M2 BIBS – ARN – Séance 1

Repliement *in silico* de l'ARN

Yann Ponty

CNRS
LIX/INRIA AMIB – École Polytechnique – France

6 Décembre 2010

Pré-introduction...

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

Problème : Vous vendez (très cher !) des cravates et le client ne dispose que de grosses coupures. Vous disposez de pièces de 1, 20 et 50 centimes en quantité infinie. Le client souhaite minimiser la monnaie reçue (en nombre de pièce), sinon il ne revient pas ... Comment rendre la monnaie sans perdre le client ?

Exemple : Aller au plus court (stratégie gloutonne)

$21 =$??

Pré-introduction...

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

Problème : Vous vendez (très cher !) des cravates et le client ne dispose que de grosses coupures. Vous disposez de pièces de 1, 20 et 50 centimes en quantité infinie. Le client souhaite minimiser la monnaie reçue (en nombre de pièce), sinon il ne revient pas ... Comment rendre la monnaie sans perdre le client ?

Exemple : Aller au plus court (stratégie gloutonne)

$21 =$ $+$

55??

Pré-introduction...

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

Problème : Vous vendez (très cher !) des cravates et le client ne dispose que de grosses coupures. Vous disposez de pièces de 1, 20 et 50 centimes en quantité infinie. Le client souhaite minimiser la monnaie reçue (en nombre de pièce), sinon il ne revient pas ... Comment rendre la monnaie sans perdre le client ?

Exemple : Aller au plus court (stratégie gloutonne)

$21 =$ $+$

$55 =$

60??

Pré-introduction...

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

Problème : Vous vendez (très cher !) des cravates et le client ne dispose que de grosses coupures. Vous disposez de pièces de 1, 20 et 50 centimes en quantité infinie. Le client souhaite minimiser la monnaie reçue (en nombre de pièce), sinon il ne revient pas ... Comment rendre la monnaie sans perdre le client ?

Exemple : Aller au plus court (stratégie gloutonne)

$21 =$

$55 =$

$60 =$??

Pré-introduction...

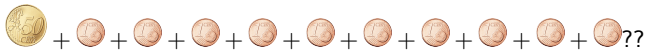
...ou comment gagner 1 million de dollars en rendant la monnaie !!

Problème : Vous vendez (très cher !) des cravates et le client ne dispose que de grosses coupures. Vous disposez de pièces de 1, 20 et 50 centimes en quantité infinie. Le client souhaite minimiser la monnaie reçue (en nombre de pièce), sinon il ne revient pas ... Comment rendre la monnaie sans perdre le client ?

Exemple : Aller au plus court (stratégie gloutonne)

$21 =$

$55 =$

$60 =$  ??

$=$!

$\Rightarrow$ ne marche pas toujours ...
Alt. : Exploration exhaustive $\Rightarrow \theta(\#\text{pièces}^N)$ calculs.

Problème difficile dans un sens bien-défini (NP-complet) $\Rightarrow$ 1 M$

Pré-introduction

Mais il existe une récurrence donnant le nombre minimal de pièce :

$$MinNbPieces(N) = \text{Min} \begin{cases} \text{1€} & \rightarrow \quad 1 + MinNbPieces(N-1) \\ \text{20€} & \rightarrow \quad 1 + MinNbPieces(N-20) \\ \text{50€} & \rightarrow \quad 1 + MinNbPieces(N-50) \end{cases}$$

Avec une mémoire extraordinaire (Deux fois N), on peut alors répondre après $N \times \#\text{Pièces}$ calculs [1].

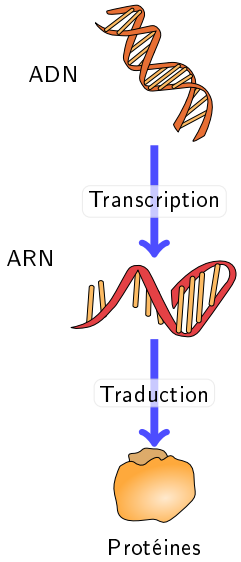
On a donc évité de reparcourir des portions déjà visitée de l'arbre des possibles :
$\Rightarrow$ Programmation dynamique.

1. Malheureusement, on n'a pas gagné le million, car le nombres de valeurs de N pour lesquels on doit calculer la récurrence ci-dessus est en général exponentielle sur le codage de N, et cet algorithme est donc toujours exponentiel au regard de la théorie de la complexité.

