

# Short Stems in RNA Pseudoknots

Baulin E., Korinevskaya A., Roytberg M.

IMPB RAS, Pushchino, Russia; NRU HSE, Moscow, Russia; MIPT, Dolgoprudny, Russia.

[baulin@lpm.org.ru](mailto:baulin@lpm.org.ru), [mroytberg@lpm.org.ru](mailto:mroytberg@lpm.org.ru)

## Introduction

We present a study of the role played by stems of length of two and three base pairs in the formation of pseudoknotted RNA structures. Such stems seem to be energetically unfavorable and therefore might be unstable. Our results show that short stems are responsible for the great variety of different types of pseudoknots. Excluding short stems from the consideration reduces almost all pseudoknots to only two simple types namely H-knot and kissing loops.

## Materials and methods

### Terminology

A **Stem** is a sequence of pairs of the form  $(i, j), (i+1, j-1), \dots, (i+k, j-k)$  where  $k > 0$  and each number denotes a nucleotide in the corresponding position of a chain and each pair  $(i, j)$  is a base pair (either Watson-Crick base pair or Wobble (GU) base pair). A **short stem** is a stem for which  $k$  equals 1 or 2, i.e. short stem is made of two or three base pairs.

An informal definition of a pseudoknot is following: a **Pseudoknot** is a structural element comprised by at least two stems such that  $i_1 + k_1 < i_2 < j_1$ . To classify pseudoknots we employ a two stage reduction process that includes (i) removing all nested stems and (ii) collapsing all base pairs of consecutive stems into one arc. Each end of each arc is then assigned with a letter; left ends are assigned with small letters and right ends are assigned with corresponding capital letters. A **pseudoknot signature** is a word composed of the letters assigned to the arcs' ends in order of its positions on the chain (see Fig.1). For example, the signature of an H-knot is **abAB** and the signature of kissing loops is **abAcBC**.

For exact definitions see:

