

Prédiction de structures d'ARN

Informatique Génomique - Master 1

Guillaume Blin

IGM-LabInfo UMR 8049,
Bureau 4B066
Université de Marne La Vallée
gblin@univ-mlv.fr
<http://igm.univ-mlv.fr/~gblin>

2007-08

Plan

Structure des ARN

L'approche thermodynamique

L'approche comparative

Approche hybride

Bibliographie

Plan

Structure des ARN

L'approche thermodynamique

L'approche comparative

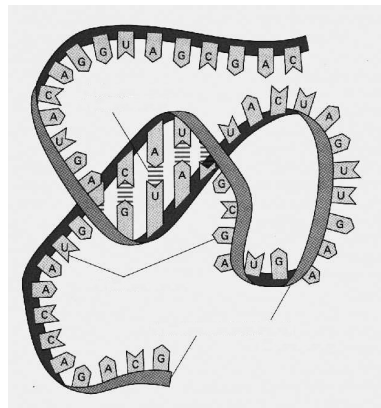
Approche hybride

Bibliographie

Rappels sur l'ARN

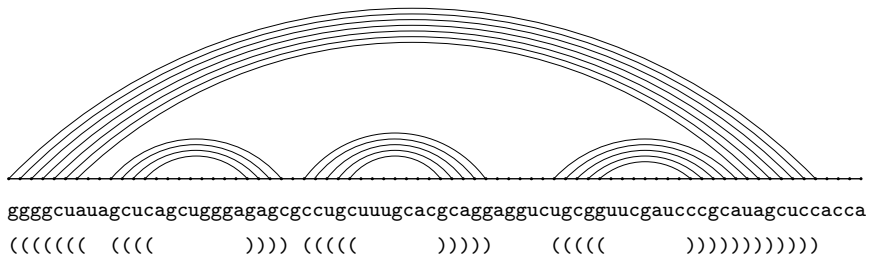
Acide RiboNucléique

- ▶ Séquence : mot sur $\{A, U, C, G\}$ orienté de 5' en 3'
- ▶ Structure : formation de liaisons hydrogènes entre deux nucléotides
 - ▶ Watson-Crick : A-U, C-G
 - ▶ Plus faible : G-U, U-C, G-A ,
...
- ▶ Pas de croisement entre les appariements
- ▶ La structure est fonctionnellement importante



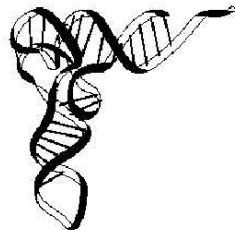
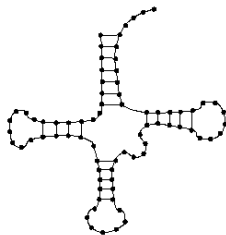
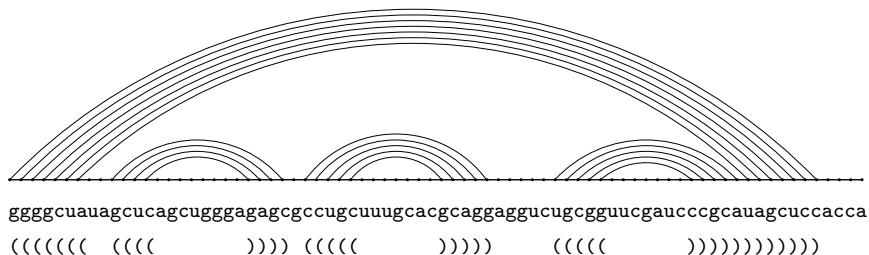
Rappels sur l'ARN

L'ARN de transfert (alanine - E. Coli)



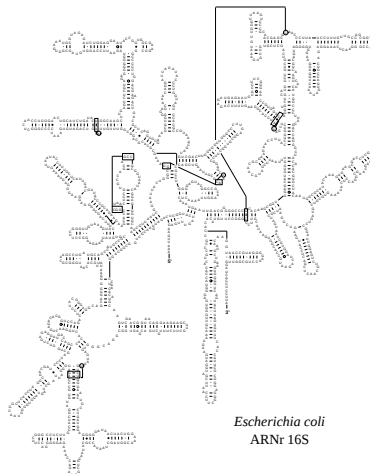
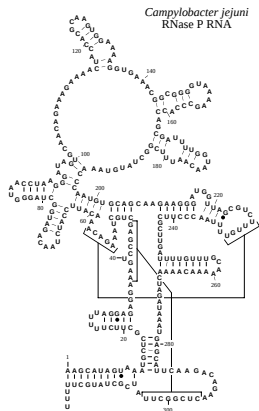
Rappels sur l'ARN

L'ARN de transfert (alanine - E. Coli)



Rappels sur l'ARN

Représentation usuelle



Comment déterminer la structure d'une molécule ?

