Part I

RNA FOLDING

You can put some informational part preamble text here.

INTRODUCTION TO RNA FOLDING

The function of non-coding RNAs is largely determined by their three-dimensional structure [4]. For instance, we can analyze the catalytic function of ribozymes in terms of basic structural motifs, e.g. hammerhead or hairpin structures [11]. Other RNAs, like riboswitches, involve changes between alternative structures [65]. Therefore, understanding the relation sequence and structure is a central challenge in molecular biology. In the last 20 years, many different methods for determining the RNA structures of molecules have emerged: from experimental lab methods to computational approaches. For experimental lab methods, X-ray crystallography and the nuclear magnetic resonance (NMR) are the most accurate approaches to offer structural information at a single base-pair resolution. Both experimental methods are often characterized by high experimental cost and low throughput. In addition to those limitations, RNA molecules are volatile and difficult to crystallize. Despite the development of more sophisticated techniques to infer the state of nucleotides in RNA molecules using enzymatic [26, 64] or chemical probes [61, 66] coupled with next-generation sequencing [1, 60], most of them can only capture RNA structures *in vitro* which mostly differ from the *in vivo* structure conformations. Experimentally, only a tiny fraction of known ncRNAs has been determined [43]. Because measuring the structure of RNAs experimentally is very difficult and expensive, computational approaches play a central role in the analysis of natural RNAs [13, 53], and are an important alternative to experimental approaches. In nature, RNAs fold into secondary structures before folding into higher-level (tertiary and quaternary) structures [3, 62]. This separation of time scales justifies focusing on the prediction of secondary structure; evidence suggests that the resulting high-level structures are indeed largely determined by the RNA's secondary structures. This chapter provides an overview of computational folding methods of RNA molecules into a secondary structure. Two techniques will be reviewed: statistical approaches such as machine learning and score-based methods.

2.1   STABILITY AND PREDICTION OF RNA SECONDARY STRUCTURES

The mapping from RNA sequences to their corresponding secondary structure defines the folding of RNA molecules. RNA folding is, therefore, a process by which a linear RNA sequence acquires a secondary structure through intra-molecular interactions. The nature of those interactions defines the thermodynamic stability of the secondary structure. The thermodynamic stability ΔG_σ of a structure σ is the free energy difference with respect to the completely unfolded state. Therefore, the free energy function defines the mapping of an RNA sequence to its corresponding free energy. Most computational methods search for structures that minimize this free energy function to predict biologically relevant structures. Structures are decomposed into components called loops to compute the free energy (See Definition 4). The loop decomposition allows building the basis of the standard energy model for RNA secondary structures called the *nearest neighbour* model [63]. This model associates tabulated free energy values to loop types and nucleotide compositions; the `Turner2004` [35] and the `Turner1999` [36] are the most widely used parameter sets. The total free energy of a secondary structure is assumed to be a sum over its constituent loops according to the additivity principle [9]. This structure decomposition allows an efficient dynamic programming (DP) algorithm to determine the minimum free energy (MFE) pseudoknot-free structure of a sequence ϕ in the structure space Σ_ϕ.

The DP technique is one of the most widely used score-based methods. The first DP algorithm that finds the structure with the maximum base pairs was proposed by Nussinov, and Jacobson [40]. A few years after, Zucker and Stieger [76] extended Nussinov's algorithm to a more realistic scoring model based on free energy, the NN model. Almost all score-based methods rely on the same DP algorithm, but the decomposition scheme and the scoring model could defer from one to another. When predicting structures with non-canonical base pairs, some other scoring schemes are used, such as nucleotide cyclic motifs score system [8, 42, 55] or equilibrium partition function [56]. Besides the score-based methods, we have the comparative sequence analysis methods, the most computationally accurate for determining the RNA secondary structure [17, 33]. Using the set of homologous structures, This method allows finding base pairs that covary to maintain WC and wobble bases of a given sequence ϕ [18].

The first comparative method predicting a common secondary structure conserved in the given homologous sequence set was developed by Han and Kim in the early nineteenth, and it was based on the comparative phylogenetic analysis. When neglecting the special base pairs and the weak interactions, the running time of both approaches (sore-based and comparative analysis) is usually $O(L^3)$ (Where L is the RNA sequence length). Many other comparative analysis methods and variations of score-based methods were also proposed to improve the computational time. More recently, a heuristic method such as `LinearFold` allows achieving good RNA folding performance in a linear time ($O(L)$).

