Research Article **Open Access**

Wei Tian, Xue Lei, Louis H. Kauffman, and Jie Liang*

A Knot Polynomial Invariant for Analysis of Topology of RNA Stems and Protein Disulfide Bonds

DOI 10.1515/mlbmb-2017-0002
Received January 4, 2017; accepted March 30, 2017

Abstract: Knot polynomials have been used to detect and classify knots in biomolecules. Computation of knot polynomials in DNA and protein molecules have revealed the existence of knotted structures, and provided important insight into their topological structures. However, conventional knot polynomials are not well suited to study RNA molecules, as RNA structures are determined by stem regions which are not taken into account in conventional knot polynomials. In this study, we develop a new class of knot polynomials specifically designed to study RNA molecules, which considers stem regions. We demonstrate that our knot polynomials have direct structural relation with RNA molecules, and can be used to classify the topology of RNA secondary structures. Furthermore, we point out that these knot polynomials can be used to model the topological effects of disulfide bonds in protein molecules.

Keywords: Topological knot, RNA molecules, Knot polynomials

1 Introduction

A topological knot may exist in a closed curve embedded in $\mathbb{R}^3$. As topological invariants, knot polynomials are used to detect the existence of knots and to classify them [1, 2, 4]. Knot polynomials can be computed for a DNA or a protein molecule: the trace of the DNA duplex or the protein backbone is first closed off such that the topology of the molecule is otherwise unaffected [3]. Smoothing operation, such as the one implemented in the KMT algorithm [5, 6], is then applied to simplify the closed curve without altering its topology. This closed curve is then projected as a diagram in $\mathbb{R}^2$ (Fig. 1). A polynomial is then computed from the oriented crossings of the projected diagram by recursively applying a skein relation, which defines a linear relationship among the edges that form the crossing. Fig. 2 shows an example of the skein relation for Conway polynomial. With different definitions of the skein relation, different types of knot polynomials can be generated [7–9].

A significant number of DNA knots are found in biological species, such as *Escherichia coli* cells harboring mutations in the GyrB or GyrA genes [10], bacteriophages P2 and P4 [11], and cauliflower mosaic viruses [12]. Knots in DNAs can lead to building-ups of stress during the replication process, although such stress can be resolved once topoisomerases nick and then close one strand of the duplex DNA [13]. Knots in proteins appear in about 1% of the structures in the Protein Data Bank (PDB) [14], where the simplest knot is the trefoil knot, and the most complex knot is the Stevedore knot that has six crossings [15]. An example of a protein knot

***Corresponding Author: Jie Liang:** Department of Bioengineering, University of Illinois at Chicago, 1200 West Harrison St., 60607, Chicago, IL, USA, E-mail: jliang@uic.edu
Wei Tian, Xue Lei: Department of Bioengineering, University of Illinois at Chicago, 1200 West Harrison St., 60607, Chicago, IL, USA
These authors contributed equally to this work.
Louis H. Kauffman: Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 1200 West Harrison St., 60607, Chicago, IL, USA

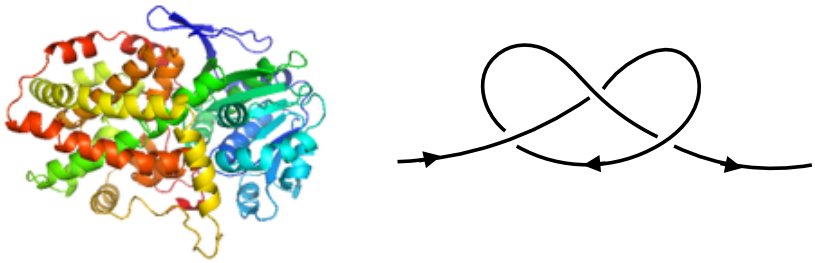

Figure 1: The structure of chain I of the protein plant acetohydroxy acid isomeroreductase (PDB code: 1YVE) and the corresponding topology of its backbone.

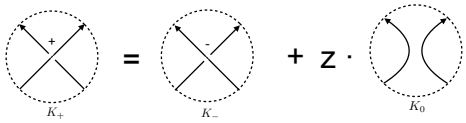

Figure 2: The skein relation for computing the Conway polynomial. There are three possible oriented connections among the four edges that form the crossing: one with a positive crossing (K_+), one with a negative crossing (K_-), and one with no crossing (K_0).

is shown in Fig. 1 [16]. Although it is hypothesized that knots in proteins may increase protein stability and prevent proteolytic degradation [17], the structure–function relationship as well as the folding pathway of the knotted proteins are generally unknown.

Compared to DNAs and proteins, our knowledge of RNA knots are more limited. An RNA molecule is 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 phosphodiester bonds. 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. RNA stems can combine in different ways, giving rise to different secondary structures of RNA molecules.

