

Cours M2 BIBS

Repliement *in silico* de l'ARN

Yann Ponty

Bioinformatics Team
École Polytechnique/CNRS/INRIA AMIB

PIMS Vancouver
Simon Fraser University
Canada

<http://www.lix.polytechnique.fr/~ponty/index.php?page=bibscasm2014>

18 Novembre 2013

Avant propos ...

...ou comment gagner 1 million de dollars en rendant la monnaie !!

Problème : Vous disposez de pièces de **1**, **20** et **50** centimes. Le client souhaite minimiser la monnaie reçue (en nombre de pièces).
Comment rendre **N** en monnaie sans perdre un client ?

Stratégie 1 : Commencer par les *grosses* pièces puis compléter avec les *petites*.

$$21 = \text{50c} + \text{1c}$$

55??

Avant propos ...

...ou comment gagner 1 million de dollars en rendant la monnaie !!

Problème : Vous disposez de pièces de **1**, **20** et **50** centimes. Le client souhaite minimiser la monnaie reçue (en nombre de pièces).
Comment rendre **N** en monnaie sans perdre un client ?

Stratégie 1 : Commencer par les *grosses* pièces puis compléter avec les *petites*.

$$21 = ??$$

Avant propos ...

...ou comment gagner 1 million de dollars en rendant la monnaie !!

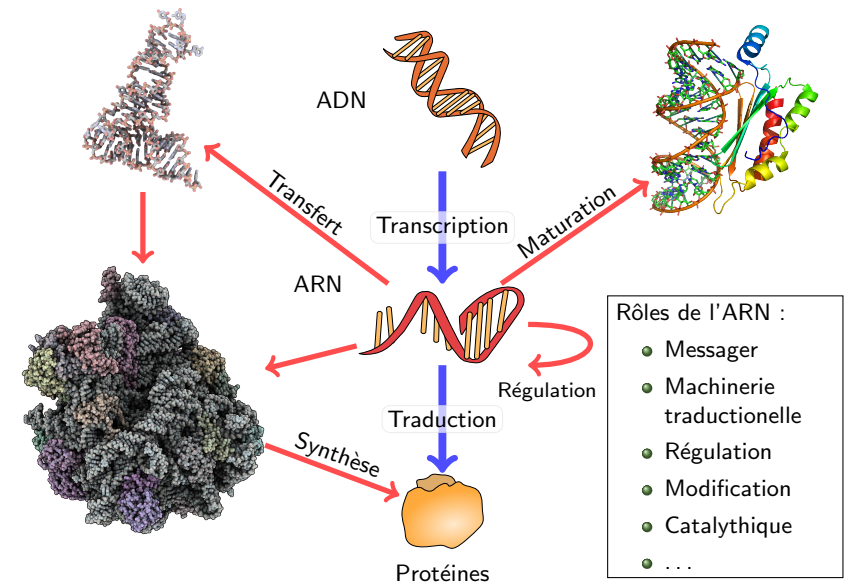
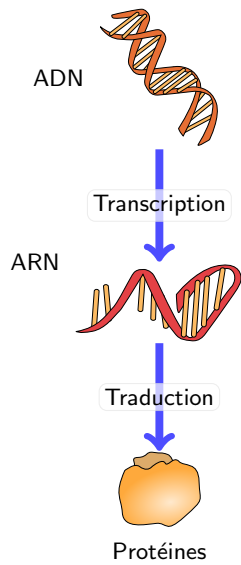
Problème : Vous disposez de pièces de **1**, **20** et **50** centimes. Le client souhaite minimiser la monnaie reçue (en nombre de pièces).
Comment rendre **N** en monnaie sans perdre un client ?

Stratégie 1 : Commencer par les *grosses* pièces puis compléter avec les *petites*.

$$21 = \text{50c} + \text{1c}$$

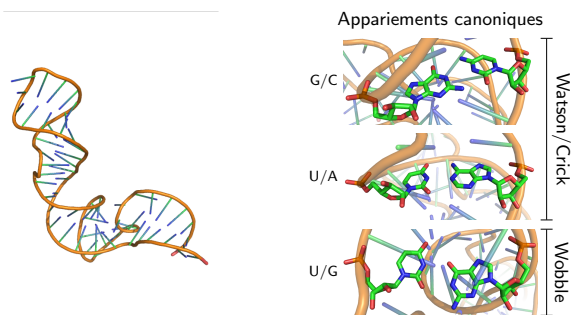
$$55 = \text{50c} + \text{1c} + \text{1c} + \text{1c} + \text{1c} + \text{1c}$$

60??



Repliement de l'ARN

ARN = Biopolymère composé de nucléotides A,C,G et U
 A : Adénosine, C : Cytosine, G : Guanine et U : Uracile



Repliement de l'ARN = Processus stochastique continu dirigé par (résultant en) un appariement des nucléotides.

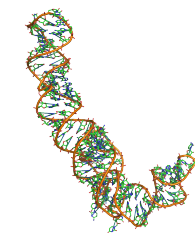
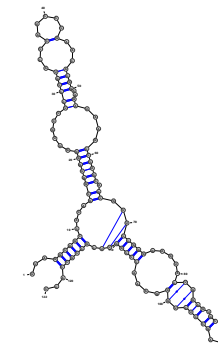
Comprendre le repliement des ARN aide à comprendre et prédire leur fonction.

Structure de l'ARN

Trois¹ niveaux de représentation :

```
UUAGGCGGCCACAGC
GGUGGGUUGCCUCC
CGUACCAUCCCGAA
CACGGAAGAUAGCC
CACCAGGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGAAA
CCGGUUCGCCCA
CC
```

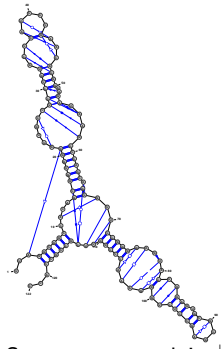
Structure primaire



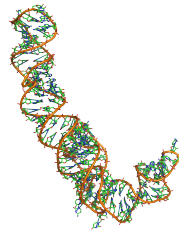
Source : 5s rRNA (PDB 1K73 :B)

Trois¹ niveaux de représentation :

```
UUAGGGGCCACAGC
GGUGGGUUGCCUCC
CGUACCCAUCCGAA
CACGGAAGAUAGCC
CACCAGGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCGGUUCGCCGCA
CC
```



Structure primaire



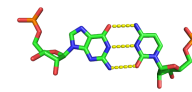
Structure secondaire⁺

Structure tertiaire

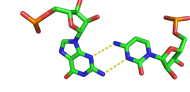
Source : 5s rRNA (PDB 1K73 :B)

1. Enfin, presque ...

- Appariements non-canoniques
 - Toute paire de base **autre que** {(A-U), (C-G), (G-U)}
 - Ou interagissant sur un bord non-standard (WC/WC-Cis) [LW01].

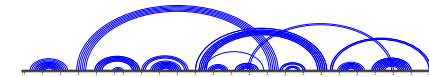


Paire CG canonique (WC/WC-Cis)



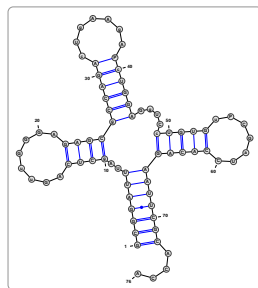
Paire CG non canonique (Sucre/WC-Trans)

- Pseudonoeuds



Structure pseudonoeud d'un Ribozyme du Groupe I (PDBID : 1Y0Q :A)

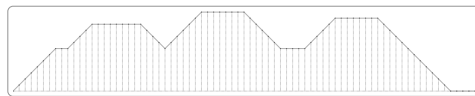
Plus expressif, mais repliement général *in silico* avec pseudonoeud :
 ⇒ NP-Complet [LP00] ... polynomial pour certaines classes [CDR⁺04].



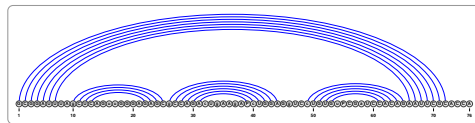
Grappe planaire (outer planar)

(((((((.....))))))((((.....)))).....((((.....)))))).....

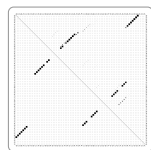
Expression bien parenthésée



Mountain view



Linéaire



Dot plot

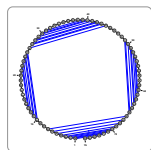
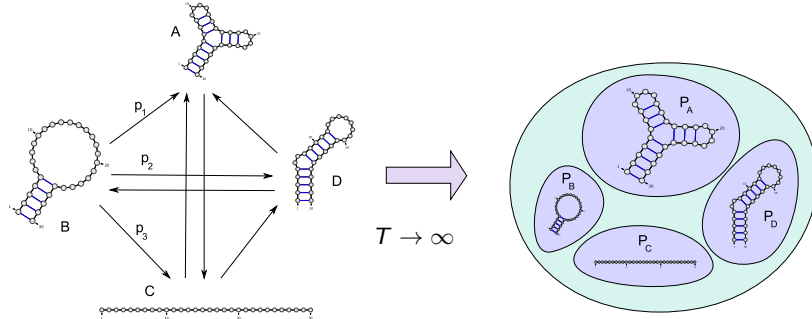


Diagramme de Feynman

Représentation différentes et équivalentes
 ⇒ Aide l'intuition algorithmique
 + Propriétés algébriques sympathiques
 ⇒ Algorithmique efficace !

- 1 Introduction
 - Fonction(s) de l'ARN
 - Repliement et structure
 - Représentations de la structure secondaire
- 2 Formalisation du repliement et outils disponibles
 - Aparté thermodynamique
 - Programmation dynamique : Rappels
- 3 Minimisation de l'énergie libre
 - Modèle de Nussinov
 - Modèle de Turner
 - MFold/Unafold
 - Performances et approches comparatives
 - Vers une prédiction ab-initio 3D
- 4 Ensemble de Boltzmann
 - Ensemble de Boltzmann
 - Nussinov : Minimisation ⇒ Comptage
 - Calcul de la fonction de partition
 - Échantillonnage statistique

A l'échelle nanoscopique, la structure de l'ARN *fluctue*.



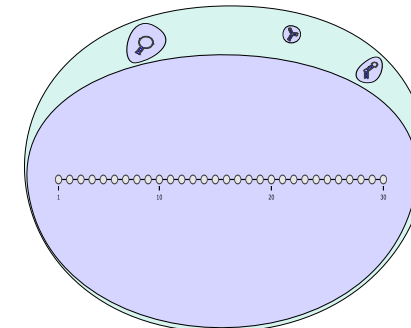
Convergence vers une **distribution stationnaire** de probabilité, l'équilibre de Boltzmann, où la probabilité est exponentiellement faible sur l'énergie libre.
Corollaire : La conformation initiale est sans d'importance.

