École polytechnique

Laboratoire d'informatique

Conception d'ARN induisant une séparation de phase liquide-liquide pour une libération programmable

Mostafa ELRAIES

Supervisor: Ariel LINDNER, Sarah BERKEMER & Yann PONTY

A report submitted as a fulfilment of the requirements of the doctoral school for the comite du suivi held by
Diego Javier ZEA & Zoher GUEROUI

Periode couverte par le compte-rendu: 1er février 2025 - Aujourd'hui
February 4, 2026

Chapter 1

Introduction

1.1 Introduction

One fundamental problem of cell biology is how biochemical processes are organized in space and time. One way to achieve this is through compartmentalization: concentrating reaction components in compartments known as organelles. However, in recent years, other mechanisms that answer this problem have been discovered that often do not rely on membrane-enclosed compartments. One such mechanism is **liquid–liquid phase separation** (LLPS), which enables the formation of dynamic bio-molecular condensates such as stress granules, nucleoli, and P-bodies. These condensates exhibit liquid-like behavior and allow the cell to concentrate locally RNAs and proteins in a reversible and tunable fashion. Compared to classic organelles, **membraneless organelles** do not have restricting membranes that separate the organelles' content from the rest of the cell. This gives flexibility and speed for biochemical reactions through rapid diffusion in an environment separated by a phase boundary.

Unlike DNA, **RNA** molecules are single stranded molecules that can form structures through base-pairing interactions between complementary regions. A given region of an RNA molecule could have a complementary motif to another region on the same strand or to a different one. This constitutes a form of competition between **intra-molecular** and **inter-molecular** interactions. Predicting the outcome of this competition remains a major open question, with broad implications for numerous phenomena observed in both in vivo and in vitro settings. The impact of this competition is especially pronounced in biological condensates, where RNA–RNA interactions play a significant, yet largely understudied role Kimchi et al. (2023). Although condensates usually form through RNA–protein interactions, researchers have observed that certain transcripts associated with repeat expansion disorders can aggregate solely through RNA-RNA interactions under both in vivo and in vitro conditions Kurokawa et al. (2023).

RNA has emerged as a key driver and regulator of LLPS. Not only can it scaffold interactions through its length and structure, but specific RNA motifs, including repeat expansions and structural features like hairpins and bulges, have been shown to modulate the physicochemical behavior of condensates. Recent advances in cryo-EM and in vitro reconstitution have revealed how *multivalency* and secondary structure enable RNA to promote or inhibit phase separation, suggesting exciting possibilities for biomolecular engineering Wang et al. (2025).

This PhD project investigates how **engineered RNAs** can be used to induce and control LLPS for synthetic biology applications, particularly in designing RNA constructs capable of *programmable molecular release*. The core objective is to understand the rules governing RNA-mediated phase behavior and use them to design systems that drive or inhibit droplet formation by varying parameters such as temperature, ions, or biochemecial environment.

From a computational perspective, we are developing a dynamic programming algorithm to model both intra- and inter-molecular RNA secondary structures, extending the classical thermodynamic model of Turner and Mathews (2009). This algorithm incorporate base-pairing energetics and loop penalties to predict interaction landscapes that are favorable to condensate formation. We leverage and interface with existing software libraries such as the ViennaRNA package to ensure compatibility with established realistic frameworks while enabling greater flexibility for our synthetic design goals Lorenz et al. (2011).

This report presents the progress achieved on:

- developing an algorithm to evaluate multistrand RNA interactions with realistic energy models,
- interfacing with an established package for extracting the energy parameters,
- and laying the groundwork for in vitro testing of engineered RNA droplets.

The project lies at the intersection of RNA bioinformatics, thermodynamic modeling, and synthetic biology, with the aim of advancing RNA-based molecular engineering via predictive design and experimental feedback.

1.2 Background

RNA molecules are increasingly recognized not only as genetic messengers but also as key regulators of cellular function through structural and spatial mechanisms. A growing body of evidence highlights their capacity to mediate the formation of membrane-less organelles via *liquid-liquid phase separation* (LLPS). This phenomenon enables the dynamic compartmentalization of biomolecules which play roles in cellular stress response and gene expression regulation, and may be involved in neurological function.

