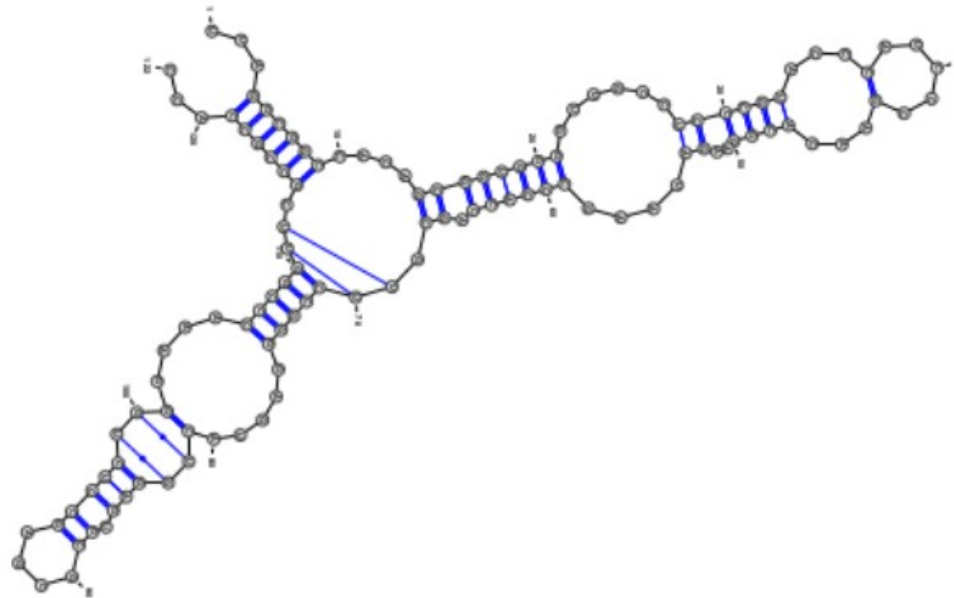


Master Bioinformatique parcours MISO 2022-23

Méthodes pour l'Analyse Bioinformatique des Séquences

# Projet : structure secondaire des ARN

## Introduction et notions élémentaires



Mathieu GENETE – [mathieu.genete@univ-lille.fr](mailto:mathieu.genete@univ-lille.fr)

# Les Acides RiboNucléiques (ARN)

- molécules composées de nucléotides **A**dénine, **C**ytosine, **G**uanine et **U**racile
- transcrits comme des copies de portions de l'ADN

## Tailles très variables :

- de 20 nucléotides (nts) à 3 000 nts environ dans la cellule
- pouvant même atteindre jusqu'à 30 000 nts pour les génomes entiers (virus)

## Grande variété de rôles joués par l'ARN au sein de la cellule :

- Médiateur de l'information génétique (ARN messagers)
- Partie-prenante de la machinerie traductionnelle (ARN ribosomaux, ARN de transfert)
- Acteur de la régulation (interférence par ARN)
- ...

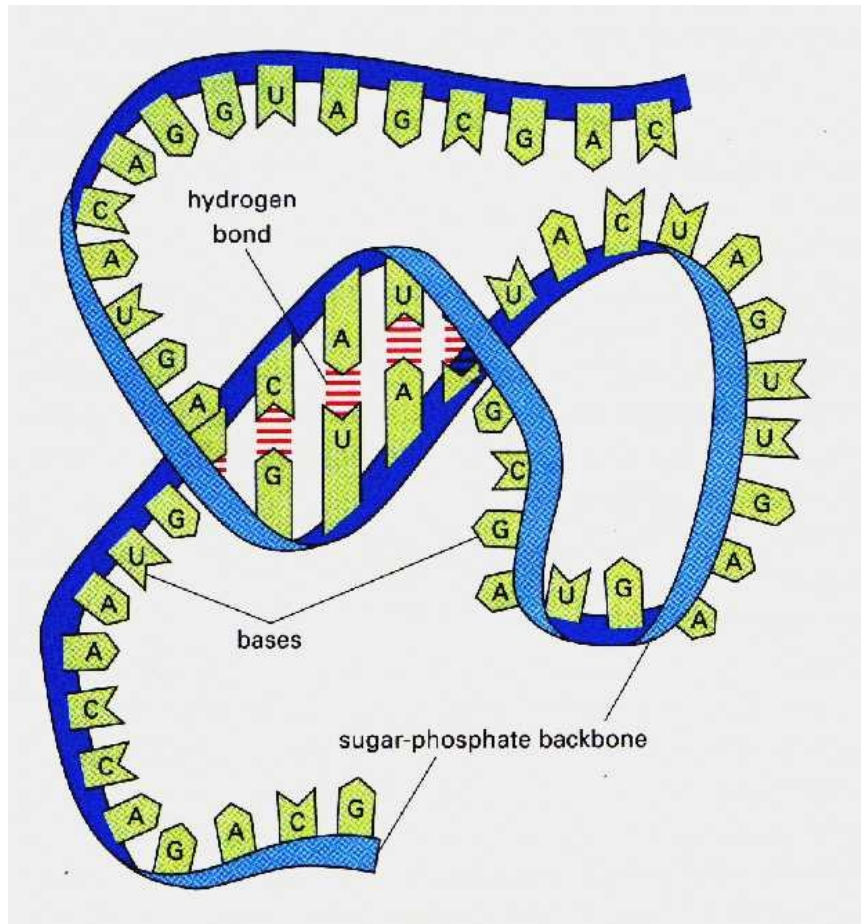
### ARN codants

information nécessaire à la synthèse d'une protéine

### ARN non-codants

se replient sur eux-mêmes pour adopter une conformation spatiale qui **détermine leur fonction**

# Structure des ARN



## Copie simple-brin

**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

faible : G-U  
U-C, G-A , . . .

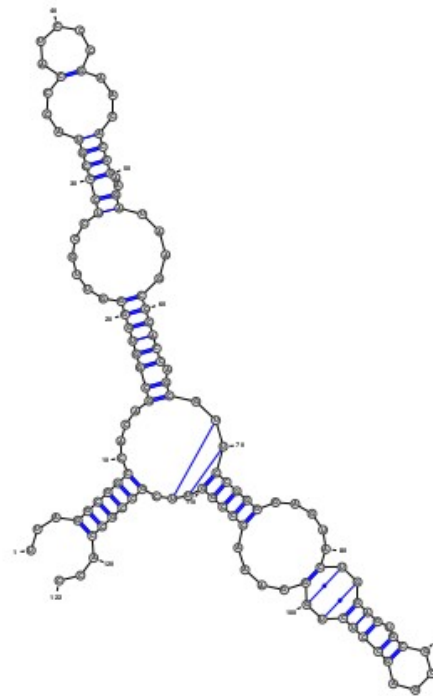
**Pas** de croisement entre les appariements

**La** structure est fonctionnellement importante

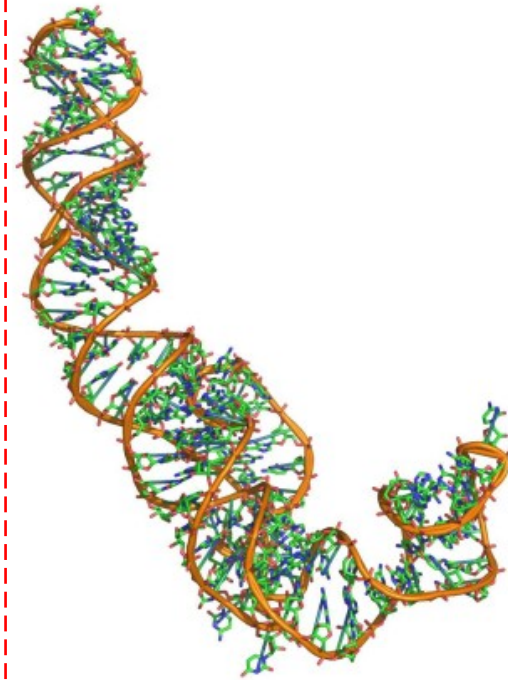
# Structure des ARN 1/4

UUAGGCGGCCACAGC  
GGUGGGGUUGCCUCC  
CGUACCCAUCCCGAA  
CACGGAAGAUAAAGCC  
CACCAGCGUUCCGGG  
GAGUACUGGAGUGCG  
CGAGCCUCUGGGAAA  
CCCGGUUCGCCGCCA  
CC

Structure Primaire  
(Séquence)



Structure Secondaire  
(Couplage partiel)

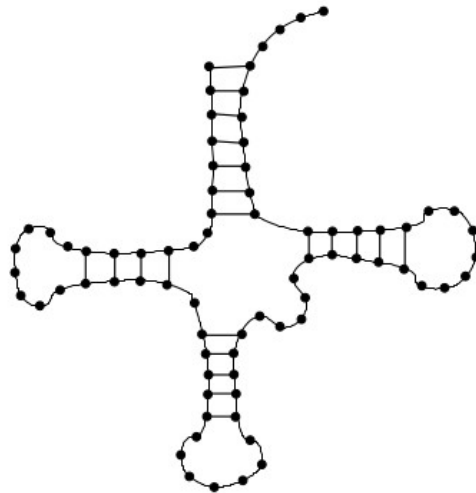
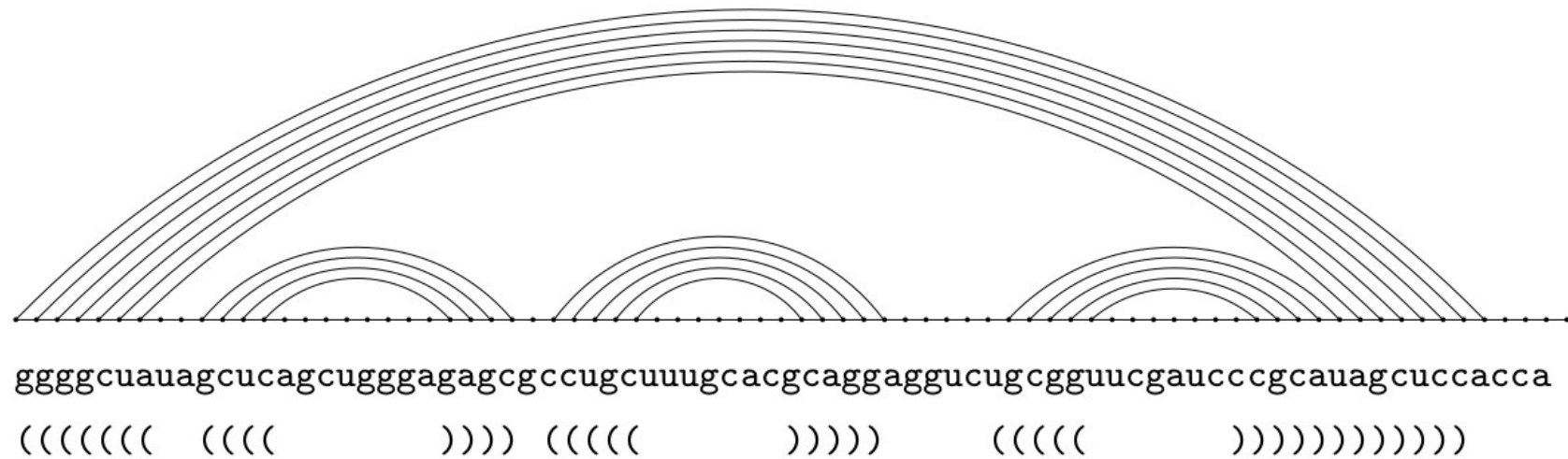