Les approches

- ▶ Structure primaire : séquençage
- ▶ Structure secondaire et tertiaire
 - ▶ Expérimentalement : cristallographie par diffraction à rayons X, résonance magnétique nucléaire (RMN) – Long, difficile et coûteux
 - ▶ Par bio-informatique : algorithmes de prédiction de structures secondaires
 - ▶ *Méthode 1* : approche thermodynamique
 - ▶ *Méthode 2* : approche comparative



Plan

Structure des ARN

L'approche thermodynamique

L'approche comparative

Approche hybride

Bibliographie

Comment déterminer la structure d'une molécule ?

Approche thermodynamique

- ▶ Trois hypothèses :
 - ▶ À chaque configuration de la molécule correspond une quantité d'énergie libre.
 - ▶ La configuration la plus stable est celle qui minimise l'énergie libre.
 - ▶ La molécule, en se repliant, adopte la configuration la plus stable.
- ▶ On s'est ramené à un problème combinatoire : trouver la structure dont l'énergie est optimale.



Comment déterminer la structure d'une molécule ?

Modèle initial (Nussinov - 1978)

- ▶ L'énergie de la molécule est la somme des énergies de chaque paire de bases.
- ▶ $\alpha(r_i, r_j)$: énergie libre de l'appariement (r_i, r_j)

$$\alpha(r_i, r_j) < 0 \quad \text{si } j - i > 3 \text{ et } r_i \leftrightarrow r_j$$

$$\alpha(r_i, r_j) = 0 \quad \text{si } i = j$$

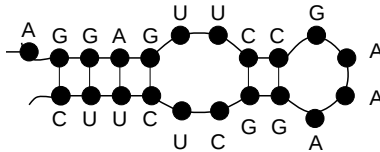
$$\alpha(r_i, r_j) = +\infty \quad \text{sinon}$$

- ▶ Énergie libre de la structure secondaire $\mathcal{S}$

$$E(\mathcal{S}) = \sum_{(r_i, r_j) \in \mathcal{S}} \alpha(r_i, r_j)$$

Comment déterminer la structure d'une molécule ?

Exemple



Fonction d'énergie :

$$\alpha(A, U) = -2$$

$$\alpha(C, G) = -3$$

$$\alpha(G, U) = -1$$

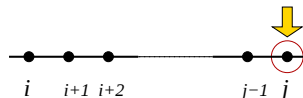
Energie totale : -15

- Comment calculer la structure optimale ? par programmation dynamique

Comment déterminer la structure d'une molécule ?

Programmation dynamique - Nussinov'78

- Etape 1 : décomposition du problème en instances plus petites



- $S_{i,j}$: structure secondaire optimale pour la sous-séquence $i \dots j$
- Trois possibilités pour r_j :
 1. r_j ne s'apparie pas avec la sous-chaîne $r_i \dots r_{j-1}$:
 $E(S_{i,j}) = E(S_{i,j-1})$
 2. r_j s'apparie avec r_i : $E(S_{i,j}) = E(S_{i+1,j-1}) + \alpha(r_i, r_j)$
 3. r_j s'apparie avec r_k pour un $i < k < j$:

$$E(S_{i,j}) = \min\{E(S_{i,k-1}) + \alpha(r_k, r_j) + E(S_{k+1,j-1}), k \in]i, j[\}$$

- $E(S_{i,j}) = \min$ cas 1, 2 et 3

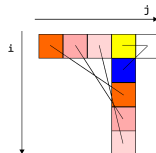
Comment déterminer la structure d'une molécule ?