RNA topology can be studied using conventional knot polynomials. A survey showed that knotted RNA molecules are rare among currently known RNA structures [18]: Only 3 knotted RNA chains, with 3, 4 and 16 crossings, respectively, are found among 2,863 RNA structures contained in the PDB. However, conventional knot polynomials only consider phosphodiester bonds in an RNA molecule, and completely ignores the topological effects of hydrogen bonds in stem regions, the basis of the formation of RNA secondary structures.

In this study, we develop a new class of knot polynomials specifically designed to study RNA stem topology. These polynomials are constructed based on how stems form RNA structures, with consideration of topological effects of both hydrogen bonds and phosphodiester bonds.

2 Methods

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. A stem creates an “entanglement” between the two parts of the backbone, and therefore can be viewed as a topological rigid vertex [19, 20], which has no counterpart in DNA duplexes or proteins, as disulfide bonds in proteins and strand-strand interactions in DNAs are not usually the subject of topological analysis.

A rigid vertex is subjected with a set of local moves (Fig. 3), which are the generalization of the Reidemeister moves [21] in the knot theory. Moves I and II allow a rigid vertex to twist vertically and horizontally, respectively. Move III allows an edge to slide under or over a vertex without affecting the other edges. These basic moves provide the ambient isotopies for RNA diagrams with rigid vertices [22–24].

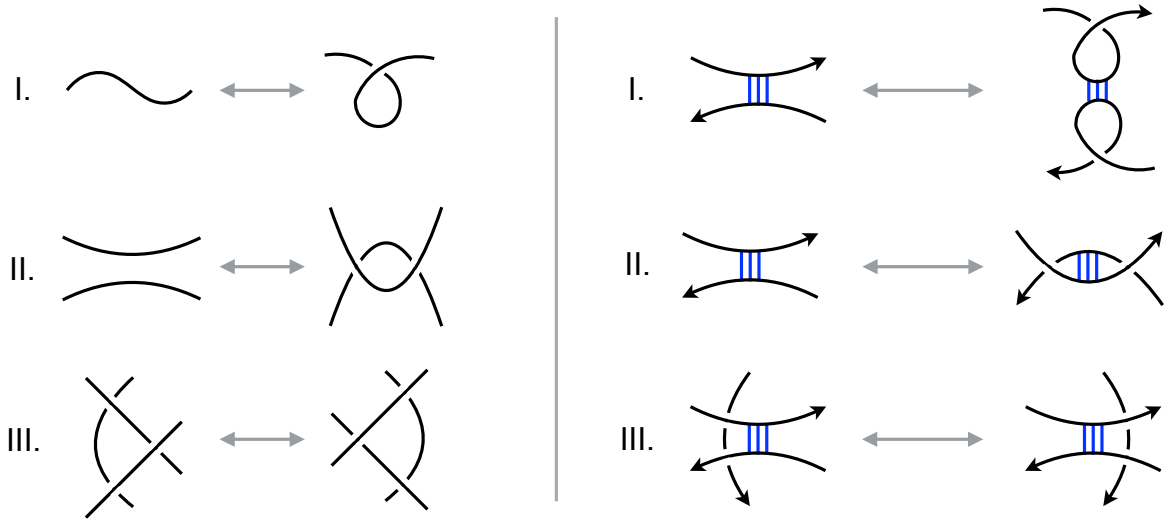

Figure 3: The left panel is the classic Reidemeister moves. The right panel is the local moves of rigid vertices.

Here, we develop a model of knot polynomials to characterize RNA structures, with topological rigid vertices representing the stem region explicitly considered. We first close off RNA strands into circles. For RNA structures with a single strand, the backbone trace is closed off by connecting the 3' and the 5' ends virtually. For RNA structures with multiple strands, the backbone trace can be closed either by connecting the 3' and the 5' ends of one strand to itself, or by connecting the 5' end of one strand and the 3' end of another strand. Different choice of closure leads to different RNA diagrams, thus possibly different polynomials. To avoid ambiguity in closure of RNA structures with three or more strands, we always close RNA strands using the self–closure approach.

We classify RNA stems into two types according to their orientation, namely, the L–type and the R–type, and define a skein relation to translate these two types of RNA stems into knot diagrams (Fig. 4). Since the number of nucleotides that form a stem is immaterial of topological consequences, we ignore the length of stems and treat stems of different sizes equally.