When pseudoknots are considered, the loop decomposition of a secondary structure and the energy rules break down. Although we can assign reasonable free energies to the helices in a pseudoknot and even to possible coaxial stacking between them, it is impossible to estimate the effects of the new kinds of loops created. Base triples pose an even greater challenge because the exact nature of the triple cannot be predicted in advance, and even if it could, we have no data for assigning free energies. Nevertheless, there are existing techniques that approximate the energies of pseudoknot loops and allow the dynamic programming technique to tackle the RNA folding with pseudoknots. However, the time complexity still reminds the main problem. Using the DP technique for the pseudoknot structure prediction, the time complexity goes up to $O(L^6)$ for the exact prediction. But for heuristic methods such as `IPKnot` [52] and `Hotknots` [45], the running time can be reduced down to $O(L^4)$.

Despite the advanced development of computational tools for RNA folding, it's challenging to understand the folding mechanism fully. In contrast to score-based and comparative analysis methods, machine learning methods are data-driven methods that require no knowledge of the folding mechanism. Nevertheless, the requirement of ML-based methods is a large amount of training data on which they can learn. In the last few decades, ML methods have been used for many aspects of RNA secondary structure prediction methods to improve the prediction performance and overcome the limitations of existing methods. However, they did not replace the mainstream score-based methods with respect to accuracy and generalization. In addition to some overfitting concerns, ML-based methods cannot give dynamic information on the RNA folding process since little data are available on structural dynamics. In addition, the training data used in ML-based methods are mostly obtained through phylogenetic

analyses. Consequently, their prediction may be biased due to the *in vivo* third elements. The following subsections provide a detailed description of some of the recent ML-based and score-based tools for secondary structure prediction.

2.1.1   *MFE prediction tools for pseudoknot-free RNA sequences using a score-base method*

The score-based methods often assume that the native or biological RNA structure is the one that minimizes/maximizes the overall total score, depending on the hypotheses made on the RNA folding mechanism. In the pseudoknot-free MFE prediction, where the special and weak interactions are neglected, the folding problem is less complex, and the scoring model is the free energy. Hence, the issue of RNA secondary structure prediction becomes an optimization problem that aims at finding the best-scoring structure $\mathcal{S}^{MFE}$ by minimizing a scoring function ΔG.

$$\mathcal{S}^{MFE} = \mathbf{argmin}_{\mathcal{S} \in \Sigma_{phi}} \Delta G(\mathcal{S}, \phi) \qquad (2.1)$$

Where Σ_ϕ is the set of all possible pseudo-knot free secondary structure for the sequence ϕ of lenght L and, $\Delta G(\mathcal{S}, \phi)$ the free energy of the structure $\mathcal{S}$ evaluated for the sequence ϕ.

Since each possible structure can be uniquely and recursively decomposed into smaller components (or loops) with independent free energy contributions, the DP is best suited for most of the following tools presented here.

- Unfold [75, 76]: It is the successor of the original mfold program which was the first realistic implementation of the DP for secondary structure predictions with a score based on the loop energy parameters and a worse case time complexity of $O(L^3)$. The initial version was an improvement of the simplest DP for secondary structure prediction known as the *maximum circular matching problem*. The authors demonstrated that the loop-based energy model is also amenable to the same algorithmic ideas. With McCaskill's algorithm [38], for computing the partition function of the equilibrium ensemble of RNA molecules, more efficient implementations of the initial program with accurate thermodynamic modelling have been provided. The latest implementation is known as Unfold.

- `RNAStructure` [37, 46]: The software first appeared in 1998 as a reimplementation of the program `mfold` with improved thermodynamic parameters. In its initial version, four major changes were made in `mfold`: (1) an improvement on the methods for forcing base pairs; (2) a filter that removed isolated WC or wobble base pairs has been added; (3) the energy parameter for interior, internal and hairpin loops were incorporated; (4) a new model for coaxial stacking of helices. It predicts the lowest free energy structure and a set of low energy structures. The new implementation also provided a user-friendly graphical interface for Windows operating system. Subsequently, the first implementation was extended to include biomolecular folding; an algorithm that finds low free energy structures common to two sequences; the partition function algorithm and all free energy structures, and the constraints with enzymatic data and chemical mapping data. The recent version includes the partition function computation for secondary structures common to two sequences and can perform stochastic sampling of common structures [19]. Additionally, it contains `MaxExpect`, which finds maximum expected accuracy structures [32], and a method for removal of pseudoknots, leaving behind the lowest free energy pseudoknot-free structure.

