

# M2 BIM/STRUCT - Lecture 3

## Advanced dynamic programming and alignment

Yann Ponty

AMIBio Team  
École Polytechnique/CNRS

## Two main routes for 2D prediction

### 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 routes for 2D prediction

### 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.

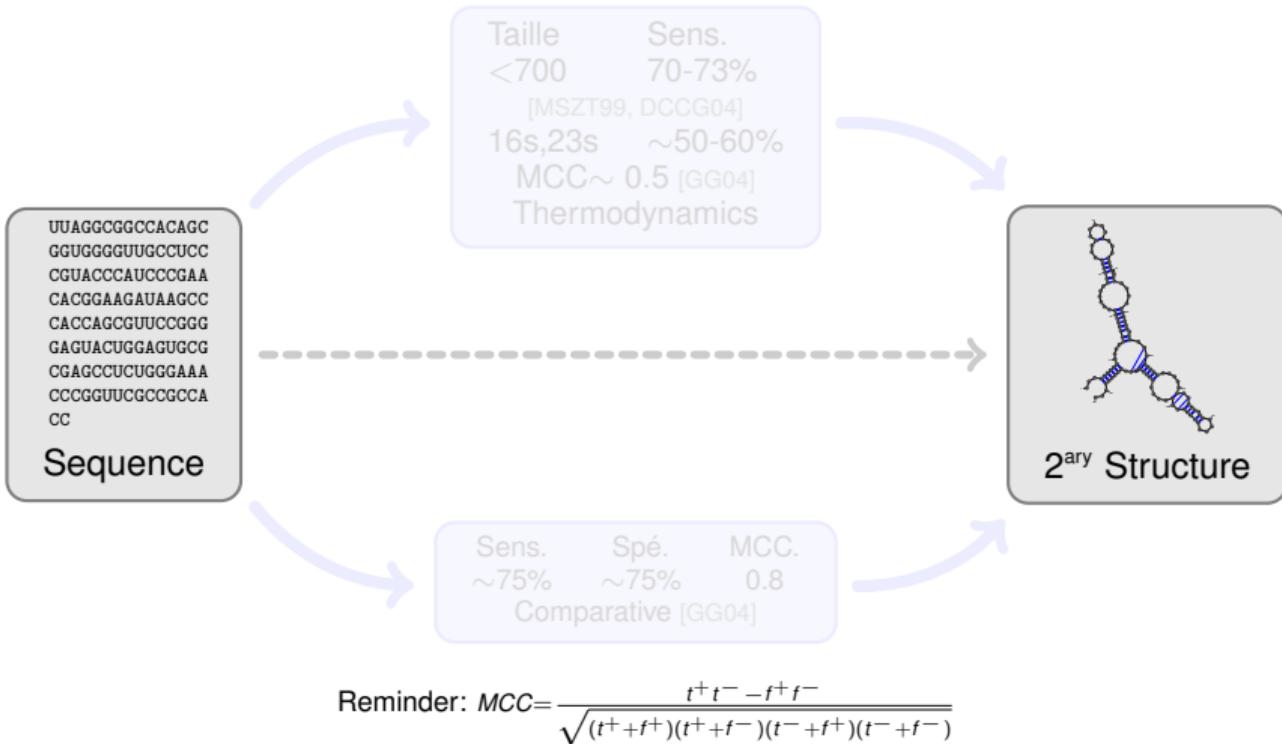
#### Avantages :

- ▶ Better performances
- ▶ (Limited) cooperativity
- ▶ Self-improving

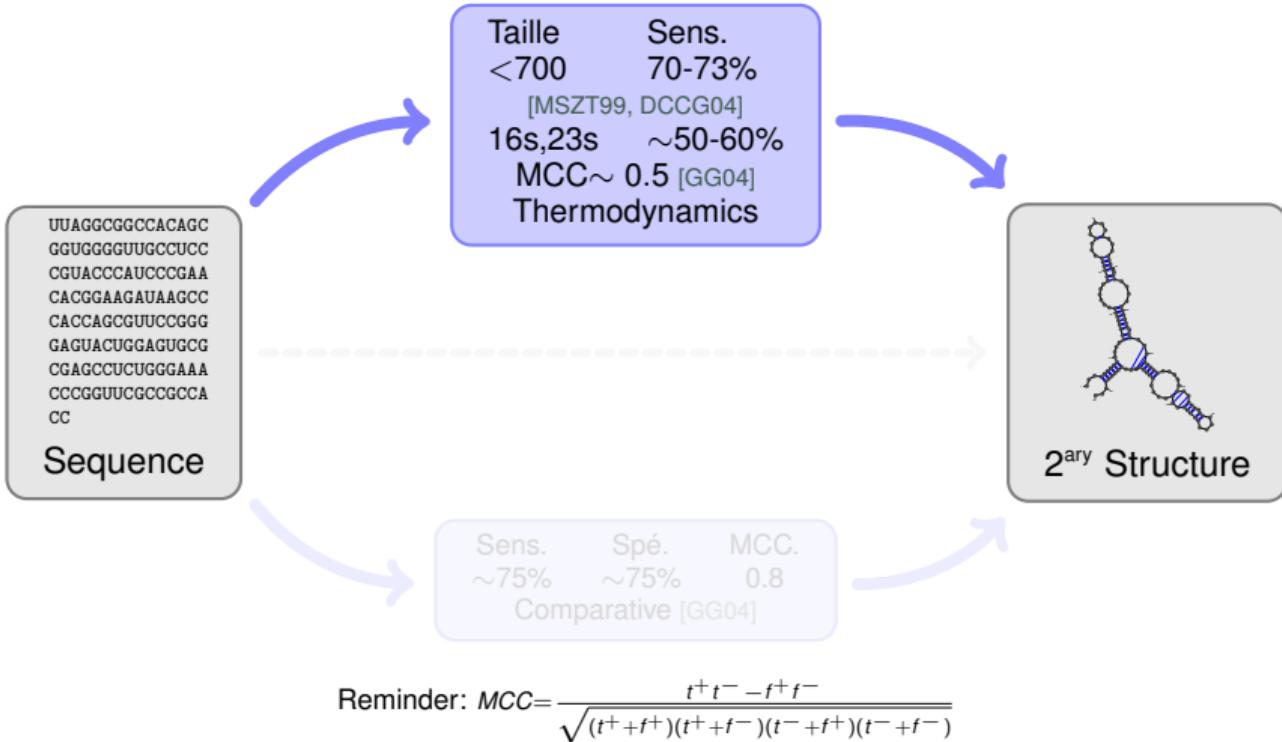
#### Limitations

- ▶ Easily unreasonable complexity
- ▶ Non exhaustive search
- ▶ Captures *transient* structures

