOI: 10.14529/jcem220201**

A GENERALISATION OF THE KAUFFMAN BRACKET POLYNOMIAL TO DETERMINE AND ANALYSE STRUCTURAL ELEMENTS IN A RNA SECONDARY STRUCTURE

A. A. Akimova, South Ural State University, Chelyabinsk, Russian Federation, akimovaaa@susu.ru

In this work, we use methods of knot theory to describe and analyze structural elements of a RNA secondary structure by construction of a new generalisation of the classical Kauffman bracket polynomial, which factorisation characterises these structural elements. To this end, we develop a mathematical model of RNA endowed with a topology invariant ($\mathcal{RKB}$ polynomial) that allows to determine a type, number and characteristics of standard structural elements that form the RNA secondary structure. In order to define the $\mathcal{RKB}$ polynomial, we introduce a new skein relation to smooth hydrogen bonds and a new technique to color nucleotides, and use the known skein relation of the Kauffman bracket polynomial. Note that the proposed technique to color nucleotides allows to take into account the positional relationship of structural elements, which can be used to investigate properties of RNA. Invariance of the $\mathcal{RKB}$ polynomial is shown. Computation of the $\mathcal{RKB}$ polynomial by the given Dot-Bracket notation is implemented as a small Mathematica program. Using $\mathcal{RKB}$ polynomials calculated by our program, we analyze some RNA secondary structures presented in the bpRNA-1m database. The obtained results agree with the real data.

Keywords: RNA; secondary structure; generalised Kauffman bracket polynomial; knot; invariant.

Dedicated to anniversary of Professor A.L. Shestakov

Introduction

The availability of huge amount of biological data opened a new direction in genomic analysis and structural prediction of DNA, RNA, and proteins in recent years. Many of these data, such as completely sequenced genomes, ribonucleic acids (RNAs), and proteins, led to an absolute demand for specialized tools to view, analyze, and predict the biological significance of the data. Throughout the last few decades, researchers pay significant attention to determining the RNA structure, since it is one of the key issues in understanding the genetic diseases and creating new drugs, which also helps the biologists to understand the role of the molecule in the cell [1–4].

RNA secondary structure prediction and classification are two important problems in the field of RNA biology. We pay attention to a part of the second one, while the first problem is widely developed (see, for example, [5, 6] for overview of RNA secondary structure prediction techniques) and is beyond the scope of our interests. Therefore, in this paper, we assume that the secondary structure is known in advance (for example, by its Dot-Bracket notation).

Behavior of RNA can be understood by finding patterns in its secondary structure in which it folds. This folding can be cumulative result of many different known and unknown biological, chemical, thermodynamic and mathematical parameters. In this work, we consider mathematical aspect of RNA secondary structures. There exist different types of techniques to classify RNA secondary structure data from mathematical point of view. Among them, we note the works that use permutations [5] and knot theory [7]. Note that knot theory has wide applications in many branches of sciences [8]. Its application in biology is explained by the existence of knotted structures in DNAs, proteins, and RNAs [9], see also [7] for perfect review of different applications of knot theory in biology.

In this work, we propose a new solution to the problem considered in [7]: use methods of knot theory to describe and analyze structural elements of a RNA secondary structure by construction of a new generalisation of the classical Kauffman bracket polynomial [10], which factorisation characterises these structural elements. To this end, we develop a mathematical model of RNA endowed with a topology invariant ($\mathcal{RKB}$ polynomial that is a new generalization of the Kauffman bracket polynomial) that allows to determine a type, number and characteristics of standard structural elements that form the RNA secondary structure. In order to define the $\mathcal{RKB}$ polynomial, we introduce a new skein relation to smooth hydrogen bonds and a new technique to color nucleotides, and use the known skein relation of the Kauffman bracket polynomial. Let us note that the proposed technique to color nucleotides allows to take into account the positional relationship of structural elements, which can be used in investigation of properties of RNA. We analyze the factorization of the proposed invariant and demonstrate that the factors of $\mathcal{RKB}$ polynomial reflect the structural features of RNAs. We show invariance of the $\mathcal{RKB}$ polynomial. Computation of the $\mathcal{RKB}$ polynomial by the given Dot-Bracket notation is implemented as a small Mathematica program. Using the $\mathcal{RKB}$ polynomials calculated by our program, we analyze some RNA secondary structures presented in the bpRNA-1m database [11]. The obtained results agree with the real data.

We consider application of the proposed generalization ($\mathcal{RKB}$ polynomial) of the Kauffman bracket polynomial in the case of RNA, however, the similar ideas can be used to investigate DNA and proteins.

The paper is organized as follows. Section 1 gives some required information on RNA as an object of study, in particular, on RNA secondary structure and its structural elements. In Section 2, we describe problem statement and present main result. Section 3 presents interpretation of RNA as a graph in $\mathbb{R}^3$ and its secondary structure as a diagram of such a graph on the plane. In Section 4, we define the $\mathcal{RKB}$ polynomial, which properties are discussed in Section 5. Finally, in Section 6, we present some computational examples: using the proposed invariant calculated by our program, we analyze some RNA secondary structures presented in the bpRNA-1m database.

1. Concept of RNA

RNA (Ribonucleic acid) is a polymeric molecule that belongs to one of the four major classes of biological macromolecules (carbohydrates, lipids, proteins, and nucleic acids), which is essential for all existing models of human life. Indeed, RNA molecules plays pivotal role in many biological functions: rebuild and transport genetic data [12], drive chemical reactions [13] and administer gene expressions [14]. Some RNA molecules play an active

role within cells by catalyzing biological reactions, or sensing and communicating responses to cellular signals. One of these active processes is protein synthesis, a universal function in which RNA molecules direct the synthesis of proteins on ribosomes. Many viruses encode their genetic information using an RNA genome. RNA molecule's capability to perform biomolecular computation through nanotechnology explains its importance among researchers from various fields [15], see also [5] for list of corresponding works.

RNA has three different types of structure: primary, secondary and tertiary ones by analogy with DNA and protein taking into account a dimension. Primary structure is a simple one dimensional sequence of nucleotides whereas secondary and tertiary structures are nothing but two dimensional and three dimensional representation of that sequence, respectively. Fig. 1 shows an example of primary and secondary structures of RNA [11].

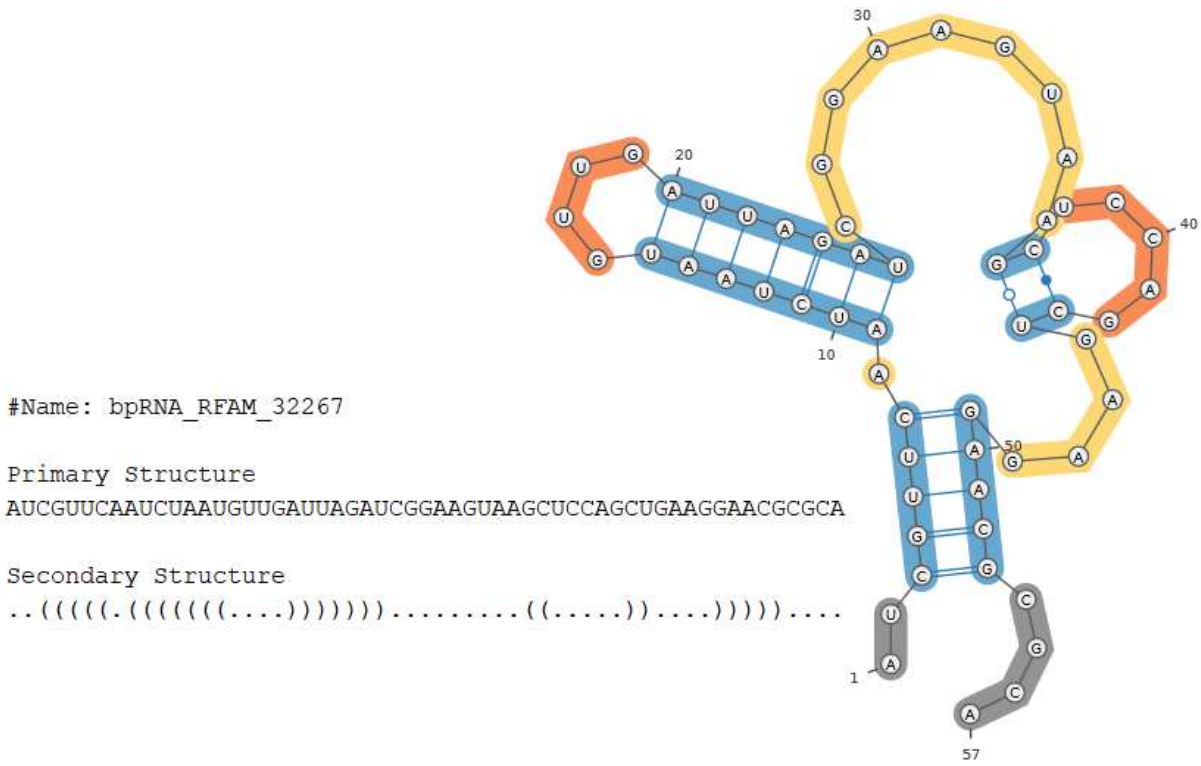


Fig. 1. Primary and secondary structures of RNA bpRNA_RFAM_32267 [11]

An RNA molecule can be considered as a chain (a random linear sequence) composed of four types of nucleotides, namely, adenine (A), uracil (U), cytosine (C), and guanine (G), with each nucleotide connected to its adjacent neighbor via a backbone. A nucleotide in one single strand RNA can pair through hydrogen bonds with another nucleotide, either from the same or from a different RNA molecule. These nucleotide pairs are called base pairs.