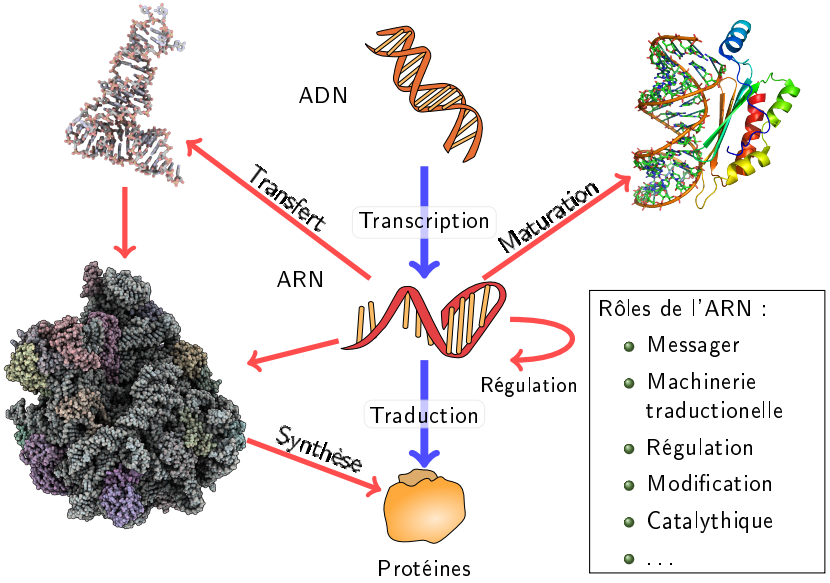
Résumé

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

Dogme fondamental de la biologie moléculaire

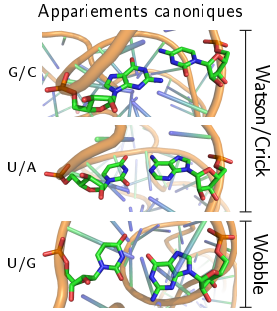


Dogme fondamental de la biologie moléculaire



Repliement de l'ARN

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



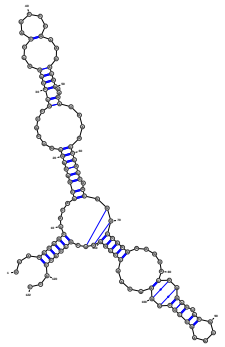
Repliement de l'ARN = Processus stochastique continu dirigé par (résultant en) un appariement des nucléotides.

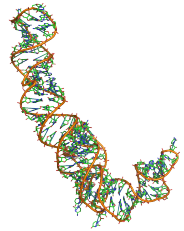
Comprendre le repliement des ARN aide à comprendre et prédire leur fonction.

Structure de l'ARN

Trois[2] niveaux de représentation :

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC
```





Structure primaire | Structure secondaire | Structure tertiaire

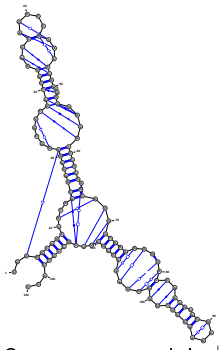
Source : 5s rRNA (PDB 1K73 :B)

2. Enfin, presque ...

Structure de l'ARN

Trois[2] niveaux de représentation :

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC
```



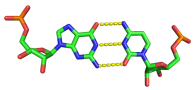
Structure primaire | Structure secondaire[+] | Structure tertiaire

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

2. Enfin, presque ...

Ignorés par la structure secondaire

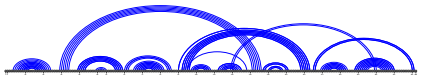
- Appariements non-canoniques
  Toute paire de base autre que {(A-U), (C-G), (G-U)}
  Ou interagissant sur un bord non-standard (WC/WC-Cis) [LW01].



Paire CG canonique (WC/WC-Cis)

Paire CG non canonique (Sucre/WC-Trans)

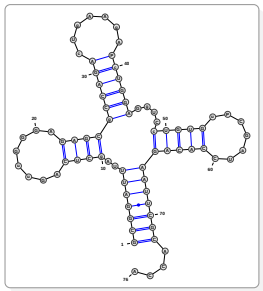
- Pseudonoeuds



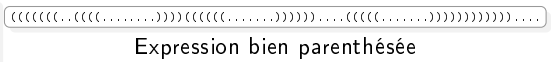
Structure pseudonoeud d'un Ribobozyme du Groupe I (PDBID : 1Y0Q :A)

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

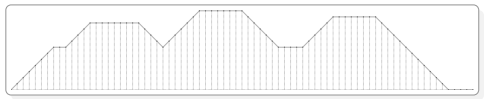
Diversité de représentations



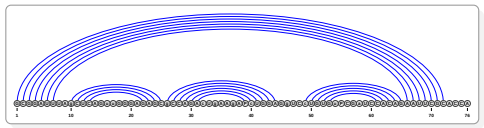
Graphe planaire (outer planar)