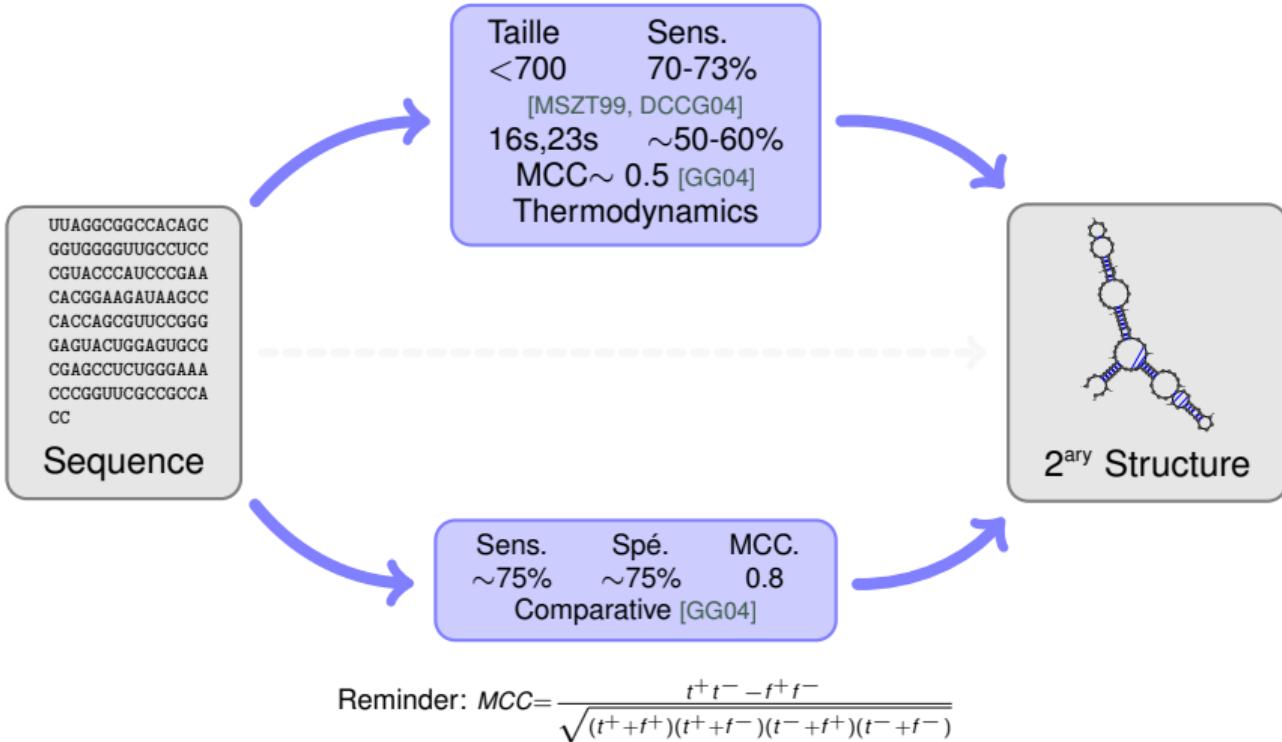
# Performances



## Performances

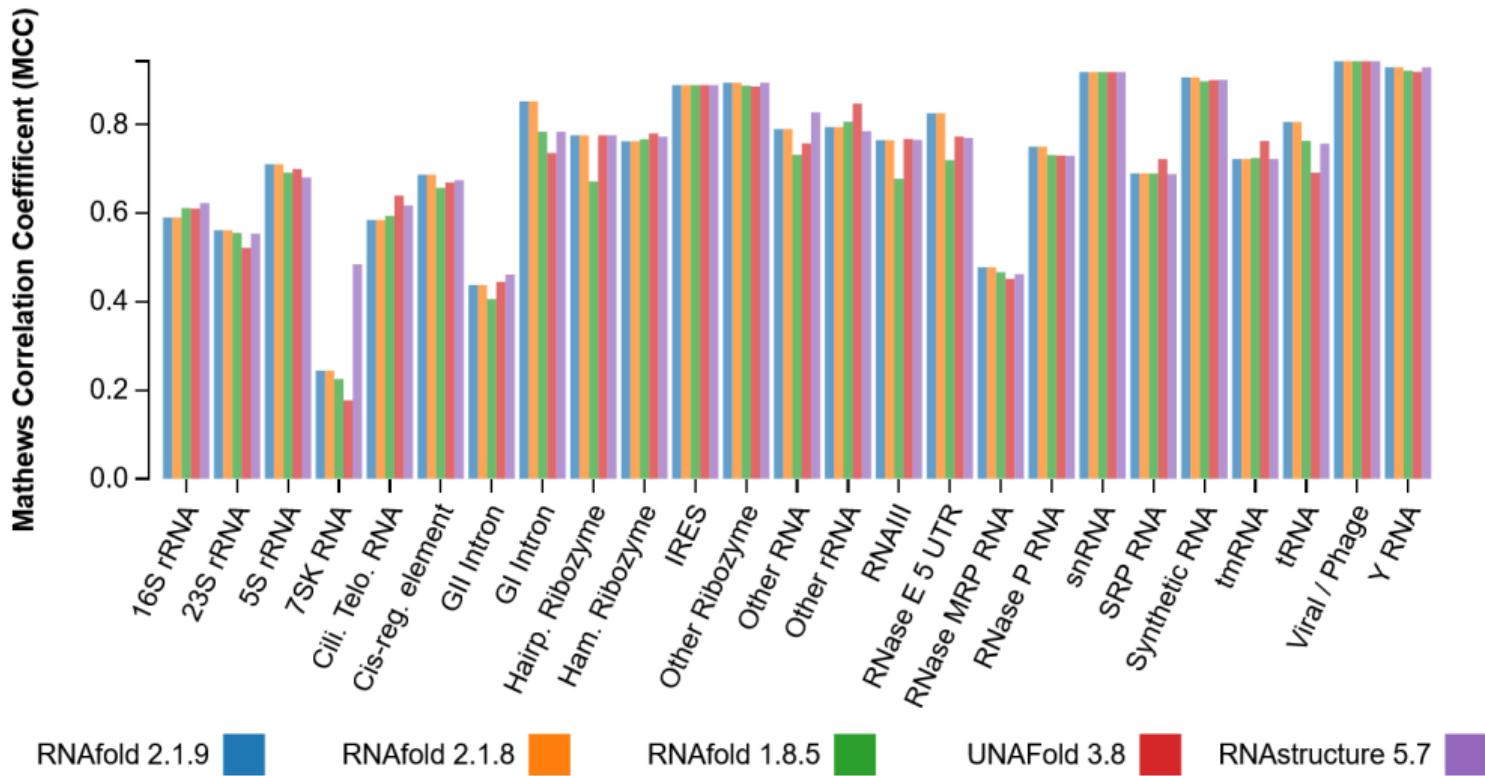


## Performances



# Performances

## Performance Benchmark (by RNA class)



# Towards a 3D ab-initio prediction

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

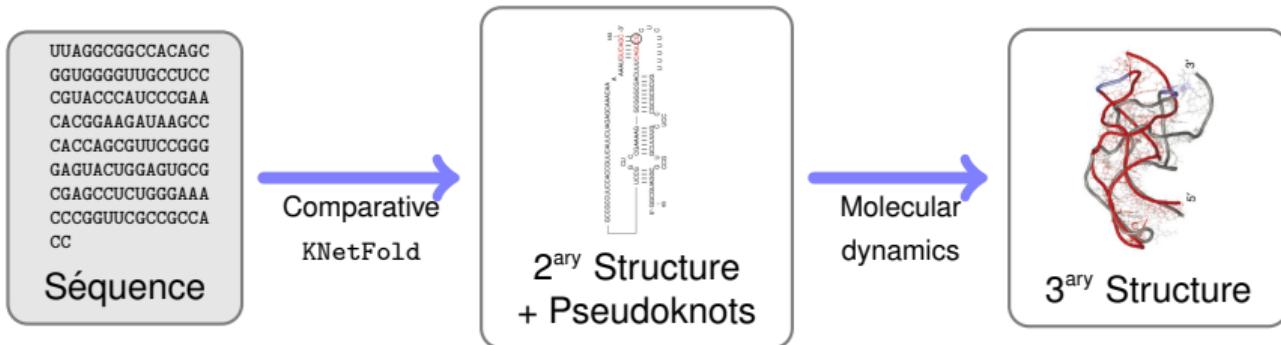
- ▶ Comparative models + Molecular dynamics: RNA2D3D [SYKB07]
- ▶ Pipeline MC-Fold/MC-sym [PM08]



# Towards a 3D ab-initio prediction

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

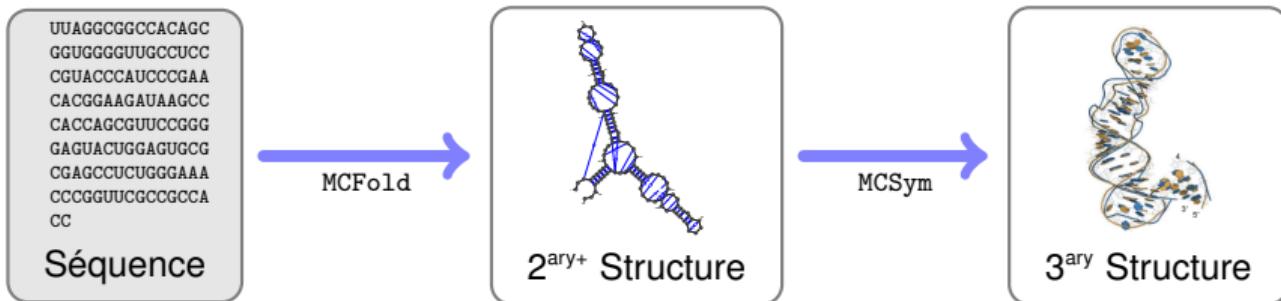
- ▶ Comparative models + Molecular dynamics: RNA2D3D [SYKB07]
- ▶ Pipeline MC-Fold/MC-sym [PM08]



# 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]



## Suboptimal structures

Prob.: Simplified energy model (no pseudoknots, only canonical BPs)

⇒ Native structure (functional) could be **overthrown**.

⇒ Investigate suboptimal structures (RNASubopt [WFHS99]),

i.e. build all structures within  $\Delta \text{ KCal.mol}^{-1}$  of MFE:

- ▶ Compute minimum free-energy matrices
- ▶ Backtrack on any contribution within  $\Delta$  of MFE;
- ▶ Update  $\Delta$  such that future backtracks create  $\geq 1$  struct.
- ▶ Recursively generate subopts and combine (brute-force ou Sort)

$$\mathcal{M}'_{1,n,\Delta} = \min \left( \begin{array}{l} \rightarrow a + c + \text{Min} \left( \mathcal{M}_{i+1,k_0-1} + \mathcal{M}^1_{k_0,j-1} \right) \\ \rightarrow a + c + \text{Min} \left( \mathcal{M}_{i+1,k_1-1} + \mathcal{M}^1_{k_1,j-1} \right) \\ \rightarrow a + c + \text{Min} \left( \mathcal{M}_{i+1,k_2-1} + \mathcal{M}^1_{k_2,j-1} \right) \end{array} \right)$$
$$E_0 - \mathcal{M}'_{1,n} = \varepsilon_0 \leq \Delta$$
$$E_1 - \mathcal{M}'_{1,n} = \varepsilon_1 > \Delta$$
$$E_2 - \mathcal{M}'_{1,n} = \varepsilon_2 \leq \Delta$$

## Suboptimal structures

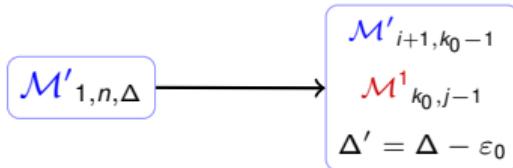
Prob.: Simplified energy model (no pseudoknots, only canonical BPs)

⇒ Native structure (functional) could be **overthrown**.

