



# Approche ab initio

## Prédiction de structure d'ARN

Séquence → Structure de plus faible énergie libre  
(selon un modèle d'énergie donné)

`CGCGAATTTAGCCDCAGTGGGAGAGCCGCGCAGACGSAANATCCDTAGGGCCCTGGTTCGATCCACAGTATTCGTCGCA`

## Modèle thermodynamique

10.2 RNA secondary structure prediction

4 nt loop +5.9	U-U	-1.1 terminal mismatch of hairpin
	G-C	-2.9 stack
1 nt bulge +3.3	A-C	-2.9 stack (special case of 1 nt bulge)
	G-C	-1.8 stack
	U-A	-0.9 stack
	A-U	-1.6 stack
	C-G	-2.1 stack
5' angle -0.3	A-U	
unstructured single strand 0.0	A-U	

overall  $\Delta G = -4.6$  kcal/mol

Figure 10.10 An example  $\Delta G$  calculation for an RNA stem loop (the wild type R17 coat protein binding site).

Pour des structures secondaires sans pseudo-noeuds

Biological sequence analysis  
Durbin, Eddy, Krogh, Mitchison  
Cambridge Univ. Press 1998

## Algorithme pour un modèle plus simple

Prédiction de structure

- sans pseudo-noeud,
- en maximisant le nombre de nucléotides appariés.

$\gamma(i,j)$  = nombre d'appariements dans la structure entre les  $i^{\text{ème}}$  et  $j^{\text{ème}}$  nucléotides.

$\delta(i,j) = 1$  si  $i$  et  $j$  sont appariables, 0 sinon.

## Algorithme de Nussinov (1978)

Les 4 façons possibles de construire la meilleure structure entre  $i$  et  $j$ , connaissant les meilleures sous-structures entre  $i+1$  et  $j-1$ .

## Algorithme de Nussinov (1978)

Attention, cette décomposition est ambiguë : plusieurs cas différents peuvent être valables simultanément. Il existe des décompositions non ambiguës pour le même problème

### Algorithme de Nussinov (1978)

$\gamma(i,j) = \text{Max} \{$

1.  $\gamma(i+1,j)$       2.  $\gamma(i,j-1)$       3.  $\gamma(i+1,j-1)+\delta(i,j)$

4.  $\text{Max}_{i < k < j} \{ \gamma(i,k) + \gamma(k+1,j) \}$  }

### Principe : programmation dynamique

$\gamma(i,j) = \text{Max} \{$

- $\gamma(i+1,j)$
- $\gamma(i,j-1)$
- $\gamma(i+1,j-1)+\delta(i,j)$
- $\text{Max}_{i < k < j} \{ \gamma(i,k) + \gamma(k+1,j) \}$

1                                  i                                  j                                  n

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0						
C <sub>2</sub>	0	0						
G <sub>3</sub>		0	0					
G <sub>4</sub>			0	0				
C <sub>5</sub>				0	0			
A <sub>6</sub>						0	0	
U <sub>7</sub>							0	0
G <sub>8</sub>								0

$\gamma(i,j) = \text{Max} \{$

- $\gamma(i+1,j)$
- $\gamma(i,j-1)$
- $\gamma(i+1,j-1)+\delta(i,j)$
- $\text{Max}_{i < k < j} \{ \gamma(i,k) + \gamma(k+1,j) \}$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0						
C <sub>2</sub>	0	0	1					
G <sub>3</sub>		0	0					
G <sub>4</sub>			0	0				
C <sub>5</sub>				0	0			
A <sub>6</sub>						0	0	
U <sub>7</sub>							0	0
G <sub>8</sub>								0

$\gamma(i,j) = \text{Max} \{$

- $\gamma(i+1,j)$
- $\gamma(i,j-1)$
- $\gamma(i+1,j-1)+\delta(i,j)$
- $\text{Max}_{i < k < j} \{ \gamma(i,k) + \gamma(k+1,j) \}$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0						
C <sub>2</sub>	0	0	1					
G <sub>3</sub>		0	0	0				
G <sub>4</sub>			0	0	1			
C <sub>5</sub>				0	0	0		
A <sub>6</sub>						0	0	1
U <sub>7</sub>							0	0
G <sub>8</sub>								0