Problèmes soulevés :

Étant donnés des modèles pour l'ensemble des conformations et l'énergie libre.

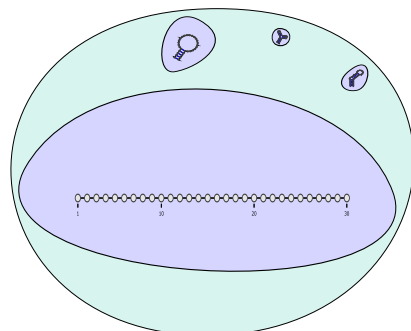
- A. Déterminer la structure la plus probable (= Energie libre minimale) à l'équilibre
- B. Déterminer des propriétés moyennes de l'ensemble de Boltzmann

Transcription : ARN synthétisé sans appariement (Sauf exception)



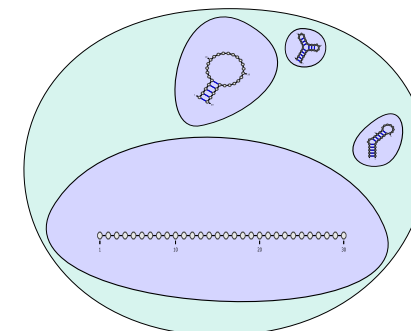
$T = 0$

Transcription : ARN synthétisé sans appariement (Sauf exception)



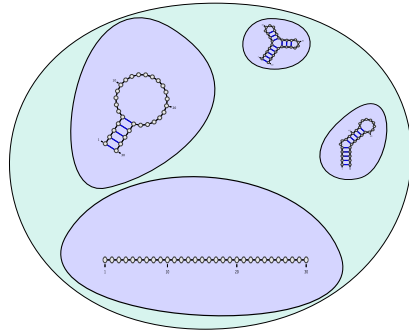
$T = 1h$

Transcription : ARN synthétisé sans appariement (Sauf exception)



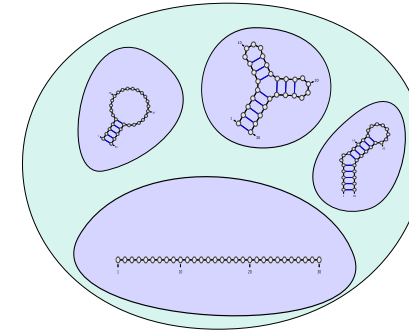
$T = 2h$

Transcription : ARN synthétisé sans appariement (Sauf exception)



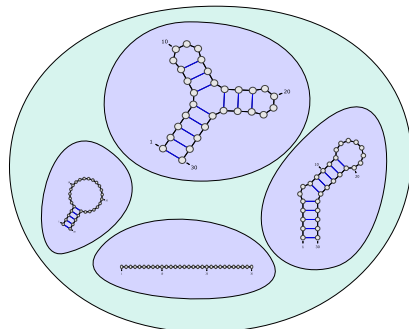
$T = 5h$

Transcription : ARN synthétisé sans appariement (Sauf exception)



$T = 10h$

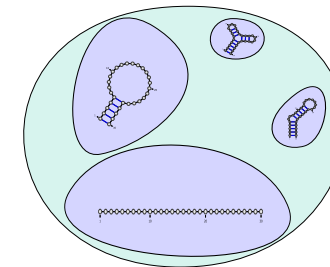
Transcription : ARN synthétisé sans appariement (Sauf exception)



$T \rightarrow \infty$

Mais majorité des ARNm dégradés avant 7h (Org. : Souris [SSN⁺09]).

Transcription : ARN synthétisé sans appariement (Sauf exception)



$T = 10h$

Mais majorité des ARNm dégradés avant 7h (Org. : Souris [SSN⁺09]).

- A. Déterminer la structure la plus probable (= Energie libre min.) à l'équilibre
- B. Déterminer des propriétés moyennes de l'ensemble de Boltzmann
- C. Déterminer la structure la plus probable à temps T .
(c.f. H. Isambert par simulation, NP-complet en déterministe [MTSC09])

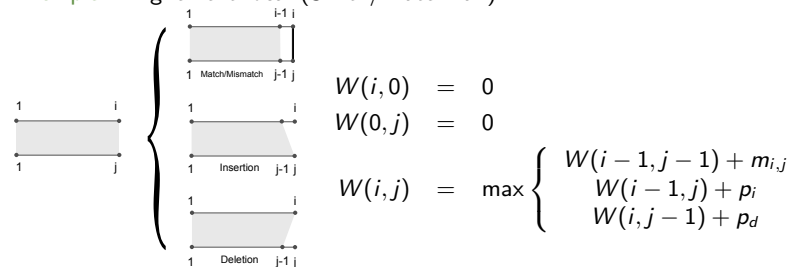
- 1 Introduction
 - Fonction(s) de l'ARN
 - Repliement et structure
 - Représentations de la structure secondaire
- 2 Formalisation du repliement et outils disponibles
 - Aparté thermodynamique
 - Programmation dynamique : Rappels
- 3 Minimisation de l'énergie libre
 - Modèle de Nussinov
 - Modèle de Turner
 - MFold/Unafold
 - Performances et approches comparatives
 - Vers une prédiction ab-initio 3D
- 4 Ensemble de Boltzmann
 - Ensemble de Boltzmann
 - Nussinov : Minimisation \Rightarrow Comptage
 - Calcul de la fonction de partition
 - Échantillonnage statistique

Programmation dynamique = Technique générale pour l'optimisation.
 Condition : Solution optimale pour P peut être reconstruite à partir de solutions pour des sous-problèmes strictes de P .

Bioinformatique :

- Espace de solutions *discret* (alignements, repliements)
- + Fonction objectif *additive* (score, énergie)
- \Rightarrow Schéma de programmation dynamique efficace.

Exemple : Alignement local (Smith/Waterman)



Détails algorithmiques

Un schéma fait intervenir des *classes* de sous-problèmes dont on sait calculer le score du *champion*.

Étant donné un schéma, deux étapes :

- Calcul matrices : Sauvegarde des meilleurs scores sur classes de sous-problèmes (Ordre inverse de celui induit par les dépendances).
- Remontée : Reconstitue le parcours ayant mené au meilleur score. (Parcours = Instance)

Complexité du calcul dépend alors :

- Taille de l'espace des sous-problèmes
- Nombre de sous-problèmes considérés (#Termes décomposition)

Exemple S/W :

$i : 1 \rightarrow n + 1 \Rightarrow \Theta(n)$

$j : 1 \rightarrow m + 1 \Rightarrow \Theta(m)$

Trois opération pour chaque sous-calcul

$\Rightarrow \Theta(m.n)$ temps/mémoire



Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

$W(i, 0) = 0$

$W(0, j) = 0$

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

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

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

| | A | C | A | C | A | C | T | A |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 2 | | | | | | |
| G | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

| | A | C | A | C | A | C | T | A |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 2 | 1 | | | | | |
| G | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

| | A | C | A | C | A | C | T | A |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 2 | 1 | 2 | | | | |
| G | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

| | A | C | A | C | A | C | T | A |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 2 | 1 | 2 | 1 | | | |
| G | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |
| C | 0 | | | | | | | |
| A | 0 | | | | | | | |

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

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

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

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

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

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

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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

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

Exemple complet

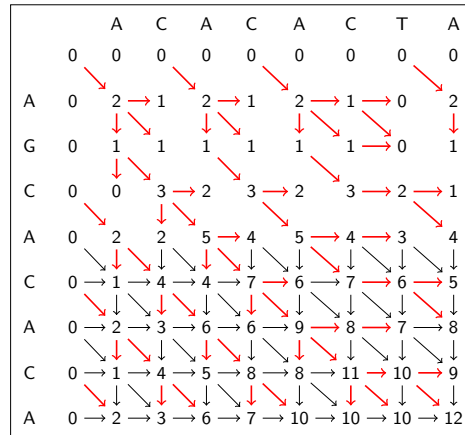
Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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



Exemple complet

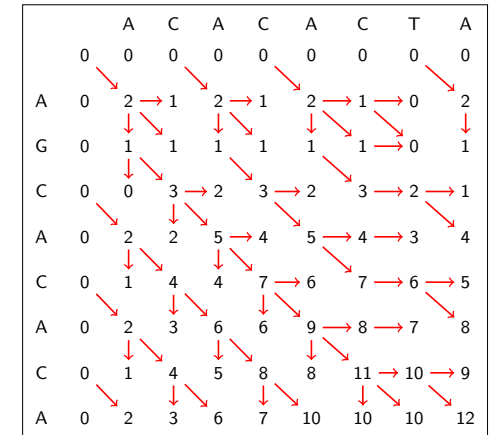
Exemple : Alignement local de séquences AGCACACA et ACACACTA

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

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

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

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



Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

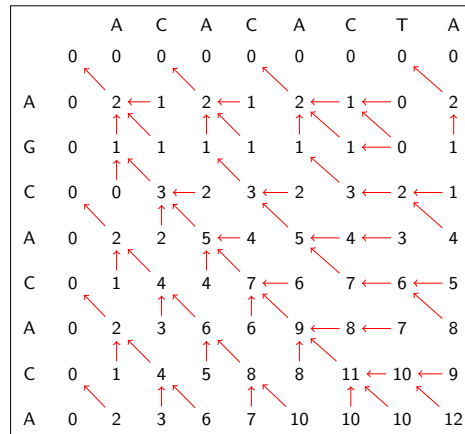
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

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

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

Meilleur alignement



Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA

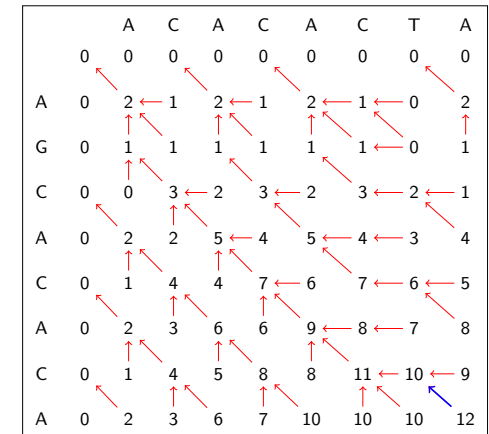
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

