M2 BIM - Lecture 1
Folding RNA *in silico*

Yann Ponty

Bioinformatics Team
École Polytechnique/CNRS/INRIA AMIB – France


December 1\textsuperscript{st}, 2015
Outline

1 Introduction
- Dynamic programming 101
- Why RNA?
- RNA folding
- RNA Structure(s)
- Some representations of RNA structure

2 Some flavours of folding prediction
- Thermodynamics vs Kinetics
- Dynamic programming: Reminder

3 Free-energy minimization
- Nussinov-style RNA folding
- Turner energy model
- MFold/Unafold
- Performances and the comparative approach
- Towards a 3D ab-initio prediction
... or how to make a million bucks by giving change parsimoniously!!

**Problem:** You have access to unlimited amount of 1, 20 and 50 cents coins. A client prefers to travel light, i.e. to minimize the #coins. How to give N cents back in change without losing a customer?

**Strategy #1:** Start with heaviest coins, and then complete/fill-up with coins of decreasing value.

\[
21 = ??
\]

55

60
…or how to make a million bucks by giving change parsimoniously!!

**Problem:** You have access to unlimited amount of 1, 20 and 50 cents coins. A client prefers to travel light, i.e. to **minimize the #coins**. How to give N cents back in change without losing a customer?

**Strategy #1:** Start with *heaviest* coins, and then complete/fill-up with coins of *decreasing* value.

\[
21 = \boxed{20} + \boxed{1} \\
55??
\]

\[
60
\]
...or how to make a million bucks by giving change parsimoniously!!

**Problem:** You have access to unlimited amount of 1, 20 and 50 cents coins. A client prefers to travel light, i.e. to **minimize the #coins**. How to give N cents back in change without losing a customer?

**Strategy #1:** Start with *heaviest* coins, and then complete/fill-up with coins of *decreasing* value.

\[
21 = \text{20} + \text{1} \\
55 = \text{50} + \text{5} + \text{5} + \text{5} + \text{5} + \text{5} \\
60??
\]
... or how to make a million bucks by giving change parsimoniously!!

**Problem:** You have access to unlimited amount of 1, 20 and 50 cents coins. A client prefers to travel light, i.e. to **minimize the #coins**. How to give \( N \) cents back in change without losing a customer?

**Strategy #1:** Start with *heaviest* coins, and then complete/fill-up with coins of *decreasing* value.

\[
21 = \bigcirc + \bigcirc \\
55 = \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc \\
60 = \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc + \bigcirc \]

Problem *a priori* (?!?) non-solvable using such a *greedy* approach, as a (simpler) problem is already NP-complete (thus Efficient solution \( \Rightarrow 1 \text{M$} \)).
or how to make a million bucks by giving change parsimoniously!!

**Problem:** You have access to unlimited amount of 1, 20 and 50 cents coins. A client prefers to travel light, i.e. to minimize the #coins. How to give N cents back in change without losing a customer?

**Strategy #1:** Start with heaviest coins, and then complete/fill-up with coins of decreasing value.

\[
\begin{align*}
21 &= 1 \times 20 + 1 \times 1 \\
55 &= 1 \times 50 + 3 \times 5 + 1 \times 1 \\
60 &= 1 \times 50 + 5 \times 5 + 1 \times 1 \\
\end{align*}
\]

Problem *a priori* (?) non-solvable using such a *greedy* approach, as a (simpler) problem is already NP-complete (thus Efficient solution ⇒ 1M$).
Strategy #2: Brute force enumeration $\rightarrow$ #Coins$^N$ (Ouch!)

Strategy #3: The following recurrence gives the minimal number of coins:

$$Min\#\text{Coins}(N) = \begin{cases} 1 + Min\#\text{Coins}(N - 1) \\ 1 + Min\#\text{Coins}(N - 20) \\ 1 + Min\#\text{Coins}(N - 50) \end{cases}$$

With some memory ($N$ intermediate computations), the minimum number of coins can be obtained after $N \times \#\text{Coins}$ operations. An actual set of coins can be reconstructing by tracing back the choices performed at each stage, leading to the minimum.

Remark: We still haven’t won the million, as $N$ has exponential value compared to the length of its encoding, so the algorithm does not qualify as efficient (i.e. polynomial).

Still, this approach is much more efficient than a brute-force enumeration: $\Rightarrow$ Dynamic programming.
Strategy #2: Brute force enumeration → #Coins^N (Ouch!)

Strategy #3: The following recurrence gives the minimal number of coins:

\[
\text{Min}\#\text{Coins}(N) = \begin{cases} 
1 + \text{Min}\#\text{Coins}(N - 1) \\
1 + \text{Min}\#\text{Coins}(N - 20) \\
1 + \text{Min}\#\text{Coins}(N - 50)
\end{cases}
\]

With some memory (N intermediate computations), the minimum number of coins can be obtained after N \times \#Coins operations. An actual set of coins can be reconstructing by tracing back the choices performed at each stage, leading to the minimum.

Remark: We still haven’t won the million, as N has exponential value compared to the length of its encoding, so the algorithm does not qualify as efficient (i.e. polynomial).

Still, this approach is much more efficient than a brute-force enumeration: ⇒ Dynamic programming.
Fundamental *dogma* of molecular biology

DNA
{A, C, G, T}*

{A, C, G, T}
Fundamental *dogma* of molecular biology

- **DNA**: \{A, C, G, T\}*
- **RNAs**: \{A, C, G, U\}*
- **Pol**: Polymerase
- **RNAs**: \{A, C, G, U\}*

The **code** (genes) 

The **machine** (enzymes)
Fundamental *dogma* of molecular biology

DNA
\{A, C, G, T\}*

RNAs
\{A, C, G, U\}*

THE CODE (genes)

THE MACHINE (enzymes)

Amino acids

Pol

Yann Ponty

M2 BIM - Lecture 1 - RNA folding
Fundamental *dogma* of molecular biology

DNA
\{A, C, G, T\}*

RNAs
\{A, C, G, U\}*

THE CODE (genes)

THE MACHINE (enzymes)

\[\text{Yann Ponty} \quad \text{M2 BIM - Lecture 1 - RNA folding}\]
Fundamental *dogma* of molecular biology

DNA
{A, C, G, T}*

RNAs
{A, C, G, U}*

\[\begin{align*}
\text{ATGGTTACCCAT} \\
\text{Pol} \\
\text{AUG} \\
\end{align*}\]
Fundamental *dogma* of molecular biology

DNA
\{A, C, G, T}\ *

RNAs
\{A, C, G, U\} *

THE CODE (genes)
THE MACHINE (enzymes)

\[\text{ATGGTTACCCAT}\]

Pol

\[\text{AUUGGUUUACCCAU}\]
Fundamental *dogma* of molecular biology

DNA
\{A, C, G, T\}*

RNAs
\{A, C, G, U\}*

THE CODE (genes)
THE MACHINE (enzymes)
Fundamental *dogma* of molecular biology

- **DNA** \{A, C, G, T\}*
- **RNAs** \{A, C, G, U\}*
- **Proteins** \{Ala, Arg, . . . , Val\}*
  
  20+ Amino acids

\[\text{ATGGTTACCCAT} \Rightarrow \text{AUUGGUUAACCCACAU} \]

\[\text{Ribosome} \]
Fundamental *dogma* of molecular biology

**DNA**
{A, C, G, T}*

**RNAs**
{A, C, G, U}*

**Proteins**
{Ala, Arg, . . ., Val}*

20+ Amino acids

THE CODE (genes)

THE MACHINE (enzymes)

\[\text{Ribosome} \]

\[\text{Met} \]

\[\text{ATGGTTACCCCAT} \]

\[\text{TACCAATGGGTA} \]

\[\text{AUGGGUUACCCAU} \]
Fundamental dogma of molecular biology

DNA
\{A, C, G, T\}^

RNAs
\{A, C, G, U\}^

Proteins
\{Ala, Arg, \ldots, Val\}^

\(20^+\) Amino acids

\[\text{Ribosome} \]

\[\text{Met} \quad \text{Val} \]

\[\text{The code (genes)} \]

\[\text{The machine (enzymes)} \]

\[\text{Yann Ponty} \]

\[\text{M2 BIM - Lecture 1 - RNA folding} \]
Fundamental *dogma* of molecular biology

**DNA**
\{A, C, G, T\}*

**RNAs**
\{A, C, G, U\}*

**Proteins**
\{Ala, Arg, \ldots, Val\}*

\(20^+\) Amino acids

\[\text{THE CODE (genes)}\]
\[\text{THE MACHINE (enzymes)}\]

\[\begin{array}{c}
\text{ATG}\ 	ext{GGT}
\end{array}\]

\[\begin{array}{c}
\text{TA}\ 	ext{ACC}\ 	ext{CAA}
\end{array}\]

\[\begin{array}{c}
\text{ATGGG}\ 	ext{GTA}
\end{array}\]

\[\begin{array}{c}
\text{AUUGGUUAA}
\end{array}\]

\[\begin{array}{c}
\text{ACCCCAAU}
\end{array}\]

Ribosome

\[\begin{array}{c}
\text{Met}
\end{array}\]

\[\begin{array}{c}
\text{Val}
\end{array}\]

\[\begin{array}{c}
\text{Thr}
\end{array}\]
Fundamental *dogma* of molecular biology