Recent experimental work by our collaborators has demonstrated that engineered RNAs can act as modular scaffolds for intracellular organization, recruiting and concentrating target molecules into programmable condensates Guo et al. (2022). These addressable phase-separated compartments offer a powerful paradigm for synthetic biology, enabling the control of reaction environments and molecular interactions within the cell.

However, the biophysical and computational modeling of such RNA-driven condensates remains an open challenge. Earlier work by our team proposed a first-generation model to explore sequence-encoded rules for RNA condensation using simplified constraints on base-pairing and repeat interactions Boehmer et al. (2025). While this study provided valuable biological insights, it did not integrate a full thermodynamic framework and did not account for the nuanced energy landscape of RNA folding and duplex formation.

1.2.1 Base Pair Maximization Model

The base pair maximization model predicts RNA secondary structure by maximizing the number of Watson–Crick (or wobble) base pairs in the fold. In this simplified model, each possible base pair contributes an equal "score" (typically $+1$ or a fixed energy change), and the algorithm seeks the structure with the greatest total pairs. The biological assumption is that more base pairs = more stability, treating all pairs equally and ignoring finer energetic details. This means a GC pair is valued the same as an AU or GU pair, and stacking interactions or loop lengths are not considered. Fig 1.1 shows a secondary structure predicted from such algorithms. As a result, the model can form unrealistic structures with many small loops, since it doesn't penalize large loops or reward contiguous stacks beyond the count of pairs. Often a minimum hairpin loop length is imposed as a constraint (since extremely short loops are physically impossible), but in its basic form the model does not require a minimum stem-loop size. Despite its simplicity, this model provides a fast baseline for secondary structure prediction and was historically the first efficient DP approach for RNA folding.

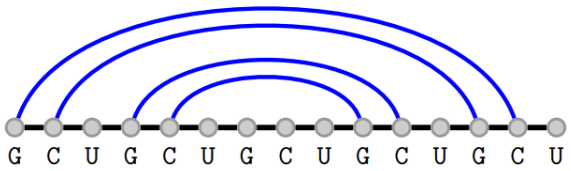

Figure 1.1: An optimal secondary structure based on base pair maximization model.

1.2.2 Turner Loop-Based Energy Model (Thermodynamic Nearest-Neighbor Model)

The Turner loop-based energy model predicts RNA secondary structure by minimizing the free energy of the fold, using an additive thermodynamic scoring of loops and base pair stacks. In this model (also known as the nearest-neighbor model), the free energy of an RNA structure is defined as the sum of contributions from individual loop motifs (hairpin loops, internal loops, bulges, multibranch loops) and stacked base pairs, based on experimentally measured parameters. Fig 1.2 shows the different types of loop motifs: hairpin loops (a stretch of unpaired bases closed by one base pair; yellow), stacked base pairs (two consecutive base pairs with no unpaired bases between them; blue), an interior loop (two base pairs separated by unpaired bases on both sides of the loop; purple), a bulge loop (two base pairs separated by unpaired bases on only one side of the loop; orange), a multiloop (three or more base pairs; green), and an exterior loop (the loop containing the two ends of the strand; gray).

Biologically, this model assumes that an RNA will adopt the structure with the minimum Gibbs free energy (MFE) at equilibrium, and that the nearest-neighbor interactions (loops and stacks) dominate the energetics of secondary structure. It also encodes known constraints (e.g. a hairpin loop must be at least 3 nt long for stability, since shorter loops are extremely unfavorable and are assigned a very high positive energy or disallowed entirely). Like the simpler model, crossing base pairs (pseudoknots) are excluded (the energy model is typically applied only to pseudoknot-free structures, which allows a neat loop decomposition). Overall, the Turner model is far more realistic, as it captures

the fact that not all base pairs are equal – context (neighbors and loops) matters – and it significantly improves prediction accuracy (MFE structures from this model are more often correct than maximum-pair structures).