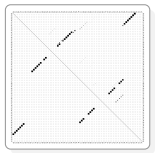
Expression bien parenthésée



Mountain view



Linéaire



Dot plot

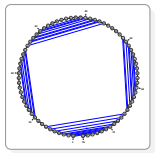


Diagramme de Feynman

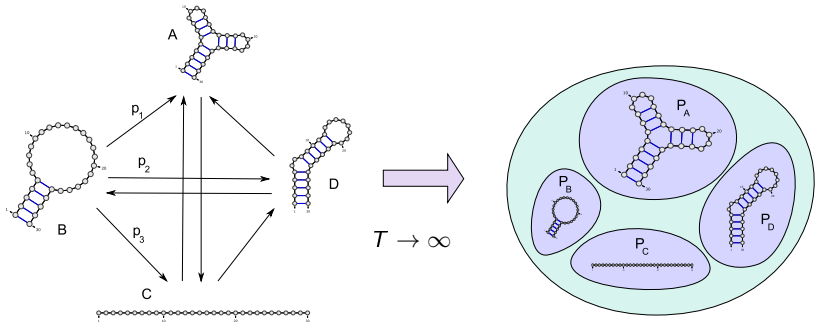
Représentation différentes et équivalentes
⇒ Aide l'intuition algorithmique
+ Propriétés algébriques sympathiques
⇒ Algorithmique efficace !

Résumé

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

Aparté thermodynamique

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



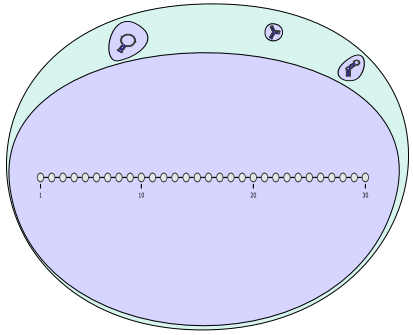
Convergence vers une distribution stationnaire de probabilité, l'équilibre de Boltzmann, où la probabilité est exponentiellement faible sur l'énergie libre.
Corollaire : La conformation initiale est sans d'importance.

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

Hors de l'équilibre

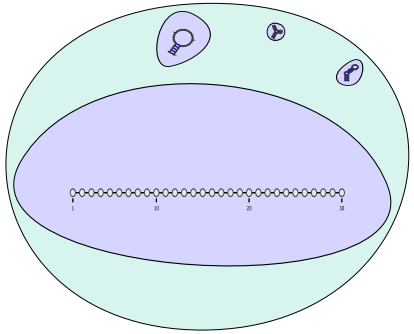
Transcription : ARN synthétisé sans appariement (Sauf exception)



$T = 0$

Hors de l'équilibre

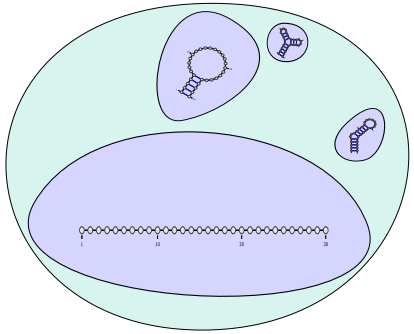
Transcription : ARN synthétisé sans appariement (Sauf exception)



$T = 1h$

Hors de l'équilibre

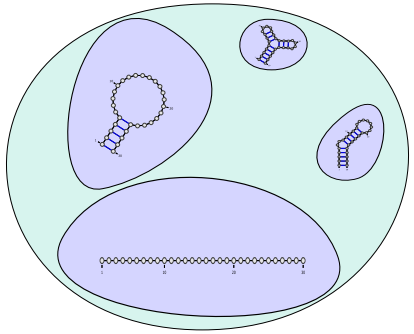
Transcription : ARN synthétisé sans appariement (Sauf exception)



$T = 2h$

Hors de l'équilibre

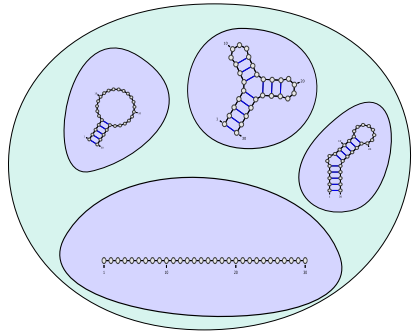
Transcription : ARN synthétisé sans appariement (Sauf exception)