- **DNA** 
  \{A, C, G, T\}*

- **RNAs** 
  \{A, C, G, U\}*

- **Proteins** 
  \{Ala, Arg, . . . , Val\}*

  **20**+ Amino acids

---

**THE CODE** (genes)  
**THE MACHINE** (enzymes)
Fundamental dogma of molecular biology

DNA
\{ A, C, G, T \}^*

RNAs
\{ A, C, G, U \}^*

Proteins
\{ Ala, Arg, \ldots, Val \}^* \\
20^+ Amino acids

THE CODE
(genes)

THE MACHINE
(enzymes)

Met Val Thr His Ile Leu His Asn
Fundamental *dogma* of molecular biology

**THE CODE** (genes)
- DNA
  - \{A, C, G, T\}*

**RNAs**
- \{A, C, G, U\}*

**Proteins**
- \{Ala, Arg, \ldots, Val\}*
  - 20+ Amino acids

\[
\begin{align*}
\text{DNA} & : \text{ATGGTTACCCCAAT} \\
\text{RNA} & : \text{AUUGGUUAACCCCAU} \\
\text{Protein} & : \text{Met Val Thr His Ile Leu His Asn}
\end{align*}
\]
**Fundamental dogma of molecular biology**

**THE CODE** (genes)
- DNA:
  - \{A, C, G, T\}*
- RNAs:
  - \{A, C, G, U\}*

**THE MACHINE** (enzymes)
- Proteins:
  - \{Ala, Arg, . . . , Val\}*
  - 20+ Amino acids

**THE CODE** (genes)
- DNA:
  - A T G G T T A C C C C A T
  - T A C C A A T G G G T A
- RNAs:
  - A U G G U U A C C C C A U

**THE MACHINE** (enzymes)
- Proteins:
  - Met Val Thr His Ile Leu His Asn
Fundamental *dogma* of molecular biology

**THE CODE**
(genes)
DNA
\{\text{A, C, G, T}\}*

**MEH...**

**THE MACHINE**
(enzymes)
Proteins
\{\text{Ala, Arg, . . . , Val}\}*

20\(^+\) Amino acids

\hspace{1cm}

\begin{align*}
\text{RNA} &: \{\text{A, C, G, U}\} \\
\text{DNA} &: \{\text{A, C, G, T}\} \\
\text{Proteins} &: \{\text{Ala, Arg, . . . , Val}\} \\
\end{align*}
Fundamental *dogma* of molecular biology

DNA → Transcription → RNA → Translation → Proteins
Fundamental dogma of molecular biology

DNA

RNA

Proteins

Carrier

Transfer

Transcription

Maturation

Participates

Regulation

RNA functions
- Messenger
- Translation
- Regulation
- Enzyme
- Catalytic
- ...

Synthesis

Yann Ponty
M2 BIM - Lecture 1 - RNA folding
Fundamental *dogma* of molecular biology

- DNA
- RNA
- Proteins
- Transcription
- Translation
- Regulation
- Maturation
- Synthesis

---

**RNA functions**

- Messenger
- Translation
- Regulation
- Enzyme
- Catalytic
- ...

---

Yann Ponty  
M2 BIM - Lecture 1 - RNA folding
RNA world: Resolving the *chicken vs egg* paradox at the origin of life...

A gene big enough to specify an enzyme would be too big to replicate accurately without the aid of an enzyme of the very kind that it is trying to specify. So the system *apparently cannot get started*.

[...] This is the RNA World. To see how plausible it is, we need to look at why proteins are good at being enzymes but bad at being replicators; at why DNA is good at replicating but bad at being an enzyme; and finally why RNA might just be good enough at both roles to break out of the Catch-22.

**R. Dawkins.** *The Ancestor’s Tale: A Pilgrimage to the Dawn of Evolution*
A gene big enough to specify an enzyme would be too big to replicate accurately without the aid of an enzyme of the very kind that it is trying to specify. So the system apparently cannot get started.

[... ] This is the RNA World. To see how plausible it is, we need to look at why proteins are good at being enzymes but bad at being replicators; at why DNA is good at replicating but bad at being an enzyme; and finally why RNA might just be good enough at both roles to break out of the Catch-22.

R. Dawkins. The Ancestor’s Tale: A Pilgrimage to the Dawn of Evolution
RNA is single-stranded and folds on itself, establishing complex 3D structures that are essential to its function(s).

RNA structures are stabilized by base-pairs, each mediated by hydrogen bonds.

Canonical base-pairs
Watson/Crick base-pairs
Wobble base-pair

G/C
U/A
U/G

Canonical base-pairs
Three\(^1\) levels of representation:

Primary structure

Secondary structure

Tertiary structure

Source: 5s rRNA (PDB 1K73:B)

\(^1\)Well, mostly...
Three\(^1\) levels of representation:

Primary structure

Secondary\(^+\) structure

Tertiary structure

Source: 5s rRNA (PDB 1K73:B)

\(^1\)Well, mostly...
Non-canonical base-pairs

Any base-pair other than \{(A-U), (C-G), (G-U)\}
Or interacting on non-standard edge (\(\neq\) WC/WC-Cis) [LW01].

Pseudoknots (PKs)

Considering PKs may lead to better predictions, **but**:

- Some PK conformations are simply unfeasible;
- Folding *in silico* with general pseudoknots is NP-complete [LP00];

Still, folding on restricted classes of conformations seems promising [CDR\(^+\)04].
Various representations for a versatile biomolecule

Outer-planar graphs
Hamiltonian-path, $\Delta(G) \leq 3$, 2-connected*

Supporting intuitions
Different representations
Common combinatorial structure

* Additional steric constraints
Various representations for a versatile biomolecule

Outer-planar graphs
Hamiltonian-path, $\Delta(G) \leq 3$, 2-connected*

Dot plots
Adjacency matrices*

Supporting intuitions
Different representations
Common combinatorial structure

* Additional steric constraints

Yann Ponty
M2 BIM - Lecture 1 - RNA folding
Various representations for a versatile biomolecule

Outer-planar graphs
Hamiltonian-path, $\Delta(G) \leq 3$, 2-connected*

Supporting intuitions
Different representations
Common combinatorial structure
* Additional steric constraints
Various representations for a versatile biomolecule

Outer-planar graphs
Hamiltonian-path, $\Delta(G) \leq 3$, 2-connected*

Motzkin words*

Supporting intuitions
Different representations
Common combinatorial structure
*Additional steric constraints
Various representations for a versatile biomolecule

Outer-planar graphs
Hamiltonian-path, $\Delta(G) \leq 3$, 2-connected*

Motzkin words*

Non-crossing arc-annotated sequences*

Supporting intuitions
Different representations
Common combinatorial structure

*Additional steric constraints
Various representations for a versatile biomolecule

- Outer-planar graphs
  - Hamiltonian-path, $\Delta(G) \leq 3, \ 2$-connected

- Motzkin words*
  - Positive 1D meanders* over $S = \{+1, -1, 0\}$

- Non-crossing arc-annotated sequences*

Supporting intuitions

- Different representations
  - Common combinatorial structure

* Additional steric constraints
1 Introduction
   - Dynamic programming 101
   - Why RNA?
   - RNA folding
   - RNA Structure(s)
   - Some representations of RNA structure

2 Some flavours of folding prediction
   - Thermodynamics vs Kinetics
   - Dynamic programming: Reminder

3 Free-energy minimization
   - Nussinov-style RNA folding
   - Turner energy model
   - MFold/Unafold
   - Performances and the comparative approach
   - Towards a 3D ab-initio prediction
At the nanoscopic scale, RNA structure fluctuates ($\approx$ Markov process).

Convergence towards a stationary distribution at the Boltzmann equilibrium, where the probability of a conformation only depends on its free-energy. **Corollary:** Initial conformation does not matter.

**Questions:** For a given conformation space and free-energy model:

- A. Determine most stable (Minimum Free-Energy) structure at equilibrium;
- B. Compute average properties of Boltzmann ensemble;
Transcription: RNA synthesized, supposedly without structure\(^2\)

But most mRNAs are degrade before 7h (Org.: Souris \([SSN+09]\)).

\(^2\)Except for co-transcriptional folding. . .
Transcription: RNA synthesized, supposedly without structure\(^2\)

\[ T = 1 \text{h} \]

But most mRNAs are degrade before 7h (Org.: Souris [SSN\(^+\)09]).

\(^2\)Except for co-transcriptional folding. . .
**Transcription:** RNA synthesized, supposedly without structure\(^2\)

\[ T = 2h \]

*But* most mRNAs are degrade before 7h (Org.: Souris [SSN+09]).

\(^2\)Except for co-transcriptional folding. . .
Transcription: RNA synthesized, supposedly without structure\(^2\)

\[ T = 5h \]

But most mRNAs are degrade before 7h (Org.: Souris [SSN\(^+\)09]).

\(^2\)Except for co-transcriptional folding...
**Transcription:** RNA synthesized, supposedly without structure\(^2\)

*But* most mRNAs are degrade before 7h (Org.: Souris [SSN\(^+\)09]).

\(^2\)Except for co-transcriptional folding. . .
Transcription: RNA synthesized, supposedly without structure\(^2\)

But most mRNAs are degrade before 7h (Org.: Souris [SSN\(^+\)09]).