- `RNAfold` [21, 30]: It is one of the most used and efficient folding tools. It computes the MFE secondary structure using an efficient DP scheme and backtraces an optimum structure. It also allows computing the partition function using McCaskill's algorithm, the matrix of base pairing probabilities, and the centroid structure. It is part of the `ViennaRNA Package`. Since its first version, it aims at suggesting an efficient implementation of Zucker's algorithm with more flexibility on the folding constraints. Many other versions have been released, including a GPU implementation. The latest stable release of the `ViennaRNA Package` is Version 2.5.0.

- `LinearFold` [23]: For many decades, the DP techniques have been the most accurate and fast at predicting pseudoknot-free structure for short input RNA sequences. But for long sequences, the prediction remains challenging because of the computational time and the lack of accurate thermodynamic energy parameters. In contrast to traditional DP methods which are often bottom-up, `LinearFold` is a left-

to-right DP. The left-to-right DP consists of scanning the input RNA sequence ϕ from left to right, maintaining a *stack* along the way and performing one of the three actions (*push, skip* or *pop*). The *stack* consists of a list of unpaired opening bracket positions and at each position $j = 1 \ldots L$, the three actions consist respectively of 1)push: opening a bracket at position j, 2) skip: unpaired nucleotide at position j and 3) pop: closing the bracket at position j. Initially, LinearFold's computational time was similar to the classical DP ($O(L^3)$) because of the *pop* action that involves three free indices (i.e. unpaired positions). But using a beam search heuristic, the time complexity was then reduced to $O(Lb \log b)$, where b is the beam size. The beam search is a popular heuristic technique used in computational linguistics. This technique allows keeping only the top b highest-scoring (or low energy) states for each prefix of the input sequences.

Although the sore-based approaches for RNA structure prediction often offer good accuracy and generalization, the non-availability of the thermodynamic energy parameters for specific loops of extended sizes presents the main challenge for predicting long sequences (i.e. $L \geq 1,000$ nucleotides). Early ML-based methods aim to improve the energy parameters by learning the underlying folding patterns from a more considerable amount of training data. In the next section of this chapter, we will present some of the recent improvements in structure prediction using ML-based methods.

2.1.2   *ML-based methods*

The ML-based methods for RNA secondary structure prediction can generally be classified into three categories according to ML's subprocess, i.e., score scheme based on ML, preprocessing and postprocessing based on ML, and prediction process on ML. All the ML-based methods in these three categories trained their models in a supervised way [73].

When using a scoring scheme based on ML, the parameter estimation in the scoring scheme is first optimized using an ML model. The estimated parameters are then used to evaluate the scores of possible conformations. Difference scoring schemes can be refined by using that approach: the free energy parameters, weights, and probabilities. The free energy parameter-refining is the most popular because several thermodynamic parameters

of the NN model have to be based on a large number of optimal melting experiments and the experiments are time and labour-consuming. In fact, not all free energy changes in structural elements can be experimentally measured because of technical difficulties. Instead of refining the free energy parameters, some ML-based approaches scream through existing data of RNA structures to extract weights that consist of different features of RNA structure elements. These weights can be used as a scoring function for DP techniques. The advantage of such a scoring function is that it decouples structure prediction and energy estimation. However, learned weights have no explanations because of the ML black box.

Another alternative for predicting RNA structures is the stochastic context-free grammars (SCFG) [12, 27, 28, 48, 50, 67]. SCFGs allow building grammar rules and induce a join probability distribution over possible RNA structure for a given sequence ϕ. In addition, the SCFG models specify probability parameters for each production rule in the grammar, which allow assigning a probability to each sequence generated by the grammar. These probability parameters are learned from datasets of RNA sequences associated with known secondary structures without carrying any external laboratory experiments [12].

Besides the ML-based methods that focus on refining the folding parameters, there are preprocessing and post-processing based on ML [20, 22, 74] and direct predicting process based on ML [29, 58, 59]. Preprocessing and postprocessing models allow for choosing the appropriate prediction method or set of prediction parameter sets and provide a means of determining the most likely structures among the possible outcomes that are useful for decision. The preprocessing and postprocessing ML tools are often based on a support vector machine (SVM).