Structure Tertiaire  
(Objet 3D)

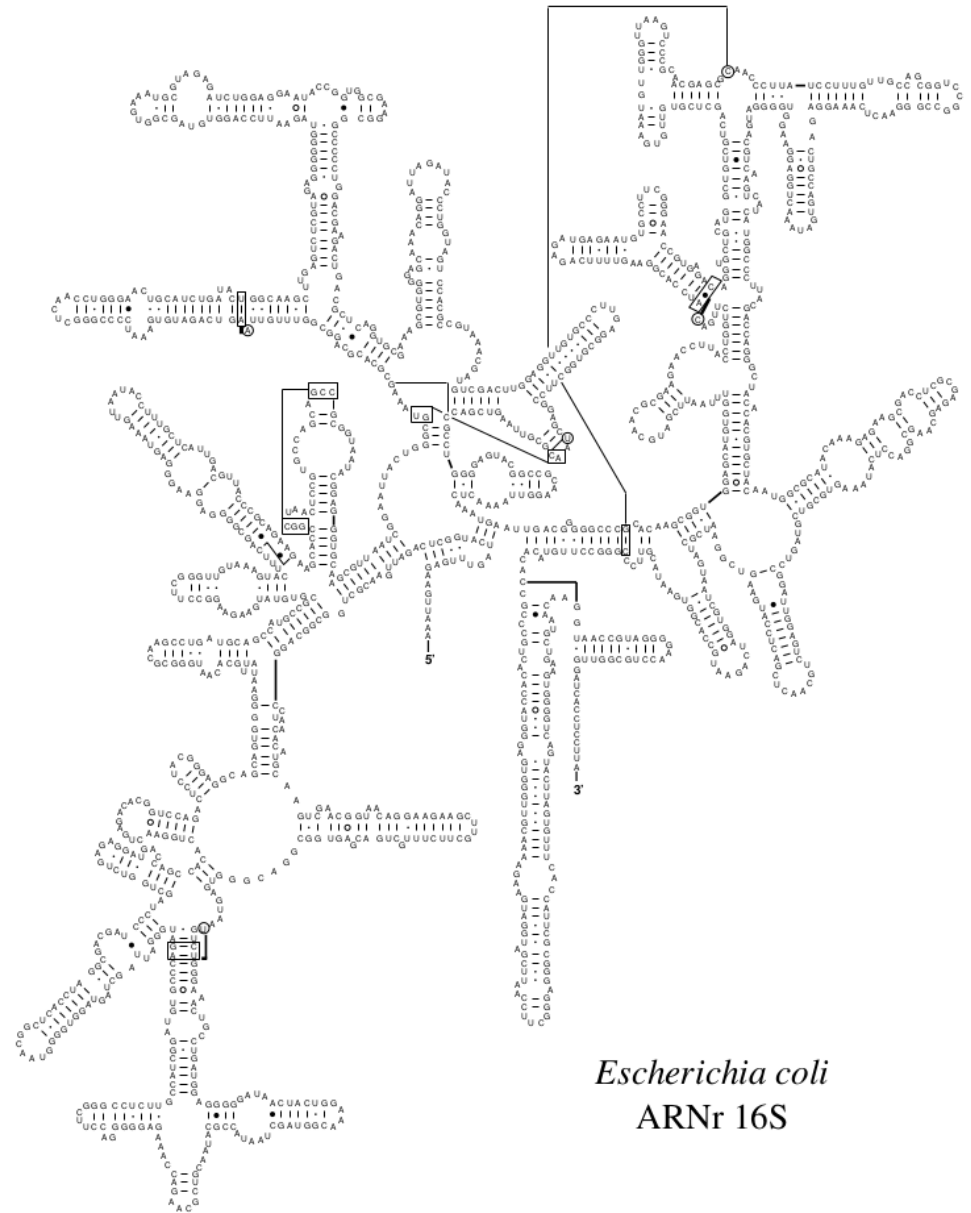
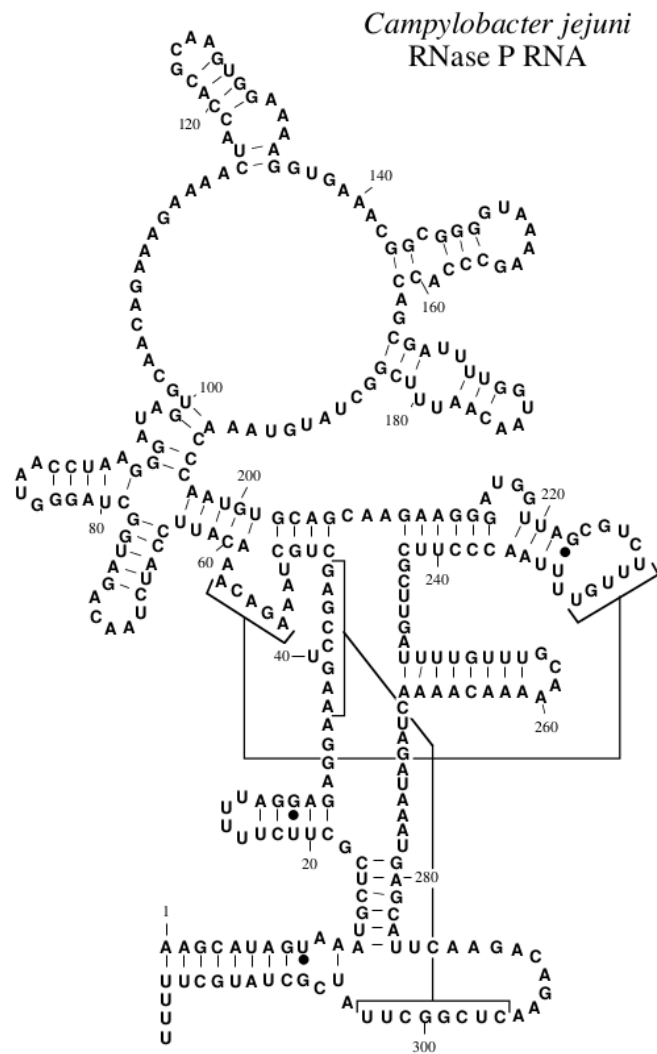
Trois principaux niveaux de représentation pour un ARN ribosomal

## Structure des ARN 2/4

### Exemple : ARN de transfert (Alanine – E. coli)



# Structure des ARN 3/4

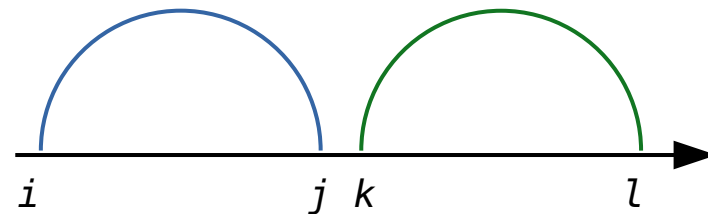
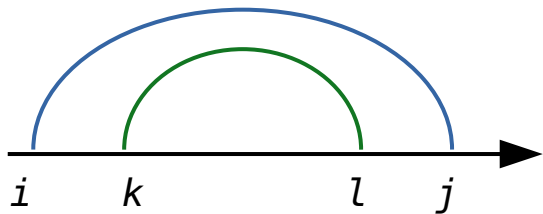


# Structure des ARN 4/4

Formellement, une **structure secondaire** est un ensemble  $S$  de paires de bases satisfaisant les contraintes suivantes :

1. **Distance minimale  $\theta$**  : Si  $(i, j) \in S$ , alors on a  $j - i > \theta$
2. **Monogamie** : Toute position est impliquée dans *au plus* une paire de  $S$
3. **Croisements interdits** : Si  $(i, j), (k, l)$  telles que  $i < k$ , alors on a

$$i < k < l < j \quad \text{ou} \quad i < j < k < l$$





# Formats de fichiers pour stocker les structures

## Format parenthésé

```
>RA7680
GGGGGCGUAGCUCAGAUgGUAGAGCGCUCGCUUgGCgUGUGAGAGGUACCGGGAUCGaUACCCGGCGCCUCCACCA
(((((((..((((.....))))).((((.....))))). .... (((((((.....))))))))))....
>sequence_test
ggggaaaccagguucguuucggucaagacaaccc
((((.....))).(((((((.....))((.....)))))).
>sequence_test2
GCAAAAAGCUUAAGGGAAAACCUCCAUAUCCCC
((((.....))..((((.....))(.))))....
>sequence_test3
CAGAGUAUGAUCACGGUUUCACCUUGGUACAGGGCGUCCACUGCACUCUG
(((((((..(((((.(((.....)).))))).((((.....)).))))).))))))
```