\(^2\)Except for co-transcriptional folding. . .
Away from equilibrium

**Transcription:** RNA synthesized, supposedly without structure\(^2\)

\[ T = 10h \]

**But** most mRNAs are degrade before 7h (Org.: Souris [SSN\(^+\)09]).

- A. Determine most stable (Minimum Free-Energy) structure at equilibrium;
- B. Compute average properties of Boltzmann ensemble;
- C. **Determine most likely structure at finite time** \( T \).
  (c.f. H. Isambert through simulation, NP-complete deterministically [MTSC09])

\(^2\)Except for co-transcriptional folding...
Dynamic programming: General principle

**Dynamic programming** = General optimization technique.

**Prerequisite:** Optimal solution for problem \( P \) can be derived from solutions to strict sub-problems of \( P \).

**Bioinformatics:**

- Discrete solution space (alignments, structures...)  
- Additively-inherited objective function (cost, log-odd score, energy...)  

⇒ Efficient dynamic programming scheme

**Example:** Local Alignment (Smith/Waterman)

\[
W(i, 0) = 0 \\
W(0, j) = 0 \\
W(i, j) = \max \left\{ \begin{array}{l}
W(i - 1, j - 1) + m_{i,j} \\
W(i - 1, j) + p_i \\
W(i, j - 1) + p_d
\end{array} \right.
\]
Dynamic programming scheme defines a space of (sub)problems and a recurrence that relates the score of a problem to that of smaller problems.

Given a scheme, two steps:

- **Matrix filling**: Computation and tabulation of best scores (Computed from smaller problems to larger ones).
- **Traceback**: Reconstruct best solution from contributing subproblems.

Complexity of algorithm depends on:

- **Cardinality** of sub-problem space
- **Number of alternatives** considers at each step (#Terms in recurrence)

Smith&Waterman example:

- **i**: $1 \to n + 1 \Rightarrow \Theta(n)$
- **j**: $1 \to m + 1 \Rightarrow \Theta(m)$
- 3 operations at each step

$\Rightarrow \Theta(m.n)$ time/memory

\[
\begin{align*}
W(i, 0) &= 0 \\
W(0, j) &= 0 \\
W(i, j) &= \max \left\{ W(i-1, j-1) + m_{i,j}, W(i-1, j) + p_i, W(i, j-1) + p_d \right\}
\end{align*}
\]
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

$$W(i,0) = 0$$
$$W(0,j) = 0$$

$$W(i,j) = \max \begin{cases} W(i-1,j-1) + m_{i,j} \\ W(i-1,j) + p_i \\ W(i,j-1) + p_d \end{cases}$$

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

|   | A | 0 |
|---|---|
| G | 0 |
| C | 0 |
| A | 0 |
| A | 0 |
| C | 0 |
| A | 0 |
| A | 0 |

A 0
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i, 0) = 0 \\
W(0, j) = 0 \\
W(i, j) = \max \left\{ \begin{array}{l}
W(i-1, j-1) + m_{i,j} \\
W(i-1, j) + p_i \\
W(i, j-1) + p_d
\end{array} \right. 
\]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Yann Ponty
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i, 0) = 0 \\
W(0, j) = 0 \\
W(i, j) = \max \left\{ \begin{array}{l}
W(i - 1, j - 1) + m_{i,j} \\
W(i - 1, j) + p_i \\
W(i, j - 1) + p_d
\end{array} \right.
\]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

W$(i,0) = 0$
W$(0,j) = 0$
W$(i,j) = \max \left\{ \begin{array}{l}
W(i - 1, j - 1) + m_{i,j} \\
W(i - 1, j) + p_i \\
W(i, j - 1) + p_d
\end{array} \right.$
Example: Local alignment of AGCACACA and ACACACTA

Costs: Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
\begin{align*}
W(i, 0) &= 0 \\
W(0, j) &= 0 \\
W(i, j) &= \max \left\{ W(i - 1, j - 1) + m_{i,j}, W(i - 1, j) + p_i, W(i, j - 1) + p_d \right\}
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

- $W(i,0) = 0$
- $W(0,j) = 0$
- $W(i,j) = \max \begin{cases} W(i-1,j-1) + m_{i,j} \\ W(i-1,j) + p_i \\ W(i,j-1) + p_d \end{cases}$

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
\begin{align*}
W(i,0) &= 0 \\
W(0,j) &= 0 \\
W(i,j) &= \max \left\{ W(i-1,j-1) + m_{i,j}, W(i-1,j) + p_i, W(i,j-1) + p_d \right\}
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

![Alignment Diagram]
Complete example

Example: Local alignment of AGCACACA and ACACACTA
Costs: Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

$W(i, 0) = 0$
$W(0, j) = 0$
$W(i, j) = \max \begin{cases} W(i - 1, j - 1) + m_{i,j} \\ W(i - 1, j) + p_i \\ W(i, j - 1) + p_d \end{cases}$
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i, 0) = 0 \\
W(0, j) = 0 \\
W(i, j) = \max \begin{cases} 
W(i - 1, j - 1) + m_{i,j} \\
W(i - 1, j) + p_i \\
W(i, j - 1) + p_d 
\end{cases}
\]
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

$$W(i,0) = 0$$
$$W(0,j) = 0$$

$$W(i,j) = \max \left\{ \begin{array}{l}
W(i-1,j-1) + m_{i,j} \\
W(i-1,j) + p_i \\
W(i,j-1) + p_d
\end{array} \right\}$$
**Example:** Local alignment of \textbf{AGCACACA} and \textbf{ACACACTA}  
**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$
Example: Local alignment of AGCACACA and ACACACTA
Costs: Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i,0) = 0 \\
W(0,j) = 0 \\
W(i,j) = \max \begin{cases} 
W(i-1,j-1) + m_{i,j} \\
W(i-1,j) + p_i \\
W(i,j-1) + p_d 
\end{cases}
\]
**Example:** Local alignment of *AGCACACA* and *ACACACTA*

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i, 0) = 0 \\
W(0, j) = 0 \\
W(i, j) = \max \left\{ W(i-1, j-1) + m_{i,j}, W(i-1, j) + p_i, W(i, j-1) + p_d \right\}
\]

**Best alignment**

```
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
```
Example: Local alignment of $\text{AGCACACA}$ and $\text{ACACACTA}$

Costs: Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

$$W(i,0) = 0$$
$$W(0,j) = 0$$

$$W(i,j) = \max \left\{ \begin{array}{l}
W(i-1,j-1) + m_{i,j} \\
W(i-1,j) + p_i \\
W(i,j-1) + p_d
\end{array} \right\}$$

Best alignment
Complete example

**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i,0) = 0 \\
W(0,j) = 0 \\
W(i,j) = \max \left\{ \begin{array}{l}
W(i-1,j-1) + m_{i,j} \\
W(i-1,j) + p_i \\
W(i,j-1) + p_d
\end{array} \right. \\
\]

Best alignment
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i,0) = \begin{cases} 0 \end{cases}
\]
\[
W(0,j) = \begin{cases} 0 \end{cases}
\]
\[
W(i,j) = \max \left\{ \begin{array}{l}
W(i-1, j-1) + m_{i,j} \\
W(i-1, j) + p_i \\
W(i, j-1) + p_d
\end{array} \right. 
\]

**Best alignment**
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i, 0) = 0 \\
W(0, j) = 0 \\
W(i, j) = \max \begin{cases}
W(i-1, j-1) + m_{i,j} \\
W(i-1, j) + p_i \\
W(i, j-1) + p_d
\end{cases}
\]

**Best alignment**

```
A   C   A   C   A   C   T   A
0   0   0   0   0   0   0   0   0
A   0   2   1   2   1   2   1   0   2
G   0   1   1   1   1   1   1   0   1
C   0   0   3   2   3   2   3   2   1
A   0   2   2   5   4   5   4   3   4
C   0   1   4   4   7   6   7   6   5
A   0   2   3   6   6   9   8   7   8
C   0   1   4   5   8   8   11  10  9
A   0   2   3   6   7   10  10  10  12
```
Example: Local alignment of AGCACACA and ACACACTA

Costs: Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

$$W(i,0) = 0$$
$$W(0,j) = 0$$
$$W(i,j) = \max \begin{cases} W(i-1,j-1) + m_{i,j} \\ W(i-1,j) + p_i \\ W(i,j-1) + p_d \end{cases}$$

Best alignment
Example: Local alignment of AGCACACA and ACACACTA

Costs: Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
W(i, 0) = 0 \\
W(0, j) = 0 \\
W(i, j) = \max \left\{ \begin{array}{c}
W(i-1, j-1) + m_{i,j} \\
W(i-1, j) + p_i \\
W(i, j-1) + p_d
\end{array} \right. 
\]

Best alignment

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Example: Local alignment of AGCACACA and ACACACTA

Costs: Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

**Best alignment**

C A C A C - A
C A C A C T A
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$


\[
W(i,0) = 0 \\
W(0,j) = 0 \\
W(i,j) = \max \left\{ \begin{array}{l}
W(i-1,j-1) + m_{i,j} \\
W(i-1,j) + p_i \\
W(i,j-1) + p_d
\end{array} \right. 
\]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Best alignment

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>A</th>
<th>C</th>
<th>-</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>
**Example:** Local alignment of AGCACACA and ACACACTA

**Costs:** Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

\[
\begin{align*}
W(i, 0) &= 0 \\
W(0, j) &= 0 \\
W(i, j) &= \max \begin{cases} \\
W(i - 1, j - 1) + m_{i,j} \\
W(i - 1, j) + p_i \\
W(i, j - 1) + p_d
\end{cases}
\end{align*}
\]

**Best alignment**

```
A G C A C A C - - A
A - - C A C A C T - A
```
Properties of DP schemes

Necessary properties:

▶ **Correctness**: $\forall$ sub-problem, the computed value must indeed maximize the objective function.