Finally, it is possible to use ML techniques to predict RNA secondary structure directly or combine it with other algorithms in an end-to-end fashion. Below are some of the most used and recent ML-based tools for RNA secondary structure prediction.

- ContraFold[10]: Using the so-called probabilistic model, the conditional log-linear model (CLLM), ContraFold appeared early in 2006 as the first probabilistic prediction tool outperforming all the existing tools including the thermodynamic tools such as RNAfold and mfold. The CLLMs is a flexible class of probabilistic models which generalize upon SCFGs by using discriminative training and feature-rich scoring. ContraFold implements a CLLM incorporat-

ing most of the features found in typical thermodynamic models allowing the tool to achieve the highest single sequence prediction accuracy to date, when compared with the currently available probabilistic models.

- ContextFold [68]: In contrast to ContraFold, ContextFold utilizes a weighted appraoch based on ML. In particular, it uses a discriminative structured-prediction learning framework combined to online learning algorithm. ContextFold uses a large training dataset of RNA sequences annotated with theirs corresponding structures to obtain a ML model made of $70,000$ free parameters which has several orders of magnitudes compared to traditional models (i.e. thermodynamic free energy parameters). At it first apparition, ContextFold's model succeeded at the error reduction of about 50% but to date, some over fitting concerns have been reported when using the tool, especially for the prediction of structures with large unpaired regions.

- Mxfold2 [51]: It is one of the most recent ML-based tools for predicting the secondary structure of RNA molecule. Its particularity is the ML technique used, a Deep Neural Network (DNN). it also belongs to the weighted approach based on ML since, the resulting model of a DNN is a set of weight parameters. MxFold2's DNN uses the max-margin framework with thermodynamic regularization, which made the folding scores predicted by MXfold2 and the free energy calculated by the thermodynamic parameters closer as possible. This method has shown robust prediction on both sequences and families of natural RNAs, which suggests that the weighted ML approaches can compensate the gaps in the thermodynamic parameter approaches.

2.1.3 *Prediction tools for pseudoknotted RNA sequences*

Folding RNA sequences with pseudoknotted interactions is computationally more expensive than a pseudoknot-free target because of the time complexity of the folding algorithms. Specifically, the time complexity of the pseudoknot-free secondary structure prediction is $O(L^3)$ when using dynamic programming approaches such as RNAfold, or less with heuristic folding methods (e.g. $O(L)$ for LinearFold and $O(L^2 \log L)$). By contrast, when considering a special class of pseudoknots, the time complexity

of folding goes up to $O(L^6)$ for an exact thermodynamic prediction using a dynamic programming approach such as [47]. When Using heuristic methods, the time complexity slows down to $O(L^4)$ (e.g. tools such as IPknot and HotKnots) or $O(L^3)$ for tool such as HFold. In this section, we introduction couple of tools for predicting RNA pseudoknotted structures.

- pKiss [25]: The program pKiss appears the first time in 2014 as an updated version of the program pknotsRG[44] which is a module of the RNA abstract shapes analysis RNAshapes [25]. Initially, the program pknotsRG was build for the prediction of some special class of pseudoknots (unknotted structures and H-type pseudoknots) and then, it was extended to predict RNA structures that exhibit kissing hairpin motifs in an arbitrarily nested fashion, requiring $O(L^4)$ time. In addition to predicting the kissing hairpin motifs, pKiss also provides new features such as: shape analysis, computation of probabilities, different folding strategies and different dangling base models.

- IPknot [52]: it was first introduced in a paper by Kengo and his collaborators in 2011 as a novel computational tool for predicting RNA secondary structure with pseudoknots using integer programming technique. IPknot uses the maximum expected accuracy (MEA) as scoring function and the maximizing expected accuracy problem is solved using an integer programming with threshold cut. In fact, IPknot decomposes a pseudoknotted structure into a set of pseudoknot-free substructures and approximates a base-pairing probability distribution that considers pseudoknots, leading to the capability of modeling a wide class of pseudoknots and running quite fast. In addition to single sequence analysis, IPknot can also predict the consensus secondary structure with pseudoknots when a multiple sequence alignment is given.