$\gamma(i,j) = \text{Max} \{$

- $\gamma(i+1,j)$
- $\gamma(i,j-1)$
- $\gamma(i+1,j-1)+\delta(i,j)$
- $\text{Max}_{i < k < j} \{ \gamma(i,k) + \gamma(k+1,j) \}$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1					
C <sub>2</sub>	0	0	1					
G <sub>3</sub>		0	0	0				
G <sub>4</sub>			0	0	1			
C <sub>5</sub>				0	0	0		
A <sub>6</sub>						0	0	1
U <sub>7</sub>							0	0
G <sub>8</sub>								0

$\gamma(i,j) = \text{Max} \{$

- $\gamma(i+1,j)$
- $\gamma(i,j-1)$
- $\gamma(i+1,j-1)+\delta(i,j)$
- $\text{Max}_{i < k < j} \{ \gamma(i,k) + \gamma(k+1,j) \}$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1					
C <sub>2</sub>	0	0	1					
G <sub>3</sub>		0	0	0				
G <sub>4</sub>			0	0	1			
C <sub>5</sub>				0	0	0		
A <sub>6</sub>					0	0	1	
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

C<sub>1</sub> C<sub>2</sub> G<sub>3</sub>

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1					
C <sub>2</sub>	0	0	1	1				
G <sub>3</sub>		0	0	0	1			
G <sub>4</sub>			0	0	1	1		
C <sub>5</sub>				0	0	0	1	
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2				
C <sub>2</sub>	0	0	1	1				
G <sub>3</sub>		0	0	0	1			
G <sub>4</sub>			0	0	1	1		
C <sub>5</sub>				0	0	0	1	
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

C<sub>1</sub> C<sub>2</sub> G<sub>3</sub> G<sub>4</sub>

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2				
C <sub>2</sub>	0	0	1	1	2			
G <sub>3</sub>		0	0	0	1			
G <sub>4</sub>			0	0	1	1		
C <sub>5</sub>				0	0	0	1	
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

C<sub>2</sub> G<sub>3</sub> G<sub>4</sub> C<sub>5</sub>

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2	2	2		
C <sub>2</sub>	0	0	1	1	2	2		
G <sub>3</sub>		0	0	0	1	1	2	
G <sub>4</sub>			0	0	1	1	2	2
C <sub>5</sub>				0	0	0	1	2
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2	2	2		
C <sub>2</sub>	0	0	1	1	2	2	3	
G <sub>3</sub>		0	0	0	1	1	2	
G <sub>4</sub>			0	0	1	1	2	2
C <sub>5</sub>				0	0	0	1	2
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

C<sub>2</sub> G<sub>3</sub> G<sub>4</sub> C<sub>5</sub> A<sub>6</sub> U<sub>7</sub>

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2	2	2		
C <sub>2</sub>	0	0	1	1	2	2	3	
G <sub>3</sub>		0	0	0	1	1	2	
G <sub>4</sub>			0	0	1	1	2	2
C <sub>5</sub>				0	0	0	1	2
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

C<sub>2</sub> G<sub>3</sub> G<sub>4</sub> C<sub>5</sub> A<sub>6</sub> U<sub>7</sub>

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2	2	2	3	
C <sub>2</sub>	0	0	1	1	2	2	3	3
G <sub>3</sub>		0	0	0	1	1	2	2
G <sub>4</sub>			0	0	1	1	2	2
C <sub>5</sub>				0	0	0	1	2
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2	2	2	3	4
C <sub>2</sub>	0	0	1	1	2	2	3	3
G <sub>3</sub>		0	0	0	1	1	2	2
G <sub>4</sub>			0	0	1	1	2	2
C <sub>5</sub>				0	0	0	1	2
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2	2	2	3	4
C <sub>2</sub>	0	0	1	1	2	2	3	3
G <sub>3</sub>		0	0	0	1	1	2	2
G <sub>4</sub>			0	0	1	1	2	2
C <sub>5</sub>				0	0	0	1	2
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

	C <sub>1</sub>	C <sub>2</sub>	G <sub>3</sub>	G <sub>4</sub>	C <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	G <sub>8</sub>
C <sub>1</sub>	0	0	1	2	2	2	3	4
C <sub>2</sub>	0	0	1	1	2	2	3	3
G <sub>3</sub>		0	0	0	1	1	2	2
G <sub>4</sub>			0	0	1	1	2	2
C <sub>5</sub>				0	0	0	1	2
A <sub>6</sub>					0	0	1	1
U <sub>7</sub>						0	0	0
G <sub>8</sub>							0	0