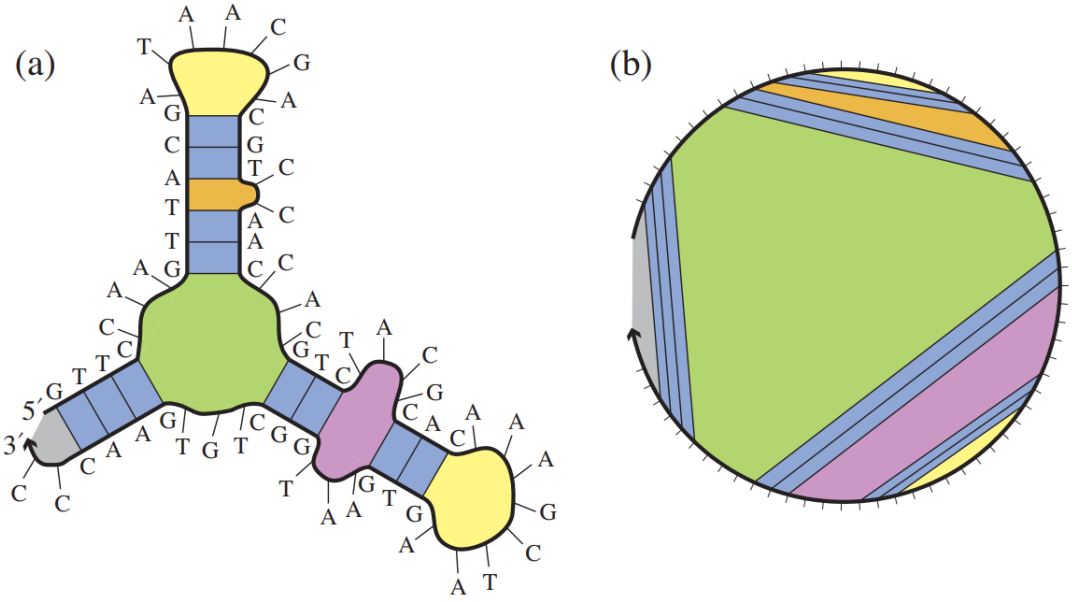

Figure 1.2: Secondary structure model for a single nucleic acid strand. (a) A sample secondary structure with the strand depicted as a directed thick line (an arrow marks the 3 end), base pairs depicted as thin lines joining complementary bases, and unpaired bases depicted as thin protruding lines. (b) An equivalent polymer graph representation, with the strand depicted as a directed thick circular arc, bases depicted as protruding tick marks, base pairs depicted as straight lines joining complementary bases, and different kind of loops colored Dirks et al. (2007).

1.3 Definitions and Problem Statement

1.3.1 Definitions

RNA Sequence and Structure. An RNA sequence is a word $s \in \{A, C, G, U\}^+$, where $|s|$ denotes its length and s_i its i-th base. A (pseudoknot-free) secondary structure over s is a set of unordered base pairs $\{i, j\}$ such that:

- Each base pair is valid: $\{s_i, s_j\} \in \mathcal{P}$, where $\mathcal{P} = \{\{A, U\}, \{C, G\}, \{G, U\}\}$;
- Each base is involved in at most one base pair;
- The structure is pseudoknot-free: if $\{i, j\}, \{k, \ell\} \in S$ and $i < k < j < \ell$, this configuration is disallowed;
- Each base pair spans at least θ unpaired bases: $|j - i| > \theta$, with $\theta := 3$ by default.

We denote the set of all such structures over s by $\Omega(s)$. Given an energy model $E : \{A, C, G, U\}^+ \times \Omega \to \mathbb{R}$, such as the Turner nearest-neighbor model, a free energy is assigned to each structure.

Multi-strand and Interaction Settings. Let $\mathcal{R} = \{r_1, \ldots, r_p\}$ be a multiset of RNA strands. A secondary structure $S \subset \mathcal{R}$ is a set of base pairs (inter- or intra-strand) that satisfies:

- All base pairings are thermodynamically valid;
- Intra-strand pairings are pseudoknot-free;
- The number of connected components in the structure (i.e., complexes) contributes an energy penalty: one K_{assoc} for each inter-strand interaction. (Note: This penalty represents the limitation that govern the competition between the intramolecular and intermolecular interactions.)

We follow a circular permutation model to to ensure that structural configurations are correctly identified regardless of strand ordering, as in Boehmer et al. (2025) (submitted), allowing general structures across permuted strand orders.

1.3.2 Problem Formulation

Building upon prior work by our team Boehmer et al. (2025), we aim to refine the computational prediction of RNA-driven liquid-liquid phase separation (LLPS) by accounting for true thermodynamic parameters. Our setting extends the classical RNA interaction models by introducing the following problem.