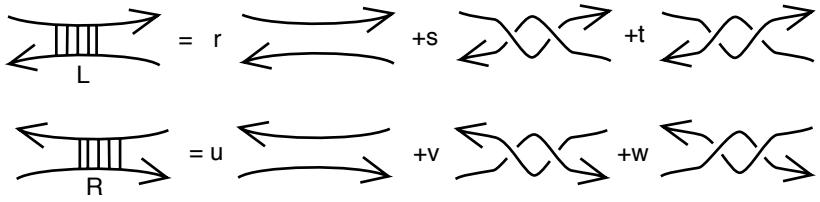

Figure 4: The skein relation for translating rigid vertices. Each type of rigid vertex can be translated into a weighted summation of three motifs without rigid vertices.

Specifically, one rigid vertex can be translated to the weighted summation of one pair of oriented antiparallel strands and two oriented double–twists. The coefficients (or weights) of these three terms are adjustable. After applying this translation to all rigid vertices of an RNA structure recursively, we obtain a summation of a group of weighted diagrams without rigid vertices. Any knot invariant, including knot polynomials, can be used to characterized resulting diagrams. Here we computed the Kauffman bracket polynomial [9] for each of the resulting diagram (see Appendix for details of the algorithm), and obtain the final polynomial for the RNA structure.

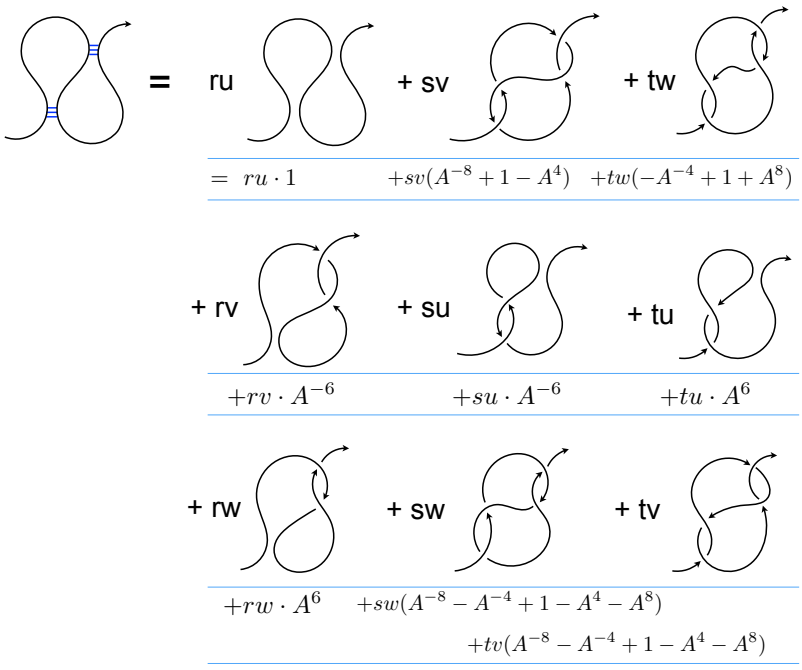

Figure 5: An RNA pseudoknot is translated into the summation of a set of knot diagram without rigid vertex after applying the skein relation in Fig. 4 recursively. The corresponding Kauffman bracket polynomial for each knot diagram is listed accordingly, along with the coefficient of each term.

3 Application

A pseudoknotted RNA molecule contains at least two stem–loop structures, in which half of one stem is intercalated between the two halves of another stem. Several important biological processes rely on RNA pseudoknots. For example, a pseudoknot in the RNA component of the telomerase is critical for its activity [25]. Pseudoknotted structure are also used by several viruses to infect host cells [26]. While pseudoknotted RNA molecules are regarded as unknotted from the usual point of view, the introduction of the rigid vertex reveals intrinsic knottedness of this class of RNA molecules. As an example, we compute the RNA polynomials for a pseudoknot with two intercalated stem–loops (Fig. 5). The resulting polynomial is

$$\begin{aligned} &A^{-8}(A^{16}(sw + t(v + w)) + A^{14}(rw + tu) - A^{12}(s(v + w) + tv) \\ &+A^8(ru + (s + t)(v + w)) - A^4(sw + t(v + w)) + A^2(rv + su) + s(v + w) + tv) \end{aligned} \tag{1}$$

As the number of the terms translated via the skein relation grows exponentially with the number of rigid vertices, the resulting RNA polynomial is fairly complex, even for this simple RNA structure.

3.1 Simplified RNA polynomial

As the coefficients in the skein relation depicted in Fig. 4 are adjustable, we can simplify the RNA polynomial by setting several coefficients to 0 to eliminate the corresponding terms. Based on the observation that the L– and R–type rigid vertices are the images of 180° flipping of each other (Fig. 6a), we need only consider the coefficients t and w in the skein relation of Fig. 4. This leads to a new skein relation that translates each rigid vertex into one oriented double–twist (Fig. 6b). This simplification fully captures the "entanglement" of a rigid vertex, while respecting the 180°–flipping relationship between the L– and R–type rigid vertices.