Programmation dynamique - Nussinov'78

► Etape 2 : construction de la table de prog. dynamique

- Une table T , de dimension 2 : $T(i, j) := E(S_{i,j})$

$$T(i, j) = \min \begin{cases} T(i, j - 1) \\ T(i + 1, j - 1) + \alpha(r_i, r_j) \\ \min\{T(i, k - 1) + \alpha(r_k, r_j) + T(k + 1, j - 1)\} \end{cases}$$



- Une table S qui stocke le devenir de r_j

► Etape 3 : construction de la structure secondaire optimale, par retour arrière

► Complexité

- Chaque case de la table nécessite $O(n)$ calculs.
- Complexité globale en $O(n^3)$.

Comment déterminer la structure d'une molécule ?

	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
C	0	0	0	0														
G		0	0	0	0													
G			0	0	0	0												
A				0	0	0	0											
U					0	0	0	0										
A						0	0	0	0									
C							0	0	0	0								
U								0	0	0	0							
U									0	0	0	0						
C										0	0	0	0					
U											0	0	0	0				
U												0	0	0	0			
A													0	0	0	0		
G														0	0	0	0	
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0

- Paire W-C = +3, Wobble +1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	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	0	0	0	0	0	0	0	0	0	0	0	0
G		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G			0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
A				0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
U					0	0	0	0	0	0	0	0	0	0	0	0	0	0
A						0	0	0	0	0	0	0	0	0	0	0	0	0
C							0	0	0	0	0	0	0	0	0	0	0	0
U								0	0	0	0	0	0	0	0	0	0	0
U									0	0	0	0	2	0	0	0	0	0
C										0	0	0	0	0	0	0	0	0
U											0	0	0	0	0	0	0	0
A												0	0	0	0	0	0	0
G													0	0	0	0	0	0
A														0	0	0	0	0
C															0	0	0	0
U																0	0	0
A																	0	0
G																		0
A																		

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	(	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	)	.
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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	(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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	(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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	(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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	(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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

	(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	7	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	1
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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

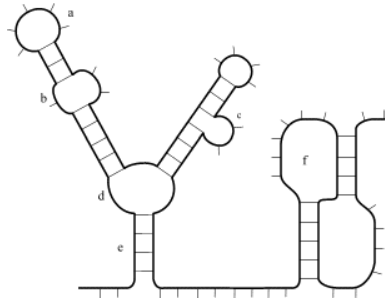
	(C	(G	(G	. A	. U	. A	) C	(U	. U	(C	. U	. U	. A	) G	) A	) C	) G	. A
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	7	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0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

- G-C=3, A-U=2, G-U=1
- Pas de lien entre i et $j < i + 4$

Comment déterminer la structure d'une molécule ?

Mfold - Zuker, Turner *et al.*

- Paramètres d'énergie
 - liaison hydrogène
 - énergie d'empilement
- Motifs structuraux



- a. épingle à cheveux (*hairpin*)
- b. boucle interne (*internal loop*)
- c. renflement (*bulge loop*)
- d. jonction
- e. tige (*duplex*)
- f. Exclut : pseudo-nœud (*pseudoknot*)

Comment déterminer la structure d'une molécule ?

Limites de l'approche thermodynamique

- ▶ Pertinence de la définition de la fonction d'énergie.
Solution : transformer l'algorithme pour obtenir un ensemble de configurations sous-optimales
- ▶ Les hypothèses biologiques ne sont pas toutes valides :
Il existe des molécules d'ARN qui se replient en formant des nœuds, ou dans lesquelles un appariement regroupent 3 nucléotides.
Solution : complexifier les algorithmes
- ▶ Pas de prise en compte d'interactions avec des molécules voisines.
- ▶ Pas de prise en compte du sens de la synthèse de l'ARN.

Plan

Structure des ARN

L'approche thermodynamique

L'approche comparative

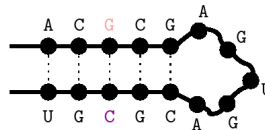
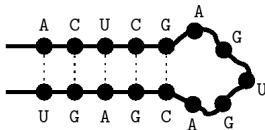
Approche hybride

Bibliographie

Comment déterminer la structure d'une molécule ?

Approche comparative

- ▶ On dispose d'une famille de séquences homologues
- ▶ Lors de l'évolution, la structure est mieux préservée que la séquence
- ▶ Les séquences homologues partagent la même structure
- ▶ Phénomène de changement de base compensatoire :
Quand une base impliquée dans un appariement mute, la base complémentaire mute également, pour préserver la structure



Comment déterminer la structure d'une molécule ?