- HotKnots [45]. In contrast to the previously mentioned tools, HotKnots implements an heuristic algorithm based on the simple idea of iteratively forming stable stems. By iteratively forming stems, the algorithm explore many alternative secondary structures using a free energy minimization for pseudoknot-free secondary structures. For each structure formed at each step, several different additions of a single substructure are considered, resulting in a

tree of candidate structures. the criterion for determining which substructures to add to partially formed structures at successive levels of the tree was also new, relative to previous algorithms: energetically favorable substructures called "hotspots" are found by a call to Zuker's algorithm, with the constraint that no base already paired may be in the structure.

2.2 RNA KINETICS

The previous section introduced how secondary structures with theirs thermodynamic properties can be predicted. However, the methods used for predictions do not tell us nothing about how the structures change overtime and how they are related to each other. In the folling section, we will discuss the folding dynamics of RNA molecule.

The folding of RNA molecules is a remarkably more complex. It is a result of the delicate balance between multiple factors: the chain entropy, ion-mediated electrostatic interactions and solvation effect, base pairing and stacking, and other non-canonical interactions [6]. It is a dynamic process which is governed by is a constant formation and/or dissolving of base pairs. In other terms, the RNA molecule navigates its structure space by following a free energy landscape. The free energy landscape here is a high-dimensional space of all possible secondary structures (Γ) weighted by their free energy $E(\Gamma)$.

As usually done, the kinetics is modelled as a continuous-time Markov chain [31], where populations of structure evolve according to transition rates. In this context, an Arrhenius formulation

$$k_{i \to j} = k_0 \exp(-\beta \Delta E^{\ddagger}_{i \to j}) \tag{2.2}$$

is commonly used to derive elementary transition from state i to state j; where $\Delta E^{\ddagger}_{i \to j}$ is the activation barrier separating i from j, and $\beta = 1/k_B T$ is the inverse thermal energy (mol/kcal). Here k_0 is the fundamental rate constant, which is solvent dependent. Three rate models describing elementary steps in the structure space are often used to study RNAs folding dynamics:

1. The base stack model [69–71]: it uses base stacks as elementary kinetic move. A move consists of an addition or a breaking of a base stack with $\Delta E^{\ddagger}_{i \to j}$ equal to the change in

the entropic free energy $T\Delta S$ and the enthalpy ΔH, respectively.

2. The base pair model [7, 14]: It uses base pair as elementary kinetic steps which gives the finest resolution, but at the cost of computation time. Here $\Delta E^{\ddagger}_{i\to j} = \Delta E/2$ where ΔE is the energy change from state i to state j or $\Delta E^{\ddagger}_{i\to j} = \Delta E$ for $\Delta E \geq 0$.

3. The helix stem model [24, 34]: the elementary move is the creation or deletion of a helix stem. It provides a coarse-grained description of the dynamics where free energy changes ($\Delta E^{\ddagger}_{i\to j}$) due to stem formation guide the folding process.

The different rate models can lead to different folding pathways. The key factor that distinguishes the different rate models is whether the barrier is determined by $(\Delta H, \Delta S)$ or by ΔE. The $(\Delta H, \Delta S)$ values for different RNA base stacks show well-separated discrete hierarchies, whereas the ΔE values show no such large separation. For two typical base stacks, 5'AU-AU3' and 5'UC-GA3', the difference $\Delta(\Delta H_{stact}, \Delta S_{stack}) = (7.4\ \text{kcal/mol}, 20\ \text{kcal/mol})$ is much larger than the difference $\Delta(\Delta E_{stack}) = 1.4\text{kcal/mol}[54]$. Because of this fact, different models can give different folding kinetics.

Depending on the rate model used, the following master-equation describe the population kinetics $p_i(t)$ for the ith state ($i = 1 \dots \Omega$, where Ω is the total number of chain conformations).