We call the polynomials calculated using this simplified skein relation the Simplified RNA Polynomial (SiRP). For instance, the SiRP of a pseudoknot is $(A^8 + 1 - A^{-4})tw$, much simpler compared to the original one (Eq 1).

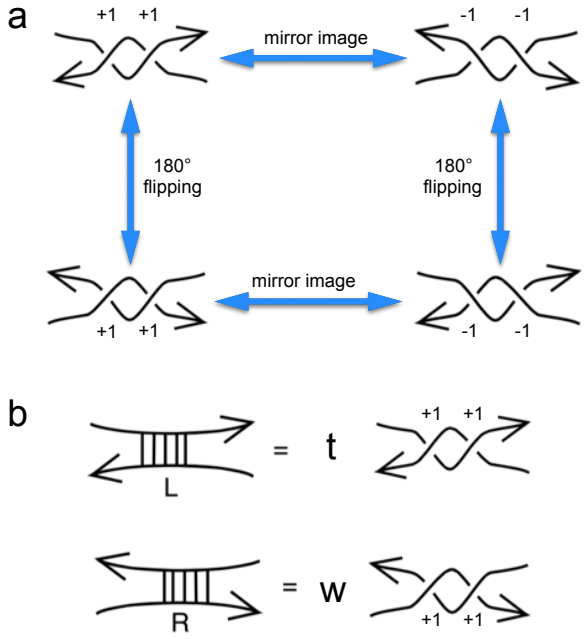

Figure 6: Oriented double–twists and simplified skein relation. (a) The relation among the four oriented double–twists in the original RNA skein relation. (b) The simplified skein relation in the computation of SiRP.

To compute the SiRP of an RNA molecule, we first follow the self–closure procedure to connect the strand(s) of the RNA molecule. We then translate the RNA diagram into a knot diagram using the simplified skein relation. Fig. 7 shows several examples of the closure and translation.

The resulting knot diagrams directly reflect the topology of RNA structures. An RNA structure with two strands connected by one stem is translated into a Hopf link (Fig. 7c), in which the double–twist reflects the "entanglement" between the original two RNA strands. To break this Hopf link, one need to flip the orientation of either of the two crossing, which corresponds to breaking the hydrogen bonds of the original RNA structure (Fig. 7c). The RNA structure in Fig. 7d can be translated into a Solomon's knot. If the orientation of one crossing of the Solomon's knot is flipped, the new knot becomes a Hopf link, reflecting the fact that the RNA structure in Fig. 7d will become the same as that in Fig. 7c if one stem is broken.

Computing SiRP can provide additional information on the secondary structures of RNAs. For example, RNAs with a stem–loop structure will have an A^6 term in their polynomials. For a structure of two RNA strands connected by one stem (Fig. 8a), its SiRP will have a factor of $1+A^8$ (Details in Appendix, Fig. 10b). If two RNA strands are connected by two or three stems (Fig. 8b and c), the SiRP will have a factor of $1 + A^8 - A^{12} + A^{16}$ or $1 + A^8 - A^{12} + A^{16} - A^{20} + A^{24}$, respectively. Generally, the SiRP of two RNAs connected by n number of consecutive stems (Fig. 8d) will have a factor of $1+A^8-(1-A^4)\Sigma_{i=2}^{n}A^{8i-4}$. Overall, the SiRP of an RNA structure can be decomposed into a product of SiRPs of more primitive motifs. Fig. 8e-g offers additional examples, demonstrating how SiRPs of complex RNA structures can be decomposed into the products of SiRPs of stem–loops and of two RNA strands connected by rigid vertices. While these examples can be decomposed into products of correspoding factors, this may not be true in cases when there are interference among stems. For example, the polynomials of the RNAs in Fig. 8(c-d) cannot be decomposed into production of polynomial of single stem.