<http://server3.lpm.org.ru/urs/struct.py?where=3>

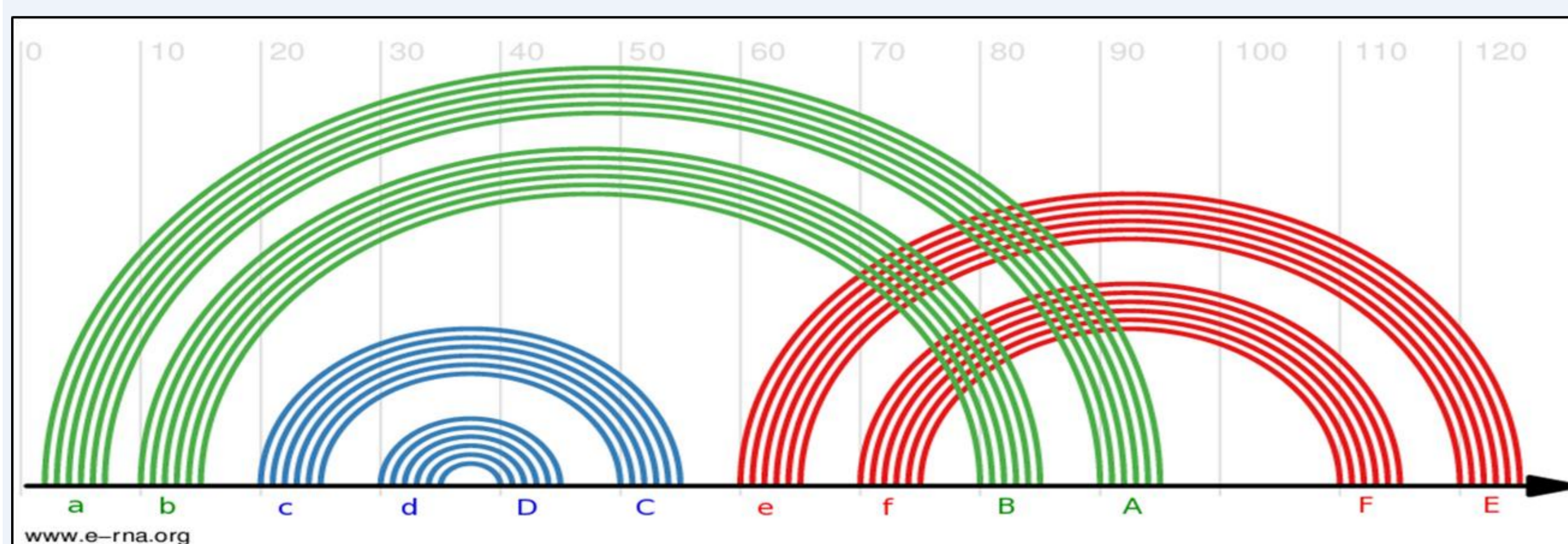


Figure 1. An example of a pseudoknot signature (abcdDCefBAFE -> abcdBADc -> abAB)

### Data

As an input data we used all RNA-containing structures from Protein Data Bank [2]. Annotations of stems and pseudoknots were obtained from URS database [1]. The overall number of considered pseudoknots – 5233 (the data is highly redundant). The overall number of unique pseudoknot signatures – 16.

### Short stems

According to the empirical thermodynamic parameters [3] the formation of a helix starts to decrease free energy of the structure only after the formation of the third stack of base pairs. With respect to this rule we believe that short stems may be unstable and should not contribute significantly to the overall free energy of folded RNA molecule. From this point it becomes interesting to look at the role of short stems in the formation of pseudoknots and see if their complexity may be reduced.

We compared stem length distributions separately for all structures and pseudoknotted structures. Short stems of length 2 have been found to be overrepresented in pseudoknotted structures (32% vs. 25% in all structures) although short stems of length 3 have not supported this trend (see Fig.2).

For indirect evidence of instability of short stems we compared stem length distributions separately for each type of respective loop. As we expected the majority of short stems were found near internal loops and bulges where they can be strengthened by longer adjacent stems (see Fig.3).