Strand Soup Minimum Free Energy (MFE) Problem

Input:

- A pool of RNA sequences $\mathcal{R} = \{r_1, \ldots, r_p\}$;
- A total number $m \in \mathbb{N}$ of interacting strands;
- A thermodynamic model E (e.g., Turner 2004).

Output:

$$E^*(R) = (m-1) \cdot K_{\text{assoc}} + \min_{r \in R - \{s\}, c \in \{0,1\}} M_{R, s_1, r_{|r|}, c} \tag{1.1}$$

This setting captures the combinatorics and thermodynamics of strand assembly in contexts where multiple copies of structured RNAs form dynamic assemblies. Our goal is to create an algorithmic framework that predicts such interactions, grounded in physical accuracy, and applicable to synthetic LLPS systems such as those studied by our experimental team.

1.4 Objectives and Approach

1.4.1 Scientific Objectives

The overarching aim of this project is to develop computational methods to accurately model RNA-driven liquid-liquid phase separation (LLPS), accounting for both the structural complexity and thermodynamic interactions of multiple RNA strands.

More specifically, we aim to:

- Develop a thermodynamically accurate model for predicting RNA secondary structures in multistrand contexts.
- Incorporate Turner-style loop-based energy models to improve on earlier simplifications used in initial work.
- Design and implement a dynamic programming algorithm capable of efficiently searching the space of RNA interaction structures in a strand soup model.
- Interface these predictions with synthetic biology experiments to validate LLPS behavior in vitro and in vivo.

1.4.2 Methodological Overview

Our methodology combines computational modeling with formal algorithmic design. We build upon prior models that used simplified base-pair counting Boehmer et al. (2025), replacing them with a Turner-style nearest-neighbor model.

To achieve this, we design a custom dynamic programming recurrence that accounts for:

- Hairpin, interior, and multibranch loop contributions using ViennaRNA low-level energy evaluation.
- A separation of folding recursions for single strands and interacting strand pairs.
- Matrix filling schemes that distinguish single-strand recursions (e.g., `C`, `M`, `F`) from multi-strand assembly evaluation.

We further aim to explore combinatorial sampling and probabilistic extensions (partition function-based models) in future phases of the project.

Chapter 2

Algorithmic Approach

To model and predict RNA interaction structures in the strand soup setting, we extend a dynamic programming algorithm originally based on simple base-pair counting to incorporate full loop-based thermodynamic energy contributions. Our model aims to compute the minimum free energy (MFE) secondary structure over a fixed number of RNA strands sampled from a multiset (or "soup"), where strand concentration and connectivity are explicitly considered.

To that end, we define a set of recurrence relations over various dynamic programming matrices. Each subproblem corresponds to a class of interaction substructures, and the solution represents the minimum energy over these conformations under the Turner energy model. Our formulation explicitly tracks pairing configurations, multiloop contributions, and handles pseudoknot-free constraints.

We describe the algorithm in increasing complexity: starting with base pair counting, we successively incorporate energy penalties for internal loops, multiloops, and exterior loops, ensuring unambiguity and correctness in the recursion structure. The energy contributions are parameterized using constants (a, b, c) corresponding to multiloop initiation, branching, and unpaired base penalties, respectively.

2.1 Definitions

Connectivity: Two bases s_i and r_j are connected if there exists a path in the polymer graph excluding cross-strand edges. A secondary structure is called *connected* if every strand in R is connected to every other. In the dynamic programming model, a bit $c \in \{0, 1\}$ is used to indicate whether the substructure formed by a subproblem is connected ($c = 1$) or not ($c = 0$). This is essential for distinguishing external loops from multiloops in the recurrence relations.

Soup setting: In the RNA soup setting, we are given a multiset R of strands from which a fixed number m are selected (with replacement) to form an interaction complex. Each structure in this setting corresponds to a connected configuration of m strands from R, and we compute the minimum free energy (MFE) across all such valid combinations.

$$\mathrm{MFE}(R, m) = \min_{t_1,\ldots,t_m \in R} \min_{S \in \Omega(\{t_1,\ldots,t_m\})} E(S) \tag{2.1}$$

Subproblem types: We associate different dynamic programming tables with structural contexts:

- $F[s, i, r, j, m, c]$ (Figure 2.1): energy of the optimal structure from position i in strand s to position j in strand r using m strands, with connectivity bit c.
- $C[s, i, r, j, m]$(Figure 2.2): optimal structure assuming the outermost base pair is (s_i, r_j) and the structure is closed.
- $M[s, i, r, j, m]$($Figure$ 2.3): structure part of a multiloop (inner base pair), containing penalties for unpaired bases and nested stems.

