

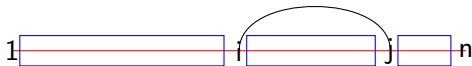
Paysage d'énergie et structures localement optimales d'un ARN

Azadeh Saffarian

April 14, 2009

- Identifier les différentes structures secondaires que peut adopter une séquence d'ARN
 - Structures sous-optimales
 - Structures localement optimales
- La vraie structure d'un ARN n'est pas toujours la structure optimale en termes d'énergie libre
- Existence de plusieurs structures alternatives (riboswitch)
- Structures avec pseudo-noeuds
- Information sur toutes les configurations possibles: paysage d'énergie
[Cupal, Hofacker, Stadler] [Flamm, Hofacker, Stadler]
- Signal (faible...) pour distinguer les ARN structurés des autres ARN [Clote, 2004]

- construction des structures sous-optimales



- pour chaque couple de positions (i, j) , calcul de la structure d'énergie minimale contenant l'appariement $i - j$
 - sélection en fonction du pourcentage de sous-optimalité
 - filtrage des structures (en utilisant un critère de distance) pour que les résultats sortis soient des structures distinctes
- au plus $n^2/2$ structures sous-optimales

[Zuker 1989, etc.]

- Exemple de résultat (100% sous-optimalité)

```
ACACAAAAGUGUGAAAAACACACAAAAGUGUAAAUCUCAAAAGAGAGAAAAACUCUCAAAAGAGA
(((...((((.....))))).))....(((...((((.....))))).)).... (-17.70)
.((((.....)))....(((.....)))....(((...((((.....))))).)).... (-15.00)
((((...((((.....))))).))....(((.....)))....(((.....))). (-14.10)
.....((((.....)))....(.((...((((.....))))).))..... (-9.20)
.....(((.....((((.....))))).))....(((.....))).... (-3.17)
```

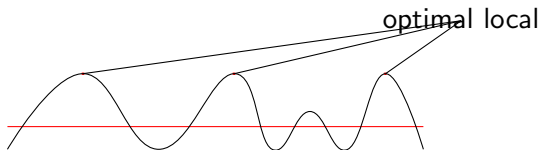
- Structure manquante

- Exemple de résultat (100% sous-optimalité)

```
ACACAAAAGUGUGAAAAACACACAAAAGUGUAAAAUCUCAAAAGAGAGAAAAACUCUCAAAAGAGA
((((...((((...))))))...))...((((...((((...))))))...)) (-17.70)
.((((...)))...((((...)))...((((...((((...))))))...)) (-15.00)
((((...((((...))))))...))...((((...)))...(((...))). (-14.10)
.....((((...)))...((...((...((((...))))))...))..... (-9.20)
.....(((.....((((...))))...((((...))))...))..... (-3.17)
```

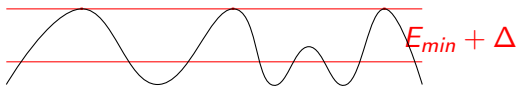
- Structure manquante

```
.((((...)))...((((...)))...((((...)))...(((...))).
```



Méthodes existantes: RNAsubopt

- Vienna Package, extension de RNAfold
- Méthode exacte qui génère toutes les structures sous-optimales dans un intervalle d'énergie libre de Δ par rapport à l'énergie libre de la structure optimale
- Le nombre de structures augmente exponentiellement avec Δ



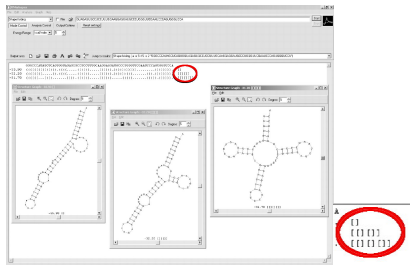
[Wuchty *et al*, 1999]