A consecutive group of base pairs forms a stem, which is the basic building block of RNA secondary structure. Here we note that a key difference between the topology of an RNA structure and that of a protein or a DNA duplex is the existence of RNA stems, which connect two regions of the RNA backbone(s) and fix their relative positions.

RNA stems can combine in different ways, giving rise to different secondary structures of RNA molecules. We consider seven recognized secondary structural elements in RNA: stem, hairpin loop, bulge loop, internal loop, multi-loop, single-stranded regions, pseudoknots. A schematic view of various structural elements is shown in Fig. 2. Any RNA secondary structure can be considered as a combination of these structural elements.

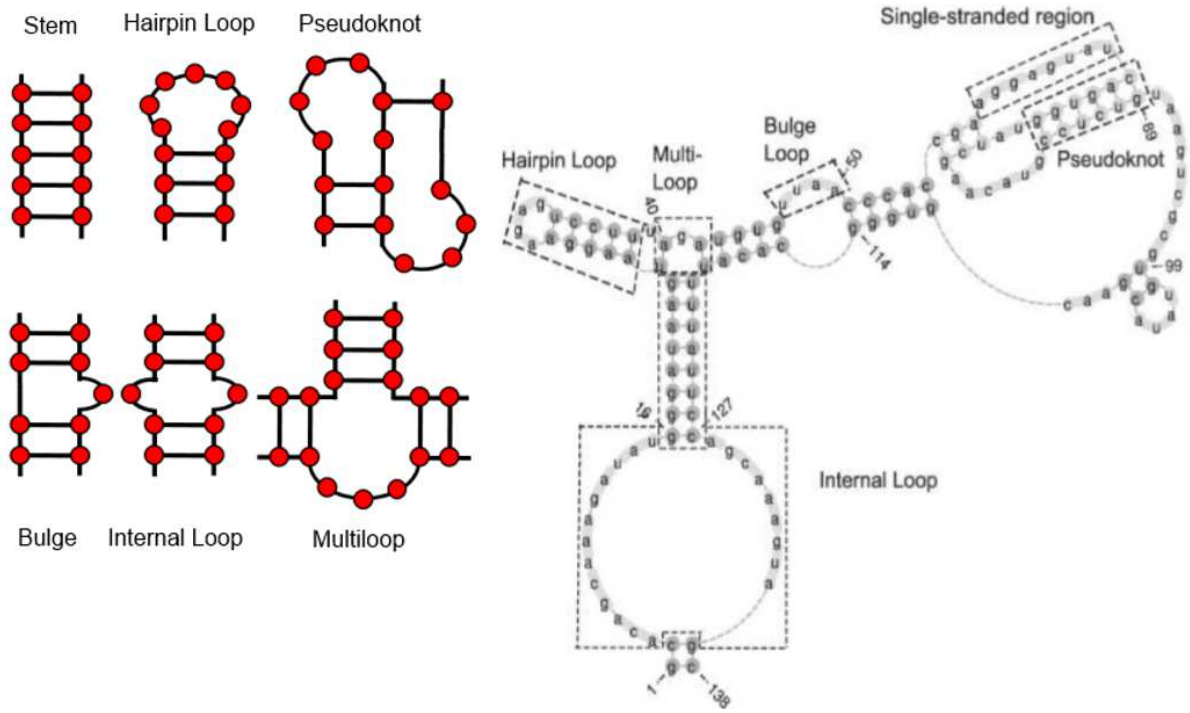


Fig. 2. Structural elements of secondary structure [5, 6]

Following [16], we consider the classification of pseudoknots per type provided in PseudoBase++ [17]: H-, HH-, HHH-, HL_out-, HL_in- and LL_in-type. Note that «H» means hairpin loop, «L» means bulge loop, «in» means internal loop or multiple internal loops, and «out» means external loop or multiple external loops. Fig. 3 shows all the six pseudoknot types [16].

For our analysis purpose, we use the bpRNA-1m database [11] (see [18] for the corresponding paper and the work [5] for list of other databases), which consists of 102,318 RNA Secondary Structures from 7 different sources. For each RNA, the bpseq file, fasta file, dot bracket file, and structure type file are provided. Moreover, the pdf file of each RNA secondary structure is available. All of the files are available to download.

Different notations are used to represent secondary structures: Dot-Bracket notation, String notation, Linked Graph notation, Circular notation, Dot Plot notation, Mountain Plot notation, Mountain Metric notation, Tree notation. In this work, we use the Dot-Bracket notation, which is just a sequence of dots and brackets in which a dot represents an unpaired nucleotide and a bracket represents a paired nucleotide, where matching brackets symbolize base pairs. In more difficult secondary structures, a more generalized version of the original Dot-Bracket notation may use additional pairs of brackets, such as <>, {}, and [], and matching pairs of uppercase/lowercase letters, etc. For more details on formats of files to represent secondary structures, see the overview included in [5].

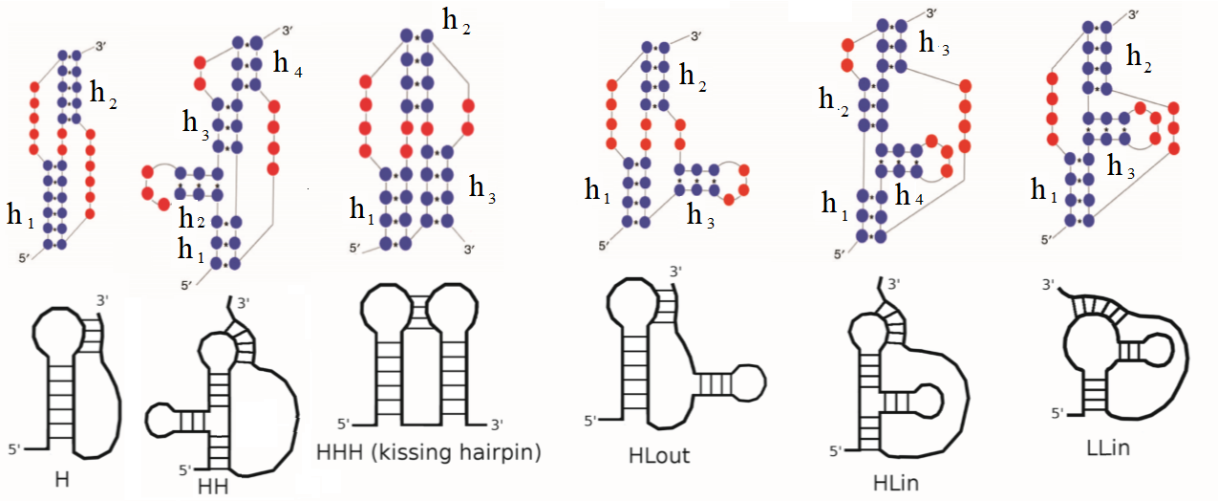


Fig. 3. RNA pseudoknot types by PseudoBase++ [16] classification: H-, HH-, HHH-, HL_out-, HL_in- and LL_in-types, see also [19] (h_i denotes number of hydrogen bonds in the i-th stem)

2. Problem Statement & Main Result

The problem is to construct a mathematical model of a RNA secondary structure to describe and analyze its structural elements. To this end, following [7], we use methods of knot theory, namely, we construct a new generalisation of the classical Kauffman bracket polynomial [10], which factorisation characterises the desired structural elements.

The following steps describe the realized construction of a mathematical model of RNA endowed with a topology invariant ($\mathcal{RKB}$ polynomial that is a new generalization of the Kauffman bracket polynomial).