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

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

Meilleur alignement



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

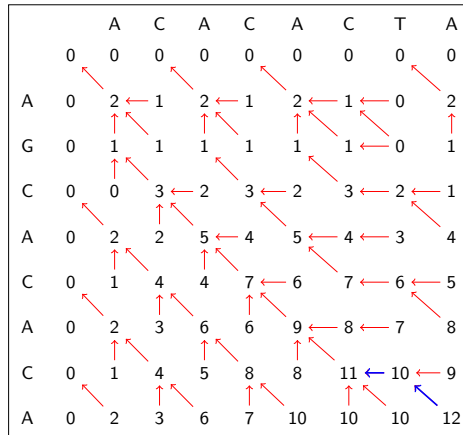
$$W(i,0) = 0$$

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

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

Meilleur aligment

- A
T A



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

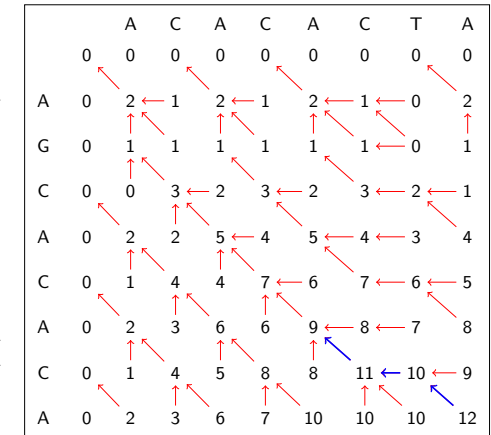
$$W(i,0) = 0$$

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

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

Meilleur aligment

C - A
C T A



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

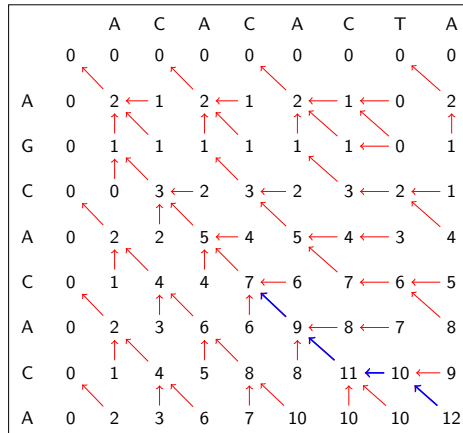
$$W(i,0) = 0$$

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

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

Meilleur aligment

A C - A
A C T A



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

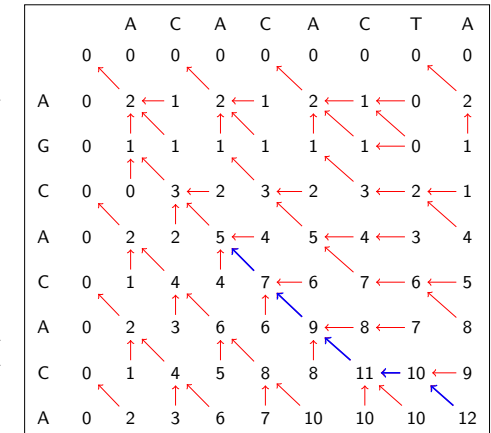
$$W(i,0) = 0$$

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

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

Meilleur aligment

C A C - A
C A C T A



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

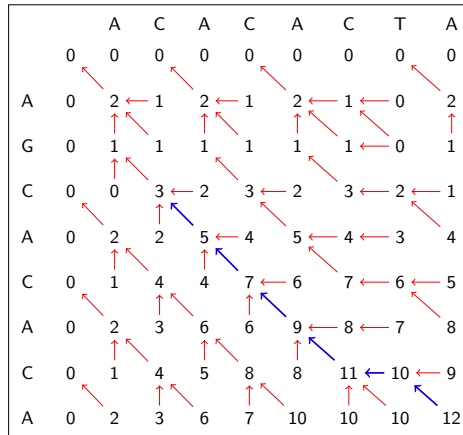
$$W(i,0) = 0$$

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

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

Meilleur aligment

A C A C - A
A C A C T A



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

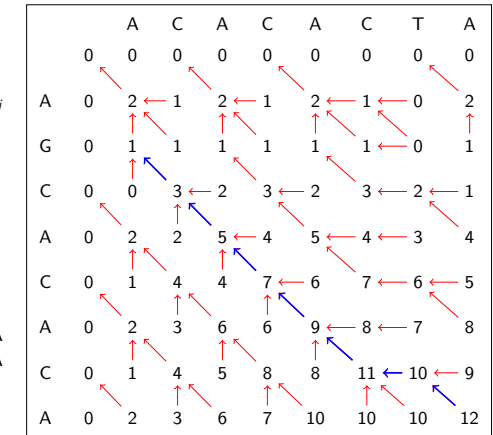
$$W(i,0) = 0$$

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

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

Meilleur aligment

C A C A C - A
C A C A C T A



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

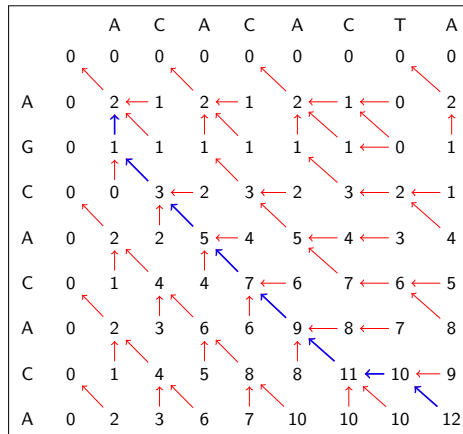
$$W(i,0) = 0$$

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

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

Meilleur aligment

G C A C A C - A
- C A C A C T A



Exemple complet

Exemple : Aligment local de séquences AGCACACA et ACACACTA

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

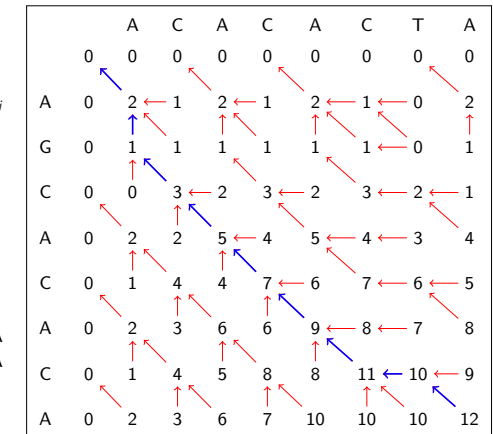
$$W(i,0) = 0$$

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

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

Meilleur aligment

A G C A C A C - A
A - C A C A C T A



Propriétés requise d'un schéma :

- **Validité** : \forall sous-problème, la valeur obtenue doit être celle de la fonction objectif.

Preuve souvent assez technique.

Propriétés souhaitables d'un schéma :

- **Complétude** : Espace des solutions engendré par la décomposition. Des astuces algorithmiques peuvent *couper des branches*...
- **Non-ambiguïté** : Chaque solution est *engendrée* au plus une fois.

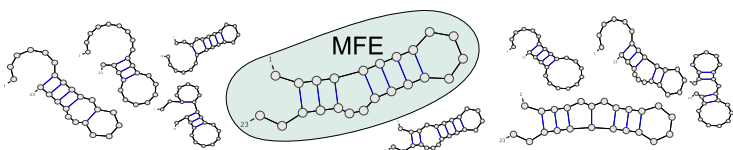
⇒ Possibilité d'énumérer l'espace des solutions.

Repliement par minimisation d'énergie

Problème A : Déterminer la structure d'énergie minimale.

Repliement *ab initio* =

Trouver structure d'un ARN ω uniquement à partir de sa séquence.



- **Conformations** : Ensemble S_ω des structures secondaires compatibles avec la structure primaire ω (contrainte d'appariements).
- **Fonction d'énergie** Énergie libre associant une valeur numérique $E_{\omega,S}$ (KCal.mol⁻¹) à tout couple séquence/conformation (ω, S) .
- **Structure native** : Conformation *fonctionnelle* de la molécule.

Remarques :

- Pas nécessairement unique (Cinétique ou structures bi-stables)
- Présence de pseudo-noeuds : Ambiguïté, quelle est la structure native ?

- 1 Introduction**
 - Fonction(s) de l'ARN
 - Repliement et structure
 - Représentations de la structure secondaire
- 2 Formalisation du repliement et outils disponibles**
 - Aparté thermodynamique
 - Programmation dynamique : Rappels
- 3 Minimisation de l'énergie libre**
 - Modèle de Nussinov
 - Modèle de Turner
 - MFold/Unafold
 - Performances et approches comparatives
 - Vers une prédiction *ab-initio* 3D
- 4 Ensemble de Boltzmann**
 - Ensemble de Boltzmann
 - Nussinov : Minimisation ⇒ Comptage
 - Calcul de la fonction de partition
 - Échantillonnage statistique

Modèle de Nussinov/Jacobson

Modèle de Nussinov/Jacobson (NJ)

Plus proche voisins simple :

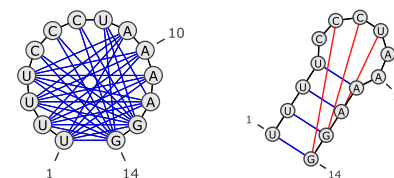
- Seuls les appariements contribuent à l'énergie
- Uniquement liaisons Watson/Crick (A/U,C/G) et Wobble (G/U)

$$\Rightarrow E_{\omega,S} = -\#Paires(S)$$

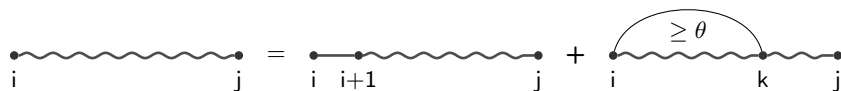
Repliement dans NJ \Leftrightarrow Maximisation du nombre de paires de bases.