$T = 5h$

Hors de l'équilibre

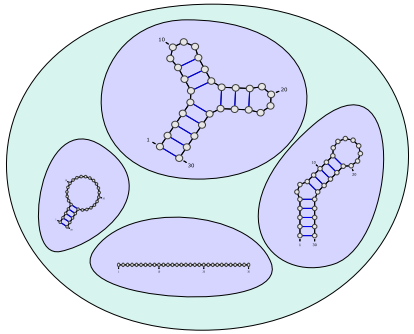
Transcription : ARN synthétisé sans appariement (Sauf exception)



$T = 10h$

Hors de l'équilibre

Transcription : ARN synthétisé sans appariement (Sauf exception)

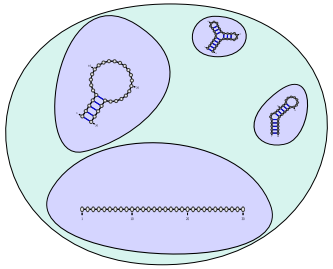


$T \to \infty$

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

Hors de l'équilibre

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])

Résumé

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

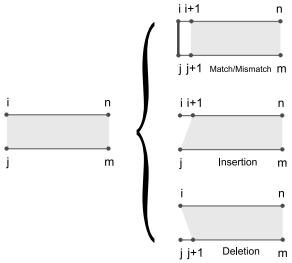
Programmation dynamique : Principe général

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 :

Espaces de solutions *discrets* (alignements, repliements)
$+$ Fonctions Objectif *additives* (score, énergie)
$\Rightarrow$ Schémas de programmation dynamique souvent efficaces.

Exemple : Alignement local (Smith/Waterman)



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

Détails algorithmiques

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

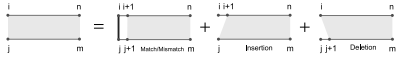
Étant donné un schéma, deux étapes :

- 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
- Nombres 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 $p_i = p_j = -1$

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

		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 $p_i = p_j = -1$

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

		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 $p_i = p_j = -1$

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

		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								

Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		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								

Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		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								

Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		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 $p_i = p_j = -1$

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

		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 $p_i = p_j = -1$

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

		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 $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

Meilleur alignement

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

Meilleur alignement

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

A
A

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

-	A	
T	A	

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

C	-	A
C	T	A

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

A	C	-	A
A	C	T	A

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

C	A	C	-	A
C	A	C	T	A

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

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

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

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

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

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

Exemple complet

Exemple : Alignement local de séquences AGCACACA et ACACACTA
Coûts : Match $m_{i,j} = +2$, Insertion/Déletion $p_i = p_j = -1$

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

		A	C	A	C	A	C	T	A
	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Meilleur alignement

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

Propriétés des schémas

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-ambiguité : Chaque solution est *engendrée* au plus une fois.

$\Rightarrow$ Possibilité d'énumérer l'espace des solutions.

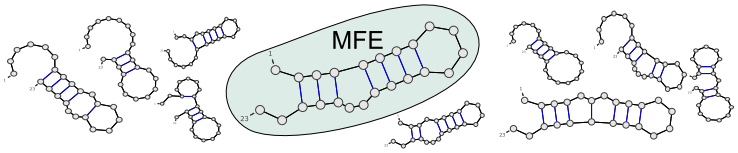
Résumé

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

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 : Ambiguité, quelle est la structure native?

Modèle de Nussinov/Jacobson

Modèle de Nussinov/Jacobson (NJ)

Plus proche voisins simple :

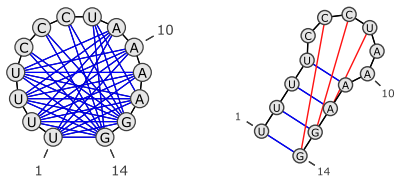
- Seuls les appariements contribuent à l'énergie
- Uniquement liaisons Watson/Crick (A/U,C/G) et Wobble (G/U)

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

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

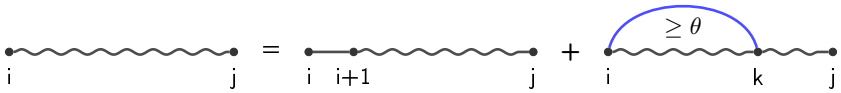
Exemple :

UUUUCCCUAAAAGG



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

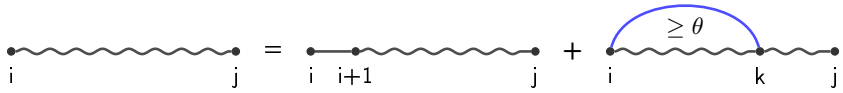
Décomposition de Nussinov/Jacobson



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

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

Décomposition de Nussinov/Jacobson



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

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

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

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

Nussinov/Jacobson

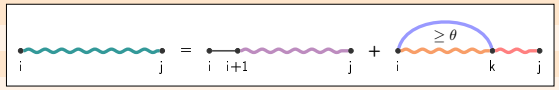
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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