La méthode comparative

► Jeu de séquences

G	A	G	C	C	C	A	G	U	U	C
	A	G	G	A	C	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U

Comment déterminer la structure d'une molécule ?

La méthode comparative

► Jeu de séquences

G	A	G	C	C	C	A	G	U	U	C
	A	G	G	A	C	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U

► Etape 1 : construction d'un alignement multiple

G	A	G	C	—	C	C	A	G	U	U	C
—	A	G	G	A	C	—	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U	—
—	A	G	G	A	C	—	U	C	U	U	C

Comment déterminer la structure d'une molécule ?

La méthode comparative

► Jeu de séquences

G	A	G	C	C	C	A	G	U	U	C
	A	G	G	A	C	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U

► Etape 1 : construction d'un alignement multiple

G	A	G	C	—	C	C	A	G	U	U	C
—	A	G	G	A	C	—	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U	—
—	A	G	G	A	C	—	U	C	U	U	C

► Etape 2 : détection des positions corrélées

Comment déterminer la structure d'une molécule ?

La méthode comparative

► Jeu de séquences

G	A	G	C	C	C	A	G	U	U	C
	A	G	G	A	C	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U

► Etape 1 : construction d'un alignement multiple

G	A	G	C	—	C	C	A	G	U	U	C
—	A	G	G	A	C	—	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U	—
—	A	G	G	A	C	—	U	C	U	U	C

► Etape 2 : détection des positions corrélées

Comment déterminer la structure d'une molécule ?

La méthode comparative

► Jeu de séquences

G	A	G	C	C	C	A	G	U	U	C
	A	G	G	A	C	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U

► Etape 1 : construction d'un alignement multiple

G	A	G	C	—	C	C	A	G	U	U	C
—	A	G	G	A	C	—	U	C	U	U	C
A	A	U	C	A	C	C	C	G	A	U	—
—	A	G	G	A	C	—	U	C	U	U	C

► Etape 2 : détection des positions corrélées

Comment déterminer la structure d'une molécule ?

Mesurer la corrélation entre deux colonnes

- $\mathcal{I}(i, j)$: information mutuelle des colonnes i et j de l'alignement multiple

$$\mathcal{I}(i, j) = \sum_{x, y=A, C, G, T} f_{x, y}^{i, j} \log_2 \frac{f_{x, y}^{i, j}}{f_x^i f_y^j}$$

- f_x^i fréquence de la base x dans la colonne i
- $f_{x, y}^{i, j}$ fréquence du couple x, y dans les colonnes i et j

$\mathcal{I}(i, j)$ varie entre 0 et 2 bits.

Quantité d'information révélée par la colonne j , la colonne i étant connue. $\mathcal{I}(i, j)$ est maximale quand i et j individuellement paraissent aléatoires ($f_x^i = f_y^j = 0.25$) et que i et j sont parfaitement corrélées.

Comment déterminer la structure d'une molécule ?

Exemples

. . A . . . U . .

. . A . . . C . .

. . A . . . U . .

. . C . . . G . .

$$\mathcal{I} = \frac{\log_2(4/3)}{2} + \frac{\log_2(4/3)}{4} + \frac{\log_2(4)}{4}$$

. . A . . . U . .

. . U . . . A . .

. . C . . . G . .

. . G . . . C . .

$$\mathcal{I} = \frac{\log_2(4)}{4} + \frac{\log_2(4)}{4} + \frac{\log_2(4)}{4} + \frac{\log_2(4)}{4} = 2$$

La corrélation entre les deux colonnes est maximale.

Comment déterminer la structure d'une molécule ?