C<sub>1</sub> C<sub>2</sub> G<sub>3</sub> G<sub>4</sub> C<sub>5</sub> A<sub>6</sub> U<sub>7</sub> G<sub>8</sub>

$$\gamma(i,j) = \text{Max} \{ \begin{array}{l} \gamma(i+1,j) \\ \gamma(i,j-1), \\ \gamma(i+1,j-1)+\delta(i,j) \\ \text{Max}_{i<k<j} \{ \gamma(i,k) + \gamma(k+1,j) \} \end{array} \}$$

### Algorithme de Zuker-Stiegler (1981)

- Modèle d'énergie plus réaliste (paramètres de Turner *et al.*)
- Même principe (programmation dynamique), avec deux matrices au lieu d'une.
- Structures secondaires toujours sans pseudonoeds.



### Trois classes de problèmes algorithmiques

**P**

- (Deterministic) Polynomial : réponse garantie en temps polynomial.

**NP**

- Non-déterministic polynomial : vérification d'un « certificat » de réponse positive en temps polynomial.

**NPC** : NP-Complet (sous-classe de NP)

- Les plus difficiles de la classe NP.
- Si  $P \neq NP$ , pas de réponse garantie en temps polynomial.

**P=NP?**

Pourtant il y a des algorithmes polynomiaux qui gèrent les pseudonoœuds avec des modèles d'énergie non triviaux

Sans pseudonoœuds		
Année	Auteurs	Complexité
1981	Zuker, Stiegler	$O(n^4)$
1999	Lingsó, Zucker, Pedersen	$O(n^3)$

Avec pseudonoœuds		
Année	Auteurs	Complexité
1999	Rivas, Eddy	$O(n^6)$
2000	Lyngsó, Pedersen	$O(n^5)$
2000	Akutsu, Uemura	$O(n^5)$
2003	Dirks, Pierce	$O(n^5)$
2004	Reeder, Giegerich	$O(n^4)$
2009	Cao, Chen	$O(n^6)$

Pourtant il y a des algorithmes polynomiaux qui gèrent les pseudonoœuds avec des modèles d'énergie non triviaux

Avec pseudonoœuds			Classe
1999	Rivas, Eddy	$O(n^6)$	R&E
2000	Lyngsó, Pedersen	$O(n^5)$	L&P
2000	Akutsu, Uemura	$O(n^5)$	A&U
2003	Dirks, Pierce	$O(n^5)$	D&P
2004	Reeder, Giegerich	$O(n^4)$	R&G
2009	Cao, Chen	$O(n^6)$	C&C

Tous ces algorithmes sont basés sur la programmation dynamique. Ils ne considèrent que des **classes restreintes** de structures avec pseudonoœuds.  
(La classe d'un algorithme est l'ensemble des structures qui peuvent être prédites théoriquement par cet algorithme.)

### Rstrictions sur les pseudonoœuds

- Dans toutes les classes sauf une (R&E), les pseudonoœuds sont *simples*:

- Et même parfois encore plus simples (pseudonoœuds de type H)

### Relations d'inclusion entre les classes

[Condon, Davy, Rastegary, Zhao, Tarrant 2004]+[Saule et al. 2011]

$$\begin{array}{c}
 \text{PKF} \subset \text{L\&P} \subset \text{D\&P} \subset \text{A\&U} \subset \text{R\&E} \\
 \text{O}(n^3) \quad \text{O}(n^6) \quad \text{O}(n^4) \quad \text{O}(n^6)
 \end{array}$$

$O(n^5)$

### Cardinalités des classes

[Saule et al. 2011]

Class	asympt.	$\alpha$	$\omega$	Compl.	Remark
PKF	$\frac{\alpha}{2\sqrt{\pi} n^{3/2}} \omega^n$	2	4	$O(n^3)$	Catalan numbers
L&P *	$\frac{1}{2} \omega^n$	-	4	$O(n^5)$	Closed formula
C&C *	$\frac{\alpha}{2\sqrt{\pi} n^{3/2}} \omega^n$	1,6651	5,857	$O(n^6)$	
R&G *	$\frac{\alpha}{2\sqrt{\pi} n^{3/2}} \omega^n$	0,1651	6,576	$O(n^4)$	
D&P *	$\frac{\alpha}{2\sqrt{\pi} n^{3/2}} \omega^n$	0,7535	7,315	$O(n^5)$	
A&U *	$\frac{\alpha}{2\sqrt{\pi} n^{3/2}} \omega^n$	0,6575	7,547	$O(n^5)$	
R&E	open	-	-	$O(n^6)$	
All	$\sqrt{2} \cdot 2^n \cdot (\frac{n}{e})^n$	-	-	NPC	Involutions with no fixed points