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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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

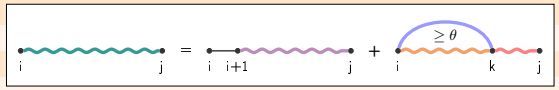
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



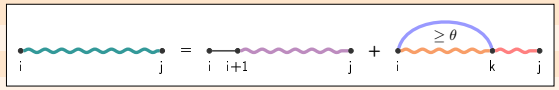
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



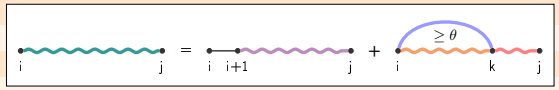
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



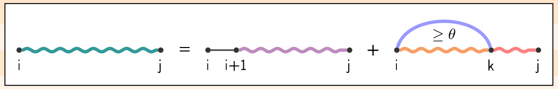
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



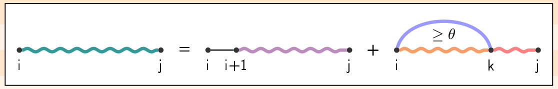
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



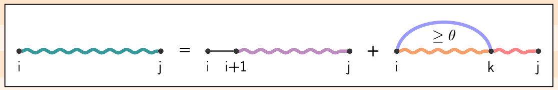
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



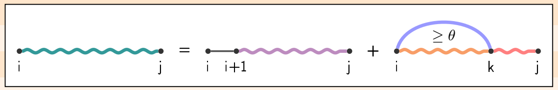
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



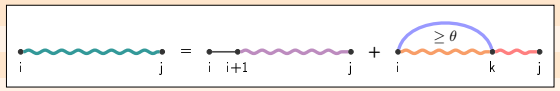
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



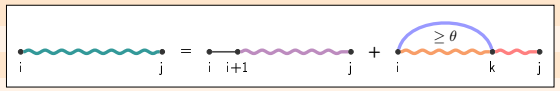
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



Nussinov/Jacobson

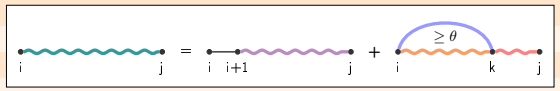
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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



Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Nussinov/Jacobson

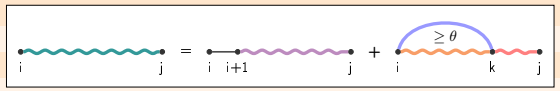
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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



Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Nussinov/Jacobson

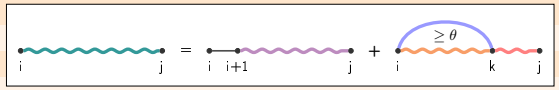
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	.	.	.	.	.	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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



Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

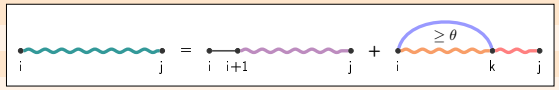
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



Nussinov/Jacobson

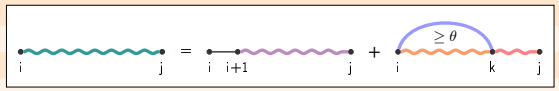
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	.	.	.	.	.	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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



Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Nussinov/Jacobson

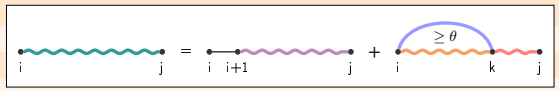
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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



Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Nussinov/Jacobson

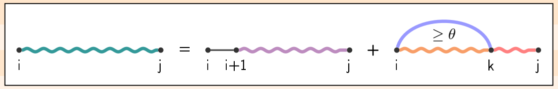
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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



Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

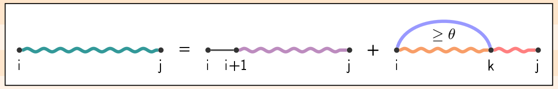
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



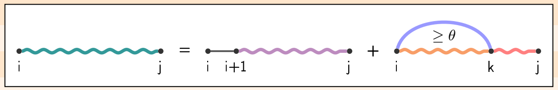
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



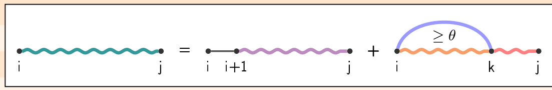
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