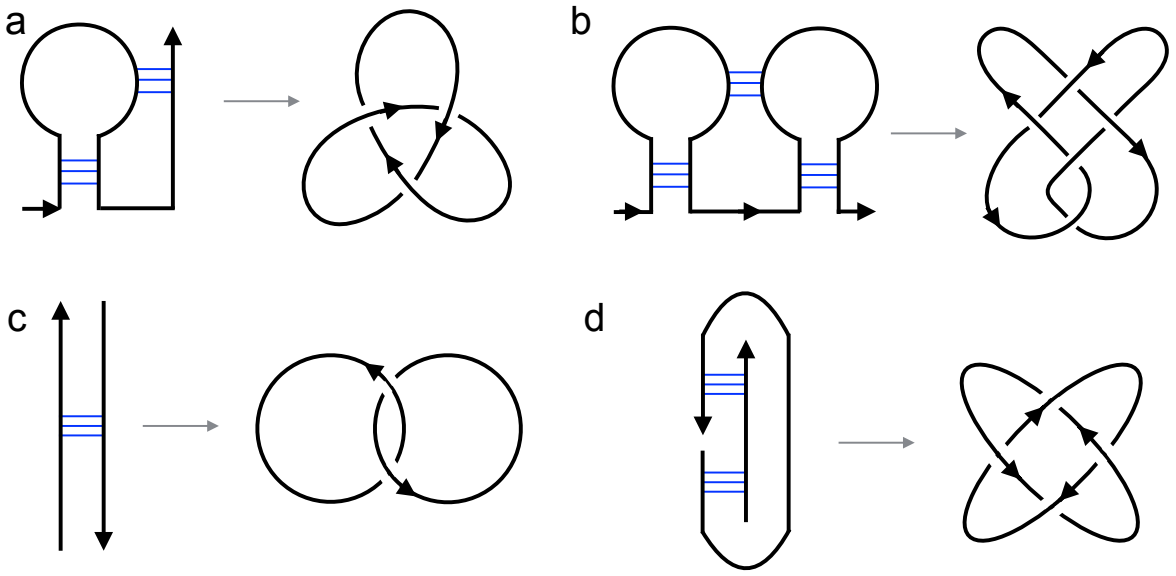

Figure 7: Examples of RNA structures before and after the self–closure and the application of the simplified skein relation.

4 Discussion

In this study, we have developed a new model of knot polynomials specifically designed for RNA molecules, which takes into account the hydrogen bonded stem regions. The resulting RNA polynomial can be used to classify topology of RNA secondary structures. Using a simplified version of the RNA polynomial, we demonstrate with examples that these new polynomials are directly related to the structures of the RNA molecules.

We used the Kauffman bracket polynomial as the intermediate knot invariant to calculate the RNA polynomial in this paper. Since the Kauffman bracket polynomial is not an invariant under the Reidemeister move I, the writhe number of the diagrams in principle can be used to further normalize the results, so that the resulting polynomials become invariant under all three moves. Details of this normalization can be found in Ref [4].

When computing knot polynomials of DNA or protein molecules, the strand of amino acids or nucleotide base pairs can be directly treated as a closed curve embedded in $\mathbb{R}^3$ after closure. However, this simple treatment is inadequate for RNA molecules, as RNA secondary structures are determined by their stem regions, which can be effectively treated as rigid vertices that fix the relative position of two RNA strands or two parts of one RNA strand. If we neglect these stems and directly compute conventional knot polynomials such as Alexander polynomials, the central elements of RNA topology will be neglected. For example, an RNA pseudoknot will be treated the same as an unpaired RNA strand. Indeed, a survey on RNA molecules based on Alexander polynomials found nothing of particular interest [18].

The new RNA polynomials introduced here, particularly the simplified ones (SiRPs), have direct relation with the topology of RNA structures. We observed that the factors of SiRPs reflect the structural features of RNAs, and demonstrated in several cases that the SiRP of an RNA can be decomposed into the product of the siginature factors of corresponding structural motifs of the RNA. Further systematical study and theory development of the decomposition of SiRPs will likely further simplify the calculation of the RNA polynomial and aid in gaining new insights into understanding RNA structures.

We mention that the idea of rigid vertices can be used to describe disulfide bonds in proteins, as they are similar to RNA stems in fixing the relative positions of distant parts of protein backbones and creating "entanglements". Once the projected diagrams of backbone curves with disulfide bonds are generated after closure as described in Ref [3], identical procedures can be followed to generate SiRP-like polynomials for protein molecules, which consider the topological effects of disulfide bonds.