## Approche comparative

Avec alignement préalable des séquences

## Détection des covariations

i            j

GCCUUCGGGC

GACUUCGGUC

GVCU-CGGCC

RNA-Alifold (Hofacker et al. 2000)

RNAz (Washietl et al. 2005)

## Application : tRNA Alanine

```

>Artibeus jamaicensis
AAGGCCTTAGCTTAATTAAGTAGTTGATTGGCATTACAGCAGCTGAGGATAAAGCTTGCAGTCCCTA
>Balaenoptera musculus
GAGGATTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Bos taurus
GAGGATTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Canis familiaris
GAGGCTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Ceratotherium simum
GAGGCTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Dasypus novemcinctus
GAGGCTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Equus asinus
AAGGCCTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Erinaceus europaeus
GAGGATTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Felis catus
GAGGCTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Hippopotamus amphibius
AAGGCCTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
>Homo sapiens
AAGGCCTTAGCTTAATTAAGTGTGGTGGTTCGATTCATTGATGATAGATATAGTCTGCAGTCCCTA
    
```

## RNA-alifold

## Application : tRNA *H.sapiens*

```

>Homo sapiens Arg, True Structure
TGGTATATAGTTTAAACAAACGAGTATTTCGACTCATTAAATATGATAATCATATTACCAA
(((L..L..(((L.....))))..(((L.....))))..(((L.....))))..))))).
>Homo sapiensArg
TGGTATATAGTTTAAACAAACGAGTATTTCGACTCATTAAATATGATAATCATATTACCAA
>Homo sapiensArg
TAGATTGAGGCGATGATTAGGGTGTAGCTGTAACTAAGTGTGGGGTTAAGTCCATTGGCTAG
>Homo sapiensAsp
AAGGATTTAGAAAACCTTTCATAACTTTGCAAGTAAATATAGCTAAATCTATATATCTTA
>Homo sapiensCys
ASCTCCAGGAGTATTTTCATATGATTCGAATTCGAAAGCAGAGCTTCAACCTGCCGGGGCTT
>Homo sapiensGln
TAGGATGGGGTGTGATAGTGGCCAGGAGAAATTTGGATTCAGGATGGGTGGATTCATAGTCTTAG
>Homo sapiensGlu
GTCCTGTAGTGGAAACACAGCATGGTTTTTCATATCATTTGGTGTGGTGTGATGCTGCGAGATA
>Homo sapiensGly
ACTCTTTTAGTATAAATAGTACCTTAACCTCAATTAAGTATTTGACAACTCAAAAAGAGTA
>Homo sapiensHis
GTAATATAGTTTAAACAAACGAGTATTTCGACTCATTAAATATGATAATCATATTACCAA
>Homo sapiensIso
AGAAATATGCTGTGAAAAGAGTACTTTGATAGAGTAAATAATAGGAGCTTAAACCCCTATTCTA
>Homo sapiensLeuCun
ACTTTAAAGGATACAGCTATCCATGGTCTGACCCGCAAAAATTTGGTGAACCTCAAATAAAGTA
    
```