Exemple de l'ARN de transfert

```

GGGGAAATTAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGGTCAGCGGTTTCGATCCCGCTATTCTCCA---
GGGGCTATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTTCGATCCCGCATAGCTCCACCA
GGGGCTATAGCTCAGCT-GGGAGAGCGCTTGATGGCATGCAAGAGGTTCAGCGGTTTCGATCCCGCTTAGCTCCACCA
GGGGAAATTAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGGTCAGCGGTTTCGATCCCGCTATTCTCCA---
GGGGCCTTAGCTCAGTC-GGTAGAGCACTGCCTTTGCAAGGCAGATGTCAGGGGTTTCGATTCCCTTAGGCTCCA---
GGGGGTATAGCTCAGTT-GGTAGAGCGCTGCCTTTGCAAGGCAGAAGTCAGCGGTTTCGA.TCCGCTTACCCCCA---
GGGGCTATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTTCGATCCCGCATAGCTCCACCA
GGGGCTATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTTCGATCCCGCATAGCTCCACCA
GGGGGCATAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGT-CGTCGGTTTCGATCCCGTCTGCTCCACCA
GGGGCCATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTTCAGGAGTTTCGATCCTCCTTGCTCCACCA
GGGGGCATAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGT-CGTCGGTTTCGATCCCGTCTGCTCCACCA
GGGGCCATAGCTCAGCTGGGGAGAGCGCCTGCCTTGACGCAGGAGGTCAACGGTTTCGATCCCGTTTGCTCCA---
GGGGGCATAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGT-CGTCGGTTTCGATCCCGTCTGCTCCACCA
GGGGCATTAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTTCAGCGGTTTCGATCCCGCTATTCTCCACCA
GGGGCCATAGCTCAGTT-GGTAGAGCGCCTGCTTTGCAAGCAGGTGT-CGTCGGTTTCGAATCCGTCTGCTCCACCA
GGGGCCGTAGCTCAGCTGGG-AGAGCACCTGCTTTGCAAGCAGGGGGTCGGAGGTTTCGATCCCGTCCGGCTCCACCA
GGGGCCGTAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGGTCGTCGGTTTCGATCCCGTCCGGCTCCACCA
GGGGCCGTAGCTCAGCT-GG-AGAGCACCTGCTTTGCAAGCAGGGGGTCGTCGGTTTCGATCCCGTCCGGCTCCACCA

```

Comment déterminer la structure d'une molécule ?

Exemple de l'ARN de transfert

```

GGGGAA1TTAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGGTCAGCGGTTTCGATCCCGCTAT2TTCTCCA---
GGGGCTATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTTCGATCCCGCATAGCTCCACCA
GGGGCTATAGCTCAGCT-GGGAGAGCGCTTGATGGCATGCAAGAGGTTCAGCGGTTTCGATCCCGCTTAGCTCCACCA
GGGGAA1TTAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGGTCAGCGGTTTCGATCCCGCTAT2TTCTCCA---
GGGGCCTTAGCTCAGTC-GGTAGAGCACTGCCTTTGCAAGGCAGATGTCAGGGGTTTCGATTCCCTTAGGCTCCA---
GGGGGTATAGCTCAGTT-GGTAGAGCGCTGCCTTTGCAAGGCAGAAGTCAGCGGTTTCGA.TCCGCTTACCCCCA---
GGGGCTATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTTCGATCCCGCATAGCTCCACCA
GGGGCTATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTCTGCGGTTTCGATCCCGCATAGCTCCACCA
GGGGGCATAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGT-CGTCGGTTTCGATCCCGTCTGGCTCCACCA
GGGGCCATAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTTCAGGAGTTTCGATCCTCCTTGGCTCCACCA
GGGGGCATAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGT-CGTCGGTTTCGATCCCGTCTGGCTCCACCA
GGGGCCATAGCTCAGCTGGGGAGAGCGCCTGCCTTGACGCAGGAGGTCAACGGTTTCGATCCCGTTTGGCTCCA---
GGGGGCATAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGT-CGTCGGTTTCGATCCCGTCTGGCTCCACCA
GGGGCA1TTAGCTCAGCT-GGGAGAGCGCCTGCTTTGCACGCAGGAGGTTCAGCGGTTTCGATCCCGCTAT2TTCTCCACCA
GGGGCCATAGCTCAGTT-GGTAGAGCGCCTGCTTTGCAAGCAGGTGT-CGTCGGTTTCGAATCCGTCTGGCTCCACCA
GGGGCCGTAGCTCAGCTGGG-AGAGCACCTGCTTTGCAAGCAGGGGGTCGGAGGTTTCGATCCCGTCCGGCTCCACCA
GGGGCCGTAGCTCAGCT-GGGAGAGCACCTGCTTTGCAAGCAGGGGGTCGTCGGTTTCGATCCCGTCCGGCTCCACCA
GGGGCCGTAGCTCAGCT-GG-AGAGCACCTGCTTTGCAAGCAGGGGGTCGTCGGTTTCGATCCCGTCCGGCTCCACCA

```

Comment déterminer la structure d'une molécule ?