These are refined and recombined to score all structure types: external loops, interior loops, and multiloops.

2.2 Full loop-based energy (unambigous)

Here, we want to additionally score interaction-loops with internal structure. To illustrate with examples, consider the following interaction conformations (interaction base pairs indicated by square brackets for clarity)

```
[...[ & ].]
[...[ & ]...(    ).]
[.(  )..[ & ]..[ & ]...(   ).]
[[ & ] & (  ) ]
```

In the first case, the interaction loop is a regular interior loop, while the in the second cases, the interactions enclose internal structure forming a multi-loop with 3 stems. The third case, is a more complex ML with 5 stems / 4 inner base pairs. In the last example, the loop of the outer base pair is not closed, such that there is no energy contribution.

2.2.1 Redefinition of F, such that we can guarantee disconnect of outermost fragments.

As preparation for a non-ambiguous, correct treatment of interaction base pairs that close external loops, we want to determine the best energy of general soup sub-interactions that would be connected *only* by a pseudo-base pair connecting the first and last strand. The previous definitions Boehmer et al. (2025) were less strict, such that even for c=0 the sub-structures could be completely connected without the pseudo-base pair.

c=0 means no transitive connection between s and r, i.e. the pseudo-base pair between first and last strand is necessary to connect the substructures of the subproblem.

F![s,i,r,j,m,c] is responsible for the introduction of new strands whenever those identified by s and r have been entirely consumed:

$$F[s,i,r,j,m,c] = \min \begin{cases} F![s,i+1,r,j,m,c] & \triangleleft\ i \text{ unpaired} \\ \min_{i<k\le|s|} C_s[s,i,k] + F![s,k+1,r,j,m,c] & \triangleleft\ i \text{ pairs with first strand} \\ \min_{1\le k\le j} C[s,i,r,k,m] + F_s[r,k+1,j] \quad \text{if } c=1 & \triangleleft\ i \text{ pairs with last strand} \\ \min\limits_{2\le m'<m-1, t, 1\le k\le|t|} \begin{pmatrix} C[s,i,t,k,m'] \\ +F![t,k+1,r,j,m-m'+1,c] \end{pmatrix} & \triangleleft\ i \text{ pairs with new strand} \end{cases} \tag{2.2}$$

$$F![s,i,r,j,m,c] = \begin{cases} \begin{cases} \infty & c=1 \\ \min_t F[t,1,r,j,m-1,1] & c=0, m>2 \quad i=|s|+1 \\ F_s[r,1,j] & c=0, m=2 \end{cases} \\ \begin{cases} \infty & c=1 \\ \min_t F[s,i,t,|t|,m-1,1] & c=0, m>2 \quad j=0 \\ F_s[s,i,|s|] & c=0, m=2 \end{cases} \\ F[s,i,r,j,m,c] \qquad \text{otherwise} \end{cases} \tag{2.3}$$