Méthodes existantes: RNAsubopt

```
GC AAGAGUCCUGGCUGUGGAAAGAUACCUACAGGAUCAACAGCUCCUGGGGAUUUGGGGC -1890 500
(. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.00
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -13.90
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.00
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.90
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.20
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.30
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -13.90
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.10
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -13.90
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.00
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -13.90
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.20
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -16.10
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.40
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.30
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.20
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.00
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.50
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.50
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.60
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.60
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.40
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.20
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.10
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.20
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -13.90
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.40
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.50
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.30
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.40
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.20
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -15.30
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.00
((. . . ((((((((((.....(((.(.....))))).)))))) . . . )))) . . . ) -14.10
```

Méthodes existantes: RNAsHapes

- *abstract shapes*: partition de l'espace de repliement en différents classes de structures, en fonction de leur forme
- RNAsubopt+RNAsHapes: solution partielle, à cause de l'espace des structures à explorer exhaustivement



[Giegerich *et al.*, 2004]

- Définition énergétique [Clote, Lorenz, 2008]
 - L'énergie libre (en respectant le modèle de plus proches voisins de Turner) obtenue en supprimant ou ajoutant un paire de bases n'est pas meilleure
 - Peu opérationnel pour le calcul
- Définition topologique [Clote, 2004]
 - On ne peut plus y ajouter une paire de bases sans créer de conflit
 - Se prête bien au calcul
 - Structures peu réalistes: appariements indépendants

Les structures localement optimales

- Définition énergétique [Clote, Lorenz, 2008]
 - L'énergie libre (en respectant le modèle de plus proches voisins de Turner) obtenue en supprimant ou ajoutant un paire de bases n'est pas meilleure
 - Peu opérationnel pour le calcul
- Définition topologique [Clote, 2004]
 - On ne peut plus y ajouter une paire de bases sans créer de conflit
 - Se prête bien au calcul
 - Structures peu réalistes: appariements indépendants
- **Energétique + topologique**: structures localement optimales (au sens topologique) pour un ensemble de tiges (énergétiquement) stables

Les structures localement optimales

- $\alpha = \alpha_1 \dots \alpha_i \dots \alpha_j \dots \alpha_n$, une séquence d'ARN de longueur n
- Une structure secondaire S sur α est un ensemble des paires de bases tel que :
 - Paire ordonnée: $(i, j) \in S \Rightarrow 1 \leq i + 3 < j \leq n$
 - *Watson-Crick* ou *GU wobble*
 - Pas de pseudo-noeuds
 - Pas de triplets de bases
- Une structure secondaire est *localement optimale* si on ne peut plus ajouter des paires de bases sans violer la définition de structure secondaire.
- Clote, 2004: algorithme de comptage des structures localement optimales

- Construction de la structure optimale [Nussinov-Jacobson]

$N(i, j)$: nombre de paires de bases de la structure optimale sur la sous-séquence $\alpha[i..j]$

$$N(i, j) = \max \left\{ \begin{array}{l} \max_{k \leq j, (i, k) \text{ paire}} (1 + N(i + 1, k - 1) + N(k + 1, j)) \\ N(i + 1, j) \end{array} \right.$$

- Comptage de toutes les structures secondaires

$S(i, j)$: nombre de structures secondaires sur la sous-séquence $\alpha[i..j]$

$$S(i, j) = + \left\{ \begin{array}{l} \sum_{k \leq j, (i, k) \text{ paire}} (1 \times S(i + 1, k - 1) \times S(k + 1, j)) \\ S(i + 1, j) \end{array} \right.$$

Les structures localement optimales

- Construction de la structure optimale [Nussinov-Jacobson]

$N(i, j)$: nombre de paires de bases de la structure optimale sur la sous-séquence $\alpha[i..j]$

$$N(i, j) = \max \begin{cases} \max_{k \leq j, (i, k) \text{ paire}} (1 + N(i + 1, k - 1) + N(k + 1, j)) \\ N(i + 1, j) \end{cases}$$

- Comptage de toutes les structures secondaires

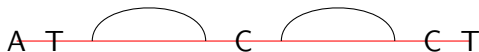
$S(i, j)$: nombre de structures secondaires sur la sous-séquence $\alpha[i..j]$

$$S(i, j) = + \begin{cases} \sum_{k \leq j, (i, k) \text{ paire}} (1 \times S(i + 1, k - 1) \times S(k + 1, j)) \\ S(i + 1, j) \end{cases} \leftarrow \text{structures pas localement optimales}$$

Comment ne compter que les structures localement optimales ?