1. Represent RNA as a mathematical object (a graph in $\mathbb{R}^3$, where one of edges is divided in half).
2. Represent RNA secondary structure as a mathematical object (a diagram on the plane).
3. Define the $\mathcal{RKB}$ polynomial by introducing a new skein relation to smooth hydrogen bonds and a new technique to color nucleotides, and using the known skein relation of the Kauffman bracket polynomial.
4. Show invariance of the $\mathcal{RKB}$ polynomial.
5. Analyze the factorization of the proposed invariant and demonstrate that the factors of $\mathcal{RKB}$ polynomial reflect the structural features of RNA.
6. Write a small Mathematica program that computes the $\mathcal{RKB}$ polynomial by the given Dot-Bracket notation of a RNA secondary structure.
7. Using the $\mathcal{RKB}$ polynomials calculated by the program, analyze some RNA secondary structures presented in the bpRNA-1m database [11] to show that the obtained results agree with the real data.

3. RNA as Graph

In this paper, from mathematical point of view, we consider RNA to be a graph $G \subset \mathbb{R}^3$, where each vertex is given by either an unpaired nucleotide  or a paired nucleotide  , and edges are given by connections between the nucleotides, which can be of two types (a backbone (blue) that consecutively goes through each nucleotide or a hydrogen bond (red) that connects exactly 2 paired nucleotides), and one of edges is divided in half (that corresponds to ends of RNA backbone).

As a result of projection of the graph $G \subset \mathbb{R}^3$ onto the plane (performed following usual rules to construct a 2-dimensional projection of a 3-dimensional knot), we obtain a diagram D with the vertices of the form  or  described above and 3 new types of classical crossings: both parts of a strand are of the same type (either backbone or hydrogen bond) or of different types (backbone and hydrogen bond). Therefore, we consider the secondary structure of a RNA to be such a diagram D.

Following [7], we close off RNA strand into a circle using the virtual closure approach: the backbone trace is closed off by connecting its ends virtually (i.e., self-intersections appeared as a result of this closure are 4-valent vertices called virtual crossings and denoted by  , see the original paper [20] for more details about virtual knot theory). Note that, since  appears as a result of virtual closure of a backbone, at least one of two parts of its strand is backbone, hence we obtain 2 types of virtual crossings only.

All 5 types of crossings described above (3 classical and 2 virtual ones) are subjected with a set of local moves, which are the obvious generalization of the Reidemeister moves $\Omega_1 - \Omega_3$ for classical crossings [21], their virtual $\Omega_1' - \Omega_3'$ and semivirtual Ω_3'' versions [20], see Fig. 4. As regards to the generalization of the Reidemeister moves for the case of 3-valent vertices of the form  , we recall that the work [22] proposes the extra vertex moves Ω_4, Ω_5 that are sufficient to generate isotopy between graphs, see Fig. 4. Together, these basic moves provide the ambient isotopy for RNA diagrams defined as plane graphs of secondary structures.

4. Definition of $\mathcal{RKB}$ Polynomial

We develop a new generalization ($\mathcal{RKB}$ polynomial) of the classical Kauffman bracket polynomial [10] specifically designed to study RNA secondary structure. Following [7], we use ideas of Kauffman bracket polynomial, but do not compress stems into rigid vertices. On the contrary, we propose one more skein relation specifically for hydrogen bonds. In more details, the work [7] ignores the length of stems and treat stems of different sizes equally, moreover, to be a rigid vertex, while we go inside each stem and take into account both number of hydrogen bonds that form the stem and a position of the stem with respect to backbone (for the latter, we introduce a new technique to color nucleotides). As a result, our $\mathcal{RKB}$ polynomial considers colored stems as complete description of structural elements of RNA secondary structure and can be factored such that to associate each factor with a structural element of RNA secondary structure described in Section 1.

Let us define the $\mathcal{RKB}$ polynomial $\{\cdot\}$, where $\mathcal{RKB}$ stands for *R*NA *K*auffman *B*racket.

Step 1 (Color Nucleotides). Let us associate each nucleotide with an ordinal color c_i,

$i = 1, 2, ...$, as follows: we go through a primary structure (backbone), color a sequence of unpaired nucleotides (i.e., single-stranded region) and the subsequent sequence of paired nucleotides by the same color and use next ordinal color before to color each single-stranded region. The following algorithm describes implementation of this idea in more details.

1. Choose any of two possible orientations of the backbone that form RNA.

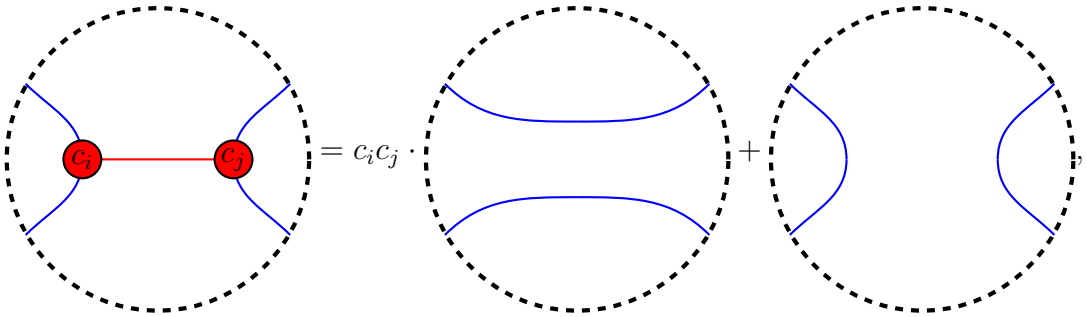
2. Color all the nucleotides in the 1-st sequence of paired nucleotides of the backbone by the same color c_1 (if there exists a single-stranded region before these sequence, then color its nucleotides by c_1 as well).

3. For each new single-stranded region, take next ordinal color.

At this moment, we omit all unpaired nucleotides.

Two colors c_i and c_{i+1} are said to be adjacent that is their ordinal numbers differ by 1. Two colors c_i and c_{i+2} are called quasiadjacent that is their ordinal numbers differ by 2.

Step 2 (Define Skein Relations). Let us define a skein relation to smooth a hydrogen bond as follows:



(1)

i.e. associate the hydrogen bond with the sum of the result of smoothing along the hydrogen

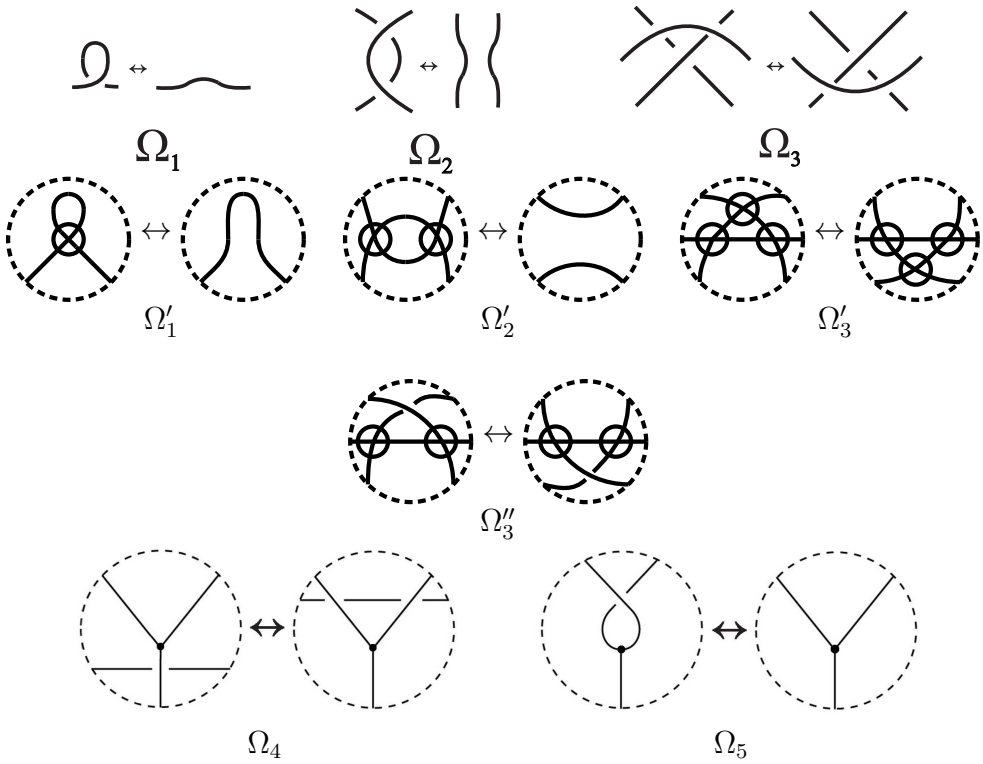


Fig. 4. Reidemeister moves: classical Ω_1 – Ω_3; virtual Ω'_1 – Ω'_3; semivirtual Ω''_3; extra Ω_4, Ω_5

bond multiplied by the product $c_i c_j$ and the result of smoothing along the backbone. Therefore, each hydrogen bond is transformed to the weighted summation of a pair of double parts of empty strains. After applying this transformation to all hydrogen bonds of an RNA secondary structure diagram recursively, we obtain a summation of a group of weighted diagrams without hydrogen bonds.