Exemple :

UUUUCCCUAAAAGG



Variante : Pondérer les paires selon leur nombre de liaisons hydrogène
 $\Delta G(G \equiv C) = -3$ $\Delta G(A = U) = -2$ $\Delta G(G - U) = -1$

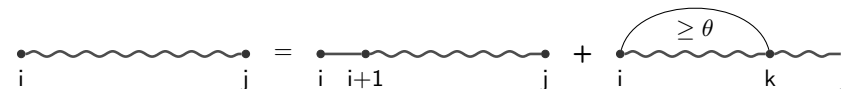


$$N_{i,t} = 0, \forall t \in [i, i + \theta]$$

$$N_{i,j} = \min \begin{cases} N_{i+1,j} & i \text{ non apparié} \\ \min_{k=i+\theta+1}^j E_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ apparié à } k \end{cases}$$

Backtrack Nussinov/Jacobson

| | 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 | 7 | 9 | 11 | 11 | 11 | |
| G | | | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 | 10 | |
| A | | | | 0 | 0 | 0 | 2 | 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 | 8 | 8 | |
| C | | | | | | | 0 | 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 | 7 | 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 | |
| G | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| A | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| C | | | | | | | | | | | | | | | | | 0 | 0 | |
| G | | | | | | | | | | | | | | | | | | 0 | |
| A | | | | | | | | | | | | | | | | | | | 0 |



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

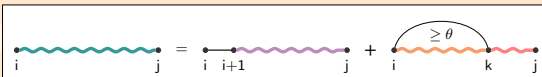
$$N_{i,j} = \min \begin{cases} N_{i+1,j} & i \text{ non apparié} \\ \min_{k=i+\theta+1}^j E_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ apparié à } k \end{cases}$$

Correction : On cherche à montrer que l'énergie de la structure d'énergie la plus faible ($MFE_{1,n}$) est bien calculée dans $N_{1,n}$. Dans toute structure secondaire restreinte à $[i, j]$ la première position i est :

- Soit non-appariée : $MFE_{i,j}$ est constituée des appariements de $MFE_{i+1,j}$.
- Soit appariée à k : $MFE_{i,j}$ contient l'appariement (i, k) et l'union des appariements de $MFE_{i+1,k-1}$ et de $MFE_{k+1,j}$. En effet, tout appariement entre les régions $[i + 1, k - 1]$ et $[k + 1, j]$ croiserait (i, k) (Pseudonoed).

Backtrack Nussinov/Jacobson

| | 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 | 7 | 9 | 11 | 11 | 11 | |
| G | | | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 | 5 | 5 | 6 | 8 | 10 | 10 | 10 | 10 | |
| A | | | | 0 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 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 | 8 | 8 | |
| C | | | | | | | 0 | 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 | 7 | 7 | |
| C | | | | | | | | | | 0 | 0 | 0 | 0 | 2 | 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 | |
| G | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| A | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| C | | | | | | | | | | | | | | | | | 0 | 0 | |
| G | | | | | | | | | | | | | | | | | | 0 | |
| A | | | | | | | | | | | | | | | | | | | 0 |



Backtrack Nussinov/Jacobson

| | 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 | 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 | 10 |
| U | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 | 10 |
| A | | | | | | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| C | | | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| U | | | | | | | | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 | 7 |
| U | | | | | | | | | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 | 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 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Diagram illustrating the recurrence relation for the Nussinov/Jacobson algorithm. It shows a sequence of nucleotides from index i to j . The value at (i, j) is equal to the maximum of two options: either pairing i with $i+1$ and then solving for the subsequence $i+1$ to j , or pairing i with k and then solving for the subsequence $i+1$ to k plus the score for pairing (i, k) and (j) . The score for pairing (i, k) and (j) is denoted as $\geq \theta$.

Backtrack Nussinov/Jacobson

| | 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 | 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 | 10 |
| U | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 | 10 |
| A | | | | | | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| C | | | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| U | | | | | | | | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 | 7 |
| U | | | | | | | | | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 | 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 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Diagram illustrating the recurrence relation for the Nussinov/Jacobson algorithm. It shows a sequence of nucleotides from index i to j . The value at (i, j) is equal to the maximum of two options: either pairing i with $i+1$ and then solving for the subsequence $i+1$ to j , or pairing i with k and then solving for the subsequence $i+1$ to k plus the score for pairing (i, k) and (j) . The score for pairing (i, k) and (j) is denoted as $\geq \theta$.

Backtrack Nussinov/Jacobson

| | 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 | 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 | 10 |
| U | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 | 10 |
| A | | | | | | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| C | | | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| U | | | | | | | | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 | 7 |
| U | | | | | | | | | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 | 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 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Diagram illustrating the recurrence relation for the Nussinov/Jacobson algorithm. It shows a sequence of nucleotides from index i to j . The value at (i, j) is equal to the maximum of two options: either pairing i with $i+1$ and then solving for the subsequence $i+1$ to j , or pairing i with k and then solving for the subsequence $i+1$ to k plus the score for pairing (i, k) and (j) . The score for pairing (i, k) and (j) is denoted as $\geq \theta$.

Backtrack Nussinov/Jacobson

| | 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 | 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 | 10 |
| U | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 5 | 7 | 7 | 8 | 10 | 10 |
| A | | | | | | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| C | | | | | | | 0 | 0 | 0 | 0 | 0 | 2 | 5 | 5 | 5 | 8 | 8 | 8 |
| U | | | | | | | | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 6 | 7 | 7 |
| U | | | | | | | | | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 5 | 7 | 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 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Diagram illustrating the recurrence relation for the Nussinov/Jacobson algorithm. It shows a sequence of nucleotides from index i to j . The value at (i, j) is equal to the maximum of two options: either pairing i with $i+1$ and then solving for the subsequence $i+1$ to j , or pairing i with k and then solving for the subsequence $i+1$ to k plus the score for pairing (i, k) and (j) . The score for pairing (i, k) and (j) is denoted as $\geq \theta$.

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Diagram illustrating sequence alignment: $i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 6 | 7 |
| C | | | | | | | | | | 0 | 0 | 0 | 0 | 2 | 3 | 5 | 5 | 7 |
| C | | | | | | | | | | | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 |
| U | | | | | | | | | | | | 0 | 0 | 0 | 0 | 2 | 2 | 3 |
| U | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 1 |
| A | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | | 0 | 0 |
| G | | | | | | | | | | | | | | | | | | 0 |
| A | | | | | | | | | | | | | | | | | | |

Diagram illustrating sequence alignment: $i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Diagram illustrating sequence alignment: $i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 2 | 3 | 5 | 5 | 7 |
| C | | | | | | | | | | | 0 | 0 | 0 | 0 | 3 | 3 | 3 | 5 |
| U | | | | | | | | | | | | 0 | 0 | 0 | 0 | 2 | 2 | 3 |
| U | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 1 |
| A | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | | 0 | 0 |
| G | | | | | | | | | | | | | | | | | | 0 |
| A | | | | | | | | | | | | | | | | | | |

Diagram illustrating sequence alignment: $i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$

Backtrack Nussinov/Jacobson

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

Backtrack Nussinov/Jacobson

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

Backtrack Nussinov/Jacobson

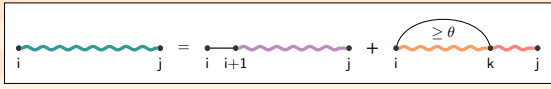
| | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|
| | (| . | . | . | . | . | . | . | . | . | . | . | . | . | . |) | . | |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 |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

Backtrack Nussinov/Jacobson

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

Backtrack Nussinov/Jacobson

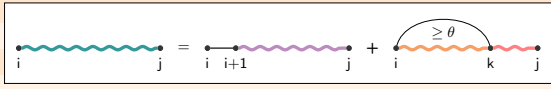
| | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|
| | (| (| . | . | . | . | . | . | . | . | . | . | . | . | . |) |) | . |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 |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



$$i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$$

Backtrack Nussinov/Jacobson

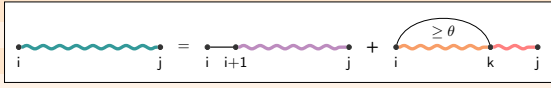
| | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|
| | (| (| . | . | . | . | . | . | . | . | . | . | . | . | . |) |) | . |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 |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



$$i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$$

Backtrack Nussinov/Jacobson

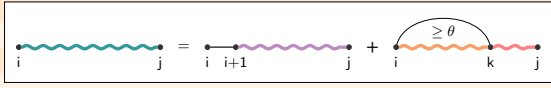
| | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|
| | (| (| . | . | . | . | . | . | . | . | . | . | . | . | . |) |) | . |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 |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



$$i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$$

Backtrack Nussinov/Jacobson

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



$$i \dots j = i \dots i+1 \dots j + \overset{\geq \theta}{i \dots k \dots j}$$

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Backtrack Nussinov/Jacobson

| | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

Backtrack Nussinov/Jacobson

| | 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 | 2 | 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 | 8 | 8 | |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 | |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | | | | 0 | 0 | |
| A | | | | | | | | | | | | | | | | | | 0 | |

Backtrack Nussinov/Jacobson

| | 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 | 2 | 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 | 8 | 8 | |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 | |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | | | | 0 | 0 | |
| A | | | | | | | | | | | | | | | | | | 0 | |

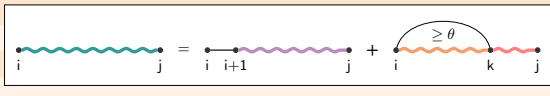
Backtrack Nussinov/Jacobson

| | 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 | 2 | 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 | 8 | 8 | |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 | |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | | | | 0 | 0 | |
| A | | | | | | | | | | | | | | | | | | 0 | |

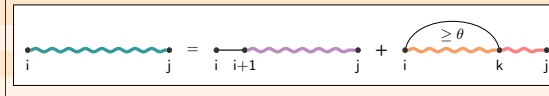
Backtrack Nussinov/Jacobson

| | 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 | 2 | 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 | 8 | 8 | |
| C | | | | | | | 0 | 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 | 0 | 1 | 2 | |
| A | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | 0 | |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 | |
| G | | | | | | | | | | | | | | | | | 0 | 0 | |
| A | | | | | | | | | | | | | | | | | | 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 | 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 2 | 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 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |

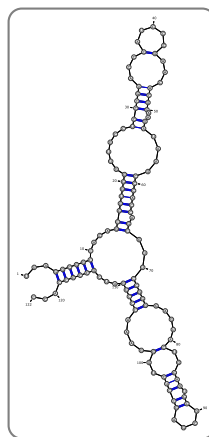


| | C | G | G | A | U | A | C | U | U | C | U | U | A | G | A | C | G | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | (| (| (| . | . | . |) | . | (| (| . | . | . |) |) |) |) | . |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 4 | 6 | 6 | 6 | 6 | 9 | 9 | 11 | 14 | 14 |
| 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 | 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 | 8 | 8 |
| C | | | | | | | 0 | 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 | 2 | 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 |
| A | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 |
| C | | | | | | | | | | | | | | | | 0 | 0 | 0 |
| G | | | | | | | | | | | | | | | | | 0 | 0 |
| A | | | | | | | | | | | | | | | | | | 0 |



Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2^{aie} :



Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

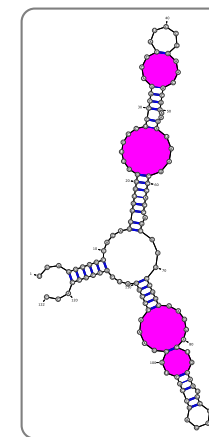
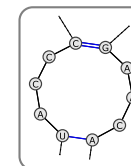
Déterminées expérimentalement
+ Interpolation pour grandes boucles.