GGCUUUUGCAUACCCUCGG  
(((.....))..(.....)).. => **OK**

GGCUUUUGCAUACCCUCGG  
((((.....))..(.)).. => **PAS OK**

  GGCUUUUGGAUACCCUCGG  
(((.....))..(.....)).. => **PAS OK**

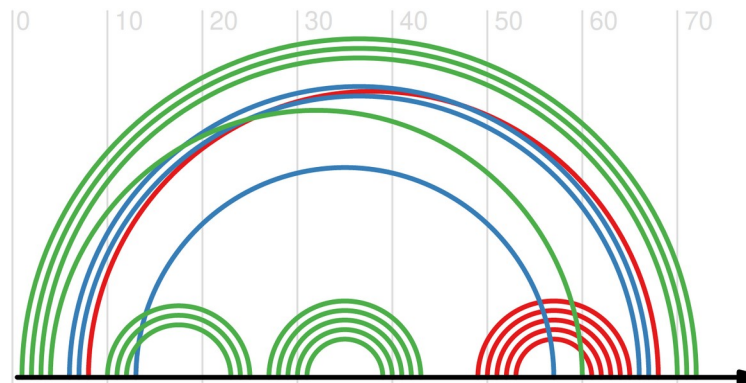
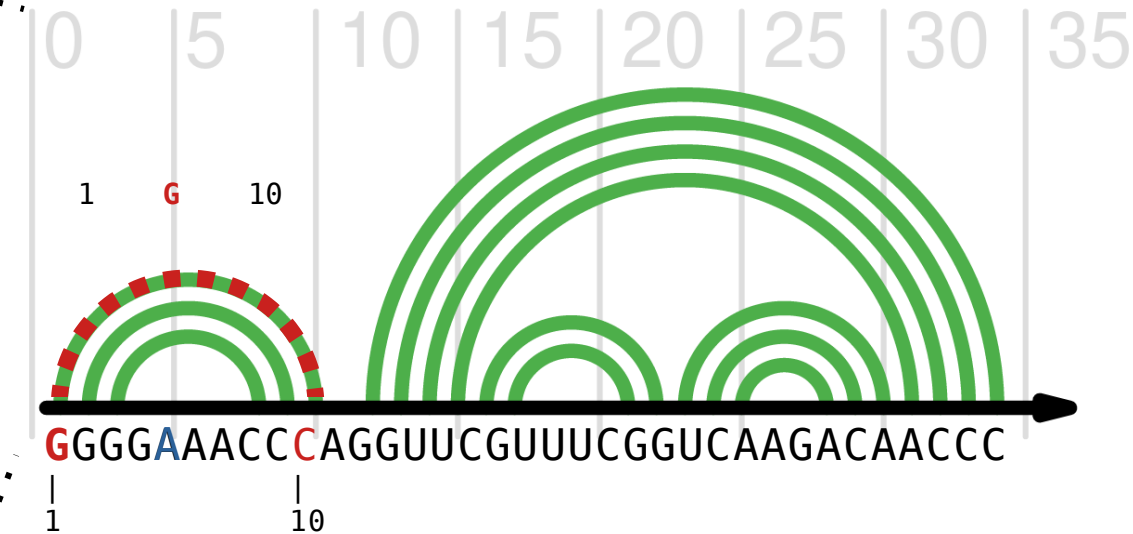
G=C  
A=U  
G=U



# Formats de fichiers pour stocker les structures

Format CT (connect)

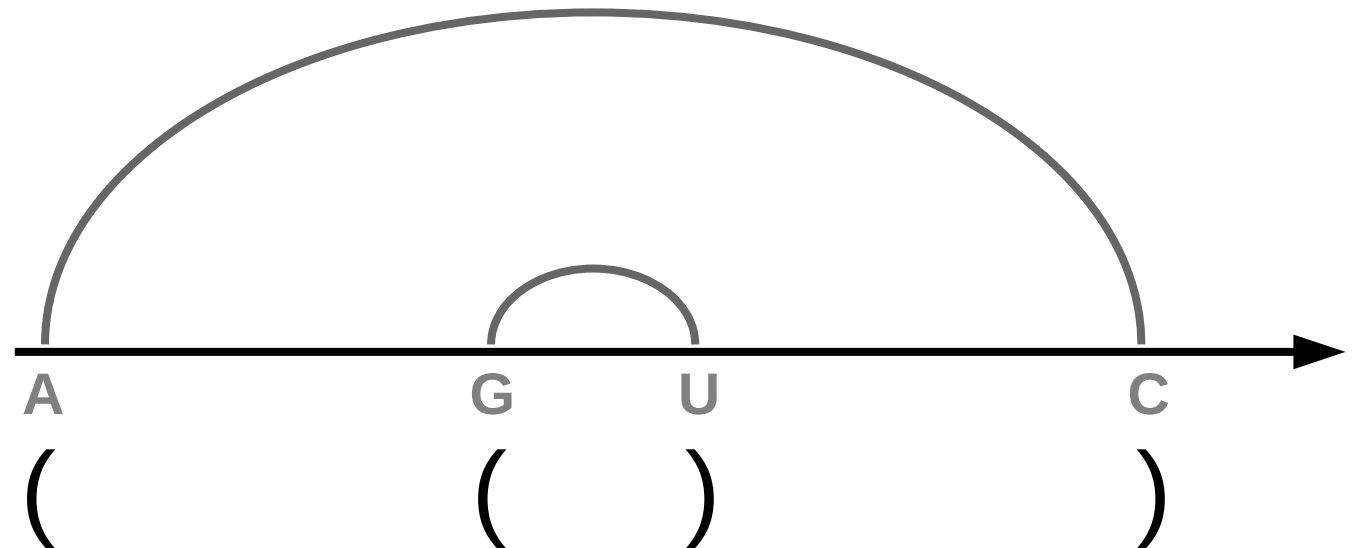
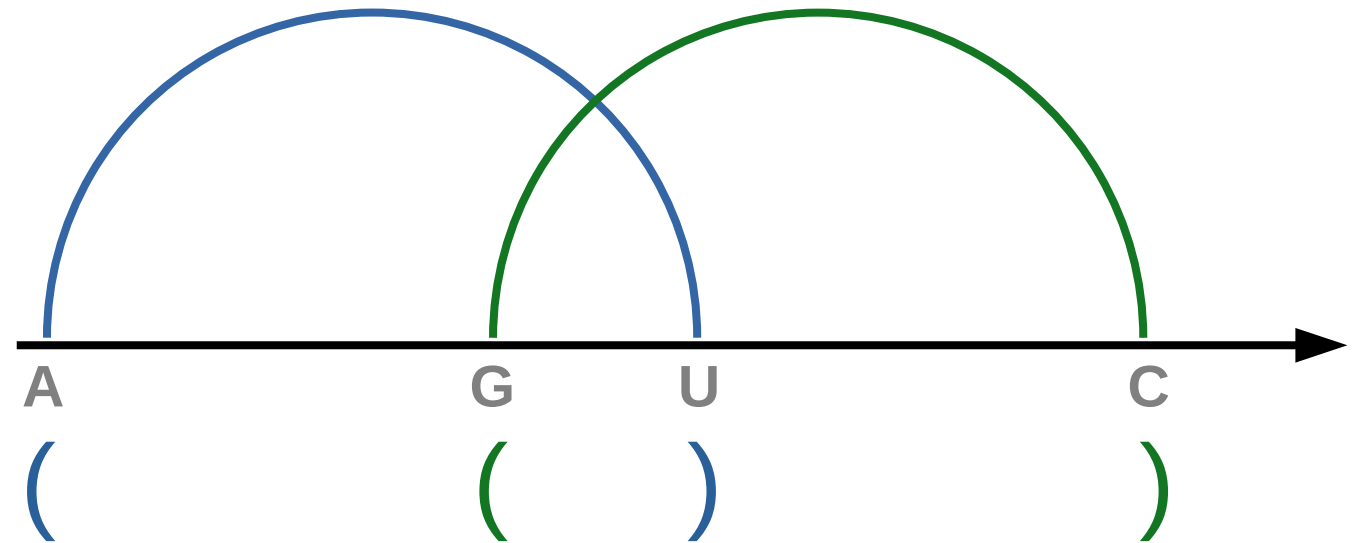
1	g	10
2	g	9
3	g	8
4	g	0
5	a	0
6	a	0
7	a	0
8	c	3
9	c	2
10	c	1
11	a	0
12	g	34
13	g	33
14	u	32
15	u	31
16	c	22
17	g	21
18	u	0
19	u	0
20	u	0
21	c	17
22	g	16
23	g	30
24	u	29
25	c	28
26	a	0
27	a	0
28	g	25
29	a	24
30	c	23
31	a	15
32	a	14
33	c	13
34	c	12
35	c	0



Croisements => **PAS OK**

Vérifier la distance minimale entre  
deux positions appariées > 3

# Formats de fichiers pour stocker les structures



# Comment déterminer la structure d'une molécule ?

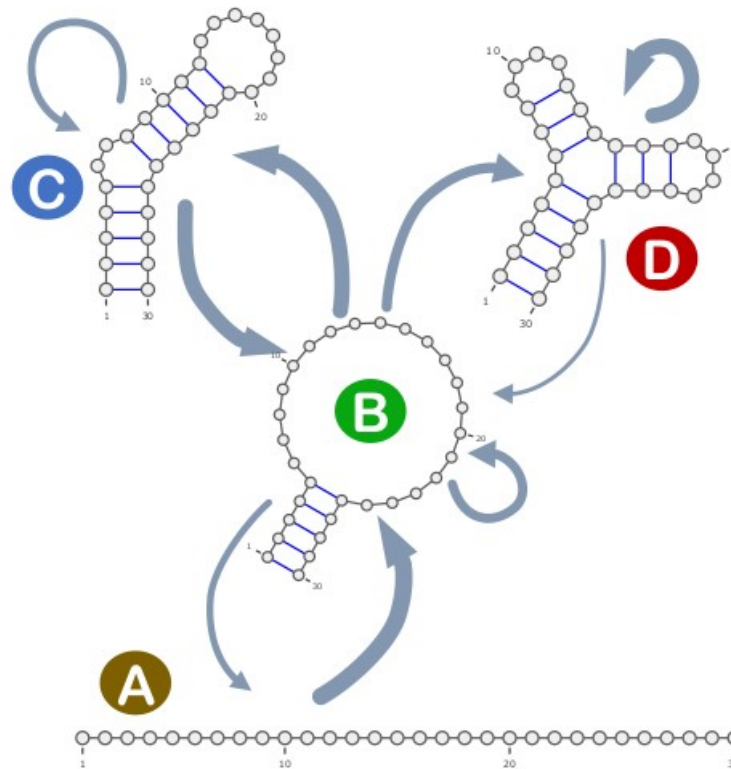
- **Structure primaire** : séquençage
- **Structure secondaire et tertiaire**
  - Expérimentalement : cristallographie par diffraction à rayons X,  
résonance magnétique nucléaire (RMN)  
  
Long, difficile et coûteux
  - Par bio-informatique: algorithmes de prédiction de structures secondaires

**Approche thermodynamique**

# Approche thermodynamique

- **Trois hypothèses :**
  - À chaque configuration de la molécule correspond un nombre de liaisons hydrogènes (*quantité d'énergie libre*).
  - La configuration la plus stable est celle qui maximise le nombre de liaisons hydrogènes (*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 le nombre de liaisons hydrogènes (*l'énergie*) est optimale.

# Approche thermodynamique



1. **A** ARN transcrit sous une forme essentiellement déstructurée
2. Fluctue alors de façon stochastique entre ses différents états
3. Le système finit par atteindre l'équilibre thermodynamique => **D**

## Modèle initial (Nussinov - 1978)

→ L'énergie de la molécule est la somme des liaisons hydrogènes (*énergies*) de chaque paire de bases.