Any knot invariant, including knot polynomials, can be used to characterize resulting diagrams. Here we use the Kauffman bracket polynomial [10] as follows. Let D be a diagram of a RNA secondary structure. Endow each angle of each crossing of D with a marker A or B according to the rule given in the center of Fig. 5. Each state s of the diagram D is defined by a combination of ways to smooth each crossing of D such as to join together either two angles endowed with a marker A, or two angles endowed with a marker B, see Fig. 5 on the left and right, respectively. Obviously, if the diagram D has n crossings, then there exist exactly 2^n states of D.

Fig. 5. A- and B-smoothings of a classical crossing

By the writhe of an oriented knot diagram D with n crossings we mean the sum over all crossings of D

$$w(D) = \sum_{i=1}^{n} \varepsilon(i),$$

where $\varepsilon(i)$ is a sign of the i-th crossing of D defined by the rules given in Fig. 6.

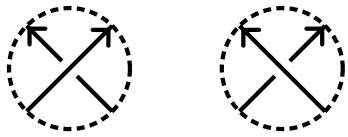


Fig. 6. Rules to define the sign $\varepsilon(i)$ of the i-th classical crossing

The exact formula of the Kauffman bracket polynomial is as follows:

$$\mathcal{X}\left(A, x\right)_D = (-A)^{-3w(D)} \left\langle A, x \right\rangle_D, \tag{2}$$

where

$$\left\langle A, x \right\rangle_D = \sum_s A^{\alpha(s)-\beta(s)} x^{\gamma(s)-1} \tag{3}$$

is the Kauffman bracket. Here $\alpha(s)$ and $\beta(s)$ are the numbers of markers A and B in the given state s, while $\gamma(s)$ is the number of curves obtained by smoothing of all crossings according to the state s (for shortness, here we denote $x = -A^2 - A^{-2}$), and $w(D)$ is the writhe of D. The sum is taken over all 2^n states of D.

Therefore, each of 3 types of classical crossings is smoothed by usual skein relation of Kauffman bracket polynomial, while virtual crossings  do not need skein relation. In

addition, we use usual normalization (consider polynomial of one of circles in each state to be 1 and polynomial of each of the rest circles to be $x = -A^2 - A^{-2}$). Finally, as usual, to obtain invariant under the Reidemeister move Ω_1, we use the writhe of the diagram.

5. Properties of $\mathcal{RKB}$ polynomial

Lemma 1. *[Invariance of $\mathcal{RKB}$ Polynomial] The $\mathcal{RKB}$ polynomial is invariant under all three Reidemeister moves Ω_1 – Ω_3 generalised on 3 types of classical crossings, both virtual Reidemeister moves Ω'_1 – Ω'_3 and semivirtual Reidemeister move Ω''_3 generalised on 2 types of virtual crossings, both extra vertex Reidemeister moves Ω_4, Ω_5 generalised on 2 types of edges (backbone and hydrogen bond), see Fig. 4.*

Proof of Lemma 1 is similar to proof of the corresponding theorem on Kauffman bracket polynomial of knots for Reidemeister moves Ω_1 – Ω_3 with classical crossings only (see, for example, [10] or [23]) and for virtual Reidemeister moves Ω'_1 – Ω'_3 and semivirtual Reidemeister move Ω''_3 with both virtual and classical crossings (see, for example, [20]).

Lemma 2. *[Factorisation of $\mathcal{RKB}$ Polynomial] The $\mathcal{RKB}$ polynomial of a RNA secondary structure with the unique backbone can be represented as a product of polynomials of stems and pseudoknots that form the RNA secondary structure.*

Proof of Lemma 2 is similar to proof of theorem on Kauffman bracket polynomial of a connected sum of knots (see, for example, [23]).

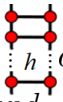
Remark 1. [Stronger $\mathcal{RKB}$ Polynomial] The $\mathcal{RKB}$ polynomial can be defined to be more stronger. To this end, we color a nucleotide associated with an opened bracket by c_i, but a nucleotide associated with a closed bracket by d_j. However, for our proposes the original definition (both types are colored by c, i.e. without using a new variable d) turns out to be enough.

Below we present connection between the $\mathcal{RKB}$ polynomial and structural elements of RNA secondary structure described in Section 1.

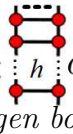
Lemma 3. *[Number of single-stranded regions] Let n_r be a number of single-stranded regions in a RNA secondary structure, n_b be a number of backbones that begin with a paired nucleotide, n_e be a number of backbones that end with an unpaired nucleotide, n_c be a number of colors in the $\mathcal{RKB}$ polynomial. Then*

$$n_r = n_c - n_b + n_e.$$

Proof of Lemma 3 is obvious by definition of $\mathcal{RKB}$ polynomial (Step 1 (Color Nucleotides)): every time then a new single-stranded region begins, take next ordinal color.

Lemma 4. *[$\mathcal{RKB}$ Polynomial of Stem] Let $\mathcal{S}^{(h)}_{(i,j)}$ be a stem with h hydrogen bonds formed by paired nucleotides colored by c_i and c_j, i.e.* 

. Denote by n_k and $\widehat{n_k}$ a pair of paired nucleotides that form the k-th hydrogen bond, $k = 1, 2, \ldots$. Assume that a dashed line that connects one of two pairs of opposite ends of the backbone of the stem $\mathcal{S}^{(h)}_{(i,j)}$,

i.e.  *cuts off a part* P *(without the stem* $\mathcal{S}^{(h)}_{(i,j)}$*) of RNA that contains only full hydrogen bonds, i.e. if* $n_k \in P$*, then* $\widehat{n_k} \in P$ *for all* $k = 1, 2, \ldots$*. Then the* $\mathcal{RKB}$ *polynomial of the stem* $\mathcal{S}^{(h)}_{(i,j)}$ *is*

$$\left\{\mathcal{S}^{(h)}_{(i,j)}\right\} = (1 + c_i c_j x)^h,$$

where $x = -A^2 - A^{-2}$ *is a variable associated with a circle without crossings.*

Proof. As a result of smoothing of all hydrogen bonds included in P by the skein relation (1), we obtain a set of states, in each of which the dashed line remains as a part of strand. This is due to the fact that P contains only full hydrogen bonds.

The last part of the proof can be performed by mathematical induction. Here we present only the base of the induction, i.e. calculations in the case of a stem formed by the unique hydrogen bond: c_i  c_j.

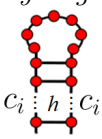
Following (1), we smooth the hydrogen bond that form the stem as follows:

$\{c_i$ $c_j\} = c_i c_j \left\{$  $\right\} + \left\{$  $\right\}$.

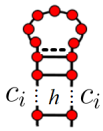
Taking into account that $\mathcal{RKB}$ polynomial of a circle without crossings is x, we obtain $(c_i c_j x + 1) \cdot \left\{$  $\right\}$.

□

Based on two colors used in a $\mathcal{RKB}$ polynomial of a stem, we can draw conclusion about structural element formed using this stem.

Lemma 5. *[*$\mathcal{RKB}$ *Polynomial of Hairpin Loop] Let* $\mathcal{L}^{(h)}_{\mathcal{H}_{(i,i+1)}}$ *be a hairpin loop with* h *hydrogen bonds formed by paired nucleotides colored by adjacent colors* c_i *and* c_{i+1}*, i.e.*  *Then*

$$\left\{\mathcal{L}^{(h)}_{\mathcal{H}_{(i,i+1)}}\right\} = \left\{\mathcal{S}^{(h)}_{(i,i+1)}\right\}.$$

Proof of Lemma 5 follows immediately from Lemma 4, since P is given by a single region:  . Also, proof can be performed by mathematical induction similar to proof of Lemma 4.