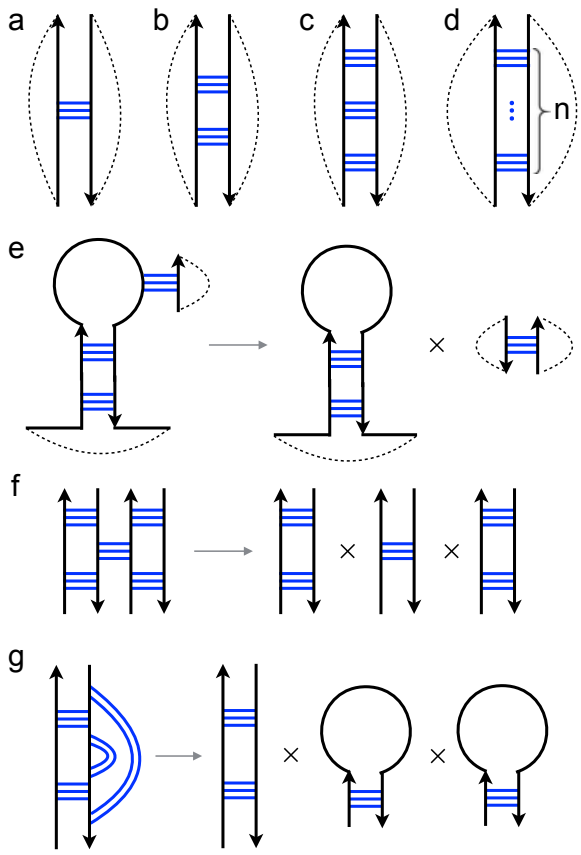

Figure 8: SiRPs are related directly to the secondary structure of RNA molecules. (a-d) two RNA strands connected by n consecutive stems have of a factor of $1 + A^8 - (1 - A^4)\Sigma_{i=2}^{n} A^{8i-4}$ in their SiRPs. (e-g) SiRPs of RNA structures can be decomposed into products of SiRPs of primitive motifs.

Knot theory has found wide applications in many branches of sciences [27]. Its application in biology has uncovered the existence of knoted structures in DNAs and proteins and provided important insight [28]. Our study further provides a new tool to study the topology of RNA molecules, which may help in further understanding the structure–relationship of RNA molecules.

A Appendix

A.1 Procedure of calculating RNA knot polynomials

After applying the skein relation of Fig. 4 to all rigid vertices (Fig. 9a), we compute the Kauffman bracket polynomial [9] of the resulting diagrams. The edges incident to a potential crossing are labeled, and these labels extend to labels on the corresponding edges of the graphs (Fig. 9b). Edges that form no crossing are computed as the product of two $\delta(\cdot)$ operators, each of which is defined as an oriented segment (Fig. 9c). The oriented crossing is computed as the $X(\cdot)$ or $Y(\cdot)$ operators, depending on the orientation (Fig. 9d). The $X(\cdot)$ and $Y(\cdot)$ operators can be further decomposed as the product of two $\delta(\cdot)$ operators.

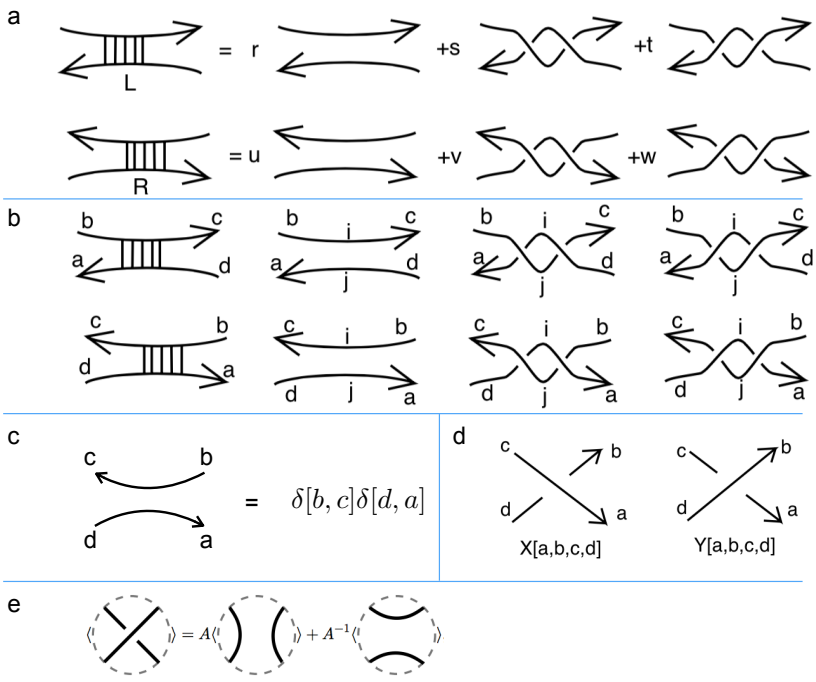

Figure 9: After the application of the RNA skein relation, the oriented crossings are labeled, and Kauffman bracket polynomial can be computed accordingly.