⇒ Investigate suboptimal structures (RNASubopt [WFHS99]),

i.e. build all structures within  $\Delta \text{ KCal.mol}^{-1}$  of MFE:

- ▶ Compute minimum free-energy matrices
- ▶ Backtrack on any contribution within  $\Delta$  of MFE;
- ▶ Update  $\Delta$  such that future backtracks create  $\geq 1$  struct.
- ▶ Recursively generate subopts and combine (brute-force ou Sort)



## Suboptimal structures

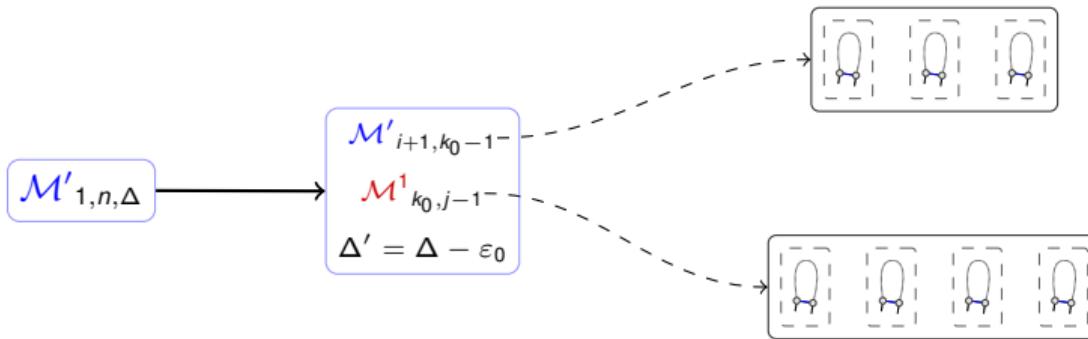
Prob.: Simplified energy model (no pseudoknots, only canonical BPs)

⇒ Native structure (functional) could be **overthrown**.

⇒ Investigate suboptimal structures (RNASubopt [WFHS99]),

i.e. build all structures within  $\Delta \text{ KCal.mol}^{-1}$  of MFE:

- ▶ Compute minimum free-energy matrices
- ▶ Backtrack on any contribution within  $\Delta$  of MFE;
- ▶ Update  $\Delta$  such that future backtracks create  $\geq 1$  struct.
- ▶ Recursively generate subopts and combine (brute-force ou Sort)



## Suboptimal structures

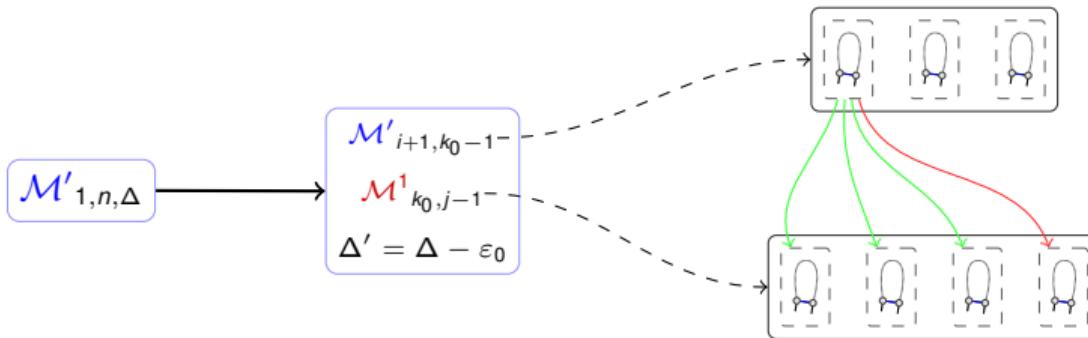
Prob.: Simplified energy model (no pseudoknots, only canonical BPs)

⇒ Native structure (functional) could be **overthrown**.

⇒ Investigate suboptimal structures (RNASubopt [WFHS99]),

i.e. build all structures within  $\Delta \text{ KCal.mol}^{-1}$  of MFE:

- ▶ Compute minimum free-energy matrices
- ▶ Backtrack on any contribution within  $\Delta$  of MFE;
- ▶ Update  $\Delta$  such that future backtracks create  $\geq 1$  struct.
- ▶ Recursively generate subopts and combine (brute-force ou Sort)



## Suboptimal structures

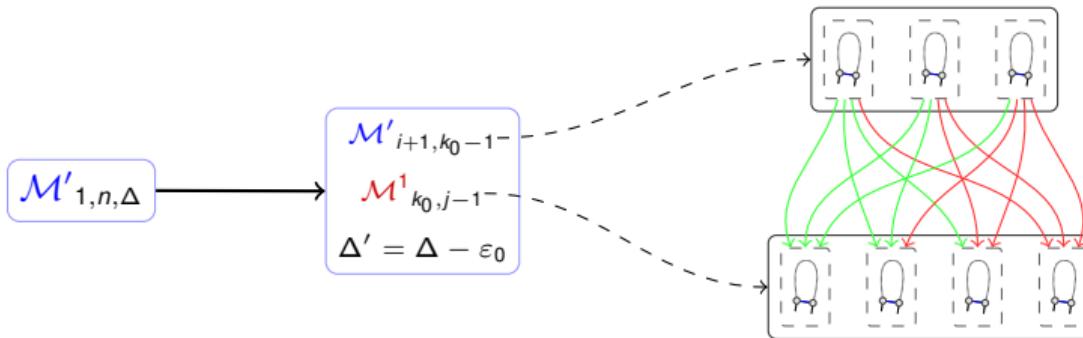
Prob.: Simplified energy model (no pseudoknots, only canonical BPs)

⇒ Native structure (functional) could be **overthrown**.

⇒ Investigate suboptimal structures (RNASubopt [WFHS99]),

i.e. build all structures within  $\Delta \text{ KCal.mol}^{-1}$  of MFE:

- ▶ Compute minimum free-energy matrices
- ▶ Backtrack on any contribution within  $\Delta$  of MFE;
- ▶ Update  $\Delta$  such that future backtracks create  $\geq 1$  struct.
- ▶ Recursively generate subopts and combine (brute-force ou Sort)



## Suboptimal structures

Prob.: Simplified energy model (no pseudoknots, only canonical BPs)

⇒ Native structure (functional) could be **overthrown**.

⇒ Investigate suboptimal structures (RNASubopt [WFHS99]),

i.e. build all structures within  $\Delta \text{ KCal.mol}^{-1}$  of MFE:

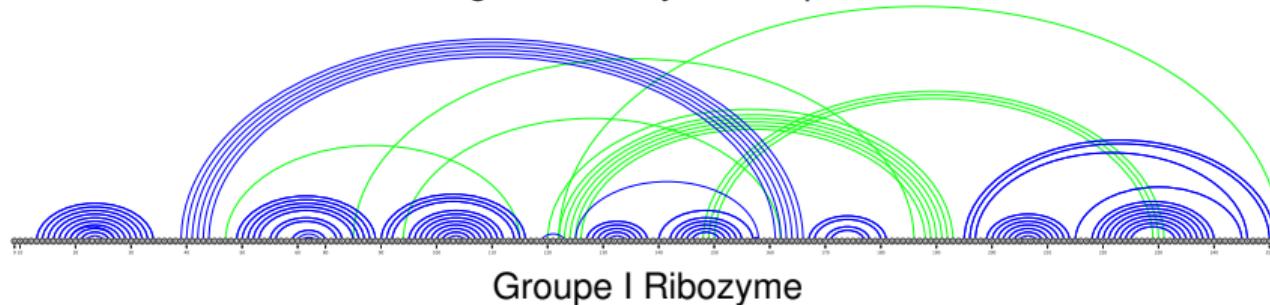
- ▶ Compute minimum free-energy matrices
- ▶ Backtrack on any contribution within  $\Delta$  of MFE;
- ▶ Update  $\Delta$  such that future backtracks create  $\geq 1$  struct.
- ▶ Recursively generate subopts and combine (**brute-force** ou **Sort**)

⇒ Time complexity (**Sort**) :  $\mathcal{O}(n^3 + n \cdot k \log(k))$

( $k$  grows exponentially fast with  $\Delta$ !)

## Predicting pseudoknotted structures

Pseudoknots are essential to the folding and activity of multiple RNA families.