□

Lemma 6. *[*$\mathcal{RKB}$ *Polynomial of Bulge Loop] Let* $\mathcal{L}^{(h_1)(h_2)}_{\mathcal{B}_{(i,j)(i,j+1)}}$ *be a bulge loop with* h_1 *hydrogen bonds formed by paired nucleotides colored by* c_i *and* c_j *and* h_2 *hydrogen bonds formed by paired nucleotides colored by* c_i *and* c_{j+1}*. Then*

$$\left\{\mathcal{L}^{(h_1)(h_2)}_{\mathcal{B}_{(i,j)(i,j+1)}}\right\} = \left\{\mathcal{S}^{(h_1)}_{(i,j)}\right\} \cdot \left\{\mathcal{S}^{(h_2)}_{(i,j+1)}\right\}.$$

Moreover, $\mathcal{RKB}$ *polynomial of total RNA secondary structure contains no factor* $(1 + c_j c_{j+1} x)$*, since the RNA secondary structure contains no stem colored by the adjacent colors* c_j *and* c_{j+1}*.*

Proof of Lemma 6 can be performed by mathematical induction similar to proof of Lemma 4.

Lemma 7. *[$\mathcal{RKB}$ Polynomial of Internal Loop] Let $\mathcal{L}_{\mathcal{I}(i,j+1)(i+1,j)}^{(h_1)(h_2)}$ be an internal loop with h_1 hydrogen bonds formed by paired nucleotides colored by c_i and c_{j+1} and h_2 hydrogen bonds formed by paired nucleotides colored by c_{i+1} and c_j. Then*

$$\left\{\mathcal{L}_{\mathcal{I}(i,j+1)(i+1,j)}^{(h_1)(h_2)}\right\} = \left\{\mathcal{S}_{(i,j+1)}^{(h_1)}\right\} \cdot \left\{\mathcal{S}_{(i+1,j)}^{(h_2)}\right\}.$$

Moreover, $\mathcal{RKB}$ polynomial of total RNA secondary structure contains no factors $(1 + c_i c_{i+1} x)$ and $(1 + c_j c_{j+1} x)$, since the RNA secondary structure contains no stems colored by the adjacent colors c_i and c_{i+1} or the adjacent colors c_j and c_{j+1}.

Proof of Lemma 7 can be performed by mathematical induction similar to proof of Lemma 4.

Lemma 8. *[$\mathcal{RKB}$ Polynomial of Multi- Loop] Let $\mathcal{L}_{\mathcal{M}(j,i)(k,i)(j+1,k)}^{(h_1)(h_2)(h_3)}$ be a multi-loop with h_1 hydrogen bonds formed by paired nucleotides colored by c_j and c_i, h_2 hydrogen bonds formed by paired nucleotides colored by c_k and c_i, h_3 hydrogen bonds formed by paired nucleotides colored by c_{j+1} and c_k. Then*

$$\left\{\mathcal{L}_{\mathcal{M}(j,i)(k,i)(j+1,k)}^{(h_1)(h_2)(h_3)}\right\} = \left\{\mathcal{S}_{(j,i)}^{(h_1)}\right\} \cdot \left\{\mathcal{S}_{(k,i)}^{(h_2)}\right\} \cdot \left\{\mathcal{S}_{(j+1,k)}^{(h_3)}\right\}.$$

Proof of Lemma 8 can be performed by mathematical induction similar to proof of Lemma 4.

Following [16], below we consider the classification of pseudoknots per type provided in PseudoBase++ [17]: *H*-, *HH*-, *HHH*-, HL_{out}-, *HL_in*- and *LL_in*-type, see Fig. 3. However, the $\mathcal{RKB}$ polynomial of any pseudoknot having another type can be constructed in the similar way.

Lemma 9. *[$\mathcal{RKB}$ Polynomial of Pseudoknot] Let $x = -A^2 - A^{-2}$ be a variable associated with a circle without crossings. Enumerate stems of a pseudoknot. Let h_i be a number of hydrogen bonds in the i-th stem, see Fig. 3.*

1. Let $\mathcal{P}_{\mathcal{H}(i,i+2)(i+1,i+3)}^{(h_1)(h_2)}$ be a H-type pseudoknot, stems of which are consecutively colored by (c_i, c_{i+2}) and (c_{i+1}, c_{i+3}). Then $\mathcal{RKB}$ polynomial of the pseudoknot $\mathcal{P}_{\mathcal{H}(i,i+2)(i+1,i+3)}^{(h_1)(h_2)}$ can be found recursively as follows:

$$\left\{\mathcal{P}_{\mathcal{H}(i,i+2)(i+1,i+3)}^{(h_1)(0)}\right\} = \left\{\mathcal{S}_{(i,i+2)}^{(h_1)}\right\};$$

$$\left\{\mathcal{P}_{\mathcal{H}(i,i+2)(i+1,i+3)}^{(h_1)(h_2)}\right\} = \left\{\mathcal{P}_{\mathcal{H}(i,i+2)(i+1,i+3)}^{(h_1)(h_2-1)}\right\} +$$

$$+c_{i+1}c_{i+3}\left\{\mathcal{S}_{(i+1,i+3)}^{(h_2-1)}\right\}\left(x + c_i c_{i+2}\sum_{j=0}^{h_1-1}\left\{\mathcal{S}_{(i,i+2)}^{(j)}\right\}\right).$$

2. Let $\mathcal{P}_{\mathcal{HH}(i,i+2)(i,i+1)(i+1,i+2)(i+2,i+3)}^{(h_1)(h_2)(h_3)(h_4)}$ be a HH-type pseudoknot, stems of which are consecutively colored by (c_i, c_{i+2}), (c_i, c_{i+1}), (c_{i+1}, c_{i+2}) and (c_{i+2}, c_{i+3}). Then

$\mathcal{RKB}$ polynomial of a pseudoknot $\mathcal{P}_{\mathcal{HH}}{}^{(h_1)(h_2)(h_3)(h_4)}_{(i,i+2)(i,i+1)(i+1,i+2)(i+2,i+3)}$ can be found recursively as follows:

$$\left\{\mathcal{P}_{\mathcal{HH}}{}^{(h_1)(0)(h_3)(0)}_{(i,i+2)(i,i+1)(i+1,i+2)(i+2,i+3)}\right\} = \left\{\mathcal{S}^{(h_1)}_{(i,i+2)}\right\} \cdot \left\{\mathcal{S}^{(h_3)}_{(i+1,i+2)}\right\};$$

$$\left\{\mathcal{P}_{\mathcal{HH}}{}^{(h_1)(h_2)(h_3)(h_4)}_{(i,i+2)(i,i+1)(i+1,i+2)(i+2,i+3)}\right\} =$$

$$\left\{\mathcal{S}^{(h_2)}_{(i,i+1)}\right\} \cdot \left(\left\{\mathcal{P}_{\mathcal{HH}}{}^{(h_1)(0)(h_3)(h_4-1)}_{(i,i+2)(i,i+1)(i+1,i+2)(i+2,i+3)}\right\} +$$

$$+c_{i+2}c_{i+3}\left\{\mathcal{S}^{(h_4-1)}_{(i+2,i+3)}\right\} \cdot \left(c_i c_{i+2} \cdot \left\{\mathcal{S}^{(h_3)}_{(i+1,i+2)}\right\} \cdot \sum_{j=0}^{h_1-1}\left\{\mathcal{S}^{(j)}_{(i,i+2)}\right\} +\right.$$

$$\left. +x + c_{i+1}c_{i+2}\sum_{j=0}^{h_3-1}\left\{\mathcal{S}^{(j)}_{(i+1,i+2)}\right\}\right)\right).$$

3. Let $\mathcal{P}_{\mathcal{HHH}}{}^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+3)(i+2,i+4)}$ be a HHH-type pseudoknot, stems of which are consecutively colored by (c_i, c_{i+2}), (c_{i+1}, c_{i+3}) and (c_{i+2}, c_{i+4}). Then $\mathcal{RKB}$ polynomial of a pseudoknot $\mathcal{P}_{\mathcal{HHH}}{}^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+3)(i+2,i+4)}$ can be found recursively as follows:

$$\left\{\mathcal{P}_{\mathcal{HHH}}{}^{(h_1)(h_2)(0)}_{(i,i+2)(i+1,i+3)(i+2,i+4)}\right\} = \left\{\mathcal{P}_{\mathcal{H}}{}^{(h_1)(h_2)}_{(i,i+2)(i+1,i+3)}\right\};$$

$$\left\{\mathcal{P}_{\mathcal{HHH}}{}^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+3)(i+2,i+4)}\right\} = \left\{\mathcal{P}_{\mathcal{HHH}}{}^{(h_1)(h_2)(h_3-1)}_{(i,i+2)(i+1,i+3)(i+2,i+4)}\right\} +$$

$$+c_{i+2}c_{i+4}\left\{\mathcal{S}^{(h_1-1)}_{(i,i+2)}\right\}\left\{\mathcal{S}^{(h_4-1)}_{(i+2,i+4)}\right\}\left(x + c_{i+1}c_{i+3}\sum_{j=0}^{h_2-1}\left\{\mathcal{S}^{(j)}_{(i+1,i+3)}\right\}\right).$$