Meilleure résultats grâce à la prise en compte de l'empilement.

Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2^{aie} :

- Boucles internes



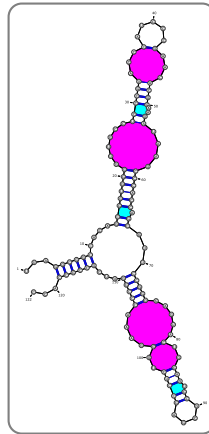
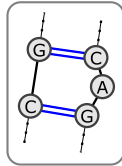
Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

Meilleure résultats grâce à la prise en compte de l'empilement.

Basée sur décomposition **non-ambiguë** en **boucles** de la structure 2^{aire} :

- Boucles internes
- Renflements



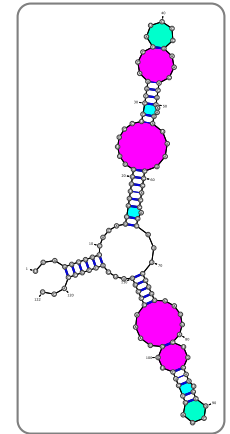
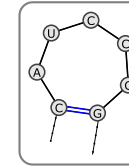
Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

Meilleure résultats grâce à la prise en compte de l'empilement.

Basée sur décomposition **non-ambiguë** en **boucles** de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales



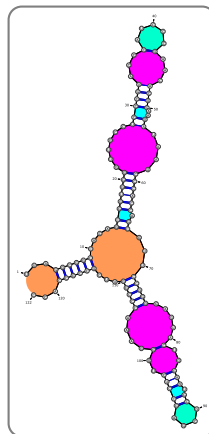
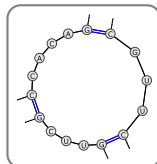
Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

Meilleure résultats grâce à la prise en compte de l'empilement.

Basée sur décomposition **non-ambiguë** en **boucles** de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples



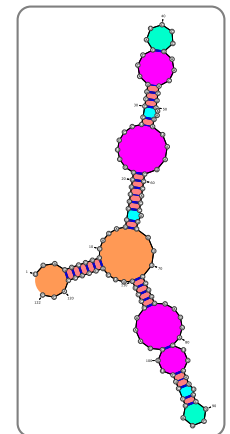
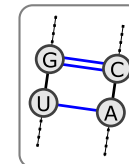
Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

Meilleure résultats grâce à la prise en compte de l'empilement.

Basée sur décomposition **non-ambiguë** en **boucles** de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples
- Empilements

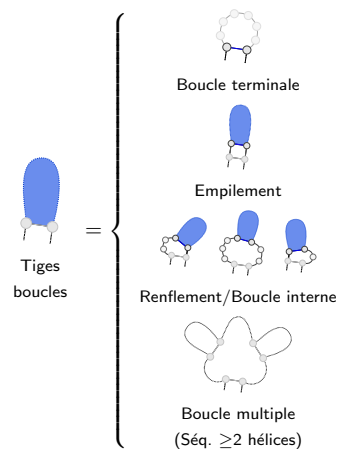


Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

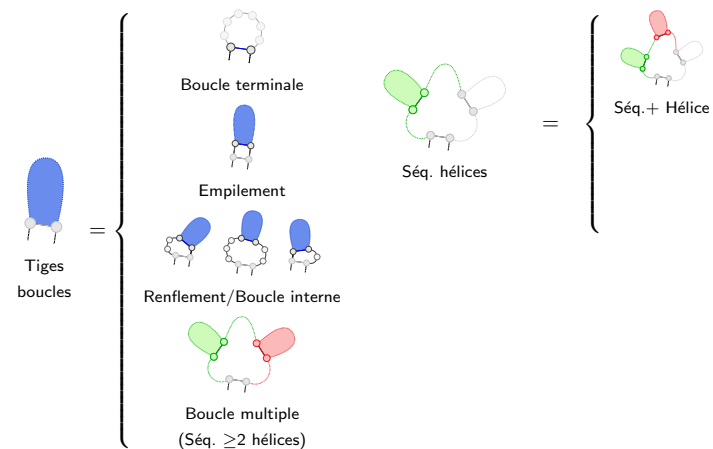
Déterminées expérimentalement
+ Interpolation pour grandes boucles.

Meilleure résultats grâce à la prise en compte de l'empilement.

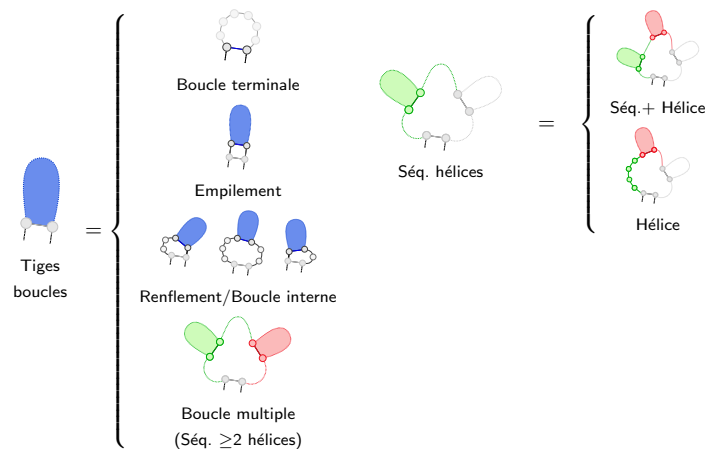
MFE DP equations



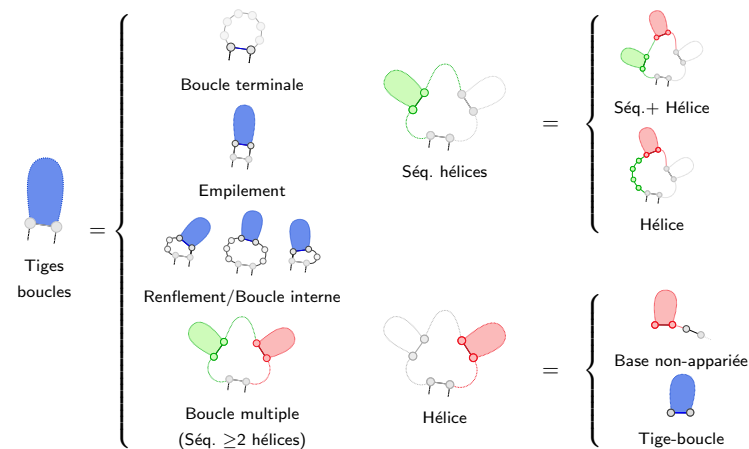
MFE DP equations



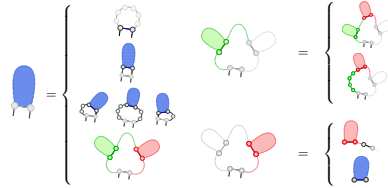
MFE DP equations



MFE DP equations



- $E_H(i, j)$: Energie de boucle terminale *fermée* par une paire (i, j)
- $E_{BI}(i, j)$: Energie de renflement ou boucle interne *fermée* par une paire (i, j)
- $E_S(i, j)$: Energie d'empilement $(i, j)/(i + 1, j - 1)$
- a, c, b : Pénalité de boucle multiple, hélice et non-appariées dans multiboucle.



Calcul des matrices

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

Reconstruction de la structure d'énergie minimale :

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

2. Avec une astuce pour les bulges/boucles internes ...

Reconstruction de la structure d'énergie minimale :

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

2. Avec une astuce pour les bulges/boucles internes ...

Reconstruction de la structure d'énergie minimale :

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

$\mathcal{O}(n)$ contributeurs potentiels au Min :
 \Rightarrow Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.

2. Avec une astuce pour les bulges/boucles internes ...

Reconstruction de la structure d'énergie minimale :

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

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

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

$\mathcal{O}(n)$ contributeurs potentiels au Min :
 ⇒ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naïf.
 Garder les meilleures contributions aux Min ⇒ Backtrack en $\mathcal{O}(n)$
 Complexités temps/mémoire en $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ pour le précalcul²

2. Avec une astuce pour les bulges/boucles internes ...

Definition (Replieement ab initio)

Partant de la séquence, trouver la conformation minimisant une fonction d'énergie.

Avantages :

- Explication mécanique
- Complexité raisonnable $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ temps/mémoire
- Exploration exhaustive

Limites :

- Pas de cinétique
- Pas d'info évolutive
- Performances limitées

Definition (Approche comparative)

Partant de plusieurs séquences homologues ou d'un alignement, trouver une conformation de score (énergie+alignement) élevé.

Avantages :

- Meilleures performances
- Affinement permanent

Limites :

- Complexité élevée
- Exploration non-exhaustive

Reconstruction de la structure d'énergie minimale :

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

$$M_{i,j} \leftarrow \text{Min}_k \left\{ \min(M_{i,k-1}, b(k-1)) + M^1_{k,j} \right\}$$

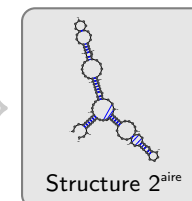
$$M^1_{i,j} \leftarrow \text{Min}_k \left\{ b + M^1_{i,j-1}, c + M'_{i,j} \right\}$$

$\mathcal{O}(n)$ contributeurs potentiels au Min :
 ⇒ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naïf.
 Garder les meilleures contributions aux Min ⇒ Backtrack en $\mathcal{O}(n)$
 Complexités temps/mémoire en $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ pour le précalcul²
 ⇒ UnaFold [MZ08] calcule la structure secondaire d'énergie minimale.

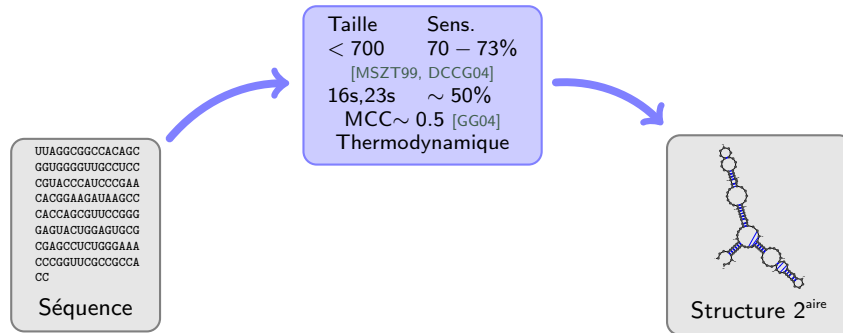
2. Avec une astuce pour les bulges/boucles internes ...