## Alignement ClustalW

CLUSTAL 2.0.1 multiple sequence alignment

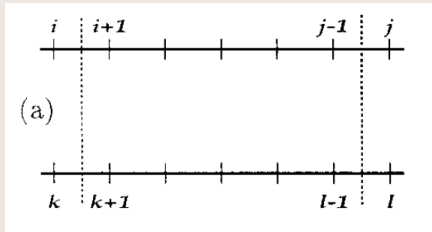
```

Homo1  ---TGGTATATAGTTTAAACAAACGAGTATTTCGACTCATTAAATATGATAATCATATA  57
Homo7  --ACTCTTTTAGTATAAATAGTACCGTAACTCCCAATTAAGTATTGACAAAC---A  54
Homo3  ---AAGTATTAGAAAACCAATTCATACCTTGTCAAGTTTAA--ATTAAAGCTTAA--  53
Homo8  --GTAATAT-AGTTTAAACAAACATCAAGTTGATGATCTGAC-AACGAGGCTTACG  55
Homo9  ----AGAATATGCTGATAAAGAGTAACTTTGATAGAGTAAATAATAGGCTTAA  55
Homo6  --GTCTCTGTAGTGAATAACAACGATGGTTTTTCA--TATCATTGGTGGTGGTGTAG  55
Homo5  TAGGATGGGGTGTGATAGTGGCCAGGAGAAATTTGATTCAGAGGAT---GGTTCGA  57
Homo2  TAGATTAAGCCAGTGAATAGGGTGTCTAGCTCTTAA--CTAAGTGTGGGGTTAA  58
Homo10 ----ACTTTAAAGGATAACAGCTATCCATGGTCTTAGGCCCAAAAATTTGGTGAAC  57
Homo4  ----AGCTCCGAGGTGATTTTCATATGATTCGAATTCGAAAGC---AGCTCAA  52
    
```



$$D_{ij,kl} = \max$$

$$D_{(i+1)(j-1),(k+1)(l-1)} + S_{ij,kl}$$

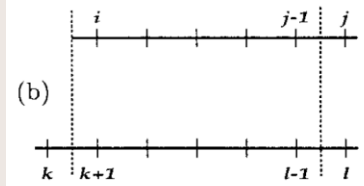


$$D_{i(j-1),(k+1)(l-1)} + S_{-j,kl}$$

$$D_{(i+1)j,(k+1)(l-1)} + S_{i-,kl}$$

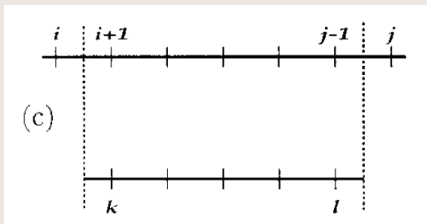
$$D_{(i+1)(j-1),k(l-1)} + S_{ij,-l}$$

$$D_{(i+1)(j-1),(k+1)l} + S_{ij,k\rightarrow}$$



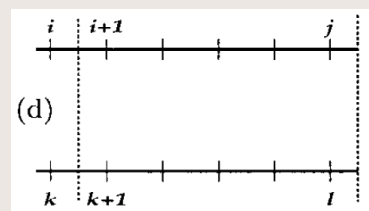
$$D_{(i+1)(j-1),kl} + S_{ij,-\rightarrow}$$

$$D_{ij,(k+1)(l-1)} + S_{-\rightarrow,kl}$$



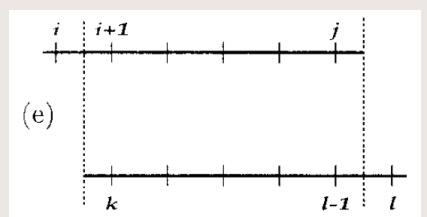
$$D_{(i+1)j,(k+1)l} + S_{i-,k\rightarrow}$$

$$D_{i(j-1),k(l-1)} + S_{-j,-l}$$



$$D_{(i+1)j,k(l-1)} + S_{i-,l}$$

$$D_{i(j-1),(k+1)l} + S_{-j,k\rightarrow}$$

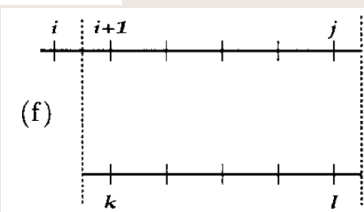


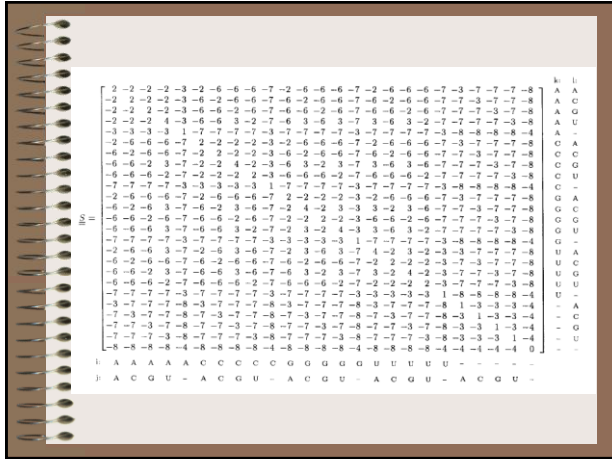
$$D_{(i+1)j,kl} + S_{i,-\rightarrow}$$

$$D_{i(j-1),kl} + S_{-j,-\rightarrow}$$

$$D_{ij,(k+1)l} + S_{-\rightarrow,k}$$

$$D_{ij,k(l-1)} + S_{-\rightarrow,l}$$





## Complexité

- Exponentielle en fonction du nombre de séquences.
- Sankoff :  $n^{3k}$  pour k séquences
- Foldalign :  $n^4$  pour deux séquences.