4. Let $\mathcal{P}_{\mathcal{HL}\_out}{}^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+4)(i+2,i+3)}$ be a HL_out-type pseudoknot, stems of which are consecutively colored by (c_i, c_{i+2}), (c_{i+1}, c_{i+4}) and (c_{i+2}, c_{i+3}). Then $\mathcal{RKB}$ polynomial of a pseudoknot $\mathcal{P}_{\mathcal{HL}\_out}{}^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+4)(i+2,i+3)}$ can be found as follows:

$$\left\{\mathcal{P}_{\mathcal{HL}\_out}{}^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+4)(i+2,i+3)}\right\} = \left\{\mathcal{P}_{\mathcal{H}}{}^{(h_1)(h_2)}_{(i,i+2)(i+1,i+4)}\right\} \cdot \left\{\mathcal{S}^{(h_3)}_{(i+2,i+3)}\right\}.$$

5. Let $\mathcal{P}_{\mathcal{HL}\_in}{}^{(h_1)(h_2)(h_3)(h_4)}_{(i,i+2)(i,i+1)(i+1,i+3)(i+1,i+2)}$ be a HL_in-type pseudoknot, stems of which are consecutively colored by (c_i, c_{i+2}), (c_i, c_{i+1}), (c_{i+1}, c_{i+3}) and (c_{i+1}, c_{i+2}). Then $\mathcal{RKB}$ polynomial of a pseudoknot $\mathcal{P}_{\mathcal{HL}\_in}{}^{(h_1)(h_2)(h_3)(h_4)}_{(i,i+2)(i,i+1)(i+1,i+3)(i+1,i+2)}$ can be found recursively as follows:

$$\left\{\mathcal{P}_{\mathcal{HL}\_in}{}^{(h_1)(h_2)(0)(0)}_{(i,i+2)(i,i+1)(i+1,i+3)(i+1,i+2)}\right\} = \left\{\mathcal{S}^{(h_1)}_{(i,i+2)}\right\} \cdot \left\{\mathcal{S}^{(h_2)}_{(i,i+1)}\right\};$$

$$\left\{\mathcal{P}_{\mathcal{HL}\_in}{}^{(h_1)(h_2)(h_3)(h_4)}_{(i,i+2)(i,i+1)(i+1,i+3)(i+1,i+2)}\right\} = \left\{\mathcal{S}^{(h_4)}_{(i+1,i+2)}\right\} \times$$

$$\times \left(\left\{\mathcal{P}_{\mathcal{HL}\_in}{}^{(h_1)(h_2)(h_3-1)(0)}_{(i,i+2)(i,i+1)(i+1,i+3)(i+1,i+2)}\right\} + c_{i+1}c_{i+3}\left\{\mathcal{S}^{(h_3-1)}_{(i+1,i+3)}\right\} \times\right.$$

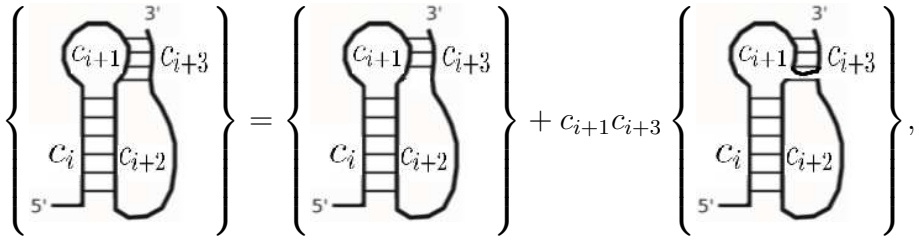
$$\left. \times \left(c_i c_{i+2}\left\{\mathcal{S}^{(h_2)}_{(i,i+1)}\right\} \cdot \sum_{j=0}^{h_1-1}\left\{\mathcal{S}^{(j)}_{(i,i+2)}\right\} + x + c_i c_{i+1}\sum_{j=0}^{h_2-1}\left\{\mathcal{S}^{(j)}_{(i,i+1)}\right\}\right)\right).$$

6. Let $\mathcal{P}_{\mathcal{LL}\_in^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+3)(i+1,i+2)}}$ *be a HL_in-type pseudoknot, stems of which are consecutively colored by* (c_i, c_{i+2}), (c_{i+1}, c_{i+3}) *and* (c_{i+1}, c_{i+2}). *Then* $\mathcal{RKB}$ *polynomial of a pseudoknot* $\mathcal{P}_{\mathcal{LL}\_in^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+3)(i+1,i+2)}}$ *can be found as follows:*

$$\left\{\mathcal{P}_{\mathcal{LL}\_in^{(h_1)(h_2)(h_3)}_{(i,i+2)(i+1,i+3)(i+1,i+2)}}\right\} = \left\{\mathcal{P}_{\mathcal{H}^{(h_1)(h_2)}_{(i,i+2)(i+1,i+3)}}\right\} \cdot \left\{\mathcal{S}^{(h_3)}_{(i+1,i+2)}\right\}.$$

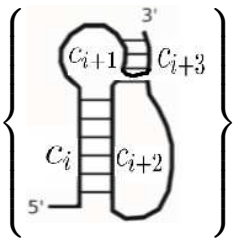
Proof. Let us show an idea of the proof of Lemma 9 on the example of the H-type pseudoknot.

We begin smooth the hydrogen bonds that form the last (second) stem as follows:

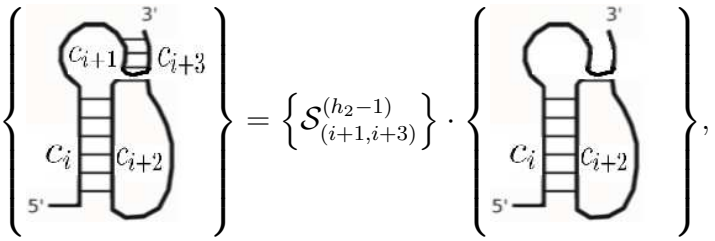


i.e.

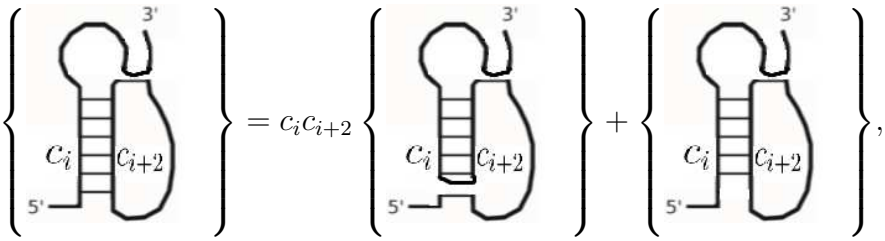
$$\left\{\mathcal{P}_{\mathcal{H}^{(h_1)(h_2)}_{(i,i+2)(i+1,i+3)}}\right\} = \left\{\mathcal{P}_{\mathcal{H}^{(h_1)(h_2-1)}_{(i,i+2)(i+1,i+3)}}\right\} + c_{i+1}c_{i+3}$$



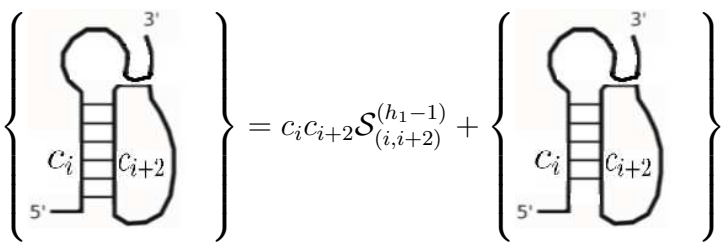
As regards to the last factor,



and, finally, we smooth the hydrogen bonds that form the first stem as follows:



i.e.



□

6. Computational Examples

Let us present some computational examples. Using a program written by means of Mathematica package, we calculate the $\mathcal{RKB}$ polynomials of some RNA presented in the database [11]. Here we consider examples without classical and virtual crossings in order to pay attention to the new skein relation and technique of color.

Example 1 (bpRNA_RFAM_14276). The Dot-Bracket notation is

$$\text{.....(|(((....((((((.....))))))..))).)....., }$$

where . stands for an unpaired nucleotide, (and) denote a paired nucleotide, and | is used to divide sequences of paired nucleotides that form different stems. The secondary structure is presented in Fig. 5 on the left.