UUAGGGGGCCACAGC
 GGUGGGGUUGCCUCC
 CGUACCCAUCCGAAA
 CACGGAAGAUAGCC
 CACCAGCGUUCGGG
 GAGUACUGGAGUGCG
 CGAGCCUCUGGAAA
 CCGGUUCCCGGCA
 CC

Séquence



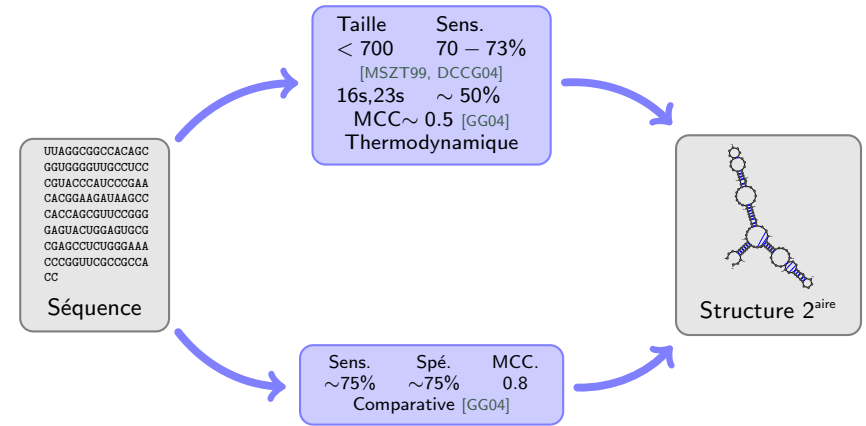
Rappel : $MCC = \frac{t^+t^- - f^+f^-}{\sqrt{(t^++f^+)(t^++f^-)(t^-+f^+)(t^-+f^-)}}$



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

Futur (proche) : Vers une prédiction 3D

But : De la séquence à des modèles tri-dimensionnels !!!

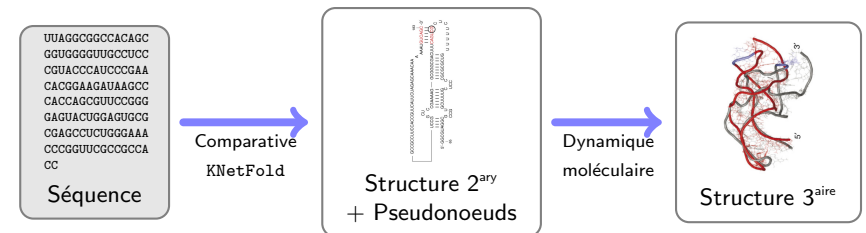


Rappel : $MCC = \frac{t^+t^- - f^+f^-}{\sqrt{(t^++f^+)(t^++f^-)(t^-+f^+)(t^-+f^-)}}$

Futur (proche) : Vers une prédiction 3D

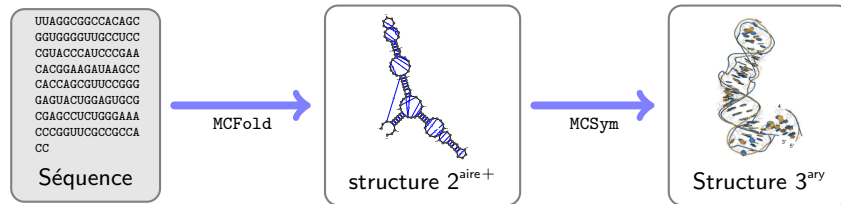
But : De la séquence à des modèles tri-dimensionnels !!!

- Modèles comparatifs + Dynamique moléculaires : RNA2D3D [SYKB07]



But : De la séquence à des modèles tri-dimensionnels !!!

- Pipeline MC-Fold/MC-sym [PM08]



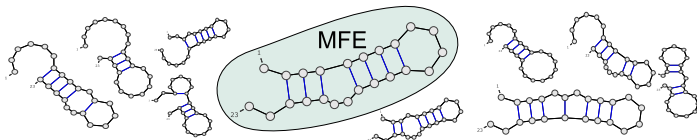
- 1 Introduction
 - Fonction(s) de l'ARN
 - Repliement et structure
 - Représentations de la structure secondaire
- 2 Formalisation du repliement et outils disponibles
 - Aparté thermodynamique
 - Programmation dynamique : Rappels
- 3 Minimisation de l'énergie libre
 - Modèle de Nussinov
 - Modèle de Turner
 - MFold/Unafold
 - Performances et approches comparatives
 - Vers une prédiction ab-initio 3D
- 4 Ensemble de Boltzmann
 - Ensemble de Boltzmann
 - Nussinov : Minimisation \Rightarrow Comptage
 - Calcul de la fonction de partition
 - Échantillonnage statistique

Ensemble canonique de Boltzmann

L'ARN *respire* \Rightarrow Il n'existe pas UNE unique conformation native.

Nouveau paradigme

Les conformations d'un ARN coexistent dans une distribution de Boltzmann.



Conséquence : La probabilité de la MFE peut être négligeable.
 \Rightarrow Comprendre les modes d'actions de l'ARN exige de prendre en considération l'ensemble des structures.

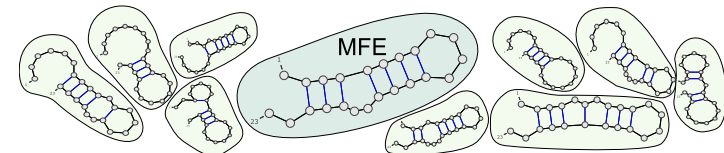
En particulier, des structures proches peuvent se *grouper* et devenir l'hypothèse la plus réaliste dans la recherche d'une conformation fonctionnelle.

Ensemble canonique de Boltzmann

L'ARN *respire* \Rightarrow Il n'existe pas UNE unique conformation native.

Nouveau paradigme

Les conformations d'un ARN coexistent dans une distribution de Boltzmann.



Conséquence : La probabilité de la MFE peut être négligeable.
 \Rightarrow Comprendre les modes d'actions de l'ARN exige de prendre en considération l'ensemble des structures.

En particulier, des structures proches peuvent se *grouper* et devenir l'hypothèse la plus réaliste dans la recherche d'une conformation fonctionnelle.

Une distribution de Boltzmann pondère chaque structure S pour un ARN ω par un **facteur de Boltzmann** $\mathcal{B}_{S,\omega} = e^{-\frac{E_{S,\omega}}{RT}}$ où :

- $E_{S,\omega}$ est l'énergie libre de S (kCal.mol^{-1})
- T est la température (K)
- R est la constante des gaz parfaits ($1.986 \cdot 10^{-3} \text{ kCal.K}^{-1}.\text{mol}^{-1}$)

Distribution renormalisée sur S_ω par la **fonction de partition**

$$\mathcal{Z}_\omega = \sum_{S \in S_\omega} e^{-\frac{E_{S,\omega}}{RT}}$$

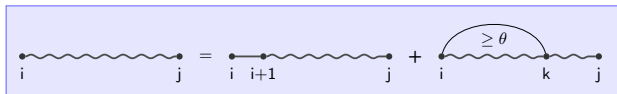
où S_ω est l'ensemble des conformations compatibles avec ω .

La **probabilité de Boltzmann** d'une structure S est alors donnée par

$$P_{S,\omega} = \frac{e^{-\frac{E_{S,\omega}}{RT}}}{\mathcal{Z}_\omega}$$

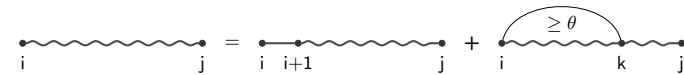
Fonction de partition

Fonction de partition = **Comptage pondéré** des structures compatibles



$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_{k=i+\theta+1}^j 1 \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \end{array} \right.$$



Récurrance sur l'**énergie minimale** d'un repliement :

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

$$N_{i,j} = \min \left\{ \begin{array}{l} N_{i+1,j} \quad (i \text{ non apparié}) \\ \min_{k=i+\theta+1}^j E_{i,k} + N_{i+1,k-1} + N_{k+1,j} \quad (i \text{ comp. avec } k) \end{array} \right.$$

Récurrance de **comptage des structures compatibles** :

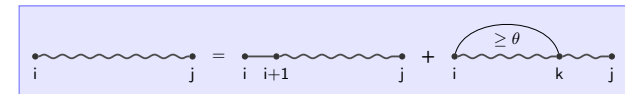
$$C_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$C_{i,j} = \sum \left\{ \begin{array}{l} C_{i+1,j} \quad (i \text{ non apparié}) \\ \sum_{k=i+\theta+1}^j 1 \times C_{i+1,k-1} \times C_{k+1,j} \quad (i \text{ comp. avec } k) \end{array} \right.$$

La décomposition est importante, le reste (MFE, comptage...) suit !