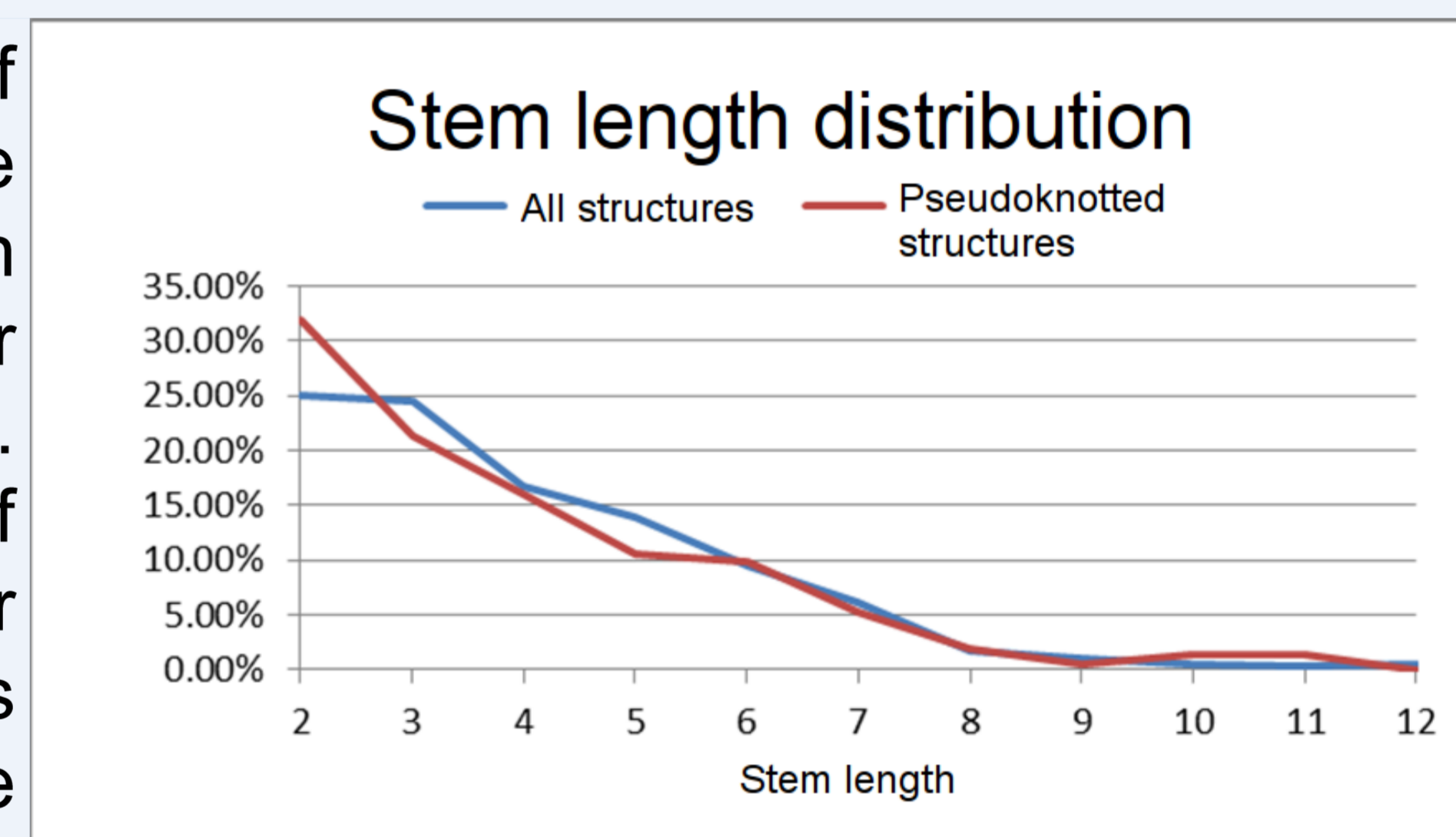


Figure 2.

## Results

The reduction has been conducted in terms of pseudoknot signatures. From each pseudoknot short stems have been excluded and its signatures have been updated via two stage reduction process (see Materials and methods section). Interchain stems have been excluded from interchain pseudoknots prior to short stems removal.

After the procedure almost all pseudoknots have been reduced to the two simplest types – either *H-knot* (signature **abAB**) or *kissing loops* (signature **abAcBC**). The only exception is a pseudoknot from Group II Intron from *Oceanobacillus iheyensis* with signature **abcdCefAFDEB**. This case will be further examined in more detail.

### Including short stems

Signature	Number of pseudoknots
abAB	3821
abAcBC	753
abAcDCDB	1
abcdBCAD	8
abcdCADB	14
abAcDBDeCE	14
abcdCABeDE	1
abAcDCeBEfDF	17
abAcDeBEfDFC	55
abcBdeDCAfEF	2
abcdCeAEfDFB	1
abcdCeBEAfDF	1
abcdCefAFDEB	6
abAcDCefDFgEGB	1
abAcDeDfgFEChBHIgI	530
abAcDeDfghiHFECjkGKIBLjMIM	8