→  $\alpha(r_i, r_j)$  : nombre de liaisons hydrogènes (*énergie libre*) de l'appariement  $(r_i, r_j)$

$$\alpha(G, C) = 3 \qquad G \equiv C$$

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

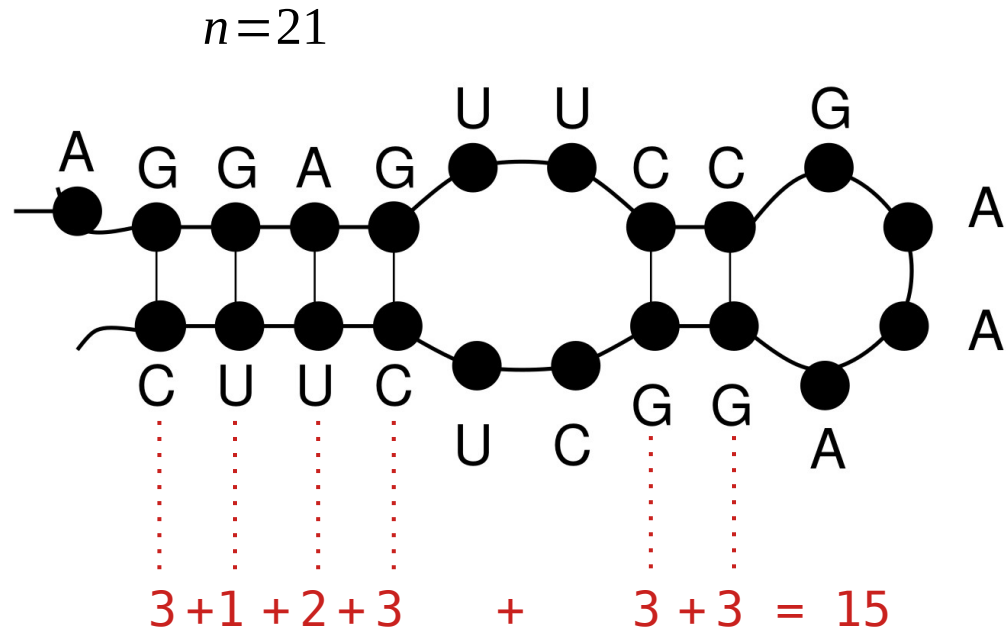
$$\alpha(G, U) = 1 \qquad G - U$$

→ Nombre de liaisons hydrogènes (énergie libre) de la structure secondaire  $S$

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

Minimisation de l'énergie  $\Leftrightarrow$  **Maximisation** du nombre de paires de bases.

## Exemple



**Nombre total de liaisons hydrogènes : 15**

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

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

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

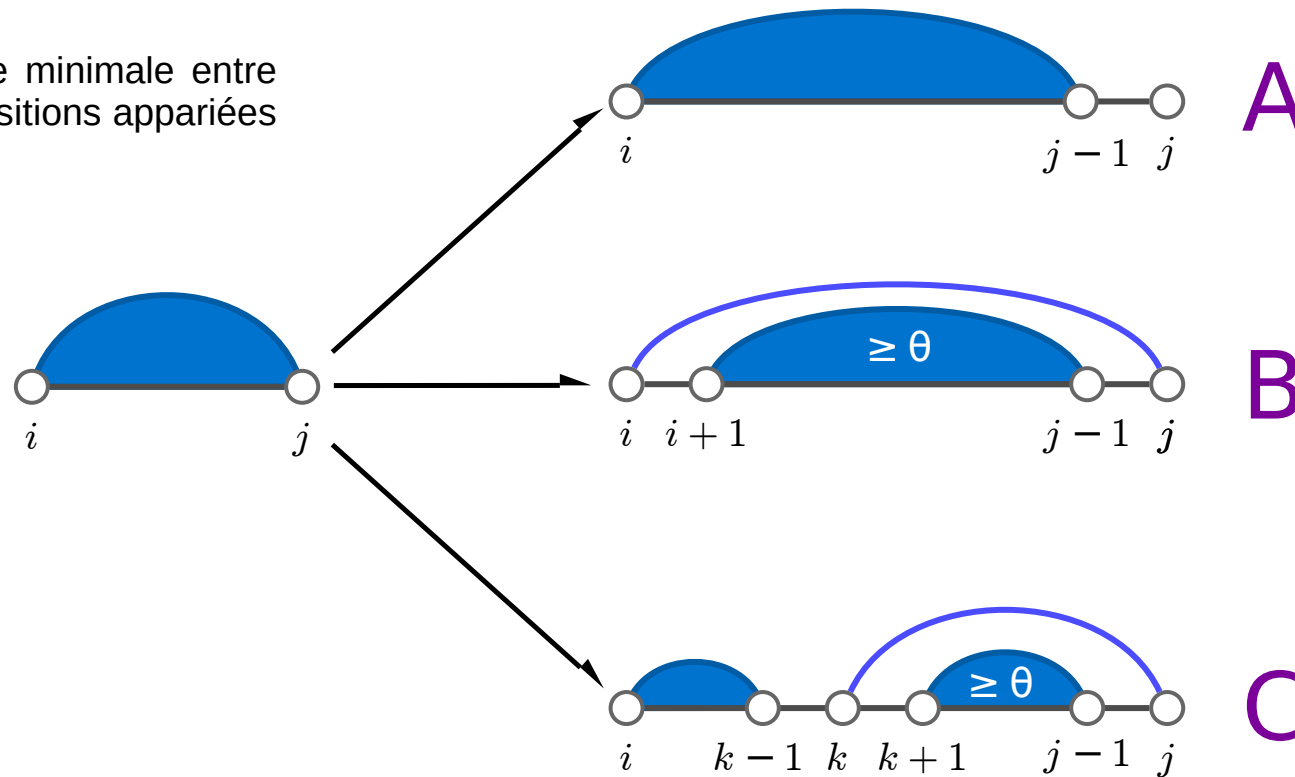
### → Comment calculer la structure optimale ?

le nombre de structures secondaires compatibles avec un ARN est, en moyenne, exponentiel sur la taille de celui-ci. **L'énumération des structures candidates est impossible !!**

**=> Programmation dynamique**

# Décomposition du problème en instance plus petites 1/2

$\theta \Rightarrow$  distance minimale entre deux positions appariées

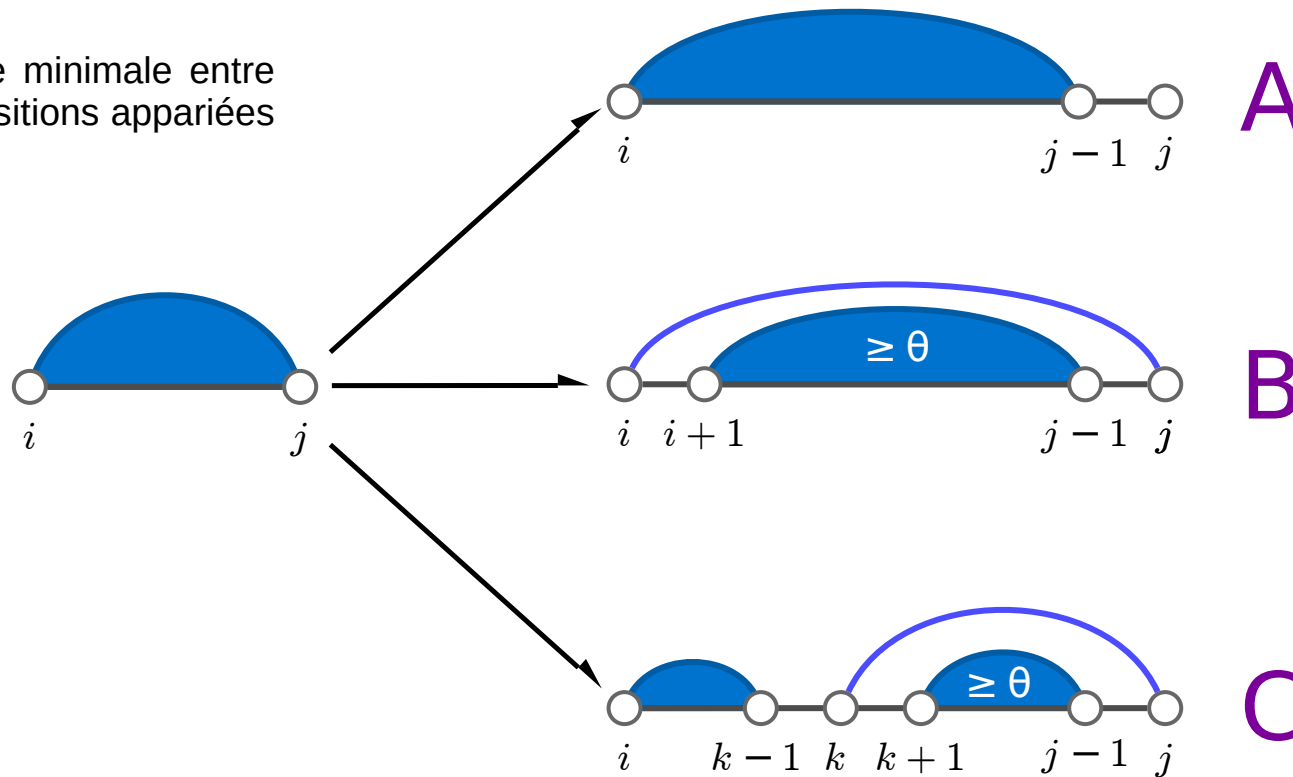


- **Cas A** Soit  $j$  libre, et précédé par une structure secondaire formée indépendamment sur la région  $[i, j-1]$
- **Cas B** Soit  $j$  est apparié à la position  $i$ ,  $j - i > \theta$ , et alors il se forme une structure secondaire sur la région  $[i + 1, j - 1]$
- **Cas C** Soit  $j$  est apparié à une position  $k$  avec  $i < k < j$ ,  $j - k > \theta$ , et des structures se forment alors dans les régions  $[i, k - 1]$  et  $[k + 1, j - 1]$ . Celles-ci sont indépendantes, du fait de l'interdiction des croisements



# Décomposition du problème en instance plus petites 2/2

$\theta \Rightarrow$  distance minimale entre deux positions appariées



- **Cas A**  $L(S_{i,j}) = L(S_{i,j-1})$
- **Cas B**  $L(S_{i,j}) = L(S_{i+1,j-1}) + \alpha(r_i, r_j)$
- **Cas C**  $L(S_{i,j}) = \max \{ L(S_{i,k-1}) + \alpha(r_k, r_j) + L(S_{k+1,j-1}), k \in ]i, j[ \}$