Fonction de partition

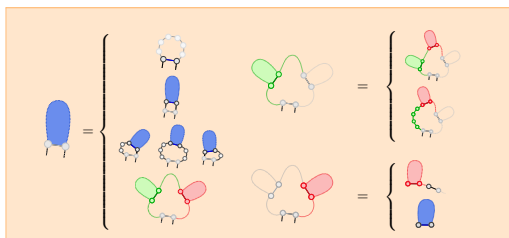
Fonction de partition = **Comptage pondéré** des structures compatibles



$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

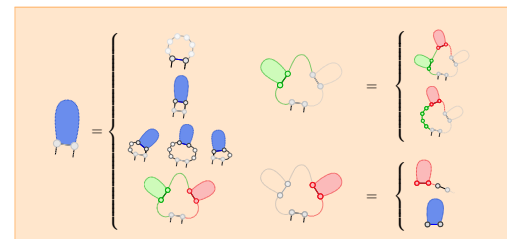
$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_{k=i+\theta+1}^j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \end{array} \right.$$

Fonction de partition = Comptage pondéré des structures compatibles



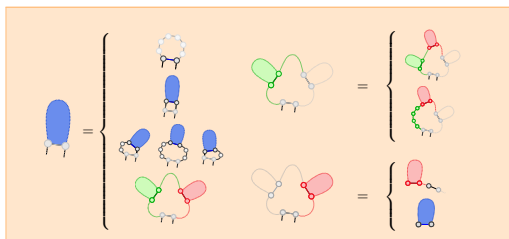
$$\begin{aligned}
 \mathcal{M}'_{i,j} &= \text{Min} \left\{ \begin{array}{l} E_H(i,j) \\ E_S(i,j) + \mathcal{M}'_{i+1,j-1} \\ \text{Min}(E_B(i,i',j',j) + \mathcal{M}'_{i',j'}) \\ a + c + \text{Min}(\mathcal{M}_{i+1,k-1} + \mathcal{M}^1_{k,j-1}) \end{array} \right\} \\
 \mathcal{M}_{i,j} &= \text{Min} \left\{ \text{Min}(\mathcal{M}_{i,k-1}, b(k-1)) + \mathcal{M}^1_{k,j} \right\} \\
 \mathcal{M}^1_{i,j} &= \text{Min} \left\{ b + \mathcal{M}^1_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
 \end{aligned}$$

Fonction de partition = Comptage pondéré des structures compatibles



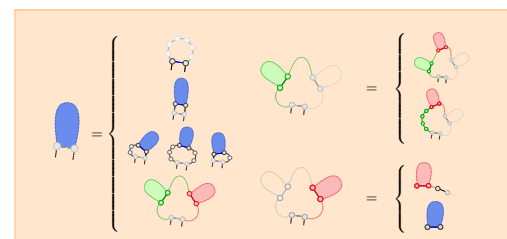
$$\begin{aligned}
 \mathcal{M}'_{i,j} &= \text{Min} \left\{ \begin{array}{l} e^{-\frac{E_H(i,j)}{RT}} \\ e^{-\frac{E_S(i,j)}{RT}} + \mathcal{M}'_{i+1,j-1} \\ \text{Min} \left(e^{-\frac{E_B(i,i',j',j)}{RT}} + \mathcal{M}'_{i',j'} \right) \\ e^{-\frac{(a+c)}{RT}} + \text{Min}(\mathcal{M}_{i+1,k-1} + \mathcal{M}^1_{k,j-1}) \end{array} \right\} \\
 \mathcal{M}_{i,j} &= \text{Min} \left\{ \text{Min}(\mathcal{M}_{i,k-1}, e^{-\frac{b(k-1)}{RT}}) + \mathcal{M}^1_{k,j} \right\} \\
 \mathcal{M}^1_{i,j} &= \text{Min} \left\{ e^{-\frac{b}{RT}} + \mathcal{M}^1_{i,j-1}, e^{-\frac{c}{RT}} + \mathcal{M}'_{i,j} \right\}
 \end{aligned}$$

Fonction de partition = Comptage pondéré des structures compatibles



$$\begin{aligned}
 \mathcal{M}'_{i,j} &= \text{Min} \left\{ \begin{array}{l} e^{-\frac{E_H(i,j)}{RT}} \\ e^{-\frac{E_S(i,j)}{RT}} \mathcal{M}'_{i+1,j-1} \\ \text{Min} \left(e^{-\frac{E_B(i,i',j',j)}{RT}} \mathcal{M}'_{i',j'} \right) \\ e^{-\frac{(a+c)}{RT}} \text{Min}(\mathcal{M}_{i+1,k-1}, \mathcal{M}^1_{k,j-1}) \end{array} \right\} \\
 \mathcal{M}_{i,j} &= \text{Min} \left\{ \text{Min} \left(\mathcal{M}_{i,k-1}, e^{-\frac{b(k-1)}{RT}} \right) \mathcal{M}^1_{k,j} \right\} \\
 \mathcal{M}^1_{i,j} &= \text{Min} \left\{ e^{-\frac{b}{RT}} \mathcal{M}^1_{i,j-1}, e^{-\frac{c}{RT}} \mathcal{M}'_{i,j} \right\}
 \end{aligned}$$

Fonction de partition = Comptage pondéré des structures compatibles



$$\begin{aligned}
 \mathcal{Z}'(i,j) &= \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i,j)}{RT}} \\ e^{-\frac{E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) \\ + \sum \left(e^{-\frac{E_B(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) \\ + e^{-\frac{(a+c)}{RT}} \sum (\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1)) \end{array} \right\} \\
 \mathcal{Z}(i,j) &= \sum \left(\mathcal{Z}(i,k-1) + e^{-\frac{b(k-1)}{RT}} \right) \mathcal{Z}^1(k,j) \\
 \mathcal{Z}^1(i,j) &= e^{-\frac{b}{RT}} \mathcal{Z}^1(i,j-1) + e^{-\frac{c}{RT}} \mathcal{Z}'(i,j)
 \end{aligned}$$

Fonction de partition = Comptage pondéré des structures compatibles

$$Z_{i,t} = 1, \forall t \in [i, i + \theta]$$

$$Z_{i,j} = \sum \left\{ \begin{array}{l} Z_{i+1,j} \\ \sum_{k=i+\theta+1}^j e^{-\frac{E_{bp}(i,k)}{RT}} \times Z_{i+1,k-1} \times Z_{k+1,j} \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma

Fonction de partition = Comptage pondéré des structures compatibles

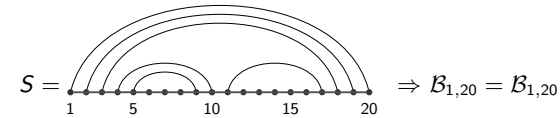
$$Z_{i,t} = 1, \forall t \in [i, i + \theta]$$

$$Z_{i,j} = \sum \left\{ \begin{array}{l} Z_{i+1,j} \\ \sum_{k=i+\theta+1}^j e^{-\frac{E_{bp}(i,k)}{RT}} \times Z_{i+1,k-1} \times Z_{k+1,j} \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
 - Correction du facteur de Boltzmann
- Facteur d'un backtrack = Produit des facteurs de ses parties
 Contributions énergétiques passent à l'exposant
 $(e^{-a/RT} \cdot Z^1 \cdot Z' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
 $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT})$

Exemple :



Fonction de partition = Comptage pondéré des structures compatibles

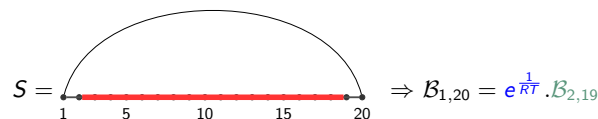
$$Z_{i,t} = 1, \forall t \in [i, i + \theta]$$

$$Z_{i,j} = \sum \left\{ \begin{array}{l} Z_{i+1,j} \\ \sum_{k=i+\theta+1}^j e^{-\frac{E_{bp}(i,k)}{RT}} \times Z_{i+1,k-1} \times Z_{k+1,j} \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
 - Correction du facteur de Boltzmann
- Facteur d'un backtrack = Produit des facteurs de ses parties
 Contributions énergétiques passent à l'exposant
 $(e^{-a/RT} \cdot Z^1 \cdot Z' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
 $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT})$

Exemple :



Fonction de partition = Comptage pondéré des structures compatibles

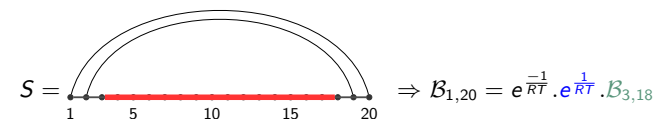
$$Z_{i,t} = 1, \forall t \in [i, i + \theta]$$

$$Z_{i,j} = \sum \left\{ \begin{array}{l} Z_{i+1,j} \\ \sum_{k=i+\theta+1}^j e^{-\frac{E_{bp}(i,k)}{RT}} \times Z_{i+1,k-1} \times Z_{k+1,j} \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
 - Correction du facteur de Boltzmann
- Facteur d'un backtrack = Produit des facteurs de ses parties
 Contributions énergétiques passent à l'exposant
 $(e^{-a/RT} \cdot Z^1 \cdot Z' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
 $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT})$

Exemple :



Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

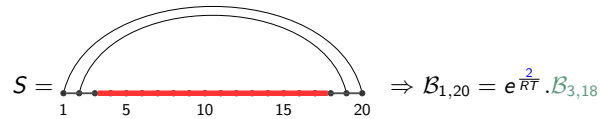
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}^j = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :



Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

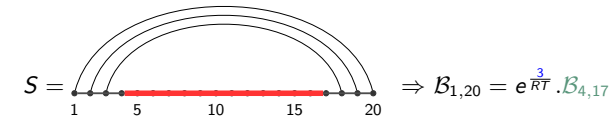
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}^j = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :



Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

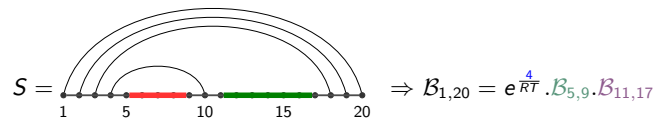
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}^j = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :



Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

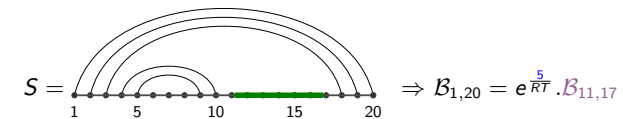
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}^j = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :



Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

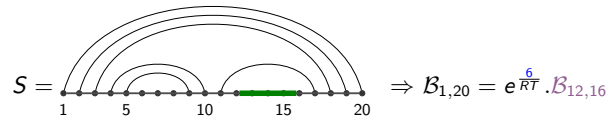
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :



Yann Ponty

Cours M2 BIBS - Algorithmique de l'ARN

Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

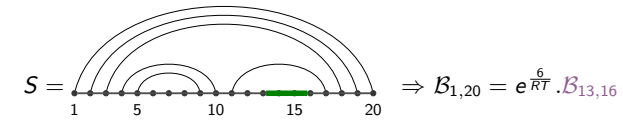
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :



Yann Ponty

Cours M2 BIBS - Algorithmique de l'ARN

Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

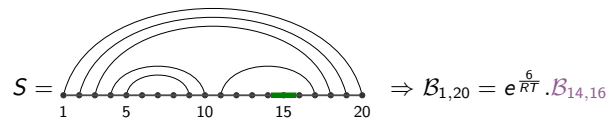
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :



Yann Ponty

Cours M2 BIBS - Algorithmique de l'ARN

Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

$$\mathcal{Z}_{i,t} = 1, \quad \forall t \in [i, i + \theta]$$

$$\mathcal{Z}_{i,j} = \sum \left\{ \begin{array}{l} \mathcal{Z}_{i+1,j} \\ \sum_j e^{-\frac{E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \\ k=i+\theta+1 \end{array} \right.$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann

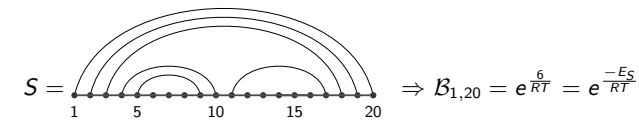
Facteur d'un backtrack = Produit des facteurs de ses parties

Contributions énergétiques passent à l'exposant

$$\left(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT} \right)$$

$$= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}$$

Exemple :

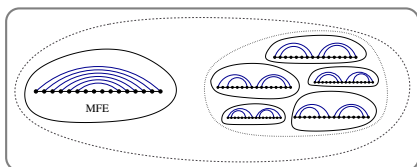


Yann Ponty

Cours M2 BIBS - Algorithmique de l'ARN

Échantillonnage statistique de structures d'ARN

La MFE (Probabilité maximale) peut être largement dominée par un ensemble \mathcal{B} de sous-optimaux structurellement similaires.
 ⇒ Conformation fonctionnelle trouvée plus probablement dans \mathcal{B} .