Les structures localement optimales

- $S = \{(x_1, y_1) \dots (x_k, y_k)\}$ une structure secondaire sur α .
- Une position i de α est *visible* si elle est à l'extérieur de tous les appariements de S



- La base b de $\{A, C, G, T\}$ est visible s'il existe une position i telle que i est visible et $b = \alpha_i$
- $Loc(i, j)[vis]$: ensemble des structures secondaires localement optimales entre les positions i et j avec les bases de vis visibles
- $Loc(i, j)$: ensemble des structures secondaires localement optimales entre les positions i et j (obtenu par union)

[Clote, 2004]

si $i \geq j$ $Loc(i, j)[vis] = 0$

sinon

$$Loc(i, j) = + \begin{cases} \sum_{k \leq j, (i, k) \text{ paire}} 1 \times Loc(i + 1, k - 1) \times Loc(k + 1, j)[vis] \\ \text{et, si la base complémentaire à } \alpha_i \text{ n'est pas dans } vis \\ Loc(i, j)[\{\alpha_i\} \cup vis] = Loc(i + 1, j)[vis] \end{cases}$$

Complexité: $O(2^4 n^3)$ [Clote, 2004]

Structures localement optimales pour un ensemble de tiges

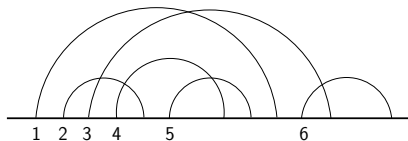
- Donnée: un ensemble de tiges potentielles
 - Tige: ensemble de paires de bases
 - Présence ou pas de renflement, de boucle interne,...
- Quatre relations possibles entre deux tiges



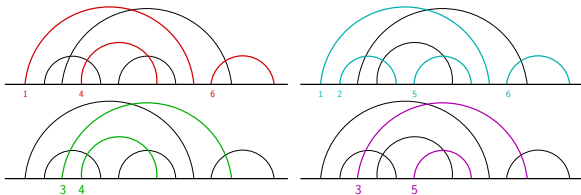
- Structure *secondaire* : ensemble de tiges deux à deux juxtaposées ou imbriquées
- Une structure secondaire S est *localement optimale* si pour toute tige f qui n'appartient pas à S , S contient une tige g telle que
 - 1 f et g sont en conflit
 - 2 ou f est incluse dans g

Structures localement optimales pour un ensemble de tiges

- Exemple: 6 tiges en entrée



- 4 structures localement optimales



- Les leçons de l'algorithme de Clote

$$Loc(i, j) = + \begin{cases} \sum_{k \leq j, (i, k) \text{ paire}} 1 \times Loc(i+1, k-1) \times Loc(k+1, j)[\text{vis}] \\ \text{et, si la base complémentaire à } \alpha_i \text{ n'est pas dans vis} \\ Loc(i, j)[\{\alpha_i\} \cup \text{vis}] = Loc(i+1, j)[\text{vis}] \end{cases}$$

- algorithme adapté au dénombrement des structures localement optimales, pas à leur construction
- la notion de visibilité est liée à la définition des appariements :
 $A - U, C - G, G - U$
- Nécessité de développer une nouvelle méthode

Construction des structures localement optimales

ensemble
de
tiges
potentielles

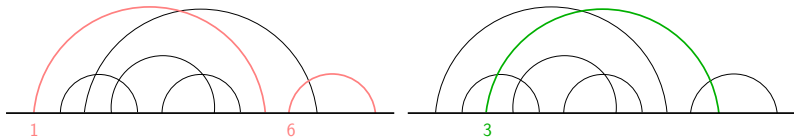
structures
maximales
par
juxtaposition

structures
localement
optimales

- Définir une famille intermédiaire de structures
 - plus riches que les tiges seules
 - moins nombreuses que les structures localement optimales
 - à partir desquelles on peut construire les structures localement optimales à la volée
- Décomposer les structures secondaires en niveaux d'imbrication