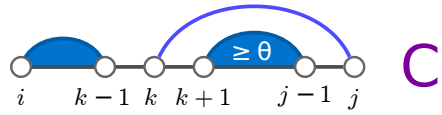
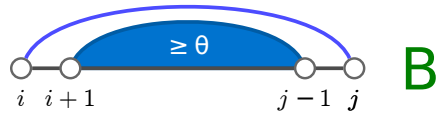
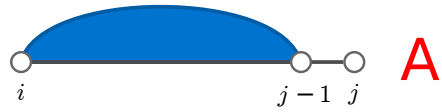
# construction de la table de programmation dynamique

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

$$T(i, j) = \max \left\{ \begin{array}{l} T(i, j-1) \\ T(i+1, j-1) + \alpha(r_i, r_j) \\ \max \{ T(i, k-1) + \alpha(r_k, r_j) + T(k+1, j-1) \} \end{array} \right.$$

**Étape suivante** : construction de la structure secondaire optimale,  
par retour arrière

$L(S_{i,j})=14$  avec  $i=0$  et  $j=17$



	j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0

• **Cas A**  $L(S_{i,j-1})=14$

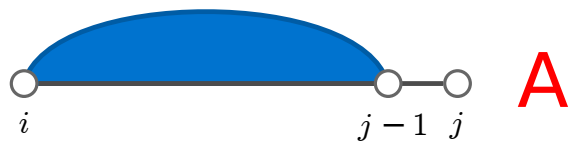
• **Cas B**  $L(S_{i+1,j-1})+\alpha(r_i,r_j)=11$

• **Cas C**  $\max\{L(S_{i,k-1})+\alpha(r_k,r_j)+L(S_{k+1,j-1}), k \in ]i, j[ \} = 11$  pour  $k=8$

																	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j

	j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0



(																)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j

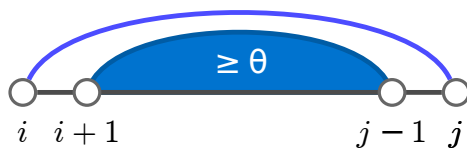
	j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0



(	(														)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j

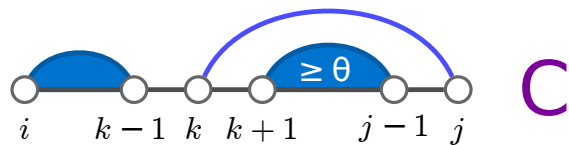
j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0



(	(						(							)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j
k

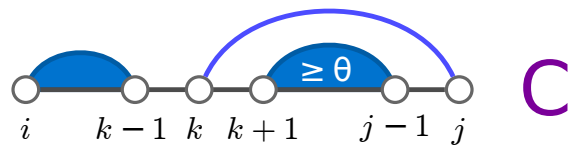
j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0



(	(						(		(				)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i  
j  
k

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0





(	(						(	.	(				)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

$i=j$

i
j
k

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	2	2	4	5	7	7	8	10	
5	A						0	0	0	0	2	2	2	5	5	5	8	8	
6	C							0	0	0	0	0	2	5	5	5	8	8	
7	U								0	0	0	0	2	3	5	5	6	7	
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0

$$j-i \leq \theta \quad \text{avec } \theta=3$$

(	(						(	.	(			.	)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j
k

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0

$$j-i \leq \theta \quad \text{avec } \theta=3$$

(	(						(	.	(		.	.	)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j
k

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0

$$j-i \leq \theta \quad \text{avec } \theta=3$$

(	(						(	.	(	.	.	.	)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

$i=j$

i
j
k

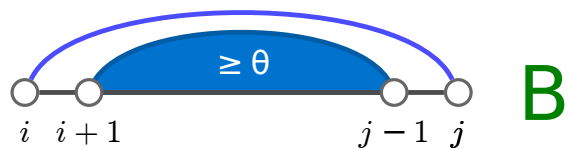
j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0

$$j-i \leq \theta \quad \text{avec } \theta=3$$

(	(	(				)	(	.	(	.	.	.	)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j
k

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	2	2	4	5	7	7	8	10	10
5	A						0	0	0	0	2	2	2	5	5	5	8	8	8
6	C							0	0	0	0	0	2	5	5	5	8	8	8
7	U								0	0	0	0	2	3	5	5	6	7	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	1	2	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0



(	(	(			.	)	(	.	(	.	.	.	)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j
k

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0

$$j-i \leq \theta \quad \text{avec } \theta=3$$

(	(	(		.	.	)	(	.	(	.	.	.	)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

i
j
k

j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	2	2	4	5	7	7	8	10	
5	A						0	0	0	0	2	2	2	5	5	5	8	8	
6	C							0	0	0	0	0	2	5	5	5	8	8	
7	U								0	0	0	0	2	3	5	5	6	7	
8	U									0	0	0	2	3	5	5	5	7	
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	1	2	
12	A													0	0	0	0	0	
13	G														0	0	0	0	
14	A															0	0	0	
15	C																0	0	
16	G																	0	
17	A																		0

$$j-i \leq \theta \quad \text{avec } \theta=3$$

(	(	(	.	.	.	)	(	.	(	.	.	.	)	)	)	)	.
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

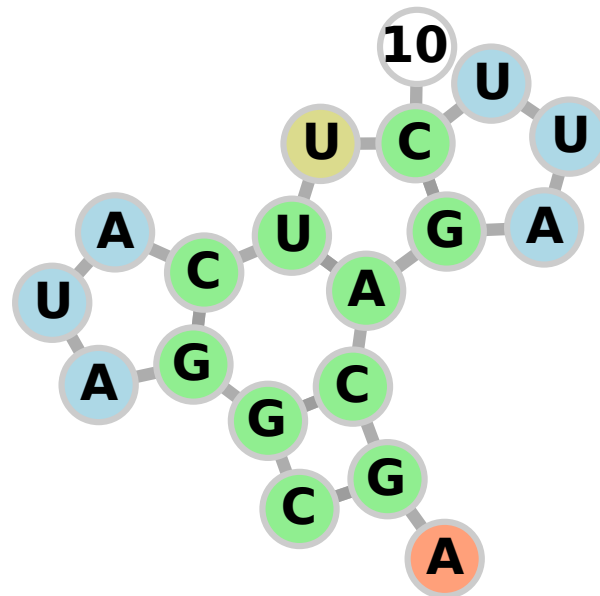
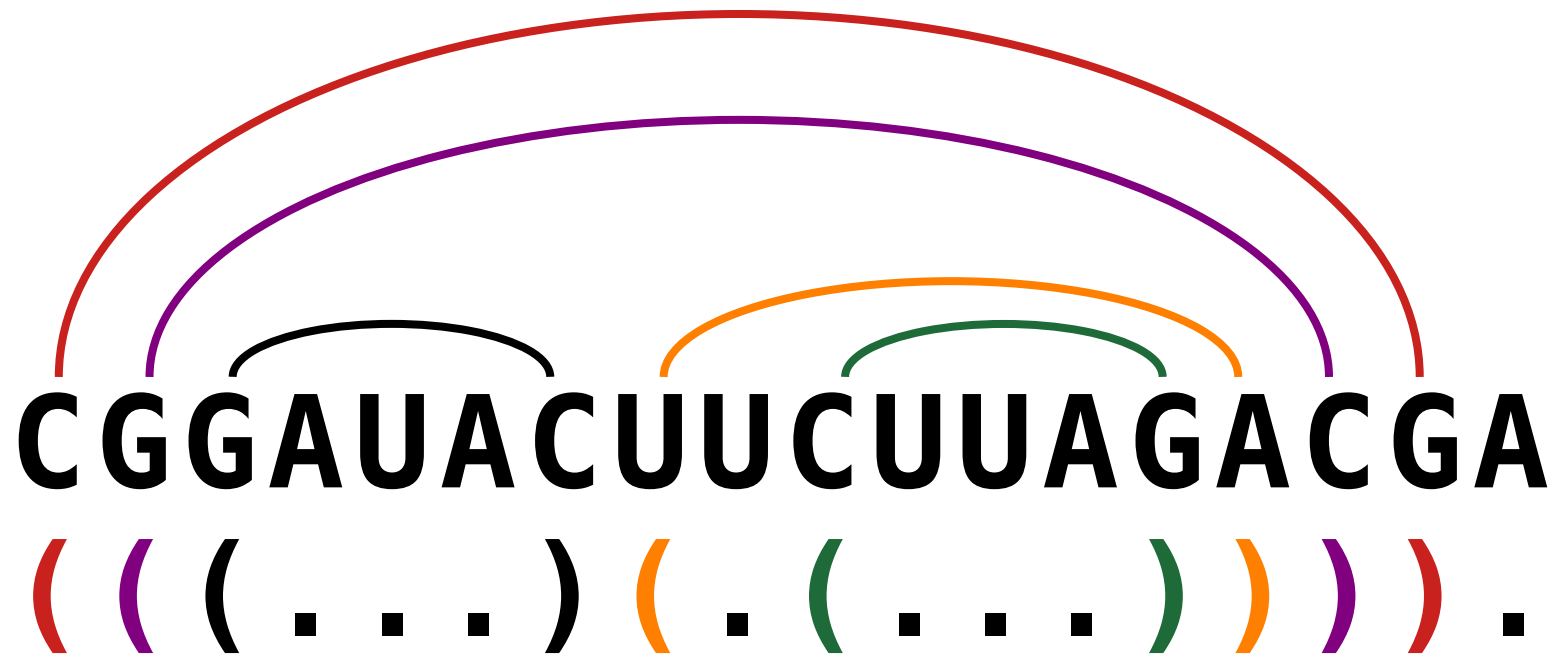
$i=j$