Eqs. (2-7) illustrate how each step in the procedure and ech operator can be calculated in a top-down order.

$$L[a,b,c,d] = r\delta[b,c]\delta[d,a] + sY[a,j,i,b]Y[c,i,j,d] + tX[a,j,i,b]X[c,i,j,d] \quad (2)$$

$$R[a,b,c,d] = u\delta[b,c]\delta[d,a] + vY[i,c,d,j]Y[j,a,b,i] + wX[i,c,d,j]X[j,a,b,i] \quad (3)$$

$$X[a,b,c,d] = A\delta[c,b]\delta[d,a] + A^{-1}\delta[c,d]\delta[a,b] \quad (4)$$

$$Y[a,b,c,d] = A^{-1}\delta[c,b]\delta[d,a] + A\delta[c,d]\delta[a,b] \quad (5)$$

$$\delta[a,b]\delta[b,c] = \delta[a,c] \quad (6)$$

$$\delta[a,b]\delta[b,a] = -A^2 - A^{-2}. \quad (7)$$

A.2 Examples of calculating SiRPs

By setting the coefficients r, s, u, and v to 0, the SiRP of an RNA can be calculated following the procedure above. As examples, we illustrate the calculation of the SiRPs of an RNA pseudoknot (Fig. 10a) and a two-strand RNA stem (Fig. 10b).

SiRP of an RNA pseudoknot. The SiRP of the pseudoknot can be calculated as the product of two terms $L(\cdot)$ and $R(\cdot)$. After the self closure and applying the skein relation in Fig. 4, the edges incident to crossings are labeled (Fig. 10a). Following the procedure above, we can write down the equations to calculate the SiRP

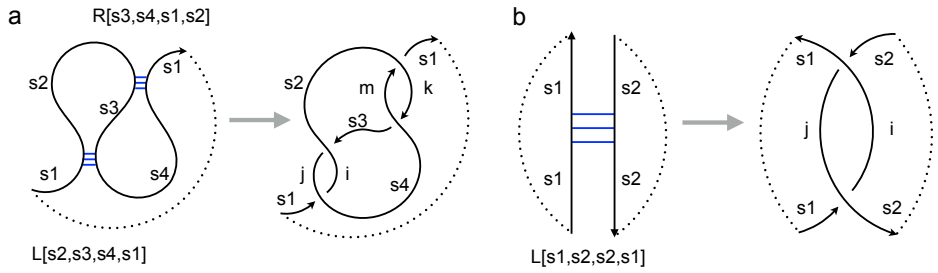

Figure 10: The translation of the rigid vertices using simplified skein relation and the labeling of edges of crossings in an RNA pseudoknot (a) and in a two-strand RNA stem (b).

(Eqs. 8-16):

$$\begin{aligned}
SiRP(\text{pseudoknot}) &= L[s2,s3,s4,s1]R[s3,s4,s1,s2] && (8)\\
L[s2,s3,s4,s1] &= tX[s2,j,i,s3]X[s4,i,j,s1] && (9)\\
R[s3,s4,s1,s2] &= wX[k,s1,s2,m]X[m,s3,s4,k] && (10)\\
X[s2,j,i,s3] &= A\delta[i,j]\delta[s3,s2]+A^{-1}\delta[i,s3]\delta[s2,j] && (11)\\
X[s4,i,j,s1] &= A\delta[j,i]\delta[s1,s4]+A^{-1}\delta[j,s1]\delta[s4,i] && (12)\\
X[k,s1,s2,m] &= A\delta[s2,s1]\delta[m,k]+A^{-1}\delta[s2,m]\delta[k,s1] && (13)\\
X[m,s3,s4,k] &= A\delta[s4,s3]\delta[k,m]+A^{-1}\delta[s4,k]\delta[m,s3] && (14)\\
\delta[s1,s2]\delta[s2,s3] &= \delta[s1,s3] && (15)\\
\delta[s1,s2]\delta[s2,s1] &= -A^2-A^{-2}. && (16)
\end{aligned}$$

By substituting Eqs (11-14) into Eqs (9) and (10), and then the results into Eq (8), we obtain a summation of products of $\delta(\cdot)$ terms. All these $\delta(\cdot)$ terms can be transformed into terms of A after repeatedly applying the rules of Eqs (15-16), and therefore the SiRP can be written at the end as

$$SiRP(\text{pseudoknot}) = (A^8+1-A^{-4})tw.$$

SiRP of a two-strand RNA stem. To compute the SiRP of a two-strand RNA stem, we apply the simplified skein relation to the RNA diagram, and then label the edges of the crossings (Fig. 10b). The SiRP can be calculated by computing $L[s1,s2,s2,s1]$. Following the aforementioned algorithm, we can write down the equations (Eqs 17-20),

$$\begin{aligned}
SiRP(\text{two-strand stem}) &= L[s1,s2,s2,s1] && (17)\\
L[s1,s2,s2,s1] &= tX[s1,j,i,s2]X[s2,i,j,s1] && (18)\\
X[s1,j,i,s2] &= A\delta[i,j]\delta[s2,s1]+A^{-1}\delta[i,s2]\delta[s1,j] && (19)\\
X[s2,i,j,s1] &= A\delta[j,i]\delta[s1,s2]+A^{-1}\delta[j,s1]\delta[s2,i] && (20)\\
& && (21)
\end{aligned}$$

By substitue Eqs (19-20) into Eq (18), and then the result into Eq (17), the SiRP of a two-strand stem can be written as

$$\begin{aligned}
SiRP(\text{two-strand stem}) &= A^6+A^2+A^{-2}+A^{-6}t\\
&= (1+A^8)(A^{-2}+A^{-6})t,
\end{aligned}$$

which has a factor $1+A^8$ as mentioned in the main text.