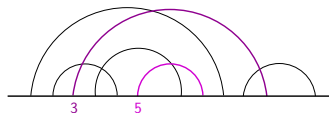
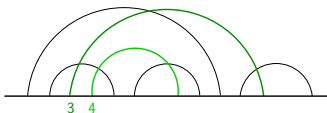
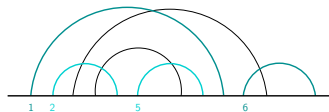
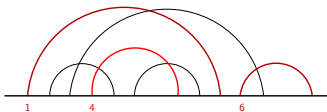
Structures maximales par juxtaposition

- une structure secondaire S est *maximale par juxtaposition* si elle satisfait
 - 1 si deux tiges f et g appartiennent à S , alors f et g sont juxtaposées
 - 2 si f est une tige qui n'appartient pas à S , alors S contient une tige g telle que
 - f et g sont conflictuelles
 - ou f est incluse dans g
 - ou f est imbriquée dans g

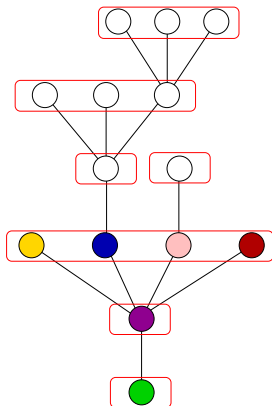
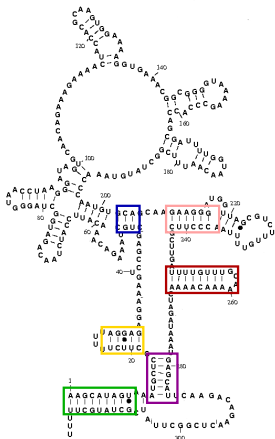


Structures maximales par juxtaposition

- une structure localement optimale se décompose en structures maximales par juxtaposition



Structures maximales par juxtaposition



Algo pour les structures maximales par juxtaposition

- les tiges sont triées par position de début, puis par position de fin
- $\text{nextJuxt}(i)$: plus petite tige ℓ telle que $i < \ell$ et ℓ est juxtaposée avec i
- $J(i, j)$: ensemble des structures maximales pour juxtaposition sur le sous-ensemble de tiges $\{i, \dots, j\}$



Algo pour les structures maximales par juxtaposition

toutes les structures juxtaposées

si i commence après j , alors $J(i, j) = \{\varepsilon\}$

sinon, si i termine après j , alors $J(i, j) = J(i + 1, j)$

sinon

$$J(i, j) = \bigcup \left\{ \begin{array}{l} \{i\} \oplus J(\text{nextJuxt}(i), j) \\ J(i + 1, j) \end{array} \right.$$

Algo pour les structures maximales par juxtaposition

toutes les structures juxtaposées/**maximales** par juxtaposition

si i commence après j , alors $J(i, j) = \{\varepsilon\}$

sinon, si i termine après j , alors $J(i, j) = J(i + 1, j)$

sinon

$$J(i, j) = \bigcup \left\{ \begin{array}{l} \{i\} \oplus J(\text{nextJuxt}(i), j) \\ \text{Filter}(i, J(i + 1, j)) \end{array} \right.$$

Filter(i, C) = structures de C qui contiennent au moins une tige en conflit avec i

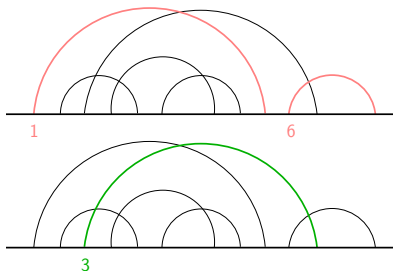
= structures de C dont la première tige non incluse et non imbriquée dans i est en conflit avec i