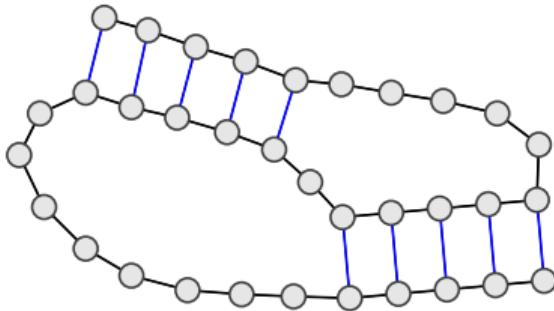
Their disregard within current folding algorithms stems both from **algorithmic** and **energetic** intricacies.

(**Pseudoknots** = Crossings  $\Rightarrow$  foldings delimited by base-pair can no longer be assumed to be independent)

Type	Complexity	Reference
Secondary structures	$\mathcal{O}(n^3)$	[MSZT99]
L&P	$\mathcal{O}(n^5)$	[LP00]
D&P	$\mathcal{O}(n^5)$	[DP03]
A&U	$\mathcal{O}(n^5)$	[Aku00]
R&E	$\mathcal{O}(n^6)$	[RE99]
Unconstrained	NP-complete	[LP00]

## Akutsu/Uemura Algorithm

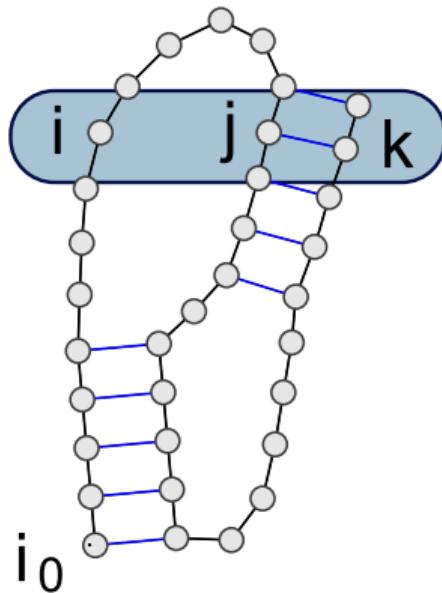
**Goal:** Capture a category of simple, yet recurrent, pseudoknots.



**Idea:** When such a PK motif is rotated, one can deduce the MFE of a triplet  $(i, j, k)$  from the MFE of triplets directly below it.

## Akutsu/Uemura Algorithm

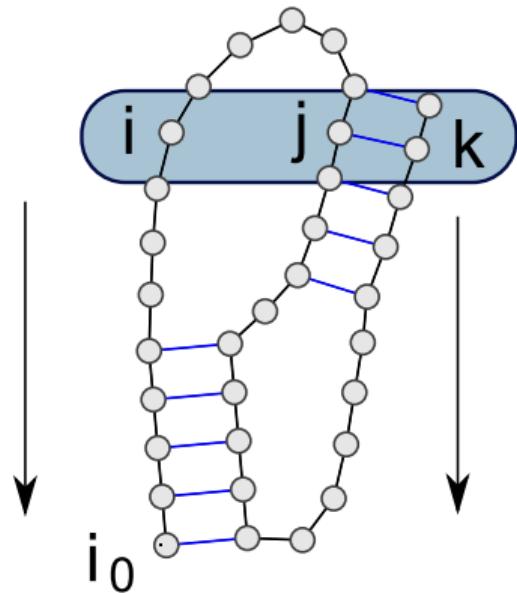
Goal: Capture a category of simple, yet recurrent, pseudoknots.



Idea: When such a PK motif is rotated, one can deduce the MFE of a triplet  $(i, j, k)$  from the MFE of triplets directly below it.

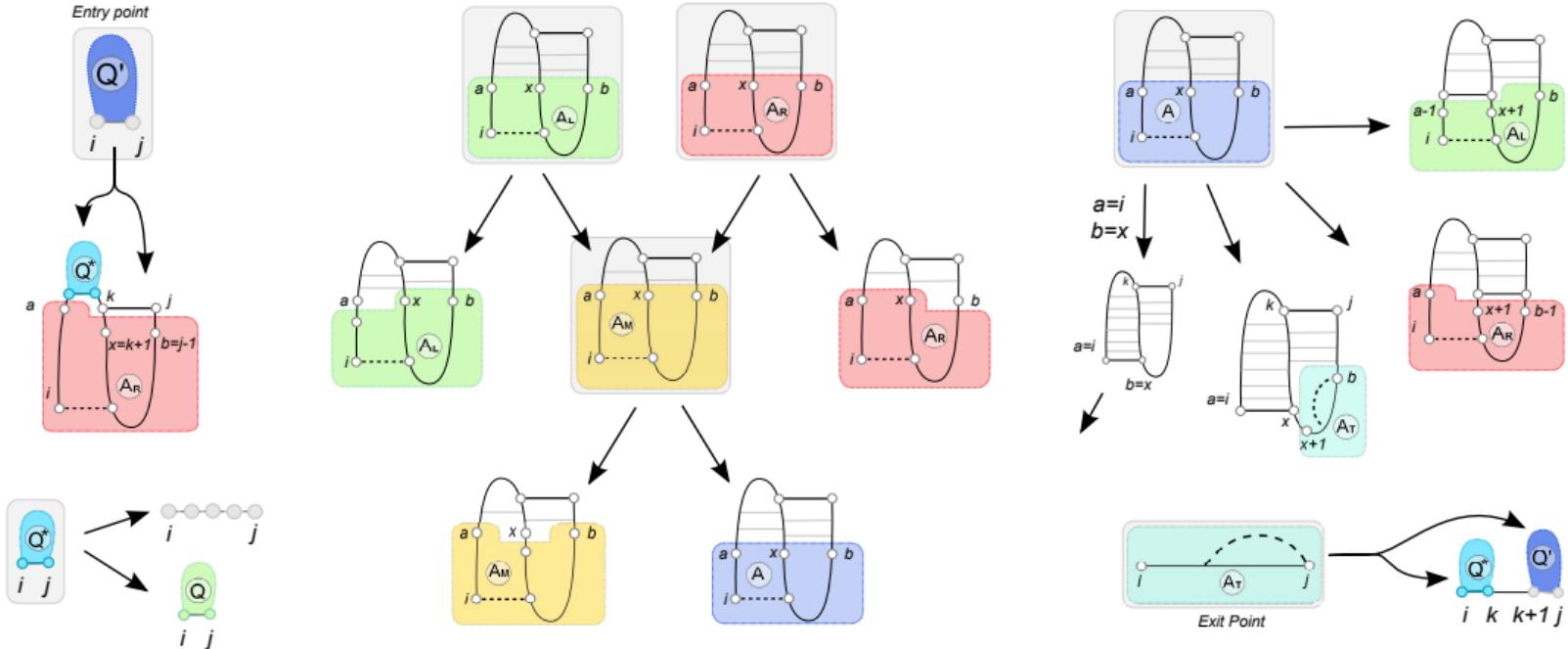
## Akutsu/Uemura Algorithm

Goal: Capture a category of simple, yet recurrent, pseudoknots.



Idea: When such a PK motif is rotated, one can deduce the MFE of a triplet  $(i, j, k)$  from the MFE of triplets directly below it.

# Akutsu/Uemura: Dynamic programming



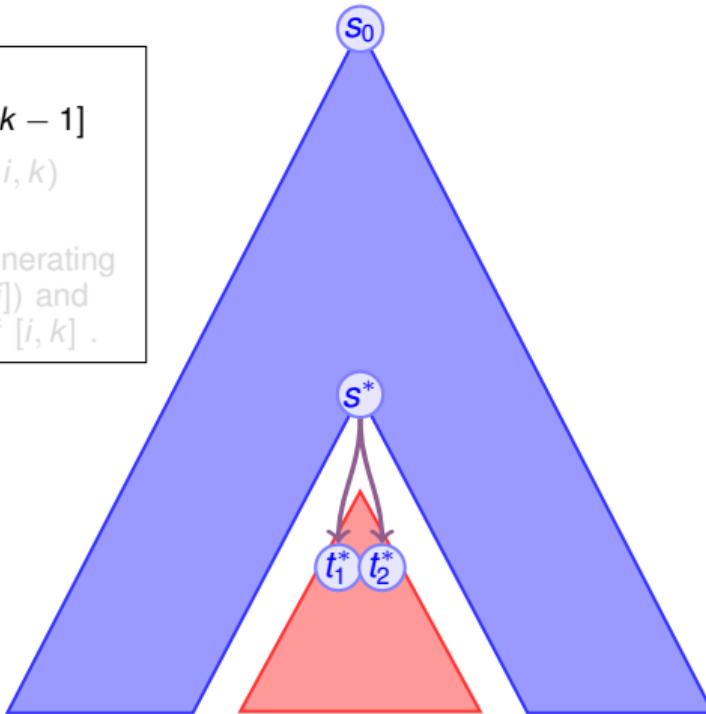
Application/Problem	Weight fun.	Time/Space	Ref.
Energy minimization	$\frac{\pi_{bp}}{RT}$	$\mathcal{O}(n^4)/\mathcal{O}(n^4)$	[Aku00]
Partition function	$e^{-\frac{\pi_{bp}}{RT}}$	$\mathcal{O}(n^4)/\mathcal{O}(n^4)$	$\Theta(n^6)$ [CC09]
BP probabilities	$e^{-\frac{\pi_{bp}}{RT}}$	$\mathcal{O}(n^4)/\mathcal{O}(n^4)$	-
Sampling ( $k$ -struct.)	$e^{-\frac{\pi_{bp}}{RT}}$	$\mathcal{O}(n^4 + kn \log n)/\mathcal{O}(n^4)$	-

Exercice: Write DP equation for MFE computation, counting and partition function.