Heuristique « à la Clustal » pour plus de 2 séquences

## Une approche heuristique : Carnac

- Recherche des tiges-boucles candidates dans chaque séquence, indépendamment
- Recherche de « points d'ancrage » : régions très conservées entre les 2 séquences
- Sélection des tiges « alignables »
- Repliement simultané « à la Sankoff » de chaque paire de tiges alignables

## Application : tRNA Alanine

```

>Artibeus jamaicensis
AAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Balaenoptera musculus
GAGGATTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Bos taurus
GAGGATTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Canis familiaris
GAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Ceratotherium simus
GAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Dasypus novemcinctus
GAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Equus asinus
AAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Erinaceus europaeus
GAGGATTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Felis catus
GAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Hippopotamus amphibius
AAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Homo sapiens
AAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
    
```

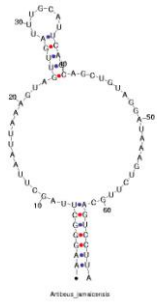
## Foldalign : pas plus de 2 séquences

```

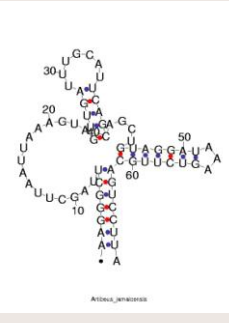
>Artibeus jamaicensis
AAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
>Balaenoptera musculus
GAGGATTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
Artibeus
AAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
Structure
(((((((.....)))))).....((((.....)))))).....((((.....)))))).....
Balaenoptera
AAGGCTTAGCTTAATTAAGTAGTTGATTTCAGCAGCTGTAGGATAAAGCTTCGACGCTCTTA
    
```

## Carnac - 3 séquences

## Carnac - 6 séquences



## Carnac - 11 séquences

Application : tRNA *H.sapiens* Arg

```
>Homo sapiens Arg, True Structure
TGGTATATAGTTTAAACAAAACGAATGATTTGACTCATTAAATATGATAATCATATTTACCAA
(((((((.....))))).((((.....))))).((((.....))))).((((.....))))).
>Homo sapiensArg
TGGTATATAGTTTAAACAAAACGAATGATTTGACTCATTAAATATGATAATCATATTTACCAA
>Homo sapiensAsn
TAGATTGAAAGGAGTGTATAGGGTCTAGCTGTTAACTAAGTGTGGGGTTAAAGTCCCATGGCTAG
>Homo sapiensAsp
AAGGTATAGAAAACCATTTCAATACTTTGCAAGTAAATATAGCTAAATCTATATATCTTA
>Homo sapiensCys
ACCTCCAGGGTATTTTCATATTGAATTGCMAATTCGAAGAGCAGCTTCAAACCTCCGGGGCTT
>Homo sapiensGln
TAGGATGGGGTGTGATAGTGGCCAGGAGAATTTGGATTCTCAGGGATGGGTTCGATTCATAGTCCTAG
>Homo sapiensGlu
GTCTCTGATGTTGAAATACACAGGATGGTTTTTCATATCAATGGTCGGTGTAGTCCGTCGAGATA
>Homo sapiensGly
ACTCTTTAGTATAAATAGTACCGTTAACTTCCAATTAAC TAGTTTTGACAAATCAAAAAGAGTA
>Homo sapiensHis
GTAATATAGTTTAAACAAAACATCAGATTGTGAATCTGACACAGAGGCTTACGACCCCTATTTFACC
>Homo sapiensIle
AGAAATATGCTGTAAAGAGATTACTTTGATAGATAAATAATAGGACTTAAACCCCTATTTFCTA
>Homo sapiensLeuCun
ACTTTTAAAGGATAACAGCTATCCATTGGTCTTAGGCCCAAAAATTTTGGTCAACTCCAATAAAAGTA
```

## Carnac - 10 séquences