Proofs usually inductive, and quite technical, but very systematic.

Desirable properties of DP schemes:

▶ **Completeness** of space of solutions generated by decomposition. Algorithmic tricks, by *cutting branches*, may violate this property.

▶ **Unambiguity**: Each solution is generated at most once.

$\Rightarrow$ Under these properties, one can *enumerate* solution space.
Introduction
- Dynamic programming 101
- Why RNA?
- RNA folding
- RNA Structure(s)
- Some representations of RNA structure

Some flavours of folding prediction
- Thermodynamics vs Kinetics
- Dynamic programming: Reminder

Free-energy minimization
- Nussinov-style RNA folding
- Turner energy model
- MFold/Unafold
- Performances and the comparative approach
- Towards a 3D ab-initio prediction
**Problem A:** Determine Minimum Free-Energy structure (MFE).

**Ab initio folding prediction =**
Predict RNA structure from its sequence $\omega$ only.

**Conformations:** Set $S_\omega$ of secondary structures compatible (w.r.t. base-pairing constraints) with primary structure $\omega$.

**Free-Energy:** Function $E_{\omega,S}$ (KCal.mol$^{-1}$), additive on motifs occurring in any sequence/conformation couple $(\omega, S)$.

**Native structure:** Functional conformation of the biomolecule.

**Remarks:**
- Not necessarily unique (Kinetics, or bi-stable structures);
- In presence of PKs $\rightarrow$ Ambiguous: Which is the native conformation?
Nussinov/Jacobson energy model (NJ)

Base-pair maximization (with a twist):
- Additive model on independently contributing base-pairs;
- Canonical base-pairs only: Watson/Crick (A/U, C/G) and Wobble (G/U)

\[ E_{\omega,S} = -\#Paires(S) \]

Folding in NJ model ⇔ Base-pair (weight) maximization

Example:

UUUUC CCCU AAAAGG

Variant: Weight each pair with −#Hydrogen bonds

\[ \Delta G(G\equiv C) = -3 \quad \Delta G(A=U) = -2 \quad \Delta G(G-U) = -1 \]
Nussinov/Jacobson energy model (NJ)

**Base-pair maximization** *(with a twist):*

- Additive model on **independently contributing** base-pairs;
- Canonical base-pairs only: Watson/Crick (A/U, C/G) and Wobble (G/U)

\[ E_{\omega,S} = -\#Paires(S) \]

Folding in NJ model ⇔ **Base-pair (weight) maximization**

**Example:**

```
UUUUCCCUA AAA GG
```

**Variant:** Weight each pair with \(-\#\)Hydrogen bonds

\[ \Delta G(G\equiv C) = -3 \quad \Delta G(A\equiv U) = -2 \quad \Delta G(G\leftarrow U) = -1 \]
\[ N_{i,t} = 0, \quad \forall t \in [i, i + \theta] \]

\[ N_{i,j} = \min \left\{ \begin{array}{ll}
N_{i+1,j} & i \text{ unpaired} \\
\min_{k=i+\theta+1} \Delta G_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ paired with } k
\end{array} \right. \]
Nussinov/Jacobson DP scheme

\[ i \quad \quad j \quad = \quad i \quad i+1 \quad j \quad + \quad i \quad k \quad j \quad \geq \theta \]

\[
N_{i,t} = 0, \quad \forall t \in [i, i+\theta]
\]

\[
N_{i,j} = \min \begin{cases} N_{i+1,j} & \text{i unpaired} \\ \min \limits_{k=i+\theta+1} \Delta G_{i,k} + N_{i+1,k-1} + N_{k+1,j} & \text{i paired with } k \end{cases}
\]

**Correctness.** Goal = Show that MFE over interval \([i, j]\) is indeed found in \(N_{i,j}\) after completing the computation. Proceed by induction:

- Assume that property holds for any \([i', j']\) such that \(j' - i' < n\).
- Consider \([i, j], j - i = n\). Let \(\text{MFE}_{i,j} := \text{Base-pairs of best struct. on } [i, j]\).
  Then first position \(i\) in \(\text{MFE}_{i,j}\) is either:
    - **Unpaired:** \(\text{MFE}_{i,j} = \text{MFE}_{i+1,j}\) \quad \rightarrow \text{free-energy} = N_{i+1,j}
    - **Paired to } k:** \(\text{MFE}_{i,j} = \{(i, k)\} \cup \text{MFE}_{i+1,k-1} \cup \text{MFE}_{k+1,j}\).
      (Indeed, any BP between \([i+1, k-1]\) and \([k+1, j]\) would cross \((i, k)\))
      \quad \rightarrow \text{free-energy} = \Delta G_{i,k} + N_{i+1,k-1} + N_{k+1,j}\]
Nussinov/Jacobson DP scheme

\[ N_{i,t} = 0, \quad \forall t \in [i, i + \theta] \]

\[ N_{i,j} = \min \left\{ \begin{array}{ll}
N_{i+1,j} & \text{i unpaired} \\
\min_{k=i+\theta+1} \Delta G_{i,k} + N_{i+1,k-1} + N_{k+1,j} & \text{i paired with k}
\end{array} \right. \]

**Correctness.** Goal = Show that MFE over interval \([i, j]\) is indeed found in \(N_{i,j}\) after completing the computation. Proceed by induction:

- Assume that property holds for any \([i', j']\) such that \(j' - i' < n\).
- Consider \([i, j], j - i = n\). Let \(\text{MFE}_{i,j} := \text{Base-pairs of best struct. on } [i, j]\).
  Then first position \(i\) in \(\text{MFE}_{i,j}\) is either:
  - **Unpaired:** \(\text{MFE}_{i,j} = \text{MFE}_{i+1,j}\) \(\rightarrow\) free-energy = \(N_{i+1,j}\)
  - **Paired to \(k\):** \(\text{MFE}_{i,j} = \{(i, k)\} \cup \text{MFE}_{i+1,k-1} \cup \text{MFE}_{k+1,j}\).
    (Indeed, any BP between \([i+1, k-1]\) and \([k+1, j]\) would cross \((i, k)\))
    \(\rightarrow\) free-energy = \(\Delta G_{i,k} + N_{i+1,k-1} + N_{k+1,j}\)
Nussinov/Jacobson DP scheme

\[ i \quad j \quad = \quad i \quad i+1 \quad j \quad + \quad i \quad k \quad j \quad \geq \theta \]

\[ N_{i,t} = 0, \quad \forall t \in [i, i + \theta] \]

\[ N_{i,j} = \min \left\{ \begin{array}{ll}
N_{i+1,j} & \text{i unpaired} \\
\min_{k=i+\theta+1} \Delta G_{i,k} + N_{i+1,k-1} + N_{k+1,j} & \text{i paired with } k
\end{array} \right. \]

Correctness. Goal = Show that MFE over interval \([i, j]\) is indeed found in \(N_{i,j}\) after completing the computation. Proceed by induction:

- Assume that property holds for any \([i', j']\) such that \(j' - i' < n\).
- Consider \([i, j]\), \(j - i = n\). Let \(\text{MFE}_{i,j} := \text{Base-pairs of best struct. on } [i, j]\). Then first position \(i\) in \(\text{MFE}_{i,j}\) is either:
  - **Unpaired**: \(\text{MFE}_{i,j} = \text{MFE}_{i+1,j}\) \(\rightarrow\) free-energy = \(N_{i+1,j}\)
  - **Paired to } k**: \(\text{MFE}_{i,j} = \{(i, k)\} \cup \text{MFE}_{i+1,k-1} \cup \text{MFE}_{k+1,j}\).
    (Indeed, any BP between \([i + 1, k - 1]\) and \([k + 1, j]\) would cross \((i, k)\))
    \(\rightarrow\) free-energy = \(\Delta G_{i,k} + N_{i+1,k-1} + N_{k+1,j}\)
\[
i = i + 1
\]

\[
\theta
\]
### RNA Folding

#### Scoring Matrix

|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| C | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 5 | 7 | 7 | 8 | 10 |   |   |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |   |   |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |   |   |   |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |   |   |   |   |   |   |
| U | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 |   |   |   |   |   |   |   |
| C | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 |   |   |   |   |   |   |   |
| U | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 3 |   |   |   |   |   |   |   |
| U | 0 | 0 | 0 | 0 | 1 | 2 |   |   |   |   |   |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |   |   |
| G | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |   |
| C | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |   |
| G | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |

#### Formulas

- \( i \leq j \)
- \( Yann Ponty \)
- M2 BIM - Lecture 1 - RNA folding
### RNA Folding Energy Matrix

The energy matrix for RNA folding is shown below. The energy matrix is used to calculate the stability of RNA secondary structures. Each entry in the matrix represents the energy cost of adding a new base pair.

The energy function is given by:

\[
E_{ij} = \begin{cases} 
0 & \text{if } i = j \\
\theta & \text{if } \forall k \neq i, j \text{ where } i < k < j, \text{ and } E_{ik} + E_{kj} \geq \theta \\
0 & \text{otherwise}
\end{cases}
\]

Where \( \theta \) is the minimum free energy of a base pair.