## Inside/outside algorithm

Structure including base pair  $(i, k)$ :

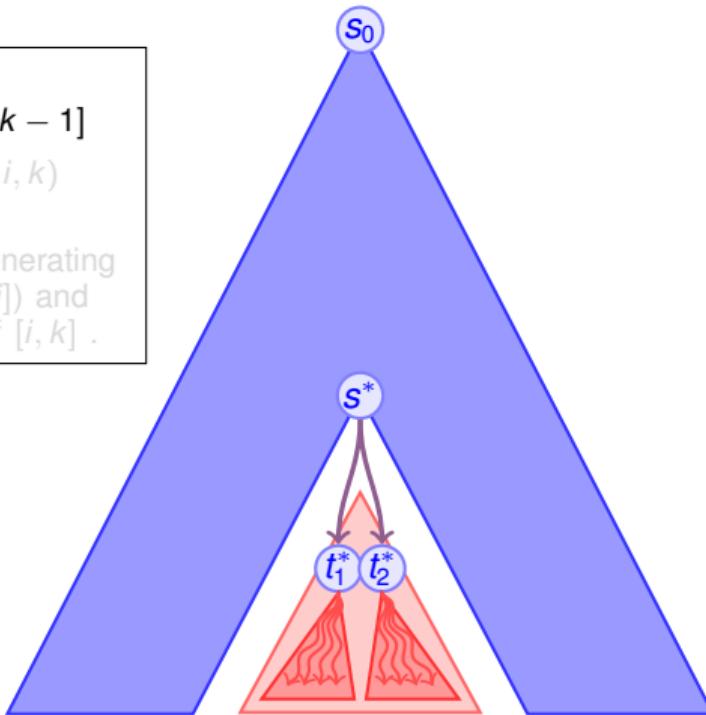
- ▶ **Inside:** Structures over  $[i + 1, k - 1]$
- ▶ **Outside:** Contexts of interval  $(i, k)$ 
  - ▶  $\forall$  interval  $[i, j], i < j \leq k$
  - ▶ Complete structure by generating brother intervals  $([k + 1, j])$  and recurse over the father of  $[i, k]$ .



## Inside/outside algorithm

Structure including base pair  $(i, k)$ :

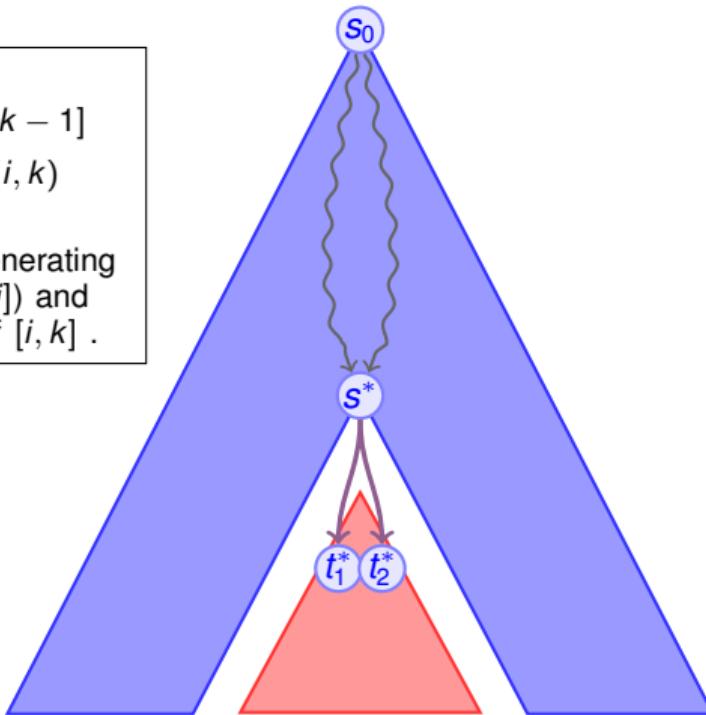
- ▶ **Inside:** Structures over  $[i + 1, k - 1]$
- ▶ **Outside:** Contexts of interval  $(i, k)$ 
  - ▶  $\forall$  interval  $[i, j], i < j \leq k$
  - ▶ Complete structure by generating brother intervals  $([k + 1, j])$  and recurse over the father of  $[i, k]$ .



## Inside/outside algorithm

Structure including base pair  $(i, k)$ :

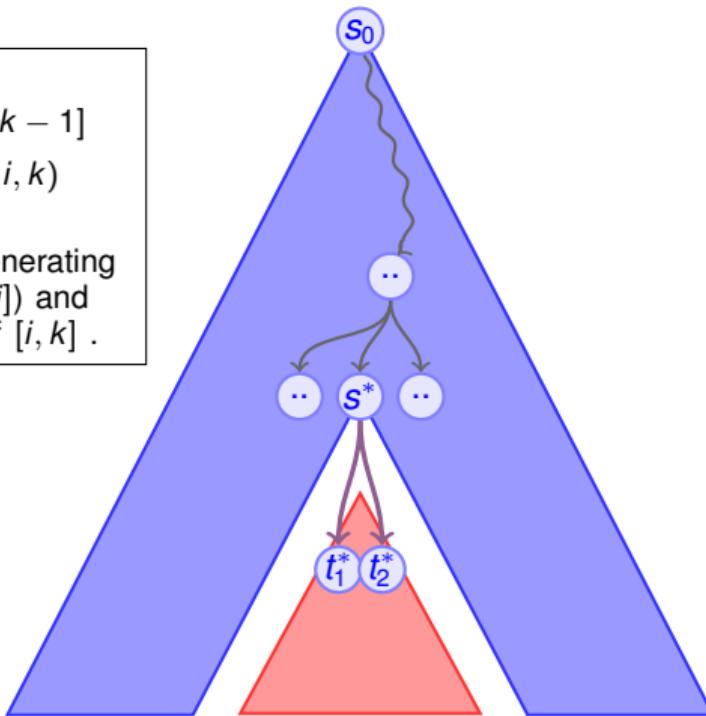
- ▶ **Inside:** Structures over  $[i + 1, k - 1]$
- ▶ **Outside:** Contexts of interval  $(i, k)$ 
  - ▶  $\forall$  interval  $[i, j], i < j \leq k$
  - ▶ Complete structure by generating **brother intervals**  $([k + 1, j])$  and recurse over the **father** of  $[i, k]$ .



## Inside/outside algorithm

Structure including base pair  $(i, k)$ :

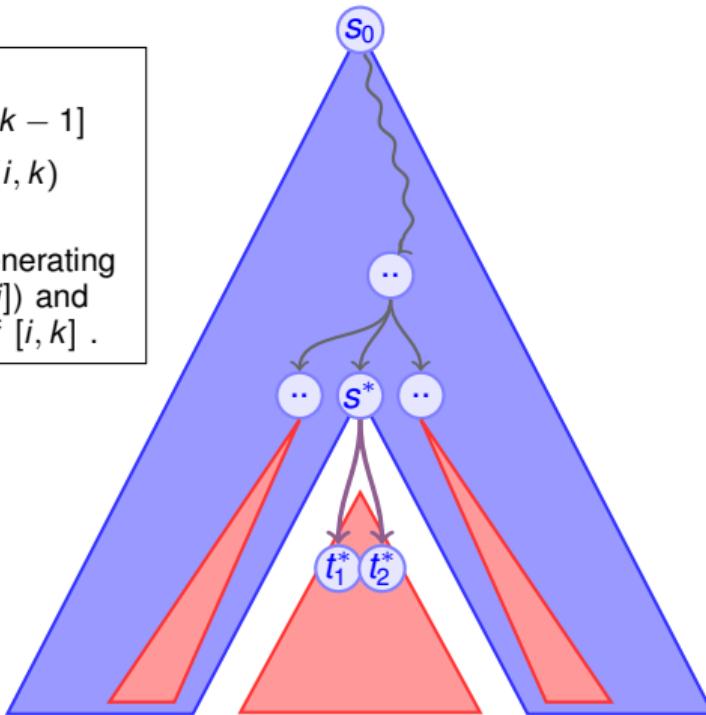
- ▶ **Inside:** Structures over  $[i + 1, k - 1]$
- ▶ **Outside:** Contexts of interval  $(i, k)$ 
  - ▶  $\forall$  interval  $[i, j], i < j \leq k$
  - ▶ Complete structure by generating brother intervals  $([k + 1, j])$  and recurse over the father of  $[i, k]$ .