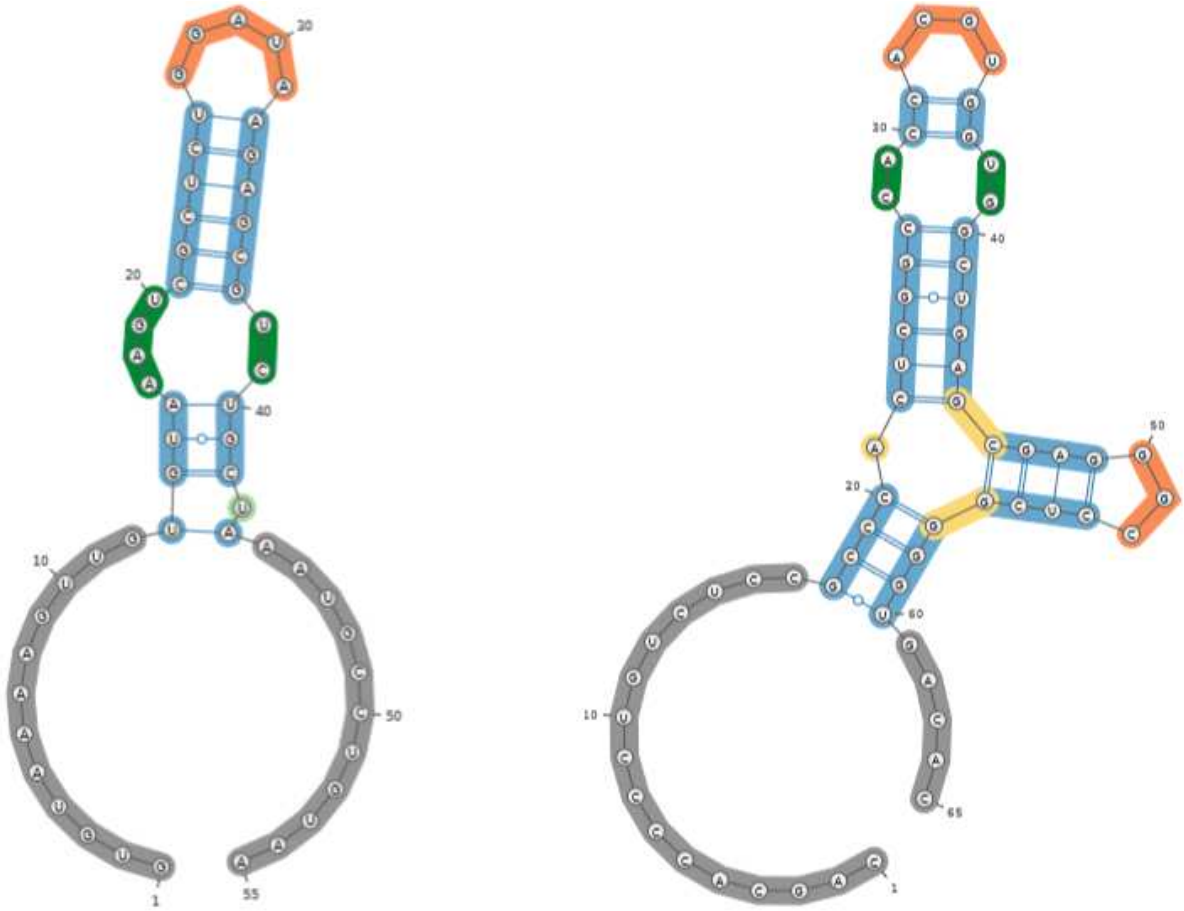


Fig. 7. Secondary structure of RNA bpRNA_RFAM_14276 (left) and bpRNA_RFAM_38844 (right) [11]

We color nucleotides involved in the given Dot-Bracket notation as follows:

$$\underbrace{\text{.....(|(((}}_{c_1}\underbrace{\text{....((((((}}_{c_2}\underbrace{\text{.....))))))}}_{c_3}\underbrace{\text{..)))}}_{c_4}\underbrace{\text{.)}}_{c_5}\underbrace{\text{.....}}_{c_6}.$$

Based on the factored $\mathcal{RKB}$ polynomial

$$(1 + c_2c_3x)^6 \cdot (1 + c_1c_4x)^3 \cdot (1 + c_1c_5x),$$

we draw the following conclusions about the structural elements, which agree with Fig. 5.

Single-stranded region. Following Lemma 3, since number of colors is 5 and both the first and the last nucleotides are unpaired, we conclude that the number of single-stranded regions is 6.

Stem. Following Lemma 4, since the number of factors of the form $(1 + c_i c_j x)^h$ is 3, we conclude that the number of stems that are not part of a pseudoknot is 3, moreover (taking into account the powers of brackets), the number of hydrogen bonds in this stems is 6, 3 and 1. In Fig. 5, the stems are denoted by blue.

Hairpin loop. Following Lemma 5, since the number of factors of the form $(1+c_i c_{i+1} x)^h$ (where c_i and c_{i+1} are adjacent colors) is 1 (namely, $(1 + c_2 c_3 x)^6$, i.e. $i = 2$), we conclude that the number of hairpin loops is 1. In Fig. 5, unpaired nucleotides of the hairpin loop are denoted by orange.

Bulge loop. Following Lemma 6, since there exists the unique factor (i.e., $(1 + c_1 c_4 x)^3 \cdot (1 + c_1 c_5 x)$) of the form $(1 + c_i c_j x)^{m_1} \cdot (1 + c_i c_{j+1} x)^{m_2}$; $i = 1$; $j = 4$, where the colors c_j and c_{j+1} are adjacent, and there exists no factor $(1 + c_j c_{j+1} x)$ (i.e., $(1 + c_4 c_5 x)$), we conclude that the number of bulge loops is 1. In Fig. 5, unpaired nucleotides of the bulge loop are denoted by reseda.

Internal loop. Following Lemma 7, since there exists the unique factor $((1 + c_1 c_4 x)^3 \cdot (1 + c_2 c_3 x)^6)$ of the form $(1 + c_i c_{j+1} x)^{h_1} \cdot (1 + c_{i+1} c_j x)^{h_2}$; $i = 1$; $j = 3$, where c_i and c_{i+1} are adjacent colors, c_j and c_{j+1} are adjacent colors, and there exist no factors $(1 + c_i c_{i+1} x)$ and $(1 + c_j c_{j+1} x)$ (i.e., $(1 + c_1 c_2 x)$ and $(1 + c_3 c_4 x)$), we conclude that the number of internal loops is 1. In Fig. 5, unpaired nucleotides of the internal loop are denoted by green.

Multi-loop. Following Lemma 8, there exist no multi-loops.

Pseudoknot. Following Lemma 9, there exist no pseudoknots.

Example 2 (bpRNA_RFAM_38844). The Dot-Bracket notation is

$$\text{.....((((.((((((..((....))..))))))((((...))))\vert))))\text{.....},}$$

where . stands for an unpaired nucleotide, (and) denote a paired nucleotide, and | is used to divide sequences of paired nucleotides that form different stems. The secondary structure is presented in Fig. 5 on the right.

We color nucleotides involved in the given Dot-Bracket notation as follows:

$$\underbrace{\text{.....((((}}_{c_1}\underbrace{\text{.((((((}}_{c_2}\underbrace{\text{..((}}_{c_3}\underbrace{\text{....))}}_{c_4}\underbrace{\text{..))))))((((...))))\vert))))}}_{c_5}\underbrace{\text{.....}}_{c_6}\underbrace{\text{.}}_{c_7}.$$

Based on the factored $\mathcal{RKB}$ polynomial

$$(1 + c_3 c_4 x)^2 (1 + c_2 c_5 x)^6 (1 + c_1 c_6 x)^4 (1 + c_5 c_6 x)^4,$$

we draw the following conclusions about the structural elements, which agree with Fig. 5.

Single-stranded region. Following Lemma 3, since number of colors is 6 and both the first and the last nucleotides are unpaired, we conclude that the number of single-stranded regions is 7.

Stem. Following Lemma 4, since the number of factors of the form $(1 + c_i c_j x)^h$ is 4, we conclude that the number of stems that are not part of a pseudoknot is 4, moreover (taking into account the powers of brackets), the number of hydrogen bonds in this stems is 2, 6, 4 and 4. In Fig. 5, the stems are denoted by blue.