i  
j  
k

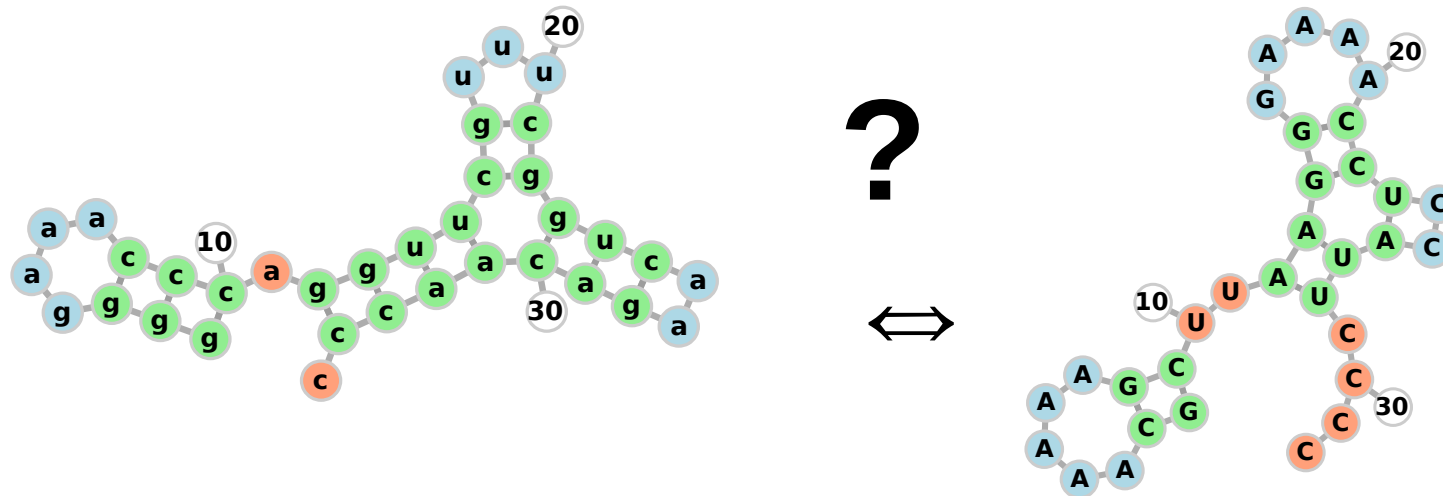
j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
i		C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
0	C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
1	G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
2	G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
3	A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
4	U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
5	A						0	0	0	0	0	2	2	2	5	5	5	8	8
6	C							0	0	0	0	0	0	2	5	5	5	8	8
7	U								0	0	0	0	0	2	3	5	5	6	7
8	U									0	0	0	0	2	3	5	5	5	7
9	C										0	0	0	0	3	3	3	5	5
10	U											0	0	0	0	2	2	2	3
11	U												0	0	0	0	0	1	2
12	A													0	0	0	0	0	0
13	G														0	0	0	0	0
14	A															0	0	0	0
15	C																0	0	0
16	G																	0	0
17	A																		0

$$j-i \leq \theta \quad \text{avec } \theta=3$$





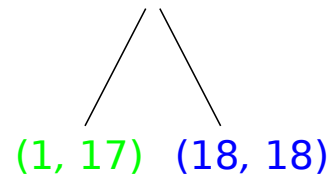
# Comparaison de structures



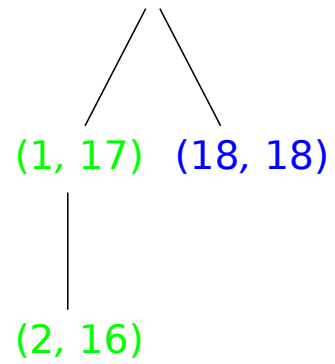
Comparer les structures :

- Avec le format parenthésé
- À l'aide d'une représentation sous forme d'arbre

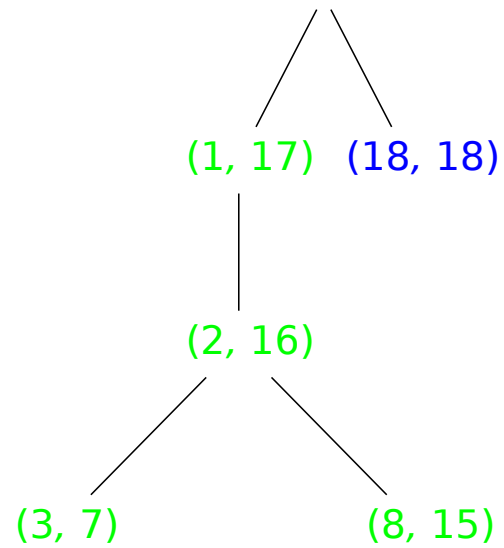
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
<b>C</b>	<b>G</b>	<b>G</b>	<b>A</b>	<b>U</b>	<b>A</b>	<b>C</b>	<b>U</b>	<b>U</b>	<b>C</b>	<b>U</b>	<b>U</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>A</b>
(	(	(	.	.	.	)	(	.	(	.	.	.	)	)	)	)	.



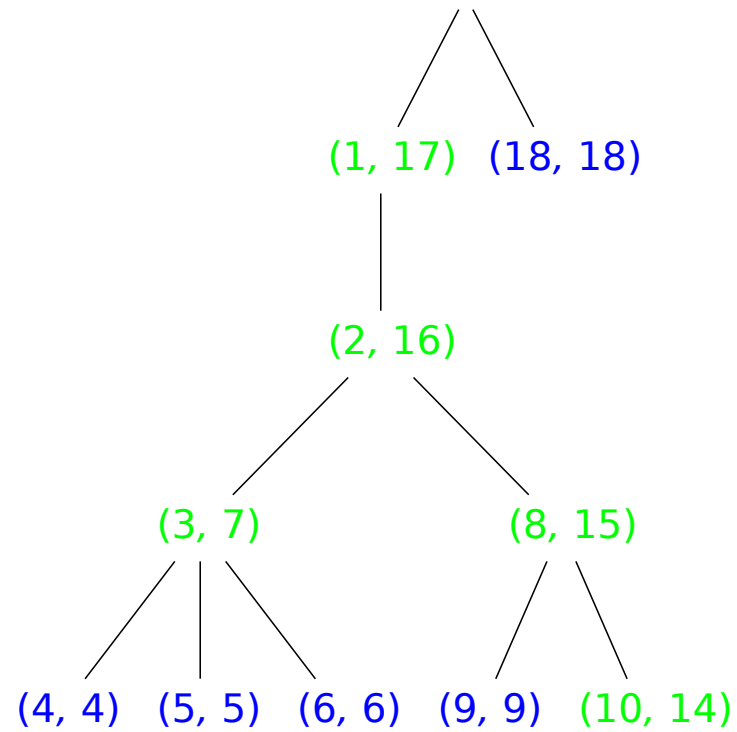
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
(	(	(	.	.	.	)	(	.	(	.	.	.	)	)	)	)	.



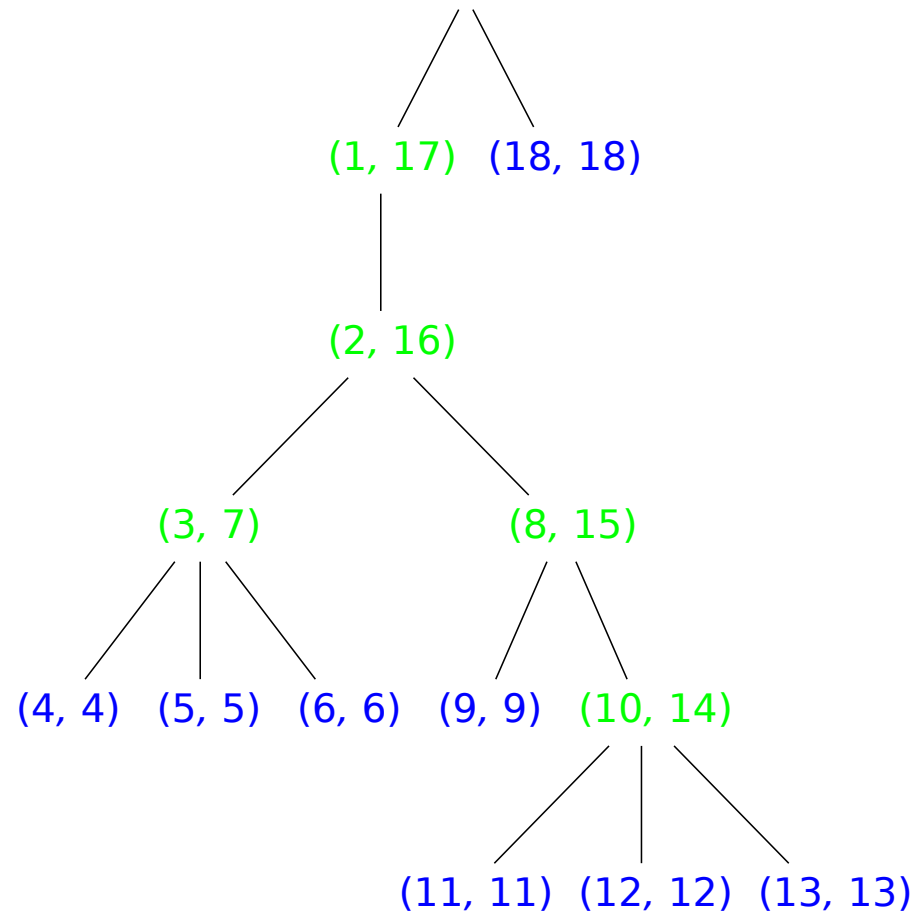
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
(	(	(	.	.	.	)	(	.	(	.	.	.	)	)	)	)	.



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
(	(	(	.	.	.	)	(	.	(	.	.	.	)	)	)	)	.



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
(	(	(	.	.	.	)	(	.	(	.	.	.	)	)	)	)	.