## Inside/outside algorithm

Structure including base pair  $(i, k)$ :

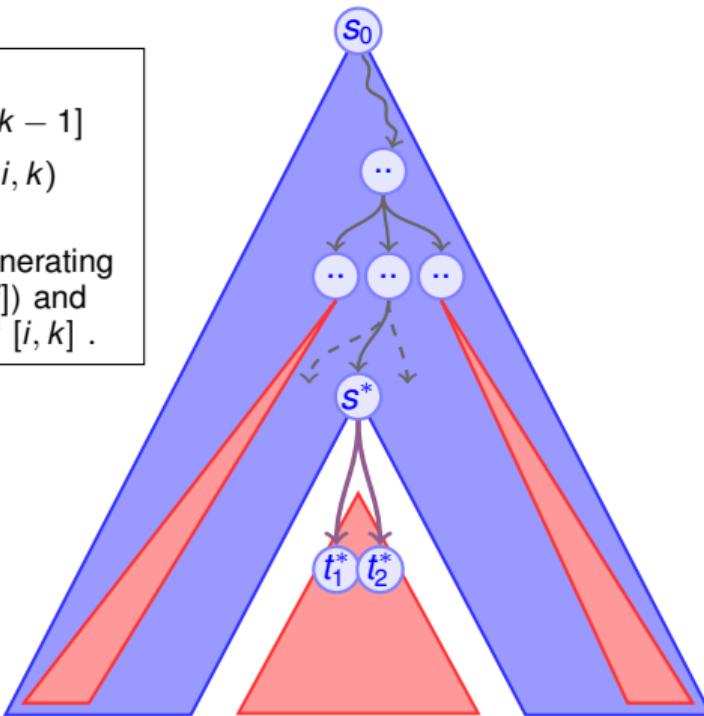
- ▶ **Inside:** Structures over  $[i + 1, k - 1]$
- ▶ **Outside:** Contexts of interval  $(i, k)$ 
  - ▶  $\forall$  interval  $[i, j], i < j \leq k$
  - ▶ Complete structure by generating brother intervals  $([k + 1, j])$  and recurse over the father of  $[i, k]$ .



## Inside/outside algorithm

Structure including base pair  $(i, k)$ :

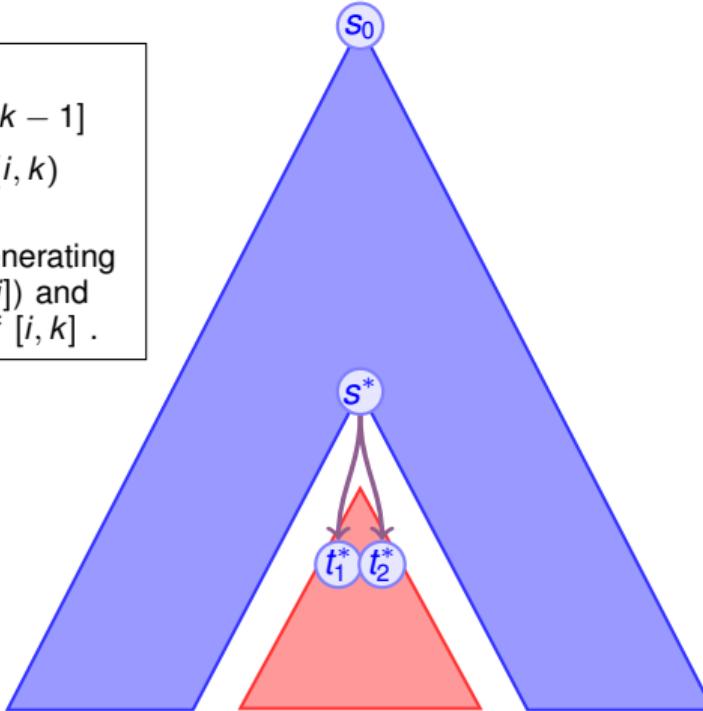
- ▶ **Inside:** Structures over  $[i + 1, k - 1]$
- ▶ **Outside:** Contexts of interval  $(i, k)$ 
  - ▶  $\forall$  interval  $[i, j], i < j \leq k$
  - ▶ Complete structure by generating brother intervals  $([k + 1, j])$  and recurse over the father of  $[i, k]$ .



## Inside/outside algorithm

Structure including base pair  $(i, k)$ :

- ▶ **Inside:** Structures over  $[i + 1, k - 1]$
- ▶ **Outside:** Contexts of interval  $(i, k)$ 
  - ▶  $\forall$  interval  $[i, j], i < j \leq k$
  - ▶ Complete structure by generating **brother intervals**  $([k + 1, j])$  and recurse over the **father** of  $[i, k]$ .



Whenever some further **technical conditions** are satisfied, this decomposition is **complete** and **unambiguous**, and implies a **simple recurrence** for computing the base pair probability matrix in  $\Theta(n^3)$ .

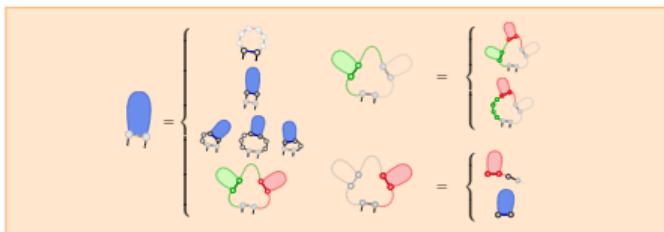
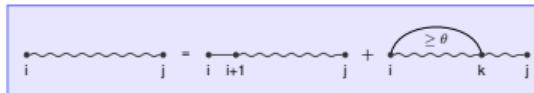
Alternatively: Duplicate sequence

## What is a good dynamic programming scheme?

## Correction of a (Ensemble) dynamic programming scheme:

Objective function correctly computed/inherited at local level

- + All the conformations can be obtained
  - ⇒ Correct algorithm (Induction)



Enumerating search space helps **but** does not constitute a proof.

Need to **show equivalence** of DP schemes, e.g. use one to simulate the other and vice versa.  
(Generating functions may help)

# What is a good dynamic programming scheme?

Correction of a (Ensemble) dynamic programming scheme:

Objective function correctly computed/inherited at local level

- + All the conformations can be obtained
- ⇒ Correct algorithm (Induction)

$$\begin{aligned}C_{i,t} &= 1, \quad \forall t \in [i, i + \theta] \\ C_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^j 1 \times C_{i+1,k-1} \times C_{k+1,j} \right.\end{aligned}$$

Homopolymer (All pairs allowed) +  $\theta = 1$   
 $\Rightarrow C_{1,n} = 1, 1, 1, 2, 4, 8, 17, 32, 82, 185, 423, \dots$



$$\begin{aligned}\mathcal{C}'_{i,j} &= \sum \begin{cases} 1 & \mathcal{C}'_{i+1,j-1} \\ \sum_{i',j'} \mathcal{C}'_{i',j'} & \sum_k C_{i+1,k-1} \times \mathcal{C}^1_{k,j-1} \end{cases} \\ \mathcal{C}_{i,j} &= \sum_k ((\mathcal{C}_{i,k-1} + 1) \times \mathcal{C}^1_{k,j}) \\ \mathcal{C}^1_{i,j} &= \mathcal{C}^1_{i,j-1} + \mathcal{C}'_{i,j}\end{aligned}$$

Homopolymer +  $\theta = 1$   
 $\Rightarrow \mathcal{C}'_{1,n} = 0, 1, 1, 2, 4, 8, 17, 32, 82, 185, 423, \dots$

Enumerating search space helps but does not constitute a proof.

Need to show equivalence of DP schemes, e.g. use one to simulate the other and vice versa.  
(Generating functions may help)

## Structural alignment: Why?

**Hypothesis:** Common evolutionary pressure = Common function .

Within certain RNA families (ex.: RNase-P), low sequence conservation **yet** high structural conservation.

**Algorithmic problems:**

- ▶ **Editing:** Compute *distance* between two secondary structures  $A$  and  $B$ .  
Find minimal cost sequence of operations to turn  $A$  into  $B$ . Already NP-complete for two secondary structures [BFRS07].
- ▶ **Alignment:** Find minimal cost super-structure.  
Generalizes sequence alignment. Polynomial ( $\mathcal{O}(n^4)$ ) for secondary structures [BDD<sup>+</sup>08], NP-complete in 3D [SZS<sup>+</sup>08].  
**Alternatives:** Local/global alignment, motifs search (aka small-in-large).
- ▶ **Superimposition:** Find solid-body geometric transform (Rotation, translation, zoom) to superimpose *as well as possible* the coordinates of two RNAs having **known matching**.  
Polynomial in 3D [McL82].

**Remark:** Algorithmic hardness stems from finding the matching (i.e. combinatorial, not geometric).

## FR3D: A geometric approach

When 3D models are available, the alignment problem can be tackled in a **purely geometric setting**.