$$C_s[s,i,j] = \text{Zuker cases as per the Vienna package} \tag{2.4}$$

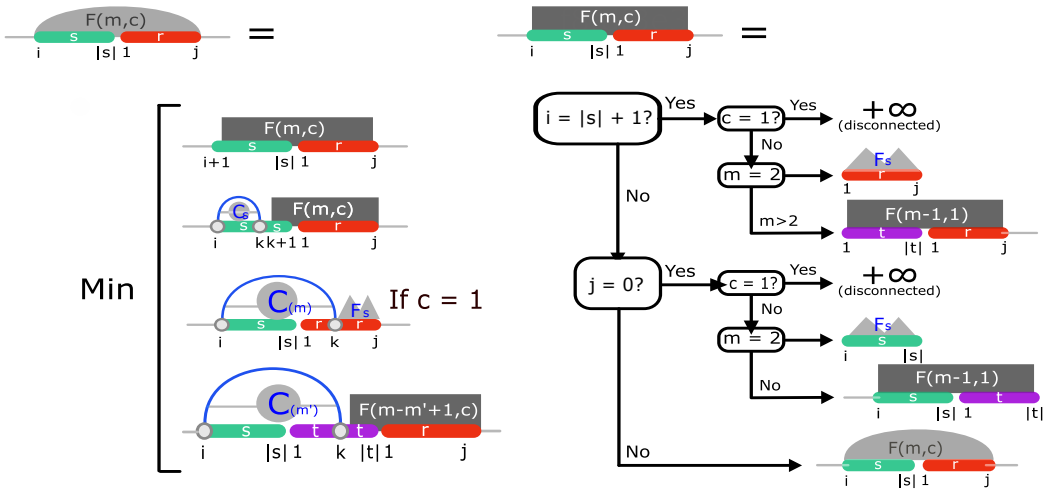

Figure 2.1: The matrix F computes the minimum free energy over all secondary structures formed by a sequence of m RNA strands. The recursion branches based on whether the first base is unpaired or forms a base pair, with energy contributions depending on the loop context. The connectivity bit c ensures the resulting structure forms a connected complex.

Closed structures (redefined). In the redefinition of $C[\cdot]$ for closed interaction structures, we need subproblems 'M' for true parts of a multi-loop (definitions follow below).

$$C[s,i,r,j,m] = \min \begin{cases} \min_{i<i'\le|s|',1\le j'<j} C[s,i',r,j',m] + \mathcal{I}(s,i,i',r,j',j) & \triangleleft \text{ interior loop} \\ \min_{i<k\le|s|} a + M^1_s[s,i+1,k] + M![s,k+1,r,j-1,m] & \triangleleft \text{ ML, leftmost stem in } s \\ \min\limits_{1\le k<j-1} \begin{pmatrix} a + M^1[s,i+1,r,k,m] \\ +M_s[r,k+1,j-1] \end{pmatrix} & \triangleleft \text{ ML, leftmost with rightmost strand} \\ \min\limits_{2\le m'<m-1,t,1\le k\le|t|} \begin{pmatrix} a + M^1[s,i+1,t,k,m'] \\ +M![t,k+1,r,j-1,m-m'+1] \end{pmatrix} & \triangleleft \text{ ML, leftmost stem with new strand} \\ F![s,i+1,r,j-1,m,0] & \triangleleft \text{ external loop (disconnect)} \end{cases} \tag{2.5}$$

$$M^1_s[s,i,j] = \min \begin{cases} \infty & i \ge j \\ M^1_s[s,i+1,j] + c & i < |s| \\ C_s[s,i,j] + b \end{cases} \tag{2.6}$$

$$M^1[s,i,r,j,m] = \min \begin{cases} M^1[s,i+1,r,j,m] + c & i < |s| \\ C[s,i,r,j,m] + b \end{cases} \tag{2.7}$$