Exemple de l'ARN de transfert

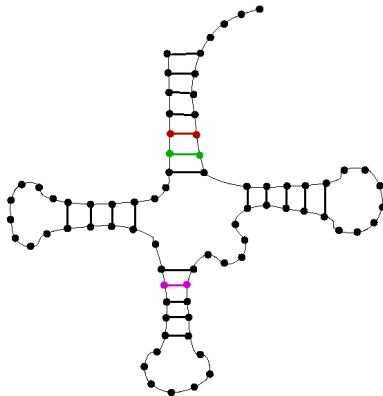
```

GGGGAA1TTAGCTCAGCT-GGGAGAGCAC2CTGCTTTGCAAGCA3GGGGGTCAGCGGTTTCGATCCCGCTA4TTCTCCA---
GGGG5CTATAGCTCAGCT-GGGAGAGCGC6CTGCTTTGCACGCA7GGAGGTCTGCGGTTTCGATCCCGCAT8AGCTCCACCA
GGGG9CTATAGCTCAGCT-GGGAGAGCGC10TTGCATGGCATGCA11AGAGGTTCAGCGGTTTCGATCCCGCTT12AGCTCCACCA
GGGGAA13TTAGCTCAGCT-GGGAGAGCAC14CTGCTTTGCAAGCA15GGGGGTCAGCGGTTTCGATCCCGCTA16TTCTCCA---
GGGG17CCTTAGCTCAGTC-GGTAGAGCAC18TGCCTTTGCAAGGC19AGATGTCAGGGGTTTCGATTCCCTTAG20GGCTCCA---
GGGG21GTATAGCTCAGTT-GGTAGAGCGC22TGCCTTTGCAAGGC23AGAAGTCAGCGGTTTCGA.TCCGCTT24ACCCCCA---
GGGG25CTATAGCTCAGCT-GGGAGAGCGC26CTGCTTTGCACGCA27GGAGGTCTGCGGTTTCGATCCCGCAT28AGCTCCACCA
GGGG29CTATAGCTCAGCT-GGGAGAGCGC30CTGCTTTGCACGCA31GGAGGTCTGCGGTTTCGATCCCGCAT32AGCTCCACCA
GGGG33G34CATAGCTCAGCT-GGGAGAGCAC35CTGCTTTGCAAGCA36GGGGT-CGTCGGTTTCGATCCCGTCT37G38CCTCCACCA
GGGG39CC40ATAGCTCAGCT-GGGAGAGCGC41CTGCTTTGCACGCA42GGAGGTTCAGGAGTTTCGATCCTCCTT43G44GCTCCACCA
GGGG45G46CATAGCTCAGCT-GGGAGAGCAC47CTGCTTTGCAAGCA48GGGGT-CGTCGGTTTCGATCCCGTCT49G50CCTCCACCA
GGGG51CC52ATAGCTCAGCTGGGGAGAGCGC53CTGCCTTGCACGCA54GGAGGTCAACGGTTTCGATCCCGTTT55G56GCTCCA---
GGGG57G58CATAGCTCAGCT-GGGAGAGCAC59CTGCTTTGCAAGCA60GGGGT-CGTCGGTTTCGATCCCGTCT61G62CCTCCACCA
GGGG63CA64TTAGCTCAGCT-GGGAGAGCGC65CTGCTTTGCACGCA66GGAGGTTCAGCGGTTTCGATCCCGCTA67TTCTCCACCA
GGGG68CC69ATAGCTCAGTT-GGTAGAGCGC70CTGCTTTGCAAGCA71GGTGT-CGTCGGTTTCGAATCCGTCT72G73GCTCCACCA
GGGG74CC75GTAGCTCAGCTGGG-AGAGCAC76CTGCTTTGCAAGCA77GGGGGTCGGAGGTTTCGATCCCGTCC78G79GCTCCACCA
GGGG80CC81GTAGCTCAGCT-GGGAGAGCAC82CTGCTTTGCAAGCA83GGGGGTCGTCGGTTTCGATCCCGTCC84G85GCTCCACCA
GGGG86CC87GTAGCTCAGCT-GG-AGAGCAC88CTGCTTTGCAAGCA89GGGGGTCGTCGGTTTCGATCCCGTCC90G91GCTCCACCA

```

Comment déterminer la structure d'une molécule ?