### Excluding short stems

Signature	Number of pseudoknots
abAB	189
abAcBC	633
abcdCefAFDEB	4

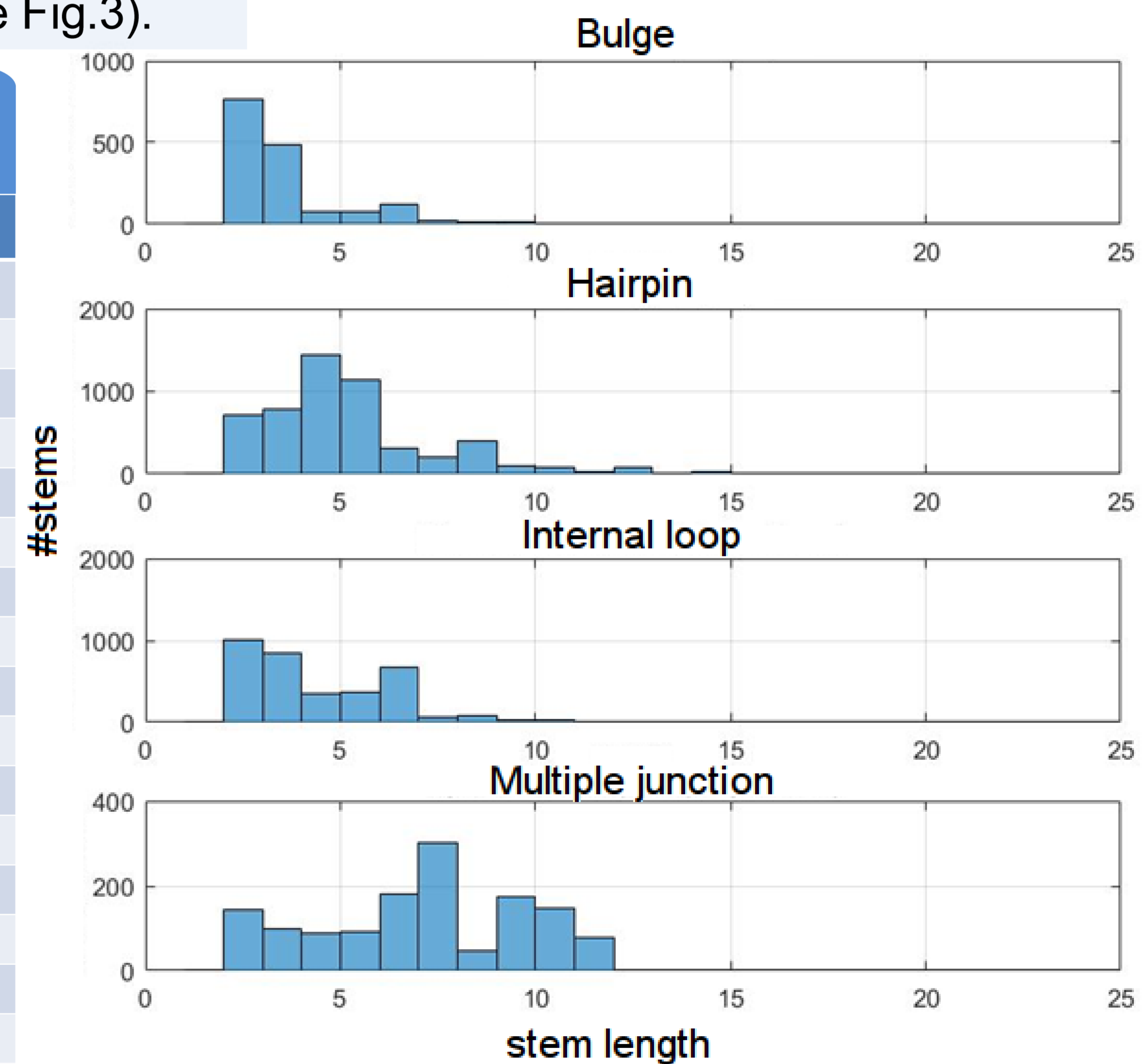


Figure 3.

## References

- [1] Baulin, Eugene, Victor Yacovlev, Denis Khachko, Sergei Spirin, and Mikhail Roytberg. "URS DataBase: universe of RNA structures and their motifs." *Database* 2016 (2016): baw085.
- [2] Berman, Helen M., John Westbrook, Zukang Feng, Gary Gilliland, Talapady N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. "The protein data bank." *Nucleic acids research* 28, no. 1 (2000): 235-242.
- [3] Tan, Zhijie, Wenbing Zhang, Yazhou Shi, and Fenghua Wang. "RNA folding: structure prediction, folding kinetics and ion electrostatics." In *Advance in Structural Bioinformatics*, pp. 143-183. Springer Netherlands, 2015.