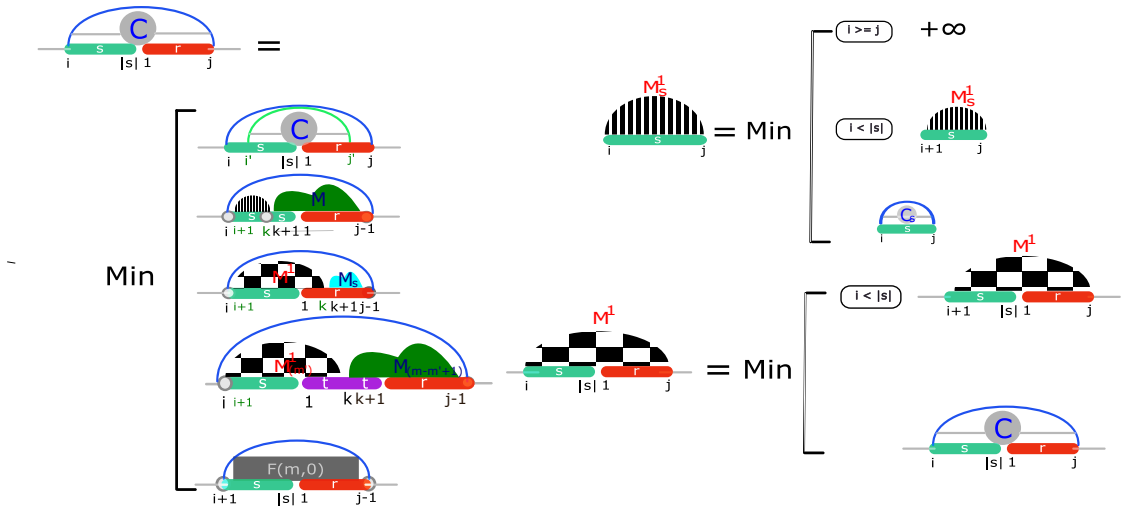

Figure 2.2: Dynamic programming scheme for C Matrix. C captures the energy of closed structures, i.e., substructures where the first and last bases are paired. It decomposes over loop types, including interior and multiloops, and serves as a key component for recursive energy evaluation in more complex models.

Parts of a multiloop. Finally, $M[\cdot]$ defines parts of a multiloop formed interaction, consecutively connected strands.

$$M[s,i,r,j,m] = \min \begin{cases} M![s,i+1,r,j,m] + c & \triangleleft\ i \text{ unpaired} \\ \min_{i<k\le|s|} C_s[s,i,k] + b + M![s,k+1,r,j,m] & \triangleleft\ i \text{ pairs with first strand} \\ \min_{1\le k<j} C[s,i,r,k,m] + b + M_s[r,k+1,j] & \triangleleft\ i \text{ pairs with last strand} \\ \min\limits_{2\le m'<m-1,t,1\le k\le|t|} \begin{pmatrix} C[s,i,t,k,m'] + b \\ +M![t,k+1,r,j,m-m'+1] \end{pmatrix} & \triangleleft\ i \text{ pairs with new strand} \end{cases} \tag{2.8}$$

With $M![\cdot]$ handling the specific border cases for $M[\cdot]$, where we guarantee that consecutive interacting RNAs must be connected:

$$M![s,i,r,j,m] = \begin{cases} \infty & i = |s|+1 \text{ or } j = 0 \\ M[s,i,r,j,m] & \text{otherwise} \end{cases} \tag{2.9}$$

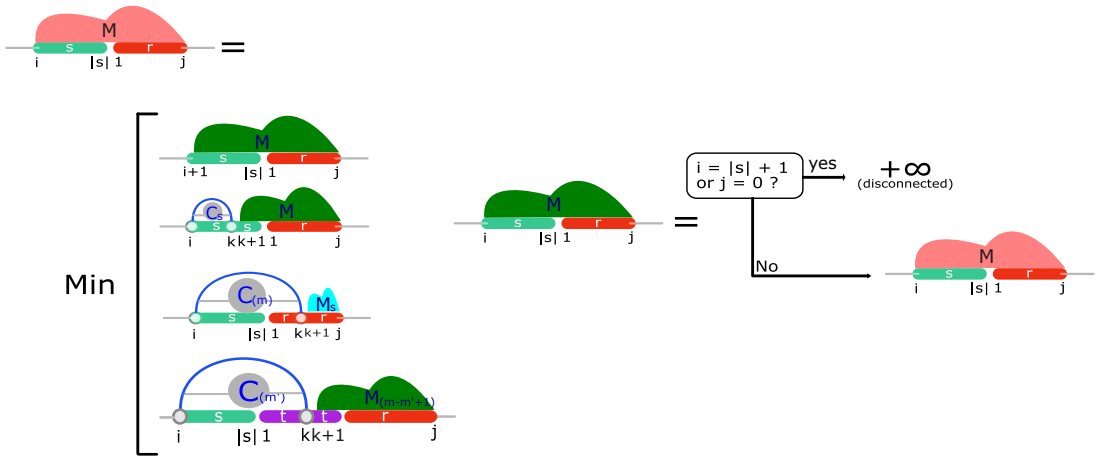

Figure 2.3: Matrix M handles multiloop components—substructures that lie within the boundaries of a multiloop closed by another pair. The recursion accounts for unpaired bases and nested stems, adding appropriate loop initiation and branch penalties.

Complexity of the algorithm MFE of Strand Soup Interaction based on Turner Energy model over m strands, taken from a collection of p sequences with n being the number of nucleotides of the longest strand, can be solved in time $\mathcal{O}\left(n^3 \cdot m^2 \cdot p^3\right)$.