Algo pour les structures maximales par juxtaposition

- Exemple (suite)

$$J(1, 6) =$$

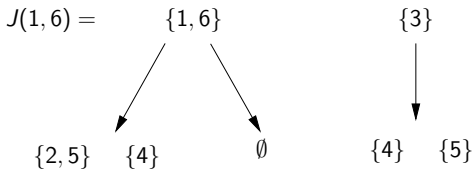
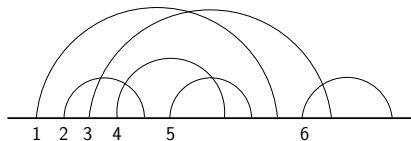
$$\cup \left\{ \begin{array}{l} \{1\} \oplus J(6, 6) = \{1, 6\} \\ \text{Filter}(1, J(2, 6)) = \text{Filter}(1, \{ \{2, 5, 6\} \times, \{3\} \checkmark \}) = \{3\} \end{array} \right.$$



- Mise en œuvre par programmation dynamique

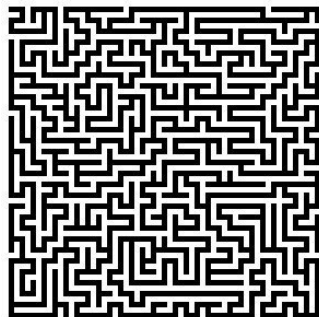
Algo pour les structures localement optimales

- $Loc(i, j)$: ensemble des structures secondaires localement optimales pour les tiges $\{i, \dots, j\}$
- $Loc(1, k) = \{S \cup_{\ell \in S} Loc(\text{succ}(\ell), \text{pred}(\ell)) \mid S \in J(1, k)\}$



Algo pour les structures localement optimales

- Construction à la volée des structures localement optimales à partir des structures maximales par juxtaposition
 - Backtracking avec une pile
 - Espace mémoire supplémentaire en $O(n)$



- Réponse à d'autres questions

quelles sont les structures localement optimales qui contiennent la tige f ?

quelles sont les structures localement optimales qui contiennent au moins x tiges ?

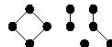
REGLISS

[home](#) | [web server](#) | [help](#) | [retrieve result with an ID](#)

Results for job Feb24200918203021475

	ggucccguaagcagucagugguuagagcguuggucuaugagccgaaggucgcggggucgagccccgccgggacca	
structure1	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-30.00)
structure2	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-27.70)
structure3	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-25.70)
structure4	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-26.80)
structure5	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-24.60)
structure6	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-27.15)
structure7	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-24.85)
structure8	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-25.53)
structure9	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-23.23)
structure10	((((((((())) ((((((())) (((((())) (((((())) ())))))))) ()))))) ())))))	(-22.34)

Visualize the energy landscape graph: [here](#)



List of all computed structures :

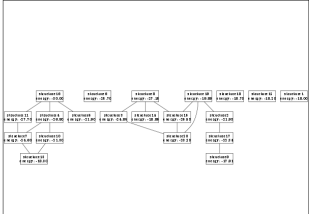
- Structure 1 :
[CT file](#) , [PS file](#) , [JPEG file](#) , [bracket notation](#)
- Structure 2 :
[CT file](#) , [PS file](#) , [JPEG file](#) , [bracket notation](#)
- Structure 3 :
[CT file](#) , [PS file](#) , [JPEG file](#) , [bracket notation](#)
- Structure 4 :
[CT file](#) , [PS file](#) , [JPEG file](#) , [bracket notation](#)
- Structure 5 :

Regliss: RNA energy landscape and secondary structures

REGLISS

home | web server | help | retrieve result with an ID

structure0: ggcuccgagucacagagggaagagcagugucuaagagcagagagagcaggaacggcucccagaccaca (-27.15)

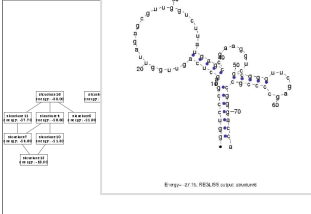


Structure diagram showing energy values for various substructures (e.g., structure0, structure1, structure2, etc.) and their associated energy values (e.g., -27.15).

REGLISS

home | web server | help | retrieve result with an ID

structure0: ggcuccgagucacagagggaagagcagugucuaagagcagagagagcaggaacggcucccagaccaca



Structure diagram showing energy values for various substructures (e.g., structure0, structure1, structure2, etc.) and their associated energy values (e.g., -27.15).

<http://bioinfo.lifl.fr/RNA/regliss>