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \sum_{j \in \Omega} k_{j\to i}p_j(t) - k_{i\to j}p_i(t), \tag{2.3}$$

where $k_{j\to i}$ and $k_{i\to j}$ are the rate constants for the respective transitions. The equivalent matrix form of Equation 2.3 is given by:

$$\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = \mathbf{M}.\mathbf{p}, \tag{2.4}$$

where $\mathbf{p} = (p_i, \dots p_{\Omega})$ is a column vector representing the frequency of structure at state $(i, \dots, \Omega)$ and, $\mathbf{M}$ is the rate matrix defined as:

$$\mathbf{M}_{ij} = \begin{cases} k_{i\to j}, & \text{if } i \neq j \\ -\sum_{j\neq i} k_{ij}, & \text{if } i = j \end{cases} \tag{2.5}$$

For a given initial folding condition $p_i(0)$, the Equation 2.4 is solvable by diagonalizing the rate matrix **M** and, the solution is the population kinetics $\mathbf{p}(t)$ for $t > 0$ is given by:

$$\mathbf{p}(t) = \sum_{m=1}^{\Omega} C_m \mathbf{n}_m \exp{-\lambda_m t} \tag{2.6}$$

where $-\lambda$ and $\mathbf{n}_m$ are the mth eigenvalue and eigenvector of the rate matrix **M**, and C_m is the coefficient that is dependent on the initial condition. The eigenvalue spectrum gives the rates of the kinetic modes of the system.

Simulating the RNA dynamics using the Equation 2.3 has some limitations. The solution to the master-equation given by Equation 2.6 can only give ensemble-average macroscopic kinetics and cannot give detailed information about the microscopic pathways [72]. Moreover the number of structures (Ω) increases rapidly with the RNA sequence length L. Therefore, the master equation is often limited to short RNA sequences. Because of these limitations a kinetics-cluster methods are alternatively used. The basic idea of the kinetic-cluster method is to classify the large structural ensemble into a much reduced systems of cluster (of macrostates) such that the overall kinetics can be represented by the inter-cluster transitions. Although both the master-equation and the kinetic-cluster methods can predict the macroscopic kinetics, and both are based on the complete conformational ensemble, the kinetic-cluster approach has the unique advantage of providing the direct information on the microscopic pathway statistics from the inter-cluster transitions [72]. An alternative approach, implemented in `kinwalker` [15], used the observation that folded intermediates are generally locally optimal conformations. Although the above mentioned theoretical models allow to simulate the dynamics of RNA folding molecules they often missed an important component, namely, the sequence-dependent conformational statics of the single-stranded coil and loop states. This important component has been suggested by several experimental studies [2, 39]. In folding experiments, Pan and coworkers observed two kinds of pathways in the free energy landscape of a natural ribozyme [41]. Firstly, the experiments revealed fast-folding pathways, in which a sub-population of RNAs folded rapidly into the native state. The second population, however, quickly reached metastable misfolded states, then slowly folded into the native structure. In some cases, these metastable states are functional. These phenomena are direct consequences of the

rugged nature of the RNA folding landscape [57]. The experiments performed by Russell and coworkers also revealed the presence of multiple deep channels separated by high energy barriers on the folding landscape, leading to fast and slow folding pathways [49]. The formal description of the above mechanism, called kinetic partitioning mechanism, was first introduced by Guo and Thirumalai in the context of protein folding [16]. In the free energy landscape, these metastable conformations form competing attraction basins in which RNA molecules are temporarily trapped. However, *in vivo*, folding into the native states can be promoted by molecular chaperones [5], which means that the active structure depends on factors other than the sequence. This may rise some discrepancy when comparing thermodynamic modelling to real data and the experimental verification of the rate model is also a challenge because the microscopic elementary processes are hidden in the ensemble averages of the measured kinetics. Many researchers believe that single-molecule experiments, with careful extrapolation to the force-free case, may provide a discerning measure. All-atom simulations with reliable force field and sampling method are highly valuable for providing detailed atomistic configurations for the transition state [6]. Alternatively, systematic theory- experiment tests as done in [72] for designed sequences can also provide critical assessment for the different rate models.

2.3 CONCLUSION

In this chapter, we have presented the RNA folding in two main steps: (1) the prediction of the secondary structure of RNA which represents the static part of the folding process; (2) the RNA kinetics which aim at modelling the dynamics of the folding. The prediction of RNA secondary structure was introduced as an optimization problem and a review on existing methods and tools have been presented. In the next chapter, we will introduce the first result of our thesis that aims at predicting efficiently RNA folding pathways using the fast Fourier transform. The predicted pathways will then allow us to derive a set of energetically suboptimal structures from which we will model the slow folding process of RNA molecules.

Part II

RNA DESIGN

Part III

DISCUSSION AND PERSPECTIVE

Part IV

APPENDIX