### Diagram

The diagram illustrates the energy calculation for adding a new base pair at position \( (i, j) \):

- The energy cost for adding a new base pair is \( \theta \) if there is no conflicting base pairs.
- The energy cost is 0 if the previous base pair is not conflicting.
- The energy cost is 0 if there is no base pair at all.

### Example

Consider the RNA sequence:

```
CGAUACUCUUAAGACGA
```

The corresponding energy matrix is:

|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 4 | 4 | 5 | 7 | 7 | 8 | 10 |   |   |   |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |   |   |   |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |   |   |   |   |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |   |   |   |   |   |   |   |
| U | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 |   |   |   |   |   |   |   |   |
| C | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 |   |   |   |   |   |   |   |   |   |
| U | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 3 |   |   |   |   |   |   |   |   |   |   |
| U | 0 | 0 | 0 | 0 | 0 | 1 | 2 |   |   |   |   |   |   |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |   |   |   |   |
| G | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| C |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| G |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| A |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

The energy cost for adding a new base pair at position \( (i, j) \) is calculated using the formula above.
<table>
<thead>
<tr>
<th>C</th>
<th>G</th>
<th>G</th>
<th>A</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>=</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>i</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>i \rightarrow i+1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>i \rightarrow k \rightarrow j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[i \leq \theta \]
\begin{array}{cccccccccccccccc}
\text{C} & \text{G} & \text{G} & \text{A} & \text{U} & \text{A} & \text{C} & \text{U} & \text{U} & \text{C} & \text{U} & \text{U} & \text{A} & \text{G} & \text{A} & \text{C} & \text{G} & \text{A} \\
\hline
\text{C} & 0 & 0 & 0 & 0 & 0 & 3 & 4 & 4 & 6 & 6 & 6 & 6 & 9 & 11 & 14 & 14 \\
\text{G} & 0 & 0 & 0 & 0 & 0 & 3 & 4 & 4 & 6 & 6 & 6 & 6 & 7 & 9 & 11 & 11 & 11 \\
\text{G} & 0 & 0 & 0 & 0 & 3 & 3 & 3 & 5 & 5 & 5 & 5 & 6 & 8 & 10 & 10 & 10 \\
\text{A} & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 4 & 4 & 5 & 7 & 7 & 8 & 10 \\
\text{U} & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 4 & 5 & 7 & 7 & 8 & 10 & & \\
\text{A} & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 5 & 5 & 5 & 5 & 8 & 8 & & \\
\text{C} & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 5 & 5 & 5 & 8 & 8 & & \\
\text{U} & 0 & 0 & 0 & 0 & 0 & 2 & 3 & 5 & 5 & 6 & 7 & & & \\
\text{U} & 0 & 0 & 0 & 0 & 2 & 3 & 5 & 5 & 5 & 7 & & & & \\
\text{C} & 0 & 0 & 0 & 0 & 3 & 3 & 3 & 5 & 5 & & & & & & \\
\text{U} & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 3 & & & & & & & \\
\text{U} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & & & & & & & \\
\text{A} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & & \\
\text{G} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & & \\
\text{A} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & & \\
\text{C} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & & \\
\text{G} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & & \\
\text{A} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & & & & & & & \\
\end{array}

\text{Yann Ponty} \\
\text{M2 BIM - Lecture 1 - RNA folding}
<table>
<thead>
<tr>
<th>C</th>
<th>G</th>
<th>G</th>
<th>A</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>()</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
i \leq j
\]

\[
i = i + 1
\]

\[
i \geq \theta
\]

\[
i \leq k
\]

\[
i = j
\]
\[
i(i, j) = i(i+1, j) + \left(\begin{array}{c}
\text{C} \\
\text{G} \\
\text{A} \\
\text{U} \\
\text{A} \\
\text{C} \\
\text{U} \\
\text{C} \\
\text{U} \\
\text{U} \\
\text{A} \\
\text{G} \\
\text{A} \\
\text{C} \\
\text{G} \\
\end{array}\right) \geq \theta
\]
\[
i \geq \theta
\]
\[
i_{ij} = i_{i+1,j} + \begin{array}{l}
i_{i,k} \quad \text{if } i < k < j, \quad \geq \theta,
i_{i,j} \quad \text{otherwise,}
\end{array}
\]
### RNA Folding Energy Matrix

The energy matrix represents the potential energy of each possible interaction in the RNA sequence. Each cell in the matrix contains the energy value of forming a base pair between the nucleotides at positions `i` and `j`. The energy function is defined as:

\[
E_{ij} = \begin{cases} 
\theta, & \text{if } i = j \text{ and } k = j-k \\
0, & \text{otherwise}
\end{cases}
\]

where `\theta` is the minimum energy required to form a cross-helix structure.

### Energy Values

The table below contains the energy values for each possible interaction in the RNA sequence:

|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
| C | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |   |
| G | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |   |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 |   |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 5 | 7 | 7 | 8 | 10 |   |   |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 |   |   |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |   |   |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |   |   |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |   |   |   |   |
| U | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 |   |   |   |
| C | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 3 |   |   |   |
| U | 0 | 0 | 0 | 0 | 1 | 2 |   |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |
| G | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |   |
| C | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |
| G | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |
| A | 0 | 0 | 0 | 0 | 0 | 0 |   |   |   |   |   |

**Diagram:**

- The diagram illustrates the energy function `E_{ij}` for forming a cross-helix structure.
- The sequence is denoted by `i` to `j`.
- The energy threshold `\theta` is indicated by the circle at the bottom right.
<table>
<thead>
<tr>
<th>C</th>
<th>G</th>
<th>G</th>
<th>A</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ i \leq j = \sum_{i \leq k < j} (i+1) \geq \theta \]
\[ i \leq j \Rightarrow Y_{i,j} = Y_{i,i+1} + Y_{j,k} \geq \theta \]
\[
C \quad G \quad G \quad A \quad U \quad A \quad C \quad U \quad U \quad C \quad U \quad U \quad A \quad G \quad A \quad C \quad G \quad A
\]

\[
\left( \begin{array}{cccccccccccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 3 & 4 & 4 & 6 & 6 & 6 & 6 & 9 & 9 & 11 & 14 & 14 \\
0 & 0 & 0 & 0 & 0 & 3 & 4 & 4 & 6 & 6 & 6 & 6 & 7 & 9 & 11 & 11 & 11 \\
0 & 0 & 0 & 0 & 3 & 3 & 3 & 5 & 5 & 5 & 6 & 8 & 10 & 10 & 10 & 10 & 10 \\
0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 4 & 4 & 5 & 7 & 7 & 8 & 10 & 10 & 10 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 4 & 5 & 7 & 7 & 8 & 10 & 10 & 10 & 10 \\
0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 5 & 5 & 5 & 8 & 8 & 8 & 8 & 8 & 8 \\
0 & 0 & 0 & 0 & 0 & 2 & 5 & 5 & 5 & 5 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\
0 & 0 & 0 & 0 & 2 & 3 & 5 & 5 & 6 & 7 & 7 & 5 & 5 & 5 & 5 & 5 & 5 & 5 \\
0 & 0 & 0 & 0 & 2 & 3 & 5 & 5 & 5 & 5 & 7 & 7 & 7 & 7 & 7 & 7 & 7 & 7 \\
0 & 0 & 0 & 0 & 3 & 3 & 3 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 & 5 \\
0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array} \right)
\]

\[
i \quad j = i \quad i+1 \quad j \quad + \quad \begin{array}{cccc}
\gamma \\
\\end{array} \quad \geq \theta
\]

Diagram:

- \(i\) to \(j\)
- \(i+1\) to \(j\)
- \(i\) to \(k\) to \(j\)
- \(i\) to \(j\)

Result:

\[
0
\]
|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 5 | 7 | 7 | 7 | 8 | 10 |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 7 | 8 | 10 |
| A | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 |
| C | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 |
| U | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 3 |
| U | 0 | 0 | 0 | 0 | 1 | 2 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 |
| A | 0 |

\[
i, j = i, i+1, j + \begin{array}{c} \geq \theta \end{array}
\]

Yann Ponty  
M2 BIM - Lecture 1 - RNA folding
\[ i \leq j \]