# Comparaison de structures

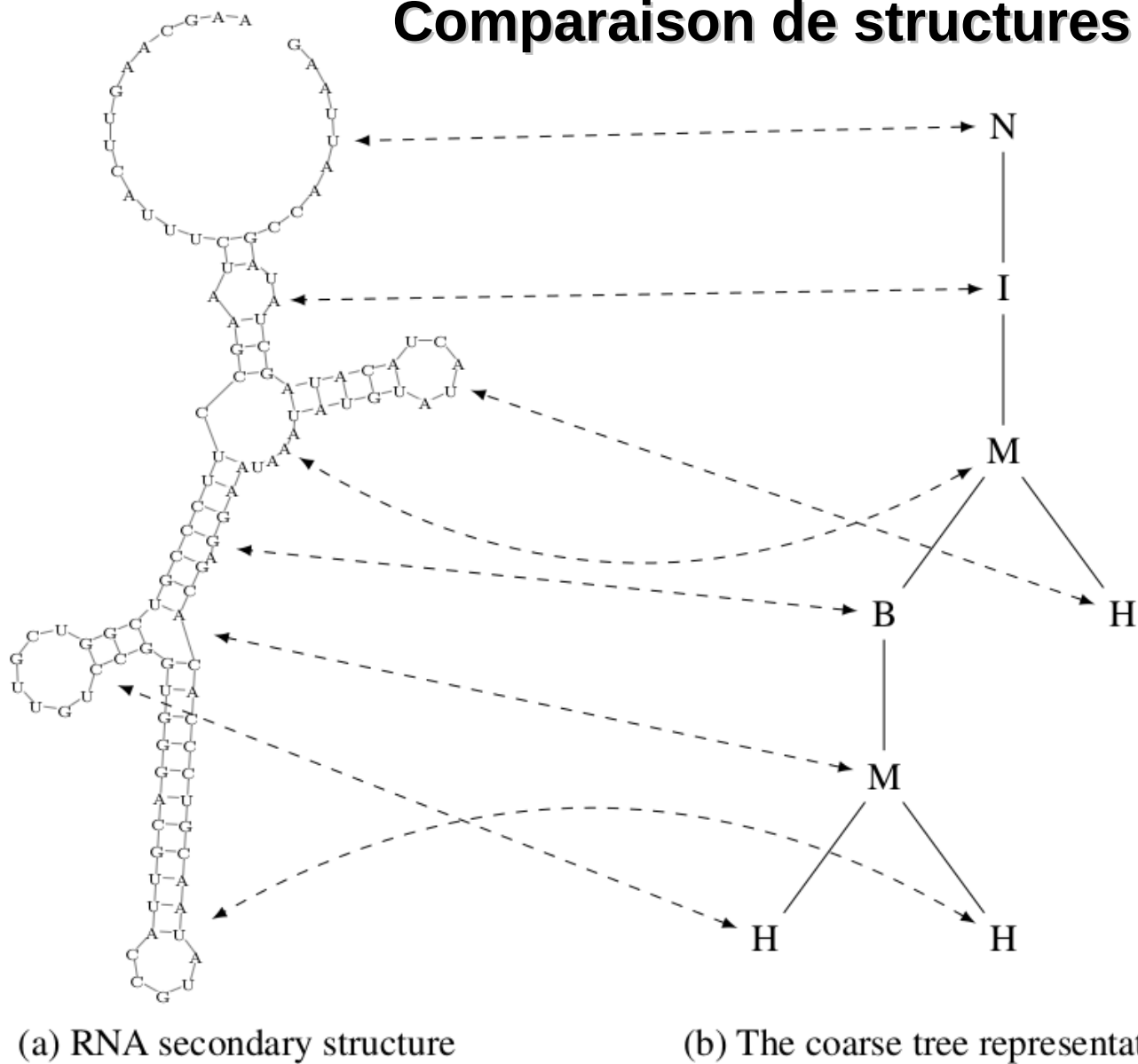
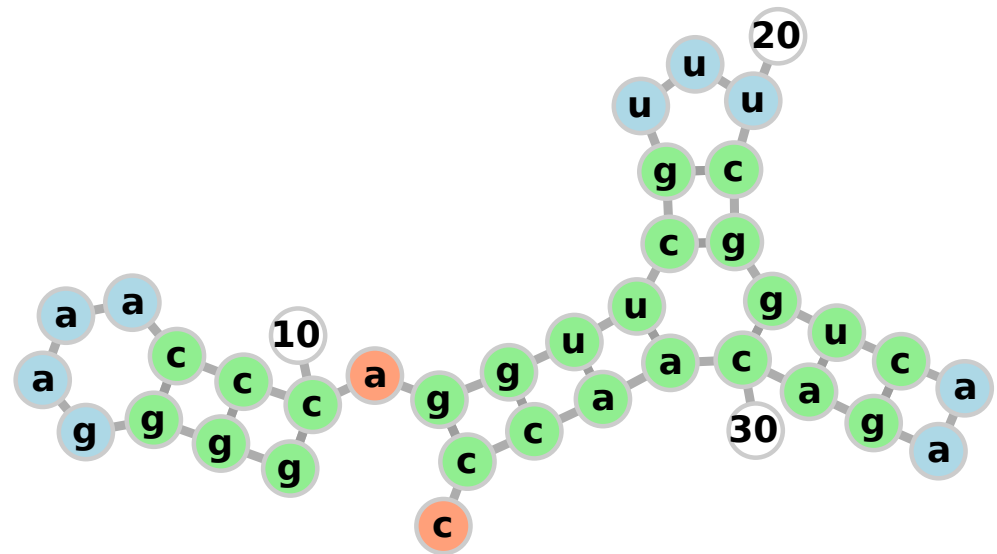
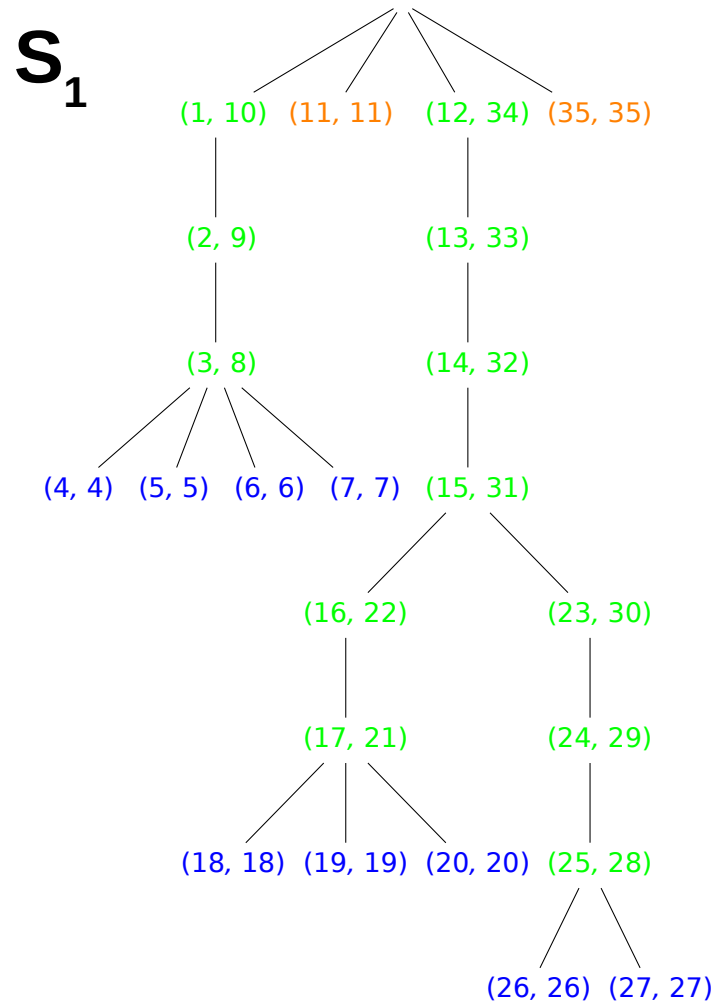


Fig. 4: Coarse grained tree representation, which represents an RNA secondary structure as a tree of structural building blocks such hairpin loops (H), multiloops (M), bulges (B), internal loops (I). Node N does not represent a structural element, it closes the secondary structure and makes sure the representation forms a tree.



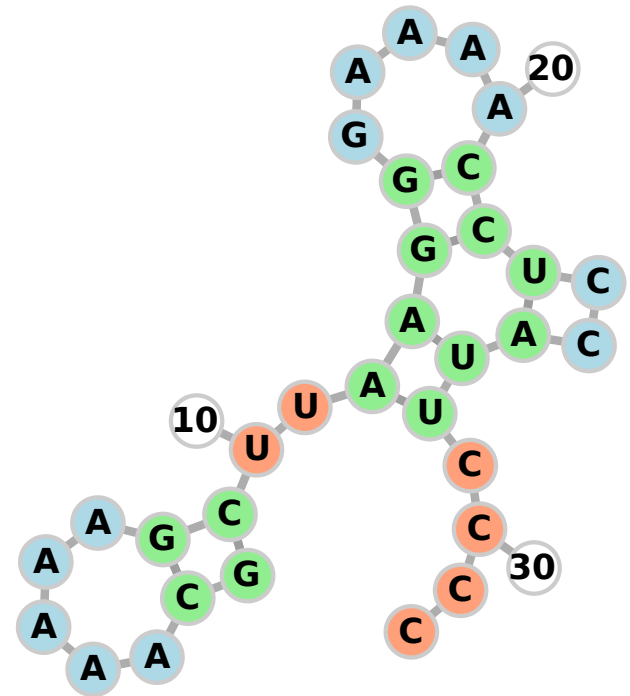
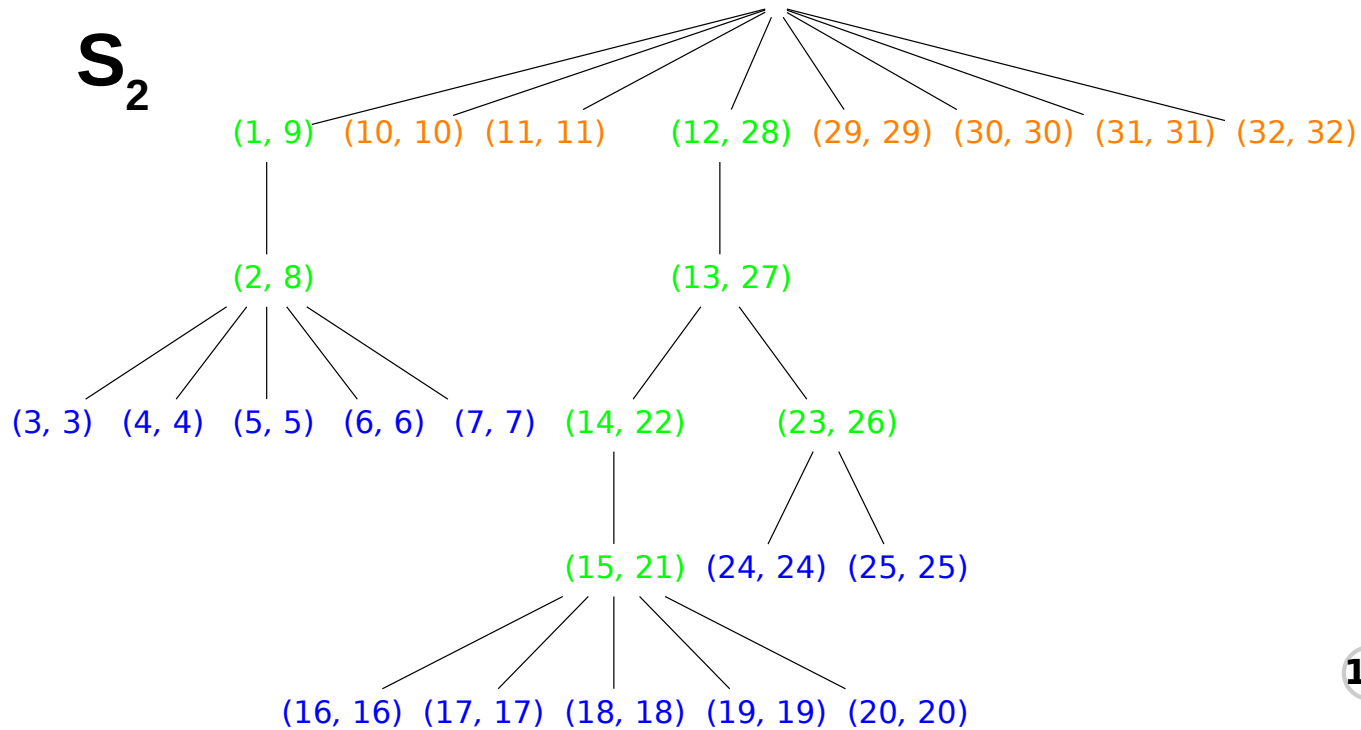
# Comparaison de structures

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
G	G	G	G	A	A	A	C	C	C	A	G	G	U	U	C	G	U	U	U	C	G	G	U	C	A	A	G	A	C	A	A	C	C	C
(	(	(	.	.	.	.	)	)	)	.	(	(	(	(	(	(	.	.	.	)	)	(	(	(	.	.	)	)	)	)	)	)	)	.



