Asymptotics of RNA Shapes:
A precise study of an alternative representation for the RNA secondary structure

Yann Ponty∗

Structural Bioinformatics Lab
Dept. of Biology – Boston College

May 2nd

Computational molecular biology is concerned with the development of mathematical models and novel algorithms to solve fundamental problems of molecular biology in the post-genome era. A central problem of structural biology concerns the algorithmic prediction of the structure of RNA and protein from only the nucleotide resp. amino acid sequence. In the context of RNA, nucleotide-level thermodynamical approaches allow for an already accurate prediction of the secondary structure. However, the native structure of an RNA is not necessarily that of Minimal Free Energy, but rather one of the suboptimals. Furthermore, the functional conformation of an RNA may not be unique, as in the case of riboswitches. Thus, approaches taking into account suboptimal structures have been developed, benefitting lately from the introduction, by Giegerich et al, of a new compact representation of the secondary structure – RNA shapes.

Giegerich et al use this representation in combination with advanced dynamic programming techniques to enumerate all the shapes compatible with a given sequence in the software RNAShapes. In order to give a bound on the applicability of the algorithm RNAshapes and as a preliminary step toward a statistical analysis on its output, we studied the asymptotic behavior of the expected number of shapes compatible with a sequence. Therefore, we used the DSV method that elegantly combines language theory modeling using grammars (at the core of successful approaches for the prediction problem, both in the context of RNA and Proteins), and singularity analysis for (almost) automatic estimates for the number of RNA Shapes, in different models. We find that the number of shapes compatible with a sequence grows exponentially with the size of that sequence, and give precise exponential constants for these growth. We also find a surprising one-to-one relationship between RNAShapes and Motzkin words.

∗ Joint work with Andy Lorenz and Peter Clote.