**RNA Folding**

Yann Ponty | M2 BIM - Lecture 1 - RNA folding
<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>G</th>
<th>A</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
</table>
|   | (  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | .  | (
| C | 0  | 0  | 0  | 0  | 0  | 3  | 4  | 4  | 6  | 6  | 6  | 6  | 6  | 9  | 9  | 11 | 14 |
| G | 0  | 0  | 0  | 0  | 0  | 3  | 4  | 4  | 6  | 6  | 6  | 6  | 7  | 9  | 11 | 11 | 11 |
| G | 0  | 0  | 0  | 0  | 3  | 3  | 3  | 5  | 5  | 5  | 5  | 5  | 6  | 8  | 10 | 10 | 10 |
| A | 0  | 0  | 0  | 0  | 2  | 2  | 2  | 2  | 4  | 4  | 5  | 7  | 7  | 8  | 10 |     |     |
| U | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 2  | 4  | 5  | 7  | 7  | 8  | 10 |     |     |
| A | 0  | 0  | 0  | 0  | 2  | 2  | 2  | 5  | 5  | 5  | 5  | 8  | 8  |     |     |     |     |
| C | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 5  | 5  | 5  | 5  | 8  | 8  |     |     |     |     |
| U | 0  | 0  | 0  | 0  | 0  | 2  | 3  | 5  | 5  | 6  | 7  |     |     |     |     |     |     |
| U | 0  | 0  | 0  | 0  | 2  | 3  | 5  | 5  | 5  | 5  | 7  |     |     |     |     |     |     |
| C | 0  | 0  | 0  | 0  | 3  | 3  | 3  | 5  | 5  | 5  |     |     |     |     |     |     |     |
| U | 0  | 0  | 0  | 0  | 2  | 2  | 2  | 2  | 3  |     |     |     |     |     |     |     |     |
| A | 0  | 0  | 0  | 0  | 0  | 1  |     |     |     |     |     |     |     |     |     |     |     |
| G | 0  | 0  | 0  | 0  | 0  |     |     |     |     |     |     |     |     |     |     |     |     |     |
| A | 0  | 0  | 0  | 0  | 0  |     |     |     |     |     |     |     |     |     |     |     |     |     |
| C | 0  | 0  | 0  | 0  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| G | 0  | 0  | 0  | 0  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| A | 0  | 0  | 0  | 0  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

\[ i \leq j \leq i + k \]

\[ i = j \]

\[ i + 1 \]

\[ \geq \theta \]
### RNA Folding

#### Energy Matrix

<table>
<thead>
<tr>
<th>C</th>
<th>G</th>
<th>G</th>
<th>A</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>(</td>
<td>(</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>)</td>
<td>)</td>
<td>)</td>
<td>)</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Equation

\[
i \leq j, \quad \Delta E_{i,j} = \Delta E_{i,i+1} + \Delta E_{i,j} + \max_{k \geq \theta} \Delta E_{i,k} + \Delta E_{k,j}
\]
\[
\begin{align*}
C & \quad G & \quad G & \quad A & \quad U & \quad A & \quad C & \quad U & \quad U & \quad C & \quad U & \quad U & \quad A & \quad G & \quad A & \quad C & \quad G & \quad A & \\
\text{( ( ( . . . ) . . . . . . . . ) ) . ) . } & \\
C & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 3 & \quad 4 & \quad 4 & \quad 6 & \quad 6 & \quad 6 & \quad 6 & \quad 6 & \quad 9 & \quad 9 & \quad 11 & \quad 14 & \quad 14 & \\
G & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 3 & \quad 4 & \quad 4 & \quad 6 & \quad 6 & \quad 6 & \quad 6 & \quad 7 & \quad 9 & \quad 11 & \quad 11 & \quad 11 & \\
G & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 3 & \quad 3 & \quad 3 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 6 & \quad 8 & \quad 10 & \quad 10 & \quad 10 & \quad 10 & \\
A & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 2 & \quad 2 & \quad 2 & \quad 2 & \quad 4 & \quad 4 & \quad 5 & \quad 7 & \quad 7 & \quad 8 & \quad 10 & \quad 10 & \quad 10 & \\
U & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 2 & \quad 2 & \quad 4 & \quad 5 & \quad 7 & \quad 7 & \quad 8 & \quad 10 & \quad 10 & \quad 10 & \quad 10 & \\
A & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 2 & \quad 2 & \quad 2 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 8 & \quad 8 & \quad 8 & \quad 8 & \quad 8 & \\
C & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 2 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 8 & \quad 8 & \quad 8 & \quad 8 & \quad 8 & \quad 8 & \\
U & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 2 & \quad 3 & \quad 5 & \quad 5 & \quad 6 & \quad 7 & \quad 6 & \quad 7 & \quad 6 & \quad 7 & \quad 6 & \quad 7 & \\
U & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 2 & \quad 3 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 7 & \quad 7 & \quad 7 & \quad 7 & \quad 7 & \quad 7 & \quad 7 & \\
C & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 3 & \quad 3 & \quad 3 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \\
U & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 2 & \quad 2 & \quad 2 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \quad 3 & \\
A & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \\
G & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \\
A & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \\
C & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \\
G & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \\
A & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \quad 0 & \\
\end{align*}
\]
\[
\begin{array}{cccccccccccccccc}
\end{array}
\]

|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|   | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 5 | 7 | 7 | 8 | 10 |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 5 | 8 | 8 | 8 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 | 8 | 8 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |
| U | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 5 | 7 |
| C | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| U | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 3 |
| U | 0 | 0 | 0 | 0 | 1 | 2 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

\[i \leq j \leq i + 1 \leq k \leq j \geq \theta\]
### RNA Folding Energy Matrix

|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
| C | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 4 | 4 | 4 | 5 | 7 | 7 | 8 | 10 |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |
| U | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 |
| C | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 |
| U | 0 | 0 | 0 | 2 | 2 | 2 | 3 |
| U | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 |

**Diagram:**

\[
i \leq j \quad \text{and} \quad i 
\]

\[
i \leq k \leq j
\]

\[
\geq \theta
\]

\[
0
\]

**Equation:**

\[
i \leq j = i \leq i+1 + j \geq \theta
\]
|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 5 | 7 | 7 | 8 | 8 | 10 |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 8 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 |
| C | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 |
| U | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 3 |
| U | 0 | 0 | 0 | 0 | 1 | 2 |
| A | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 |

\[ Yann Ponty \]

M2 BIM - Lecture 1 - RNA folding
i \leq \theta

Yann Ponty

M2 BIM - Lecture 1 - RNA folding
\[
\begin{align*}
A & \quad C & \quad G & \quad A & \quad U & \quad A & \quad C & \quad U & \quad U & \quad C & \quad U & \quad U & \quad A & \quad G & \quad A & \quad C & \quad G & \quad A \\
\end{align*}
\]

|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
| C | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 4 | 4 | 5 | 7 | 7 | 8 | 10 |
| U | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 |
| C | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 |
| U | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 |
| C | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 |
| U | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 3 |
| U | 0 | 0 | 0 | 0 | 1 | 2 |
| A | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 |
| C | i = i + 1 |
| G | j = j |
| A | \[
\begin{align*}
\text{Yann Ponty} & \quad M2 BIM - Lecture 1 - RNA folding
\end{align*}
\]
\[
i(j) = (i) + (i+1) j \geq \beta
\]
### RNA Folding

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>G</th>
<th>G</th>
<th>A</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(</td>
<td>(</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>)</td>
<td>)</td>
<td>)</td>
<td>)</td>
<td>)</td>
<td>)</td>
<td>)</td>
<td>)</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Example

\[ i < j \]

\[ e(i, j) = e(i, i+1) + e(j, k) \geq \theta \]

\[ \text{Energy} \]

\[ \theta \]

\[ \text{Minimum Energy} \]

\[ \text{RNA Structure} \]

\[ \text{Stem} \]

\[ \text{Loop} \]

\[ \text{Hairpin} \]

\[ \text{Internal Loop} \]

\[ \text{Secondary Structure} \]

\[ \text{Primary Structure} \]

\[ \text{Nussinov/Jacobson} \]

\[ \text{Yann Ponty} \]

\[ \text{M2 BIM - Lecture 1 - RNA folding} \]
### RNA Folding Energy Matrix

The energy matrix represents the energy of pairing between nucleotides in a RNA sequence. Each cell in the matrix corresponds to a specific pair of nucleotides, with the energy value indicating the stability of the pairing.

#### Matrix Values

- **C**: Cytosine
- **G**: Guanine
- **A**: Adenine
- **U**: Uracil

#### Energy Calculation

- The energy of a base pair is calculated as the sum of the standard energy of the base pair, plus a penalty term if the distance between the bases is greater than a threshold \( \theta \).

#### Example

|   | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|   | ( | ( |   | . | . | . | ) | . | ( |   |   | . | . | . | . | ) | ) | ) | ) | ) | ) | ) | ) | ) | ) | ) | ) |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| G | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 9 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| G | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 4 | 4 | 4 | 5 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| U | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 4 | 5 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 5 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| C | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 5 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| U | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| U | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 5 | 5 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| C | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| U | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| U | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| A | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| A | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
### RNA Folding Energy Matrix

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>G</th>
<th>G</th>
<th>A</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>C</th>
<th>U</th>
<th>U</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Energy Calculation

\[ i, j = i, i+1 \]

\[ i, j \geq \theta \]

\[ \geq \theta \]

\[ 0 \]
\[
\begin{array}{cccccccccccccccc}
( & ( & ( & . & . & . & ) & ) & ) & ) & ) & ) & ) & ) & ) & ) & ) & ) \\
C & 0 & 0 & 0 & 0 & 0 & 3 & 4 & 4 & 6 & 6 & 6 & 9 & 9 & 11 & 14 & 14 \\
G & 0 & 0 & 0 & 0 & 0 & 3 & 4 & 4 & 6 & 6 & 6 & 7 & 9 & 11 & 11 & 11 \\
G & 0 & 0 & 0 & 0 & 3 & 3 & 3 & 5 & 5 & 5 & 5 & 6 & 8 & 10 & 10 & 10 \\
A & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 4 & 4 & 4 & 5 & 7 & 7 & 8 & 10 \\
U & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 4 & 5 & 7 & 7 & 8 & 10 & 0 \\
A & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 5 & 5 & 5 & 5 & 8 & 8 & 8 \\
C & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 5 & 5 & 5 & 8 & 8 & 0 & 0 & 0 \\
U & 0 & 0 & 0 & 0 & 0 & 2 & 3 & 5 & 5 & 6 & 7 & 0 & 0 & 0 & 0 \\
U & 0 & 0 & 0 & 0 & 2 & 3 & 5 & 5 & 5 & 5 & 7 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 3 & 3 & 3 & 5 & 5 & 5 & 5 & 0 & 0 & 0 & 0 \\
U & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 0 & 0 & 0 & 0 \\
U & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
G & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
C & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
G & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

\[
i = \min(i,i+1) + \begin{cases} \geq \theta & \text{if stable} \\ \leq \theta & \text{if unbound} \end{cases}
\]

Yann Ponty  
M2 BIM - Lecture 1 - RNA folding
Turner energy model

Based on unambiguous decomposition of 2\textsuperscript{ary} structure into loops:

- Internal loops
- Bulges
- Terminal loops
- Multi loops
- Stackings

Free-energy $\Delta G$ of a loop depend on bases, asymmetry, dangles …

Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.
Based on unambiguous decomposition of 2\textsuperscript{ary} structure into loops:

- Internal loops
- Bulges
- Terminal loops
- Multi loops
- Stackings

Free-energy $\Delta G$ of a loop depend on bases, asymmetry, dangles ...  

Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.
Based on **unambiguous** decomposition of 2\textsuperscript{ary} structure into **loops**:

- Internal loops
- Bulges
  - Terminal loops
  - Multi loops
- Stackings

Free-energy $\Delta G$ of a loop depend on bases, asymmetry, dangles …

Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.
Based on unambiguous decomposition of 2\text{ary} structure into loops:

- Internal loops
- Bulges
- Terminal loops
  - Multi loops
  - Stackings

Free-energy $\Delta G$ of a loop depend on bases, assymmetry, dangles \ldots

Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.
Based on **unambiguous** decomposition of 2\textsuperscript{ary} structure into **loops**:

- Internal loops
- Bulges
- Terminal loops
- Multi loops
- Stackings

Free-energy $\Delta G$ of a loop depend on bases, assymmetry, dangles …

Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.
Based on **unambiguous** decomposition of 2\(^{ary}\) structure into **loops**:

- Internal loops
- Bulges
- Terminal loops
- Multi loops
- Stackings

Free-energy $\Delta G$ of a loop depend on bases, asymmetry, dangles ... 

Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.
MFE DP equations

\[
\begin{align*}
\text{Stem loop} &= \begin{cases} 
\text{Terminal loops} \\
\text{Stackings} \\
\text{Bulges/Internal loops} \\
\text{Multi loops (Sequence } \geq 2 \text{ helices)}
\end{cases}
\end{align*}
\]
MFE DP equations

Terminal loops

Stackings

Bulges/Internal loops

Multi loops (Sequence ≥2 helices)

Seq. + Helix

Helix sequence

Stem loop
MFE DP equations

\[
\text{Stem loop} = \begin{cases} 
\text{Terminal loops} \\
\text{Stackings} \\
\text{Bulges/Internal loops} \\
\text{Multi loops (Sequence } \geq 2 \text{ helices)}
\end{cases}
\]

\[
\text{Seq. + Helix} = \begin{cases} 
\text{Helix sequence} \\
\text{Helix}
\end{cases}
\]
MFE DP equations

Stem loop

Terminal loops

Stackings

Bulges/Internal loops

Multi loops (Sequence ≥ 2 helices)

Helix sequence

Seq. + Helix

Helix

Unpaired base

Stem Sloop

Seq. + Helix

Helix

Unpaired base

Stem Sloop

Seq. + Helix

Helix

Unpaired base

Stem Sloop
- $E_H(i,j)$: Energy of terminal loop *enclosed by* $(i,j)$ pair
- $E_{BI}(i,j)$: Energy of bulge or internal loop *enclosed by* $(i,j)$ pair
- $E_S(i,j)$: Energy of stacking $(i,j)/(i+1,j-1)$
- Penalty for multi loop $(a)$, and occurrences of unpaired base $(b)$ and helix $(c)$ in multi loops.

**DP recurrence**

\[
\begin{align*}
\mathcal{M}'_{i,j} &= \min \left\{ \begin{array}{l}
E_H(i,j) \\
E_S(i,j) + \mathcal{M}'_{i+1,j-1} \\
\min_{i',j'} (E_{BI}(i',i,j') + \mathcal{M}'_{i',j'}) \\
a + c + \min_k (\mathcal{M}_{i+1,k-1} + \mathcal{M}^1_{k,j-1})
\end{array} \right. \\
\mathcal{M}_{i,j} &= \min_k \left\{ \min (\mathcal{M}_{i,k-1}, b(k-1)) + \mathcal{M}^1_{k,j} \right\} \\
\mathcal{M}^1_{i,j} &= \min_k \left\{ b + \mathcal{M}^1_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
\end{align*}
\]
Backtracking to reconstruct MFE structure:

\[
\begin{align*}
\mathcal{M}'_{i,j} &= \text{Min} \begin{cases} 
E_H(i,j) \\
E_S(i,j) + \mathcal{M}'_{i+1,j-1} \\
\text{Min}_{i',j'}(E_{BI}(i,i',j',j) + \mathcal{M}'_{i',j'}) \\
a + c + \text{Min}_k(M_{i+1,k-1} + \mathcal{M}^1_{k,j-1}) \end{cases} \\
\mathcal{M}_{i,j} &= \text{Min}_k \left\{ \text{min}(\mathcal{M}_{i,k-1}, b(k-1)) + \mathcal{M}^1_{k,j} \right\} \\
\mathcal{M}^1_{i,j} &= \text{Min}_k \left\{ b + \mathcal{M}^1_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
\end{align*}
\]

Complexity:

For each min, \(O(n)\) potential contributors

\( \Rightarrow \) **Worst-case** complexity in \(O(n^2)\) for naive backtrack.

Keep best contributor for each Min \( \Rightarrow \) **Backtracking in** \(O(n)\)

\( \Rightarrow \) UnaFold [MZ08]/RNAFold [HFS\textsuperscript{+}94] compute the MFE for the Turner model in overall\(^3\) time/space complexities in \(O(n^3)/O(n^2)\)

\(^3\)Using a trick/restriction for internal loops...
Backtracking to reconstruct MFE structure:

\[
\begin{align*}
\mathcal{M}_{i,j}' & = \min \{ E_H(i, j), E_S(i, j) + \mathcal{M}_{i+1,j-1}', \min_{i',j'} (E_{BI}(i, i', j, j') + \mathcal{M}_{i',j}', a + c + \min_k (\mathcal{M}_{i+1,k-1} + \mathcal{M}_{k,j-1}')) \} \\
\mathcal{M}_{i,j} & = \min_k \{ \min (\mathcal{M}_{i,k-1}, b(k-1)) + \mathcal{M}_{k,j}' \} \\
\mathcal{M}_{i,j}' & = \min_k \{ b + \mathcal{M}_{i,j-1}', c + \mathcal{M}_{i,j}' \}
\end{align*}
\]

Complexity:

For each min, \( \mathcal{O}(n) \) potential contributors

\( \Rightarrow \) **Worst-case** complexity in \( \mathcal{O}(n^2) \) for naive backtrack.

Keep best contributor for each Min \( \Rightarrow \) **Backtracking in** \( \mathcal{O}(n) \)

\( \Rightarrow \) **UnaFold** [MZ08]/**RNAFold** [HFS+94] compute the MFE for the Turner model in **overall** \( \mathcal{O}(n^3)/\mathcal{O}(n^2) \) time/space complexities

---

3 Using a trick/restriction for internal loops...
Backtracking to reconstruct MFE structure:

\[
\mathcal{M}'_{i,j} = \text{Min} \left\{ \begin{array}{c}
E_H(i,j) \\
E_S(i,j) + \mathcal{M}'_{i+1,j-1} \\
\text{Min}_{i',j'} \left( E_{BI}(i,i',j,j) + \mathcal{M}'_{i',j'} \right) \\
a + c + \text{Min}_k \left( \mathcal{M}_{i+1,k-1} + \mathcal{M}^1_{k,j-1} \right) \end{array} \right\}
\]

\[
\mathcal{M}_{i,j} = \text{Min}_k \left\{ \min \left( \mathcal{M}_{i,k-1}, b(k-1) \right) + \mathcal{M}^1_{k,j} \right\}
\]

\[
\mathcal{M}^1_{i,j} = \text{Min}_k \left\{ b + \mathcal{M}^1_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
\]

**Complexity:**

For each min, \(O(n)\) potential contributors

⇒ *Worst-case* complexity in \(O(n^2)\) for naive backtrack.