### Problem

**Input:** Motif  $m$ , target structure  $b$  (ordered set of 3D points).

**Output:** Matching of  $m$  versus a subset of  $b$  that minimizes a notion of geometric **discrepancy**.

**Geometric discrepancy:** In FR3D [SZS<sup>+</sup>08], a **discrepancy** function  $D$  combines two error functions  $L$  et  $A$ , respectively accounting for the **superimposability** ( $L$ ) and **base orientation** ( $A$ ) of  $m$  and  $b$ .

$$L = \sqrt{\min_{R, T} \sum_{i=1}^m \|b_i - R(T(m_i))\|^2} \quad A = \sqrt{\sum_{i=1}^m \alpha_i^2} \quad D = \frac{1}{m} \sqrt{L^2 + A^2}$$

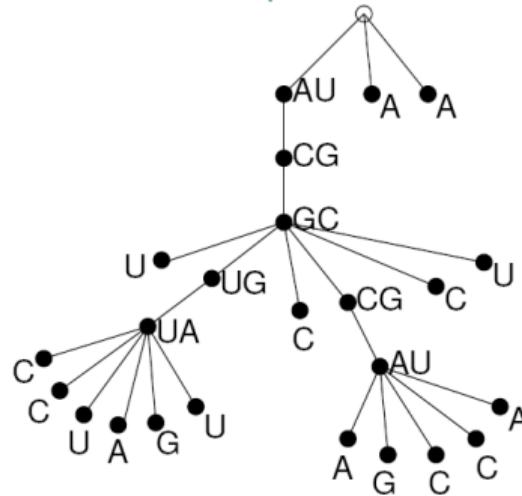
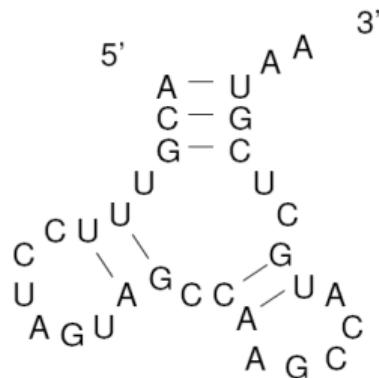
$R, T$ : Rotation and translation.  $c_i$ : Center of mass (CM) of base  $m_i$ .  $\alpha_i$ : Spread between orientation of CMs/bases in  $m_i$  et  $b_i$ .

Backtrack + Incremental pruning (Bounds on  $D$ )  $\Rightarrow$  **Combinatorial explosion!**

**But** exact search feasible for smaller motifs.

# Structures to Trees

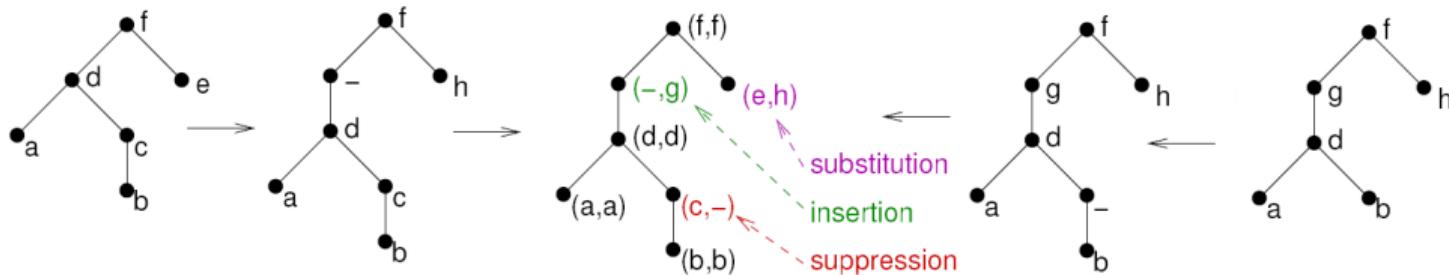
The alignment of two secondary structures is based on their tree-like representations<sup>1</sup>.



Base pairs  $\Rightarrow$  internal nodes

Unpaired bases  $\Rightarrow$  Leaves

**Alignment** = Complete matching having minimal cost.



## Aligning Trees<sup>2</sup>

$$\delta(\text{Tree}_1, \text{Tree}_2) = \min \left\{ \begin{array}{l} \delta(\text{Tree}_1, \text{Tree}_2) + \text{del}(\bullet) \\ \delta(\text{Tree}_1, \text{Tree}_2) + \text{ins}(\bullet) \\ \delta(\text{Tree}_1, \text{Tree}_2) + \text{subst}(\bullet, \bullet) \end{array} \right.$$

## Aligning Forests

$$\delta(\text{Forest}_1, \text{Forest}_2) = \min \left\{ \begin{array}{l} \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Forest}_1 \setminus \text{Tree}_1, \text{Forest}_2 \setminus \text{Tree}_2) \mid \text{Forest}_1 \setminus \text{Tree}_1 = \text{Forest}_2 \setminus \text{Tree}_2 \} \\ \quad + \text{del}(\bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Forest}_1 \setminus \text{Tree}_1, \text{Forest}_2 \setminus \text{Tree}_2) \mid \text{Forest}_1 \setminus \text{Tree}_1 = \text{Forest}_2 \setminus \text{Tree}_2 \} \\ \quad + \text{ins}(\bullet) \\ \delta(\text{Forest}_1, \text{Forest}_2) + \delta(\text{Forest}_1 \setminus \text{Tree}_1, \text{Forest}_2 \setminus \text{Tree}_2) \end{array} \right.$$

Worst-case complexity in  $\mathcal{O}(n^4)$  [JWZ94], on average in  $\mathcal{O}(n^2)$  [HDD07].

But RNA-specific operations are lacking

---

<sup>2</sup>Idem

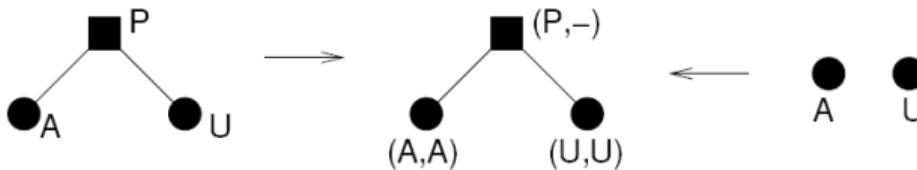
## RNAForester [HVG04]

Parametrization of operation costs, but some operations, atomic in a realistic model, must be composed from available ones.

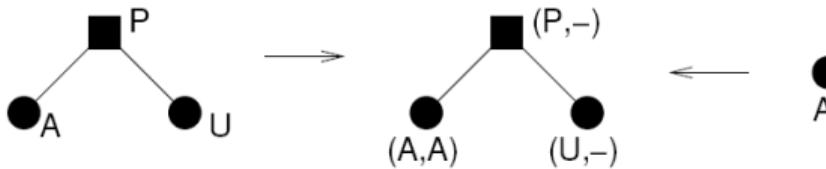
Example: To detach a base-pair, delete node (base-pair), and insert two leaves (bases).

RNAForester: Based on Jiang, Wang & Zhang algorithm  
+ Integration of RNA-specific operations<sup>3</sup>.

arc-breaking



arc-altering



<sup>3</sup>Idem

## NestedAlign [BDD<sup>+</sup>08]

$$\delta(\text{Tree}_1, \text{Tree}_2) = \min \left\{ \begin{array}{l} \delta(\text{Tree}_1, \text{Tree}_2) + \text{BDel}(\bullet) \\ \delta(\text{Tree}_1, \text{Tree}_2) + \text{BIns}(\bullet) \\ \delta(\text{Tree}_1, \text{Tree}_2) + \text{BSub}(\bullet, \bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{PDel}(\bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{PIns}(\bullet) \\ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) + \text{PSub}(\bullet, \bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{Fus}(\bullet, \bullet, \bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{Sci}(\bullet, \bullet, \bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{GAlt}(\bullet, \bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{DAlt}(\bullet, \bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{GComp}(\bullet, \bullet) \\ \min \{ \delta(\text{Tree}_1, \text{Tree}_2) + \delta(\text{Tree}_1, \text{Tree}_2) : \text{Tree}_1 = \text{Tree}_2 \} + \text{DComp}(\bullet, \bullet) \end{array} \right\}$$