Expérience : [DCL05]

- Échantillonner des structures selon une probabilité de Boltzmann
- Effectuer un clustering
- Construire structure consensus dans le plus lourd cluster

⇒ Amélioration relative pour spécificité (+17.6%) et sensibilité (+21.74%, sauf Introns du groupe II)

Problème

Comment engendrer des structures dans la distribution de Boltzmann ?

Remontée stochastique

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i+1, j-1) \quad \text{A} \\ \sum \left(e^{-\frac{E_B(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{-(a+c)}{RT}} \sum (\mathcal{Z}(i+1, k-1) \mathcal{Z}^1(k, j-1)) \quad \text{C} \end{array} \right.$$

Remontée stochastique

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} \rightarrow e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i+1, j-1) \quad \text{A} \\ \text{???} \rightarrow \sum \left(e^{-\frac{E_B(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ \rightarrow e^{-\frac{-(a+c)}{RT}} \sum (\mathcal{Z}(i+1, k-1) \mathcal{Z}^1(k, j-1)) \quad \text{C} \end{array} \right.$$

Remontée stochastique

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i+1, j-1) \quad \text{A} \\ \sum \left(e^{-\frac{E_B(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{-(a+c)}{RT}} \sum (\mathcal{Z}(i+1, k-1) \mathcal{Z}^1(k, j-1)) \quad \text{C} \end{array} \right.$$

↓
r

A₁—A₂—B_i—B_{i+1}—...—B_{j-1}—B_j—C_i—C_{i+1}—...—C_{j-1}—C_j

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i + 1, j - 1) \quad \text{A} \\ \sum \left(e^{-\frac{E_{B_l}(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{-(a+c)}{RT}} \sum (\mathcal{Z}(i + 1, k - 1) \mathcal{Z}^1(k, j - 1)) \quad \text{C} \end{array} \right.$$

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i + 1, j - 1) \quad \text{A} \\ \sum \left(e^{-\frac{E_{B_l}(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{-(a+c)}{RT}} \sum (\mathcal{Z}(i + 1, k - 1) \mathcal{Z}^1(k, j - 1)) \quad \text{C} \end{array} \right.$$

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i + 1, j - 1) \quad \text{A} \\ \sum \left(e^{-\frac{E_{B_l}(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{-(a+c)}{RT}} \sum (\mathcal{Z}(i + 1, k - 1) \mathcal{Z}^1(k, j - 1)) \quad \text{C} \end{array} \right.$$

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i + 1, j - 1) \quad \text{A} \\ \sum \left(e^{-\frac{E_{B_l}(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{-(a+c)}{RT}} \sum (\mathcal{Z}(i + 1, k - 1) \mathcal{Z}^1(k, j - 1)) \quad \text{C} \end{array} \right.$$

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$
- 3 Répéter sur les sous-structures

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i + 1, j - 1) \quad \text{A} \\ \sum \left(e^{-\frac{E_{B_l}(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{(a+c)}{RT}} \sum (\mathcal{Z}(i + 1, k - 1) \mathcal{Z}^1(k, j - 1)) \quad \text{C} \end{array} \right.$$

Correction : Tout $S \in \mathcal{S}_\omega$ engendrée de façon unique (Unambiguïté de Turner)

La probabilité d'engendrer S est donc

$$p_S = \frac{\mathcal{B}(E_1)}{\mathcal{B}(S_\omega)} \cdot \frac{\mathcal{B}(E_2)}{\mathcal{B}(E_1)} \cdot \frac{\mathcal{B}(E_3)}{\mathcal{B}(E_2)} \cdots \frac{\mathcal{B}(\{S\})}{\mathcal{B}(E_m)}$$

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$
- 3 Répéter sur les sous-structures

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i + 1, j - 1) \quad \text{A} \\ \sum \left(e^{-\frac{E_{B_l}(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{(a+c)}{RT}} \sum (\mathcal{Z}(i + 1, k - 1) \mathcal{Z}^1(k, j - 1)) \quad \text{C} \end{array} \right.$$

Correction : Tout $S \in \mathcal{S}_\omega$ engendrée de façon unique (Unambiguïté de Turner)

La probabilité d'engendrer S est donc

$$p_S = \frac{1}{\mathcal{B}(S_\omega)} \cdot \frac{1}{1} \cdot \frac{1}{1} \cdots \frac{\mathcal{B}(\{S\})}{1}$$

Algorithme (Reformulation SFo1d [DL03])

Précalcul : Calculer les matrices (\mathcal{Z} , \mathcal{Z}' , \mathcal{Z}^1) des fonctions de partition.

Remontée stochastique :

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$
- 3 Répéter sur les sous-structures

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i + 1, j - 1) \quad \text{A} \\ \sum \left(e^{-\frac{E_{B_l}(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{(a+c)}{RT}} \sum (\mathcal{Z}(i + 1, k - 1) \mathcal{Z}^1(k, j - 1)) \quad \text{C} \end{array} \right.$$

Correction : Tout $S \in \mathcal{S}_\omega$ engendrée de façon unique (Unambiguïté de Turner)

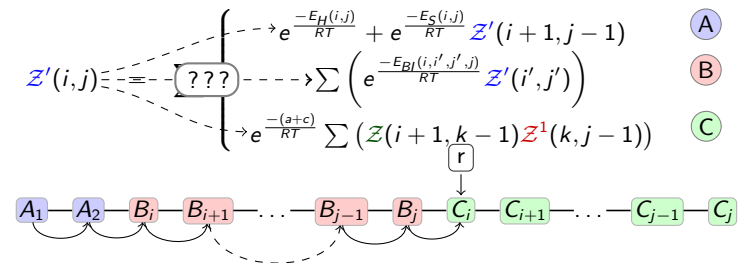
La probabilité d'engendrer S est donc

$$p_S = \frac{\mathcal{B}(\{S\})}{\mathcal{B}(S_\omega)} = \frac{e^{-E_S/RT}}{\mathcal{Z}} = P_{S, \omega}$$

Complexité ???

Algorithme (Reformulation SFo1d [DL03])

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$
- 3 Répéter sur les sous-structures



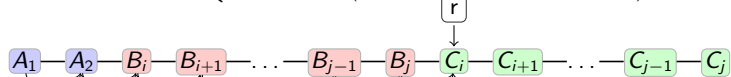
Complexité en moyenne en $\Theta(n\sqrt{n})$ dans l'hypothèse tout appariement.

Adaptation d'un parcours Boustrophedon $\Rightarrow \mathcal{O}(n \log nk)$ au pire.

Algorithme (Reformulation SFo1d [DL03])

- 1 Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i, j))$
- 2 Retirer à r les contributions à $\mathcal{Z}'(i, j)$, jusqu'à ce que $r < 0$
- 3 Répéter sur les sous-structures

$$\mathcal{Z}'(i, j) = \sum \left\{ \begin{array}{l} e^{-\frac{E_H(i, j)}{RT}} + e^{-\frac{E_S(i, j)}{RT}} \mathcal{Z}'(i+1, j-1) \quad \text{A} \\ \sum \left(e^{-\frac{E_B(i, i', j', j)}{RT}} \mathcal{Z}'(i', j') \right) \quad \text{B} \\ e^{-\frac{(g+c)}{RT}} \sum (\mathcal{Z}(i+1, k-1) \mathcal{Z}^1(k, j-1)) \quad \text{C} \end{array} \right.$$



Après $\Theta(n)$ opérations, on réitère sur un interval de taille $n-1$
 \Rightarrow Complexité du cas au pire en $\mathcal{O}(n^2 k)$ pour k échantillons

Complexité en moyenne en $\Theta(n\sqrt{n})$ dans l'hypothèse tout appariement.
 Adaptation d'un parcours Boustrophedon $\Rightarrow \mathcal{O}(n \log nk)$ au pire.

- A. Condon, B. Davy, B. Rastegari, S. Zhao, and F. Tarrant.
 Classifying RNA pseudoknotted structures.
Theoretical Computer Science, 320(1) :39–50, 2004.
- 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.
- Y. Ding, C. Y. Chan, and C. E. Lawrence.
 RNA secondary structure prediction by centroids in a boltzmann weighted ensemble.
RNA, 11 :1157–1166, 2005.
- Y. Ding and E. Lawrence.
 A statistical sampling algorithm for RNA secondary structure prediction.
Nucleic Acids Research, 31(24) :7260–7301, 2003.
- P. Gardner and R. Giegerich.
 A comprehensive comparison of comparative rna structure prediction approaches.
BMC Bioinformatics, 5(1) :140, 2004.
- 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.
- N. Leontis and E. Westhof.
 Geometric nomenclature and classification of RNA base pairs.
RNA, 7 :499–512, 2001.
- D.H. Mathews, J. Sabina, M. Zuker, and D.H. Turner.
 Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.
J Mol Biol, 288 :911–940, 1999.

- Ján Maňuch, Chris Thachuk, Ladislav Stacho, and Anne Condon.
 Np-completeness of the direct energy barrier problem without pseudoknots.
 pages 106–115, 2009.
- N. R. Markham and M. Zuker.
Bioinformatics, chapter UNAFold, pages 3–31.
 Springer, 2008.
- M. Parisien and F. Major.
 The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.
Nature, 452(7183) :51–55, 2008.
- 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.
DNA Res, 16(1) :45–58, Feb 2009.
- 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.