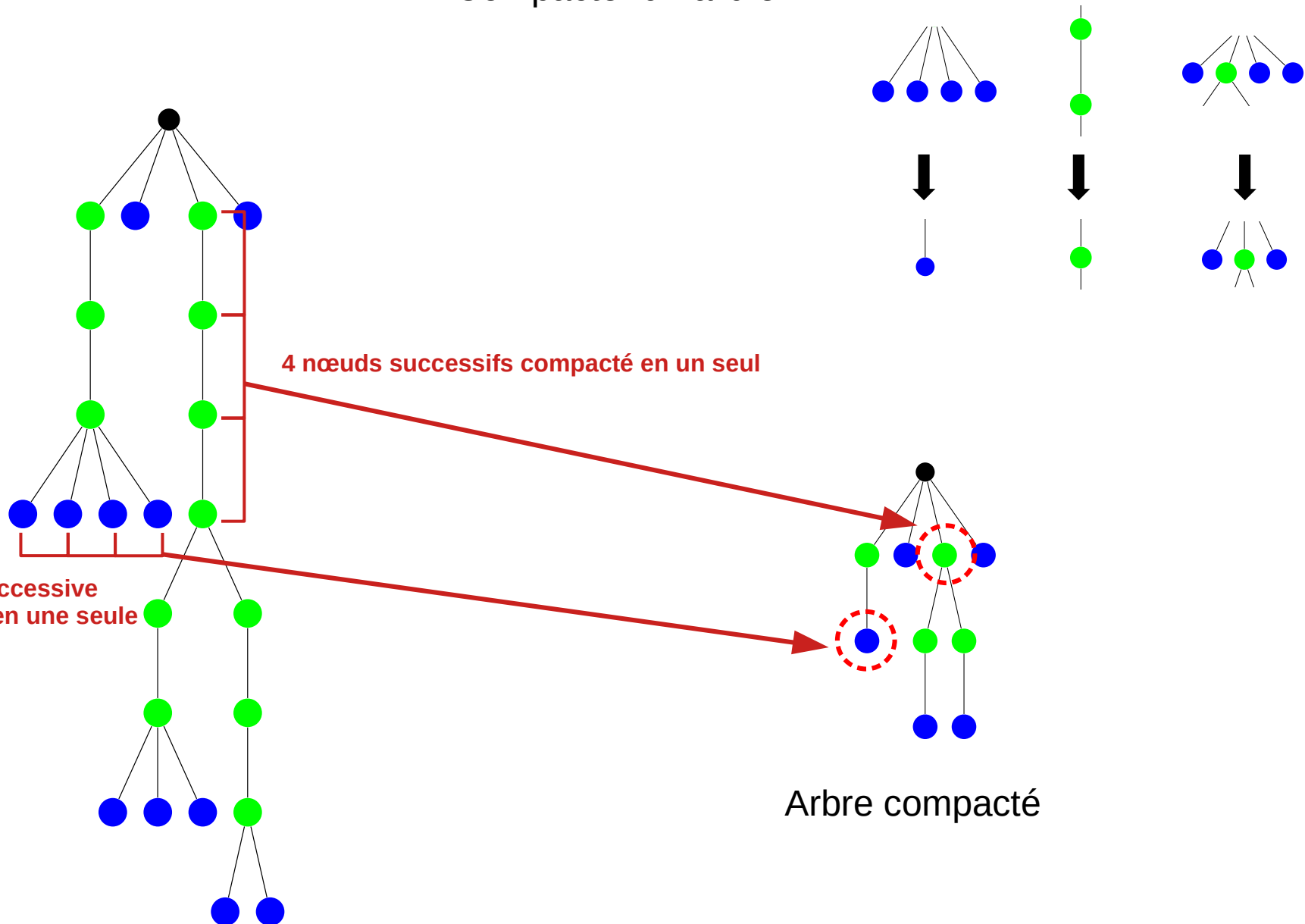
# Comparaison de structures

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
G	C	A	A	A	A	A	G	C	U	U	A	A	G	G	G	A	A	A	A	C	C	U	C	C	A	U	U	C	C	C	C
(	(	.	.	.	.	.	)	)	.	.	(	(	(	(	.	.	.	.	.	)	)	(	.	.	)	)	)	.	.	.	.



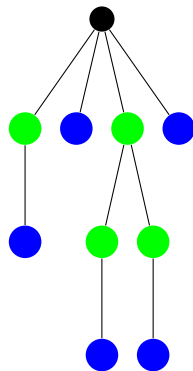
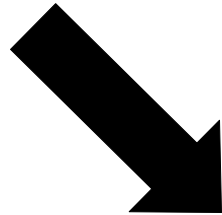
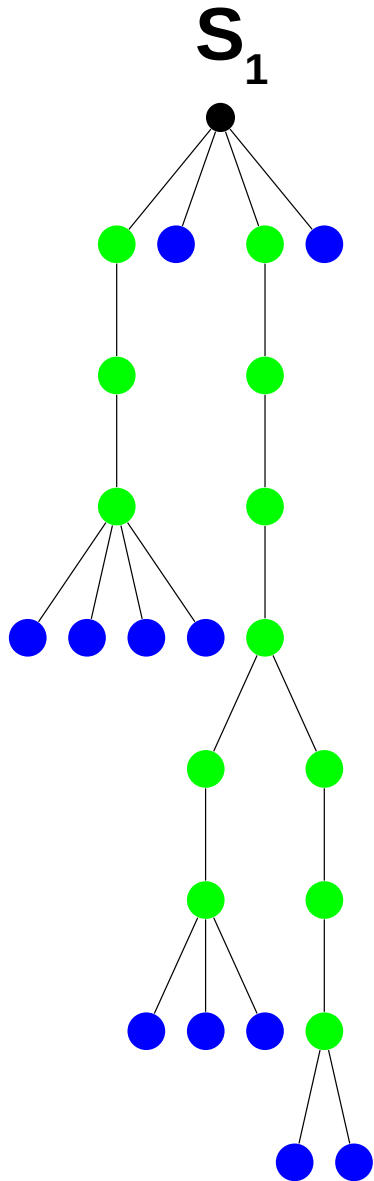
# Comparaison de structures

Compacter un arbre

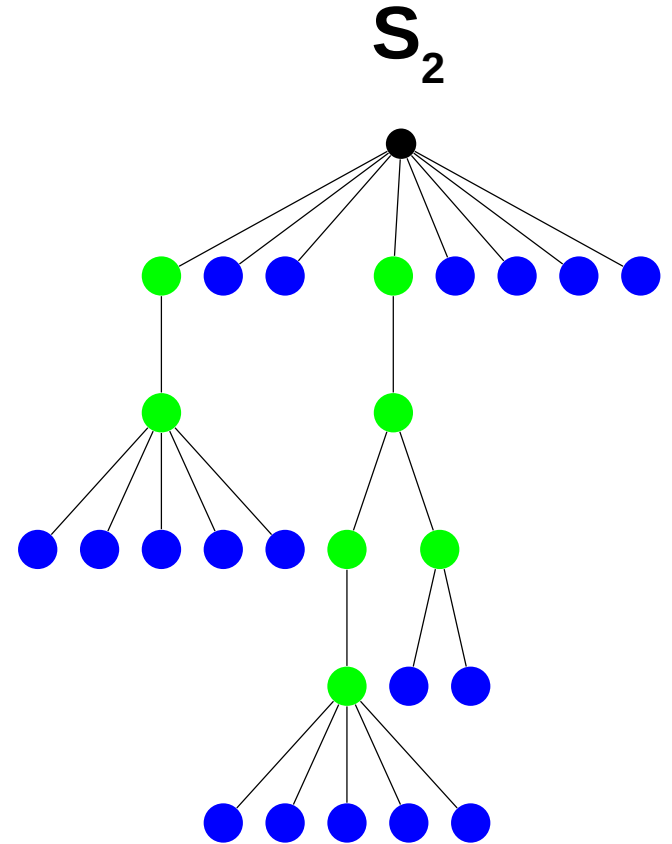
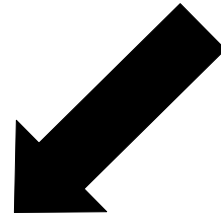


# Comparaison de structures

$$S_1 \Leftrightarrow S_2 ?$$



Architecture identiques

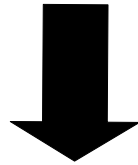


$$S_1 \Leftrightarrow S_2$$

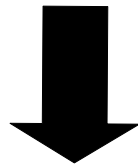
# Comparaison de structures

Compacter les structures au format parenthésé

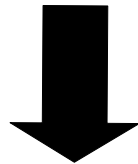
((. . . .)) . . ((((. . . .))(. .))) . . . .



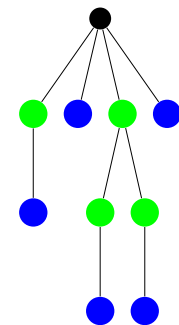
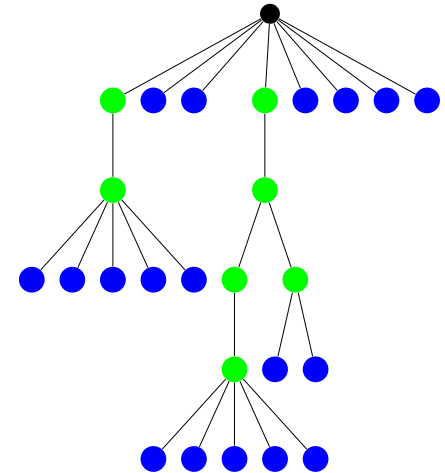
((.)) . ((((.))(.))) .



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



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

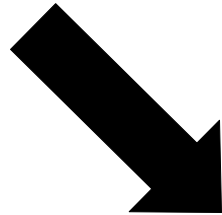


# Comparaison de structures

$$S_1 \Leftrightarrow S_2 ?$$

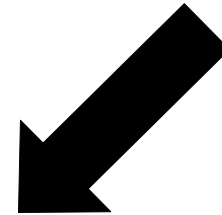
$S_1$

$(((((\dots))) \cdot ((((((\dots))) (((\dots))))))) \cdot$



$S_2$

$((\dots)) \cdot ((((((\dots))) (\dots)))) \dots$



$(\cdot) \cdot ((\cdot)(\cdot)) \cdot$

$$S_1 \Leftrightarrow S_2$$