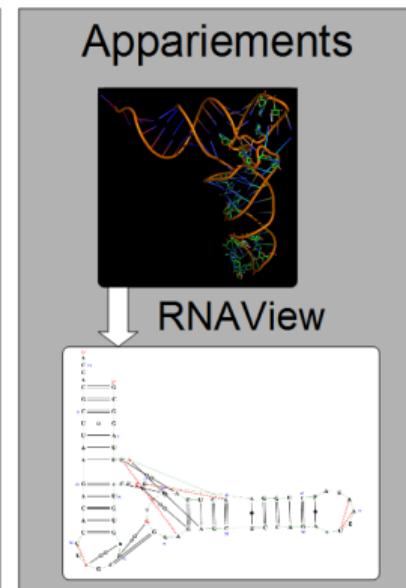
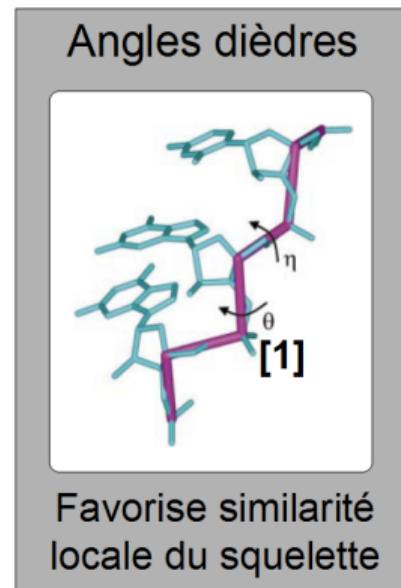
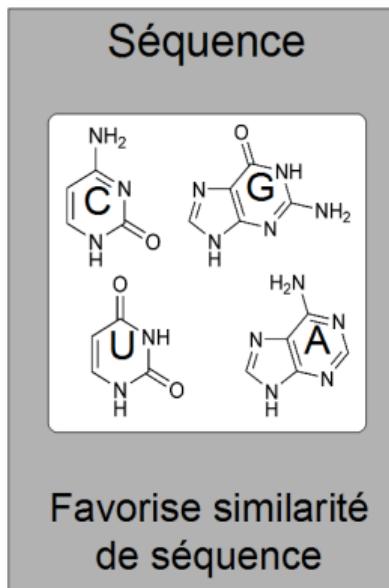
si  $\bullet$  base  
si  $\bullet$  base  
si  $\bullet$  et  $\bullet$  bases  
si  $\bullet$  paire  
si  $\bullet$  paire  
si  $\bullet$  et  $\bullet$  paires  
si  $\bullet$  paire et  $\bullet$  base  
si  $\bullet$  paire et  $\bullet$  base  
si  $\bullet$  paire et  $\bullet$  base  
si  $\bullet$  paire  
si  $\bullet$  paire et  $\bullet$  base  
si  $\bullet$  paire

## Integrative methods

DIAL [FPLC07] is an integrative method which focuses on local similarities.

Idea: RNA is flexible, meaningless local variations (even of small amplitudes) may induce large geometric discrepancies.

DIAL captures local similarities at three levels:

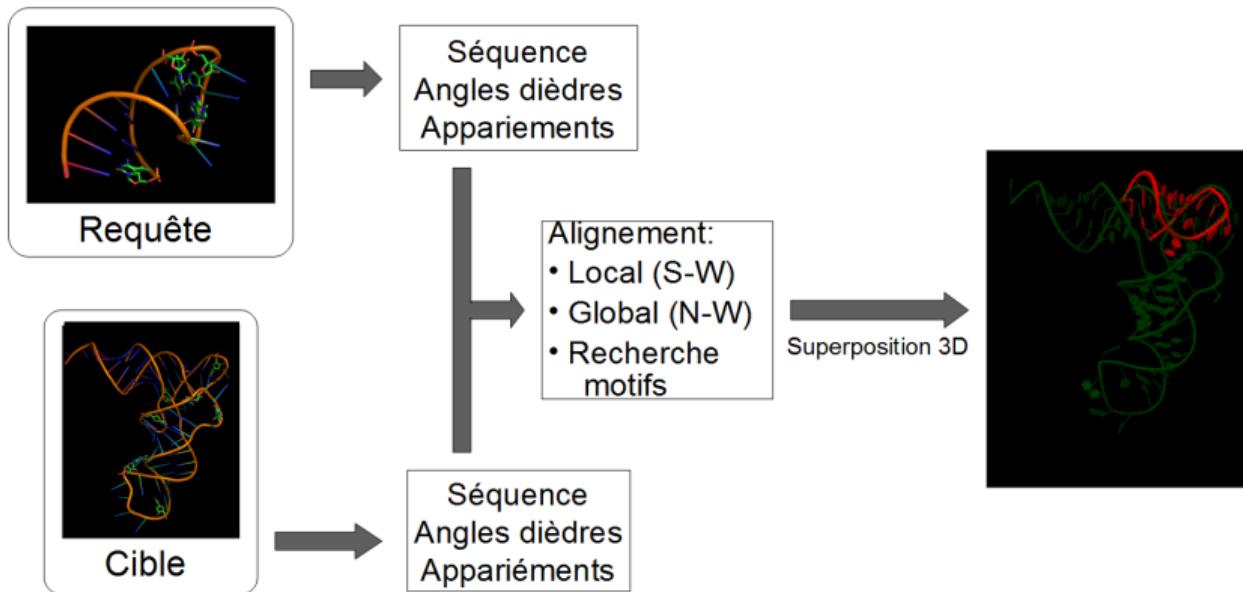


## Integrative methods

DIAL [FPLC07] is an integrative method which focuses on local similarities.

Idea: RNA is flexible, meaningless local variations (even of small amplitudes) may induce large geometric discrepancies.

A sequence alignment algorithm is then used



## References I

-  Tatsuya Akutsu.  
Dynamic programming algorithms for rna secondary structure prediction with pseudoknots.  
*Discrete Appl. Math.*, 104(1-3):45–62, 2000.
-  G. Blin, A. Denise, S. Dulucq, C. Herrbach, and H. Touzet.  
Alignment of rna structures.  
*Transactions on Computational Biology and Bioinformatics*, ...:..., 2008.  
A paraître.
-  Guillaume Blin, Guillaume Fertin, Irena Rusu, and Christine Sinoquet.  
Extending the Hardness of RNA Secondary Structure Comparison.  
In Bo Chen, Mike Paterson, and Guochuan Zhang, editors, *ESCAPE'07*, volume 4614 of *LNCS*, pages 140–151, Hangzhou, China, Apr 2007.
-  S. Cao and S-J Chen.  
Predicting structured and stabilities for h-type pseudoknots with interhelix loop.  
*RNA*, 15:696–706, 2009.
-  K. Doshi, J. J. Cannone, C. Cobaugh, and R. R. Gutell.  
Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for rna secondary structure prediction.  
*BMC Bioinformatics*, 5(1):105, 2004.

## References II

-  Robert M Dirks and Niles A Pierce.  
A partition function algorithm for nucleic acid secondary structure including pseudoknots.  
*J Comput Chem*, 24(13):1664–1677, Oct 2003.
-  F. Ferrè, Y. Ponty, W. A. Lorenz, and Peter Clote.  
Dial: A web server for the pairwise alignment of two RNA 3-dimensional structures using nucleotide, dihedral angle and base pairing similarities.  
*Nucleic Acids Research*, 35(Web server issue):W659–668, July 2007.
-  P. Gardner and R. Giegerich.  
A comprehensive comparison of comparative rna structure prediction approaches.  
*BMC Bioinformatics*, 5(1):140, 2004.
-  Claire Herrbach, Alain Denise, and Serge Dulucq.  
Average complexity of the jiang-wang-zhang pairwise tree alignment algorithm and of a rna secondary structure alignment algorithm.  
*In Proceedings of MACIS 2007, Second International Conference on Mathematical Aspects of Computer and Information Sciences*, 2007.
-  M. Hochsmann, B. Voss, and R. Giegerich.  
Pure multiple RNA secondary structure alignments: A progressive profile approach.  
*01(1):53–62*, 2004.

## References III

-  Tao Jiang, Lusheng Wang, and Kaizhong Zhang.  
Alignment of trees - an alternative to tree edit.  
In *CPM '94: Proceedings of the 5th Annual Symposium on Combinatorial Pattern Matching*, pages 75–86, London, UK, 1994. Springer-Verlag.
-  R. B. Lyngsø and C. N. S. Pedersen.  
RNA pseudoknot prediction in energy-based models.  
*Journal of Computational Biology*, 7(3-4):409–427, 2000.
-  D. McLachlan.  
Rapid comparison of protein structures.  
*Acta crystallographica A*, 38(6):871–873, 1982.
-  D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner.  
Expanded sequence dependence of thermodynamic parameters improves prediction of rna secondary structure.  
*Journal of Molecular Biology*, 288(5):911–940, May 1999.
-  M. Parisien and F. Major.  
The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.  
*Nature*, 452(7183):51–55, 2008.

## References IV

-  E. Rivas and S.R. Eddy.  
A dynamic programming algorithm for RNA structure prediction including pseudoknots.  
*J Mol Biol*, 285:2053–2068, 1999.
-  B. A. Shapiro, Y. G. Yingling, W. Kasprzak, and E. Bindewald.  
Bridging the gap in rna structure prediction.  
*Curr Opin Struct Biol*, 17(2):157–165, Apr 2007.
-  M. Sarver, C. Zirbel, J. Stombaugh, A. Mokdad, and N. B. Leontis.  
FR3D: Finding local and composite recurrent structural motifs in RNA 3D.  
*Journal of Mathematical Biology*, 56(1–2):215–252, January 2008.
-  S. Wuchty, W. Fontana, I.L. Hofacker, and P. Schuster.  
Complete suboptimal folding of RNA and the stability of secondary structures.  
*Biopolymers*, 49:145–164, 1999.