Hairpin loop. Following Lemma 5, since the number of factors of the form $(1+c_ic_{i+1}x)^h$ (where c_i and c_{i+1} are adjacent colors) is 2 (namely, $(1 + c_3c_4x)^2$ and $(1 + c_5c_6x)^4$, i.e. $i = 3$ and $i = 5$), we conclude that the number of hairpin loops is 2. In Fig. 5, unpaired nucleotides of the hairpin loop are denoted by orange.

Bulge loop. Following Lemma 6, there exist no bulge loops.

Internal loop. Following Lemma 7, since there exists the unique factor (i.e., $(1+c_2c_5x)^6 \cdot (1+c_3c_4x)^2$) of the form $(1+c_ic_{j+1}x)^{h_1} \cdot (1+c_{i+1}c_jx)^{h_2}$; $i = 2$; $j = 4$, where c_i and c_{i+1} are adjacent colors, c_j and c_{j+1} are adjacent colors, and there exist no factors $(1+c_ic_{i+1}x)$ and $(1 + c_jc_{j+1}x)$ (i.e., $(1 + c_2c_3x)$ and $(1 + c_4c_5x)$), we conclude that the number of internal loops is 1. In Fig. 5, unpaired nucleotides of the internal loop are denoted by green. Let us note that there exists the factor $(1 + c_2c_5x)^6 \cdot (1 + c_1c_6x)^4$ of the necessary form, but this factor is not associated with an internal loop due to the existence of the factor $(1+c_5c_6x)$.

Multi-loop. Following Lemma 8, since there exists the unique factor (i.e., $(1+c_1c_6x)^4 \cdot (1 + c_5c_6x)^4 \cdot (1 + c_2c_5x)^6$) of the form $(1 + c_jc_ix)^{h_1} \cdot (1 + c_kc_ix)^{h_2} \cdot (1 + c_{j+1}c_kx)^{h_3}$; $i = 6$; $j = 1$; $k = 5$, where c_j and c_{j+1} are adjacent colors, we conclude that the number of multi-loops is 1. In Fig. 5, some nucleotides of the multi- loop are denoted by yellow.

Pseudoknot. Following Lemma 9, there exist no pseudoknots.

The author is grateful to Ph.G. Korablev for useful discussions of mathematical part of the model, N.A. Manakova for problem statement and useful discussions, as well as S.A. Zagrebina for creation of perfect conditions for productive work on the paper.

The work is supported by the RFBR grant No. 20-01-00127.

References

1. Tinoco I., Bustamante C. How RNA Folds. *J. Molecular Biology*, 1999, vol. 293, no. 1, pp. 271–281.
2. Varani G., McClain W.H. The G x U Wobble Base Pair. A Fundamental Building Block of RNA Structure Crucial to RNA Function in Diverse Biological Systems. *EMBO Reports*, 2000, vol. 1, pp. 18–23.
3. Lodish H., Berk A., Zipursky S.L., Matsudaira P., Baltimore D., Darnell J. Molecular Cell Biology, sixth ed. *W.H. Freeman*, 2007.
4. Schultes E.A., Bartel D.P. One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds. *Science*, 2000, vol. 289, pp. 448–452.
5. Chheda N., Gupta M.K. *RNA as a Permutation*, available at: arXiv:1403.5477v1 (accessed on October 27, 2021).
6. Ray S.S., Pal S.K. RNA Secondary Structure Prediction Using Soft Computing. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2013, vol. 10, no. 1, pp. 2–17.
7. Tian W., Lei X., Kauffman L.H., Liang J. A Knot Polynomial Invariant for Analysis of Topology of RNA Stems and Protein Disulfide Bonds. *Mol Based Math Biol*, 2017, vol. 5, no. 1, pp. 21–30. DOI: 10.1515/mlbmb-2017-0002
8. Adams C. The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots. *American Mathematical Society*, 1994.

9. Chen S.J., Dill K.A. Symmetries in proteins: A knot theory approach. *The Journal of Chemical Physics*, 1996, vol. 104, no. 15, pp. 5964–5973.
10. Kauffman L.H. State models and the Jones polynomial. *Topology*, 1987, vol. 26, pp. 395–407.
11. bpRNA-1m database, available at: http://bprna.cgrb.oregonstate.edu/ (accessed on October 27, 2021).
12. Lolle S.J., Victor J.L., Young J.M., Pruitt R.E. Genomewide non-mendelian inheritance of extra-genomic information in arabidopsis. *Nature*, 2005, vol. 434, no. 7032, pp. 505–509.
13. Lilley D.M. Structure, folding and mechanisms of ribozymes. *Current opinion in structural biology*, 2005, vol. 15, no. 3, pp. 313–323.
14. Tomari Y., Zamore P.D. Perspective: machines for RNA. *Genes and Development*, 2005, vol. 19, no. 5, pp. 517–529.
15. Sullenger B.A. Series introduction: Emerging clinical applications of nucleic acids. *Journal of Clinical Investigation*, 2000, vol. 106, no. 8, pp. 921–922.
16. Taufer M., Licon A., Araiza R., Mireles D., Van Batenburg F., Gultyaev A.P., Leung M.Y. PseudoBase++: an extension of PseudoBase for easy searching, formatting and visualization of pseudoknots. *Nucleic Acids Res*, 2009, vol.37 (suppl. 1), pp. 127–135. DOI:10.1093/nar/gkn806
17. Legendre A., Angel E., Tahi F. *Structure Prediction with Pseudoknots*, available at: https://rnavlab.utep.edu/database (accessed on October 27, 2021).
18. Danaee P., Rouches M., Wiley M., Deng D., Huang L., Hendrix D.A. bpRNA: Large-scale Automated Annotation and Analysis of RNA Secondary Structure. *bioRxiv*, article ID: 271759.
19. Legendre A., Angel E., Tahi F. Bi-objective integer programming for RNA secondary structure prediction with pseudoknots. *BMC Bioinformatics*, 2018, vol. 19, no. 13, pp. 1–15. DOI: 10.1186/s12859-018-2007-7
20. Kauffman L.H. Virtual Knot Theory. *Europ. J. Combinatorics*, 1999, vol. 20, pp. 663–691.
21. Reidemeister K. *Knotentheorie*. Berlin, Heidelberg, Springer, 1932.
22. Kauffman L.H. Invariants of Graphs in Three-Space Trans. *Amer. Math. Soc.*, 1989, vol. 311, pp. 697–710.
23. Sacdalan A.M. Aspects of the Jones polynomial. *Theses Digitization Project John M. Pfau Library*, 2006.

Alyona A. Akimova, PhD (Math), Department of Mathematical and Computer Modelling, Department of Mathematical Physics Equations, South Ural State University (Chelyabinsk, Russian Federation), akimovaaa@susu.ru

Received May 18, 2022.

УДК 515.162 DOI: 10.14529/jcem220201

ОБОБЩЕНИЕ СКОБОЧНОГО ПОЛИНОМА КАУФМАНА ДЛЯ ОПРЕДЕЛЕНИЯ И АНАЛИЗА СТРУКТУРНЫХ ЭЛЕМЕНТОВ ВТОРИЧНОЙ СТРУКТУРЫ РНК

А. А. Акимова

В этой работе мы используем методы теории узлов для описания и анализа структурных элементов вторичной структуры РНК путем построения нового обобщения классического скобочного полинома Кауфмана, факторизация которого характеризует структурные элементы РНК. С этой целью разработана математическая модель РНК, включающая топологический инвариант ($\mathcal{RKB}$ полином) и позволяющая определить тип, количество и характеристики стандартных структурных элементов, образующих вторичную структуру РНК. Чтобы определить $\mathcal{RKB}$ полином, мы вводим новое правило сглаживания водородных связей и новую технику окрашивания нуклеотидов, а также используем известное сглаживание классических перекрестков, используемое в скобочном полиноме Кауфмана. Отметим, что предложенная методика окрашивания нуклеотидов позволяет учитывать взаимное расположение структурных элементов, что может быть использовано при изучении свойств РНК. Показана инвариантность $\mathcal{RKB}$ полинома. Вычисление $\mathcal{RKB}$ полинома по заданному коду Dot-Bracket реализовано в виде небольшой программы средствами пакета Mathematica. Используя $\mathcal{RKB}$ полиномы, рассчитанные нашей программой, мы анализируем некоторые вторичные структуры РНК, представленные в базе данных bpRNA-1m. Полученные результаты согласуются с реальными данными. *Ключевые слова: РНК; вторичная структура; обобщенный скобочный полином Кауфмана; узел; инвариант.*

Акимова Алена Андреевна, кандидат физико-математических наук, доцент кафедры математического и компьютерного моделирования, доцент кафедры уравнений математической физики, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), akimovaaa@susu.ru

Поступила в редакцию 18 мая 2022 г.