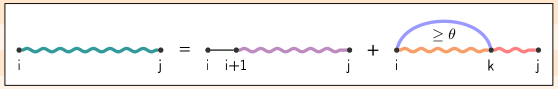
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



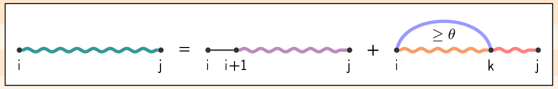
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



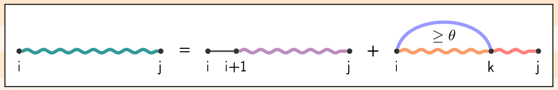
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



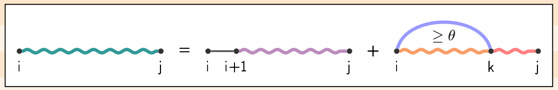
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



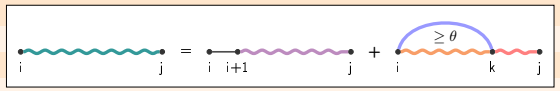
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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



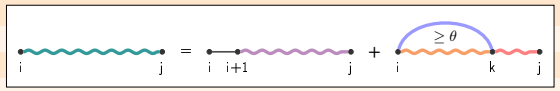
Nussinov/Jacobson

	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	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	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	0	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



Nussinov/Jacobson

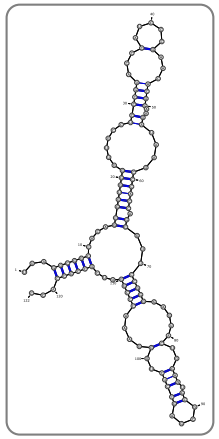
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	(	(	.	.	.	)	)	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C		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



Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Modèle de Turner

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



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

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

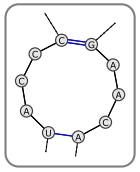
Meilleure résultats grâce à la prise en compte de l'empilement.

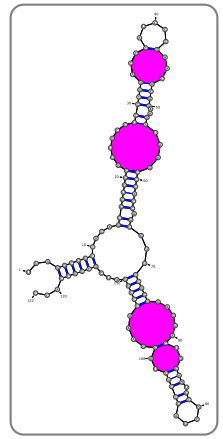
Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Modèle de Turner

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

- Boucles internes





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

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

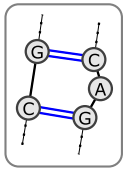
Meilleure résultats grâce à la prise en compte de l'empilement.

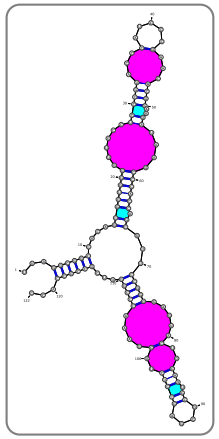
Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Modèle de Turner

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

- Boucles internes
- Renflements





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

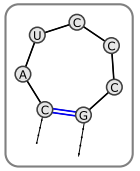
Déterminées expérimentalement
+ Interpolation pour grandes boucles.

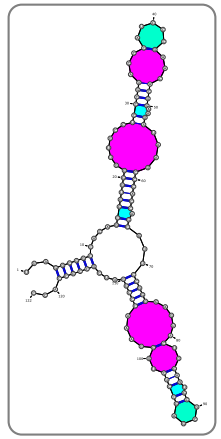
Meilleure résultats grâce à la prise en compte de l'empilement.

Modèle de Turner

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

- Boucles internes
- Renflements
- Boucles terminales





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

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

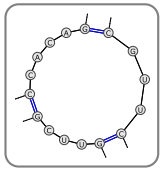
Meilleure résultats grâce à la prise en compte de l'empilement.

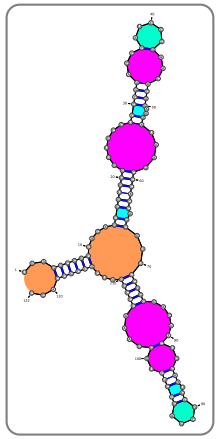
Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Modèle de Turner

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

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples





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

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

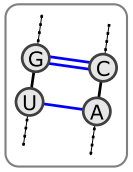
Meilleure résultats grâce à la prise en compte de l'empilement.