Chapter 3

Implementation and Future Work

The dynamic programming model for RNA strand soup interactions described in the previous section is currently almost finalized in C++, leveraging low-level energy evaluation functions from the ViennaRNA package Lorenz et al. (2011). The model emphasizes modularity and performance, separating the concerns of energy computation from matrix recurrence management. A wrapper interface is being developed to abstract access to Turner model energy parameters Turner and Mathews (2009), allowing flexible evaluation of substructures such as interior loops, multiloops, and exterior loops.

Limitations

We started the project quite late (Feb 2025) which limited us in terms of what we can achieve in the first year of PhD. We succeeded in extending the funding period, so everything shall proceed according to the original plan with the next semester being more experimental wet lab work than computational work solely.

Future Plans

In the upcoming stages of the project, the focus will shift toward:

- **Extension to probabilistic settings**: Incorporating a partition function code and ensemble statistics (e.g., base-pairing probabilities), following the Boltzmann distribution formalism.
- **Integration with experimental data**: Using synthetic RNA constructs designed by previous experimental team members to test the predictive capabilities of the algorithm in real-world droplet formation contexts (this is part of the experimental work of my PhD).
- **Visualization tools**: Developing utilities to represent predicted secondary structures and interaction topologies graphically, easing interpretation and debugging.

The long-term goal is to build a complete in silico framework that supports the design and simulation of synthetic RNA systems exhibiting liquid–liquid phase separation (LLPS), guiding both wet-lab validation and future synthetic biology applications. RNA systems will be engineered to allow programmable release of recruited molecules in response to environmental triggers, such as changes in pH, salt concentration, and temperature. This capability we aim to confirm through biophysical simulations (run by our collaborator prof. Alessandro BARDUCCI) and experimental validations. Ultimately, the project aims to establish Transcriptionally Engineered Addressable RNA Solvent (TEARS) as multifunctional synthetic organelles that can recruit both RNA modification enzymes and specific RNA molecules, facilitating targeted molecular compartmentalization and biochemical reactions within distinct cellular environments. This will enhance our understanding of phase separation dynamics and contribute significantly to advancements in synthetic biology, particularly in gene regulation and therapeutic delivery.

References

Boehmer, K., Berkemer, S. J., Will, S. and Ponty, Y. (2025), Rna triplet repeats: Improved algorithms for structure prediction and interactions. Manuscript submitted for publication.

Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E. and Pierce, N. A. (2007), 'Thermodynamic analysis of interacting nucleic acid strands', *SIAM Review* **49**(1), 65–88.
URL: *https://doi.org/10.1137/060651100*

Guo, H., Ryan, J. C., Song, X., Mallet, A., Zhang, M., Pabst, V., Decrulle, A. L., Ejsmont, P., Wintermute, E. H. and Lindner, A. B. (2022), 'Spatial engineering of *E. coli* with addressable phase-separated rnas', *Cell* **185**(20), 3823–3837.e23.
URL: *https://doi.org/10.1016/j.cell.2022.09.016*

Kimchi, O., King, E. M. and Brenner, M. P. (2023), 'Uncovering the mechanism for aggregation in repeat expanded rna reveals a reentrant transition', *Nature Communications* **14**(1), 332.
URL: *https://doi.org/10.1038/s41467-023-35803-x*

Kurokawa, R., Kurokawa, M., Mitsutake, A., Nakaya, M., Baba, A., Nakata, Y., Moritani, T. and Abe, O. (2023), 'Clinical and neuroimaging review of triplet repeat diseases', *Japanese Journal of Radiology* **41**(2), 115–130.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F. and Hofacker, I. L. (2011), 'Viennarna package 2.0', *Algorithms for Molecular Biology* **6**(1), 26.
URL: *https://doi.org/10.1186/1748-7188-6-26*

Turner, D. H. and Mathews, D. H. (2009), 'Nndb: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure', *Nucleic Acids Research* **38**(suppl_1), D280–D282.
URL: *https://doi.org/10.1093/nar/gkp892*

Wang, L., Xie, J., Gong, T., Wu, H., Tu, Y., Peng, X., Shang, S., Jia, X., Ma, H., Zou, J., Xu, S., Zheng, X., Zhang, D., Liu, Y., Zhang, C., Yongbo, L., Huang, Z., Shao, B., Ying, B. and Su, Z. (2025), 'Cryo-em reveals mechanisms of natural rna multivalency', *Science* .