Keep best contributor for each Min ⇒ Backtracking in \(O(n)\)

⇒ UnaFold [MZ08]/RNAFold [HFS+94] compute the MFE for the Turner model in overall\(^3\) time/space complexities in \(O(n^3)/O(n^2)\)

---

\(^3\)Using a trick/restriction for internal loops...
Backtracking to reconstruct MFE structure:

\[
\begin{align*}
\mathcal{M}'_{i,j} &= \min \left\{ E_H(i,j), \right.
\left. E_S(i,j) + \mathcal{M}'_{i+1,j-1}, \right.
\left. \min_{i',j'} \left( E_{BI}(i,i',j,j') + \mathcal{M}'_{i',j'} \right) \right. \\
&\quad \left. a + c + \min_k \left( \mathcal{M}_{i+1,k-1} + \mathcal{M}^1_{k,j-1} \right) \right\} \\
\mathcal{M}_{i,j} &= \min_k \left\{ \min \left( \mathcal{M}_{i,k-1}, b(k-1) \right) + \mathcal{M}^1_{k,j} \right\} \\
\mathcal{M}^1_{i,j} &= \min_k \left\{ b + \mathcal{M}^1_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
\end{align*}
\]

**Complexity:**

For each min, \(O(n)\) potential contributors

\[\Rightarrow \textbf{Worst-case} \text{ complexity in } O(n^2) \text{ for naive backtrack}.\]

Keep best contributor for each Min \(\Rightarrow \textbf{Backtracking in } O(n)\)

\[\Rightarrow \text{UnaFold [MZ08]/RNAFold [HFS+94]} \text{ compute the MFE for the Turner model in overall\(^3\) time/space complexities in } O(n^3)/O(n^2)\]

\(^3\)Using a trick/restriction for internal loops...
Backtracking to reconstruct MFE structure:

\[ M'_{i,j} = \min \begin{cases} E_H(i, j) \\ E_S(i, j) + M'_{i+1,j-1} \\ \min_{i',j'} (E_{BI}(i, i', j', j) + M'_{i',j'}) \\ a + c + \min_k (M_{i+1,k-1} + M_{k,j-1}^1) \end{cases} \]

Complexity:

For each min, \( O(n) \) potential contributors
\( \Rightarrow \) **Worst-case** complexity in \( O(n^2) \) for **naive backtracking**.
Keep best contributor for each Min \( \Rightarrow \) **Backtracking in** \( O(n) \)

\( \Rightarrow \) UnaFold [MZ08]/RNAFold [HFS+94] compute the MFE for the Turner model in **overall** time/space complexities in \( O(n^3)/O(n^2) \)

---

\( ^3 \) Using a trick/restriction for internal loops...
Backtracking to reconstruct MFE structure:

\[
\mathcal{M}'_{i,j} = \text{Min} \left\{ \begin{array}{l}
E_H(i, j) \\
E_S(i, j) + \mathcal{M}'_{i+1,j-1} \\
\text{Min}_{i',j'} (E_{BI}(i, i', j, j') + \mathcal{M}'_{i',j'}) \\
a + c + \text{Min}_k (\mathcal{M}_{i+1,k-1} + \mathcal{M}^1_{k,j-1})
\end{array} \right\}
\]

\[
\mathcal{M}_{i,j} = \text{Min}_k \left\{ \text{min} (\mathcal{M}_{i,k-1}, b(k - 1)) + \mathcal{M}^1_{k,j} \right\}
\]

\[
\mathcal{M}^1_{i,j} = \text{Min}_k \left\{ b + \mathcal{M}^1_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
\]

Complexity:

For each min, $\mathcal{O}(n)$ potential contributors
⇒ **Worst-case** complexity in $\mathcal{O}(n^2)$ for **naive backtrack**.
Keep best contributor for each Min ⇒ **Backtracking in** $\mathcal{O}(n)$

⇒ UnaFold [MZ08]/RNAFold [HFS+94] compute the MFE for the Turner model in overall $^3$ time/space complexities in $\mathcal{O}(n^3)/\mathcal{O}(n^2)$

---

$^3$Using a trick/restriction for internal loops...
Backtracking to reconstruct MFE structure:

\[
\begin{align*}
\mathcal{M}'_{i,j} &= \text{Min} \left\{ E_H(i, j) \right. \\
& \quad \left. \quad E_S(i, j) + \mathcal{M}'_{i+1,j-1} \right. \\
& \quad \left. \quad \text{Min}_{i',j'} \left( E_{BI}(i, i', j, j') + \mathcal{M}'_{i',j'} \right) \right. \\
& \quad \left. \quad a + c + \text{Min}_k \left( \mathcal{M}_{i+1,k-1} + \mathcal{M}^1_{k,j-1} \right) \right. \\
\mathcal{M}_{i,j} &= \text{Min}_k \left\{ \text{min} \left( \mathcal{M}_{i,k-1}, b(k - 1) \right) + \mathcal{M}^1_{k,j} \right\} \\
\mathcal{M}^1_{i,j} &= \text{Min}_k \left\{ b + \mathcal{M}^1_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
\end{align*}
\]

Complexity:

For each min, \( O(n) \) potential contributors
⇒ **Worst-case** complexity in \( O(n^2) \) for naive backtrack.
Keep best contributor for each Min ⇒ **Backtracking in** \( O(n) \)

⇒ **UnaFold [MZ08]/RNAFold [HFS+94]** compute the MFE for the Turner model in overall\(^3\) time/space complexities in \( O(n^3)/O(n^2) \)

\(^3\)Using a trick/restriction for internal loops...
Two main approaches

**Definition (Ab initio folding)**
Starting from sequence, find conformation that minimizes free-energy.

**Advantages:**
- Mechanical nature allows the (in)validation of models
- Reasonable complexity $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ time/space
- *Exhaustive* nature

**Limitations:**
- Hard to include PKs
- Highly dependent on energy model
- No cooperativity
- Limited performances

**Definition (Comparative approach)**
Starting from homologous sequences, postulate common structure and find best possible tradeoff between folding & alignment.

**Avantages:**
- Better performances
- (Limited) cooperativity
- Self-improving

**Limitations**
- Easily unreasonable complexity
- Non exhaustive search
- Captures *transient* structures
Two main approaches

**Definition (Ab initio folding)**
Starting from sequence, find conformation that minimizes free-energy.

**Advantages:**
- Mechanical nature allows the (in)validation of models
- Reasonable complexity \( O(n^3)/O(n^2) \) time/space
- Exhaustive nature

**Limitations:**
- Hard to include PKs
- Highly dependent on energy model
- No cooperativity
- Limited performances

**Definition (Comparative approach)**
Starting from homologous sequences, postulate common structure and find best possible tradeoff between folding & alignment.

**Avantages :**
- Better performances
- (Limited) cooperativity
- Self-improving

**Limitations**
- Easily unreasonable complexity
- Non exhaustive search
- Captures transient structures
Performances

Sequence

Taille Sens.
<700 70-73% 
[MSZT99, DCCG04]
16s,23s ~50%
MCC~ 0.5 [GG04]

Thermodynamics

Sens. Spé. MCC.
~75% ~75% 0.8
Comparative [GG04]

2^ary Structure

Reminder: \( MCC = \frac{t^+ t^- f^+ f^-}{\sqrt{(t^+ f^+)(t^+ f^-)(t^- f^+)(t^- f^-)}} \)
Performances

Sequence

Taille Sens.
<700 70-73%
[MSZT99, DCCG04]
16s, 23s ~50%
MCC~ 0.5 [GG04]

Thermodynamics

Sens. Spé. MCC.
~75% ~75% 0.8

Comparative [GG04]

Reminder: $MCC = \frac{t^+ t^- f^+ f^-}{\sqrt{(t^+ + f^+)(t^- + f^-)(t^- + f^+)(t^+ + f^-)}}$

2°ry Structure
Performances

Sequence

Taille Sens.
<700 70-73% [MSZT99, DCCG04]
16s, 23s ∼50%
MCC ∼ 0.5 [GG04]

Thermodynamics

Sens. Spé. MCC.
∼75% ∼75% 0.8

Comparative [GG04]

Reminder: \[ MCC = \frac{t^+ t^- f^+ f^-}{\sqrt{(t^+ + f^+)(t^+ + f^-)(t^- + f^+)(t^- + f^-)}} \]
Goal: From sequence to all-atom/coarse grain 3D models!!!

- Comparative models + Molecular dynamics: RNA2D3D [SYKB07]
- Pipeline MC-Fold/MC-sym [PM08]
Towards a 3D ab-initio prediction

**Goal:** From sequence to all-atom/coarse grain 3D models!!!

- Comparative models + Molecular dynamics: RNA2D3D \([SYKB07]\)

- Pipeline MC-Fold/MC-sym \([PM08]\)
Towards a 3D ab-initio prediction

**Goal:** From sequence to all-atom/coarse grain 3D models!!!

- Comparative models + Molecular dynamics: RNA2D3D \[\text{SYKB07}\]
- Pipeline MC-Fold/MC-sym \[\text{PM08}\]

```
UUAGGCGGCCCAAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC
```

**Séquence**

1. **MCFold**
2. \(2^{\text{ary+}}\) Structure
3. **MCSym**
4. \(3^{\text{ary}}\) Structure
Classifying RNA pseudoknotted structures.

Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction.

P. Gardner and R. Giegerich.
A comprehensive comparison of comparative RNA structure prediction approaches.

Fast folding and comparison of RNA secondary structures.

R. B. Lyngsø and C. N. S. Pedersen.
RNA pseudoknot prediction in energy-based models.

N. Leontis and E. Westhof.
Geometric nomenclature and classification of RNA base pairs.

Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.

Jan Manuch, Chris Thachuk, Ladislav Stacho, and Anne Condon.
NP-completeness of the direct energy barrier problem without pseudoknots.
N. R. Markham and M. Zuker.

M. Parisien and F. Major.
The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.

Lioudmila V Sharova, Alexei A Sharov, Timur Nedorezov, Yulan Piao, Nabeebi Shaik, and Minoru S H Ko.
Database for mrna half-life of 19 977 genes obtained by dna microarray analysis of pluripotent and differentiating mouse embryonic stem cells.

B. A. Shapiro, Y. G. Yingling, W. Kasprzak, and E. Bindewald.
Bridging the gap in rna structure prediction.
Exercise: Parsing/folding RNAs (Python)