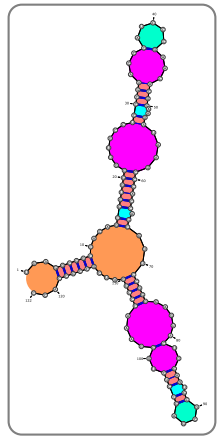
Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

Modèle de Turner

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

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





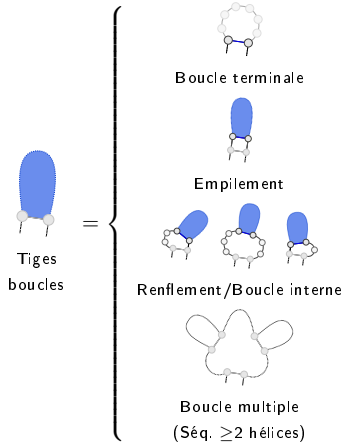
Énergies libres Δ G des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

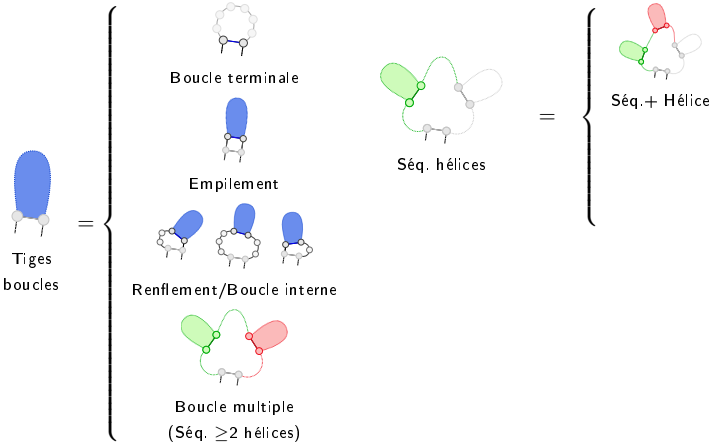
Meilleure résultats grâce à la prise en compte de l'empilement.

Yann Ponty M2 BIM - ARN – Séance 1 – Repliement de l'ARN

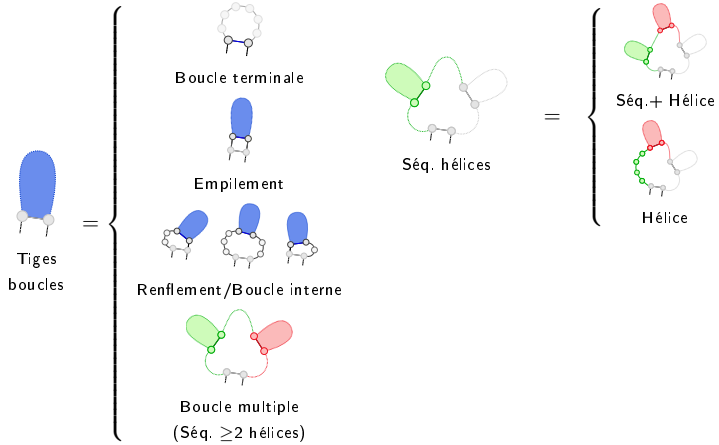
MFE DP equations



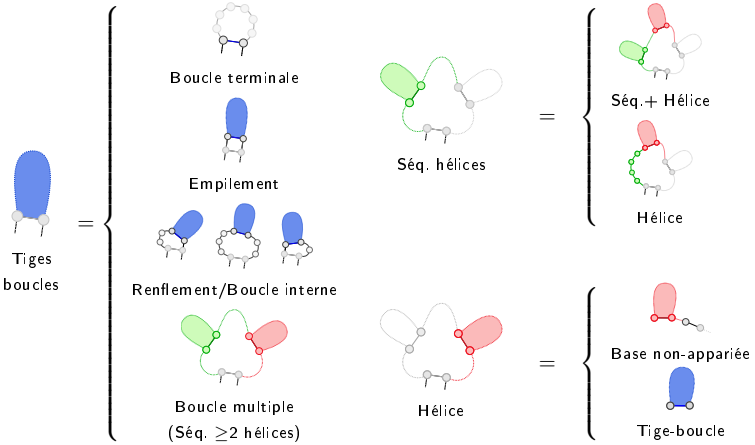
MFE DP equations



MFE DP equations

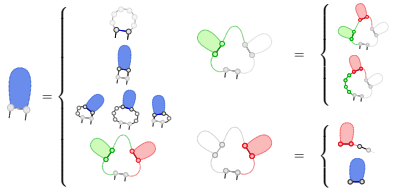


MFE DP equations



MFold Unafold

- $E_H(i,j)$: 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

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

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

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

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

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

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

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

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

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

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

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

$\mathcal{O}(n)$ contributeurs potentiels au Min :
⇒ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.
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 précalcul [3]

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

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

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

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

$\mathcal{O}(n)$ contributeurs potentiels au Min :
⇒ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.
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 précalcul [3]
⇒ `UnaFold` [MZ08] calcule la structure secondaire d'énergie minimale.

