ABSTRACT (NONO SAHA CYRILLE MERLEAU)

Since the discovery of the structure of deoxyribonucleic acid (DNA) in the early 1953s, and its double-chained complement of information hinting at its means of replication, biologists have recognized the strong connection between molecular structure and function. In the past two decades, there has been a surge of research on an ever-growing class of ribonucleic acid (RNA) molecules that are non-coding but whose various folded structures allow a diverse array of vital functions. From the well-known splicing and modification of ribosomal RNA, non-coding RNAs (ncRNAs) are now known to be intimately involved in possibly every stage of DNA translation and protein transcription, as well as RNA signalling and gene regulation processes.

Despite the rapid development and declining cost of modern molecular methods, they typically can only describe ncRNA's structural conformations *in vitro*, which differ from their *in vivo* counterparts. Moreover, it is estimated that only a tiny fraction of known ncRNA has been documented experimentally, often at a high cost. There is thus a growing realization that computational methods must play a central role in the analysis of ncRNAs. Not only do computational approaches hold the promise of rapidly characterizing many ncRNAs yet to be described, but there is also the hope that by understanding the rules that determine their structure, we will gain better insight into their function and design. Many studies revealed that the ncRNA functions are performed by high-level structures that often depend on their low-level structures, such as the secondary structure. This thesis studies the computational folding mechanism and inverse folding of ncRNAs at the secondary level.

In this thesis, we describe the development of two bioinformatic tools that have the potential to improve our understanding of RNA secondary structure. These tools are as follows: (1) `RAFFT` for efficient prediction of pseudoknot-free RNA folding pathways using the fast Fourier transform (FFT); (2) `aRNAque`, an evolutionary algorithm inspired by Lévy flights for RNA inverse folding with or without pseudoknot (A secondary structure that often poses difficulties for bio-computational detection).

The first tool, `RAFFT`, implements a novel heuristic to predict RNA secondary structure formation pathways that has two com-

ponents: (i) a folding algorithm and (ii) a kinetic ansatz. When considering the best prediction in the ensemble of 50 secondary structures predicted by RAFFT, its performance matches the recent deep-learning-based structure prediction methods. RAFFT also acts as a folding kinetic ansatz, which we tested on two RNAs: the coronavirus frameshifting stimulation element (CFSE) and a classic bi-stable sequence. In both test cases, fewer structures were required to reproduce the full kinetics, whereas known methods (such as Treekin) required a sample of 20,000 structures and more.

The second tool, aRNAque, implements an evolutionary algorithm (EA) inspired by the Lévy flight, allowing both local global search, and which supports pseudoknotted target structures. The number of point mutations at every step of aRNAque EA is drawn from a Zipf distribution. Therefore, our proposed method increases the diversity of designed RNA sequences and reduces the average number of evaluations of the evolutionary algorithm. The overall performance showed improved empirical results compared to existing tools through intensive benchmarks on both pseudoknotted and pseudoknot-free datasets.

In conclusion, we highlight some promising extensions of the versatile RAFFT's method to RNA-RNA interaction studies. We also provide an outlook of both tools' implications in studying evolutionary dynamics.