Acknowledgement: This work was supported by NIH grant GM079804.

References

[1] G Burde and H Zieschang. *Knots*, Volume 5 of *de Gruyter Studies in Mathematics*. de Gruyter, 1985.

[2] K Murasagi. *Knot Theory and Its Applications*. Birkhäuser, Boston, 1996.

[3] RC Lua and AY Grosberg. Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Computational Biology*, 2(5):350–357, 2006.

[4] LH Kauffman. *Knots and Physics (4th edition)*. World Scientific, 2012.

[5] WR Taylor and A Aszodi. *Protein geometry, classification, topology and symmetry*. Series in biophysics. CRC Press, 2004.

[6] K Koniaris and M Muthukumar. Self-entanglement in ring polymers. *J. Chem. Phys.*, 95:2873–2881, 1991.

[7] JW Alexander. Topological invariants of knots and links. *Trans. Amer. Math. Soc.*, 30:275–306, 1928.

[8] VFR Jones. A polynomial invariant for knots via von Neumann algebra. *Bull. Amer. Math. Soc. (N.S.)*, 12:103–111, 1985.

[9] LH Kauffman. State models and the Jones polynomial. *Topology*, 26:395–407, 1987.

[10] K. Shishido, N. Komiyama, and S. Ikawa. Increased production of a knotted form of plasmid pBR322 DNA in Escherichia coli DNA topoisomerase mutants. *J Mol Biol*, 195:215–8, 1987.

[11] LF Liu, JL Davis, and R Calendar. Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases. *Nucleic Acids Res*, 9:3979–89, 1981.

[12] J Ménissier, G de Murcia, G Lebeurier, and L Hirth. Electron microscopic studies of the different topological forms of the cauliflower mosaic virus DNA: knotted encapsidated DNA and nuclear minichromosome. *EMBO J*, 2:1067–1071, 1983.

[13] A Uzman, Harvey Lodish, Arnold Berk, Lawrence Zipursky, David Baltimore, and James Darnell. Volume 29, Chapter Section 1.2 The Molecules of Life. Freeman, W. H. & Company, New York, NY, 2000.

[14] R Mishra and S Bhushan. Knot theory in understanding proteins. *J Math Biol*, 65:1187–1213, 2012.

[15] D Bölinger, JI Sułkowska, HP Hsu, LA Mirny, M Kardar, JN Onuchic, and P Virnau. A Stevedore's protein knot. *PLOS Comput Biol*, 6:e1000731, 2010.

[16] WR Taylor. A deeply knotted protein structure and how it might fold. *Nature*, 406:916–919, 2000.

[17] P Virnau, LA Mirny, and M Kardar. Intricate knots in proteins: Function and evolution. *PLOS Comput Biol*, 2:e122, 2006.

[18] C Micheletti, MD Stefano, and H Orland. Absence of knots in known RNA structures. *Proc Natl Acad Sci USA*, 112:2052–2057, 2015.

[19] LH Kauffman and YB Magarshak. Chapter Vassiliev knot invariants and the structure of RNA folding. Volume 5 of *Series on Knots and Everything*. World Scientific, 1995

[20] LH Kauffman and YB Magarshak. Graph invariants and the topology of RNA folding. *J. Knot Theo. and Ramif.*, 03:233, 1994.

[21] K Reidemeister. *Knotentheorie*. Ergebnisse der Matematik und ihrer Grenzgebiete. Springer, (Alte Folge), 1932.

[22] LH Kauffman. Invariants of graphs in three-space. *Trans. AMS*, 311(2):697–710, 1989.

[23] LH Kauffman and P Vogel. Link polynomials and a graphical calculus. *J. Knot Theo. and Ramif.*, 1:59–104, 1992.

[24] LH Kauffman and R Mishra. Nodal parity invariants of knotted rigid vertex graphs. *J. Knot Theo. and Ramif.*, 22(4):1340002, 2013.

[25] JL Chen and CW Greider. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc Natl Acad Sci USA*, 102:8080–5, 2005.

[26] CW Pleij, K Rietveld, and L Bosch. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.*, 13:1717–31, 1985.

[27] C Adams. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. American Mathematical Society, 1994.

[28] SJ Chen and KA Dill. Symmetries in proteins: A knot theory approach. *The Journal of Chemical Physics*, 104(15):5964–5973, 1996.