Vérification



Structure secondaire de l'ARNt

Plan

Structure des ARN

L'approche thermodynamique

L'approche comparative

Approche hybride

Bibliographie

caRNAC

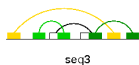
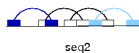
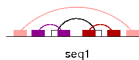
caRNAC (Perriquet Touzet - 2003)

- ▶ Trois sources d'information :
 - ▶ énergie libre
 - ▶ mutations compensatoires
 - ▶ conservation éventuelle des séquences
- ▶ Supporter des séquences et des structures
- ▶ Pas d'alignement préalable des séquences
- ▶ Permettre la gestion de longs jeux de données
- ▶ <http://bioinfo.lifl.fr/carnac>



caRNAc

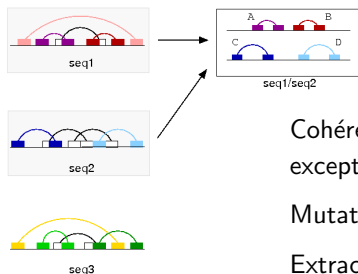
caRNAc : stratégie de prédiction



1. Identification des tiges potentielles, séquence par séquence

caRNAc

caRNAc : stratégie de prédiction



Cohérence avec les motifs
exceptionnellement conservés

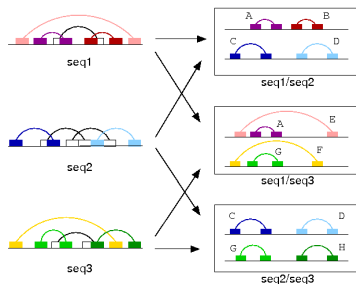
Mutations compensatoires

Extraction de la structure commune
par programmation dynamique

2. Confrontation deux à deux

caRNAc

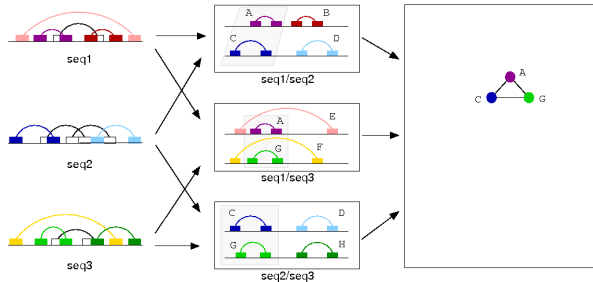
caRNAc : stratégie de prédiction



3. Combinaison des prédictions 2 à 2 : graphe des tiges

caRNAc

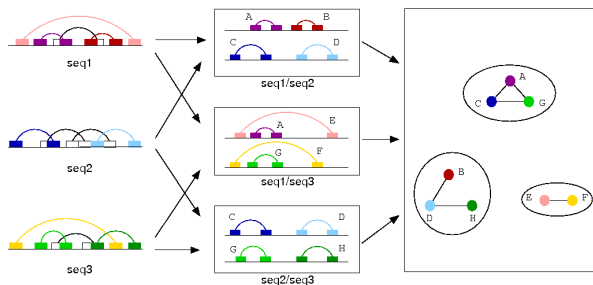
caRNAc : stratégie de prédiction



3. Combinaison des prédictions 2 à 2 : graphe des tiges

caRNAc

caRNAc : stratégie de prédiction



4. Sélection finale des tiges

Plan

Structure des ARN

L'approche thermodynamique

L'approche comparative

Approche hybride

Bibliographie

Bibliographie

Références

1. TOUZET, H AND VARRÉ, S.. *Cours*,
[http ://www2.lifl.fr/SEQUOIA/members.php](http://www2.lifl.fr/SEQUOIA/members.php).
2. TOMPA, M.. *Computational Biology*, Course in english,
[http ://www.cs.washington.edu/education/courses/527/00w](http://www.cs.washington.edu/education/courses/527/00w)
3. J. SETUBAL AND J. MEIDANIS. *Introduction to Computational Molecular Biology*. PWS Publishing Co, 1997.