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

Deux approches

Definition (Repliement 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.

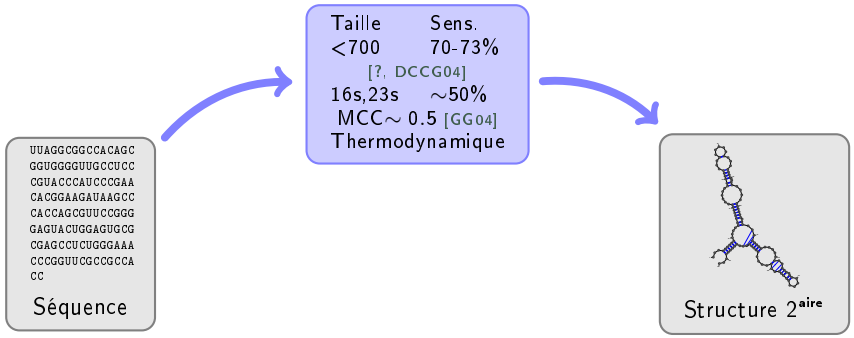
Avantages :
- Meilleures performances
- Affinement permanent

Limites :
- Complexité élevée
- Exploration non-exhaustive

Performances

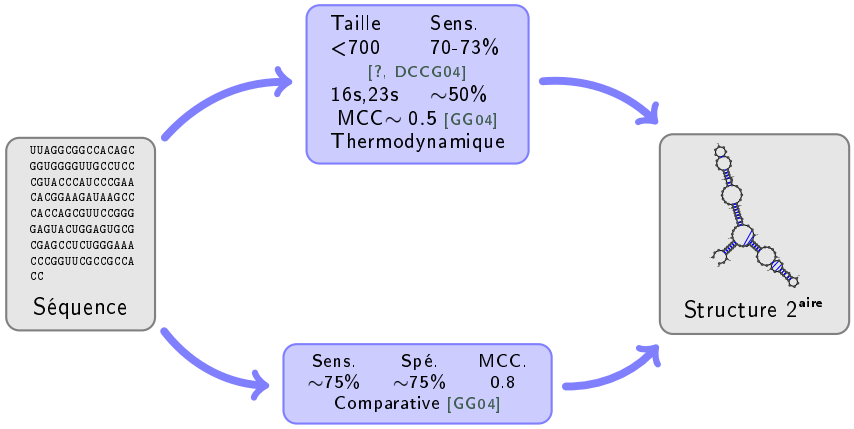
Rappel : $MCC = \frac{t^+t^- - f^+f^-}{\sqrt{(t^+ + f^+)(t^+ + f^-)(t^- + f^+)(t^- + f^-)}}$

Performances



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

Performances



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

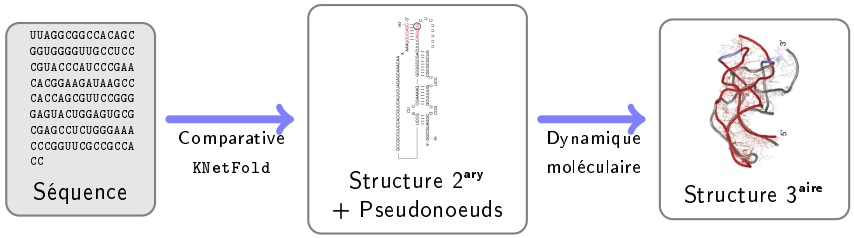
Futur (proche) : Vers une prédiction 3D

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

Futur (proche) : Vers une prédiction 3D

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

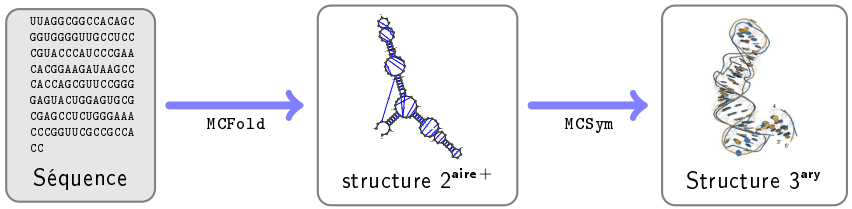
- Models comparatifs + Dynamique moléculaires : `RNA2D3D` [SYKB07]



Futur (proche) : Vers une prédiction 3D

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

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



References I

A. Condon, B. Davy, B. Rastegari, S. Zhao, and F. Tarrant.
Classifying RNA pseudoknotted structures.
Theoretical Computer Science, 320(1) :35–50, 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.

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.

Ján Manuch, 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.

References II

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.

Exercice : Parsing/repliement des structures secondaires en Python