

# NP-complétude de EDIT(Nested, Nested)

G.Blin, G. Fertin, I. Rusu, C. Sinoquet

Institut de Recherche en  
Informatique de Nantes

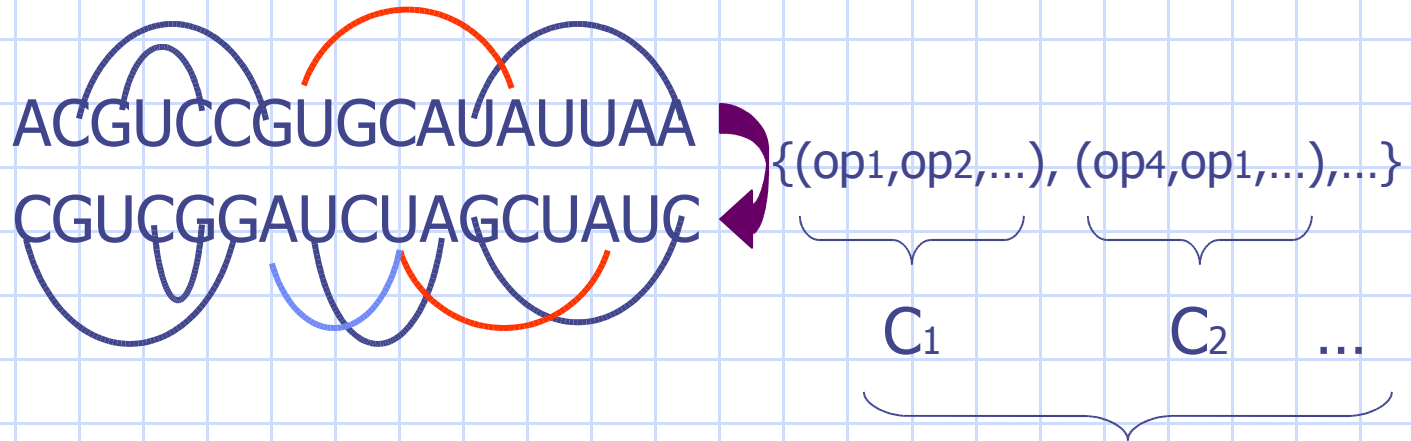
# Motivations

---

- ◆ Comparaison de structures secondaires d'ARN.
- ◆ Prise en compte simultanément des structures primaire et secondaire.
  - Études plus pertinentes
- ◆ Réponse à un problème ouvert proposé par Zhang et al en 2001

# EDIT(Nested, Nested) ?

- ◆ Distance d'édition entre deux séquences annotées par des arcs.



- ◆ Plain

- ◆ Nested

- ◆ Crossing

- ◆ Unlimited

Distance d'édition min

# EDIT(Nested, Nested) ?

## ◆ Opérations sur les bases:

- Suppression :  $W_d$ ,
- Insertion :  $W_d$ ,
- Substitution :  $W_m$

U  
-

Base-deletion  
 $W_d$

-  
U

Base-insertion  
 $W_d$

U U  
A U

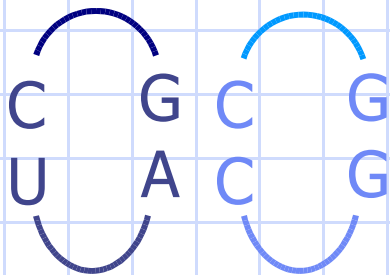
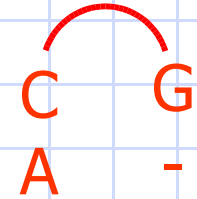
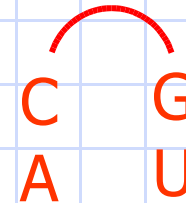
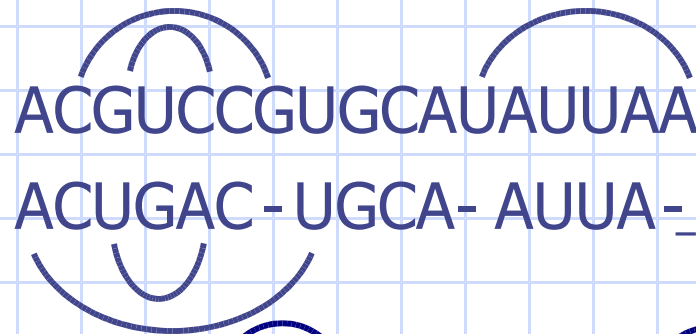
Base-mismatch  
 $W_m$

Base-match

# EDIT(Nested, Nested) ?

[Lin, Ma, Zhang – RECOMB01]

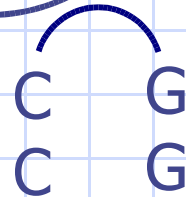
◆ Opérations sur arcs :  $W_{am}$ ,  $W_b$ ,  $W_a$ ,  $W_r$



Arc-mismatch

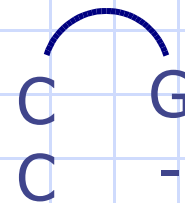
$W_{am}$

Arc-match



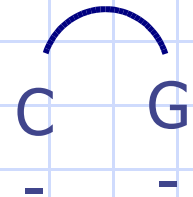
Arc-breaking

$W_b$



Arc-altering

$W_a$

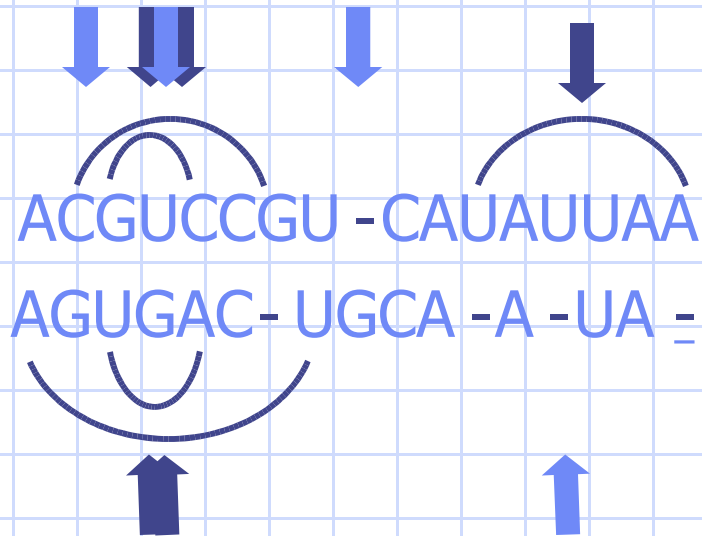


Arc-removing

$W_r$

# Exemple de calcul EDIT(Nested, Nested)

◆  $\theta$  un alignement :



$$\text{Score}(\theta) = W_a + W_r + W_b + W_{am} + 2W_m + 2W_d$$

# Complexité de EDIT

	Unlimited	Crossing	Nested	Plain
Unlimited	Max SNP-Hard			
Crossing		Max SNP-Hard		
Nested			?	$O(nm^3)$ où $m$ taille de séqu. plain
Plain				$O(nm)$

EDIT(Nested, Nested) est NP-Complet

# Problème NP-Complet

## ◆ Deux grandes classes de problèmes:

### ■ P

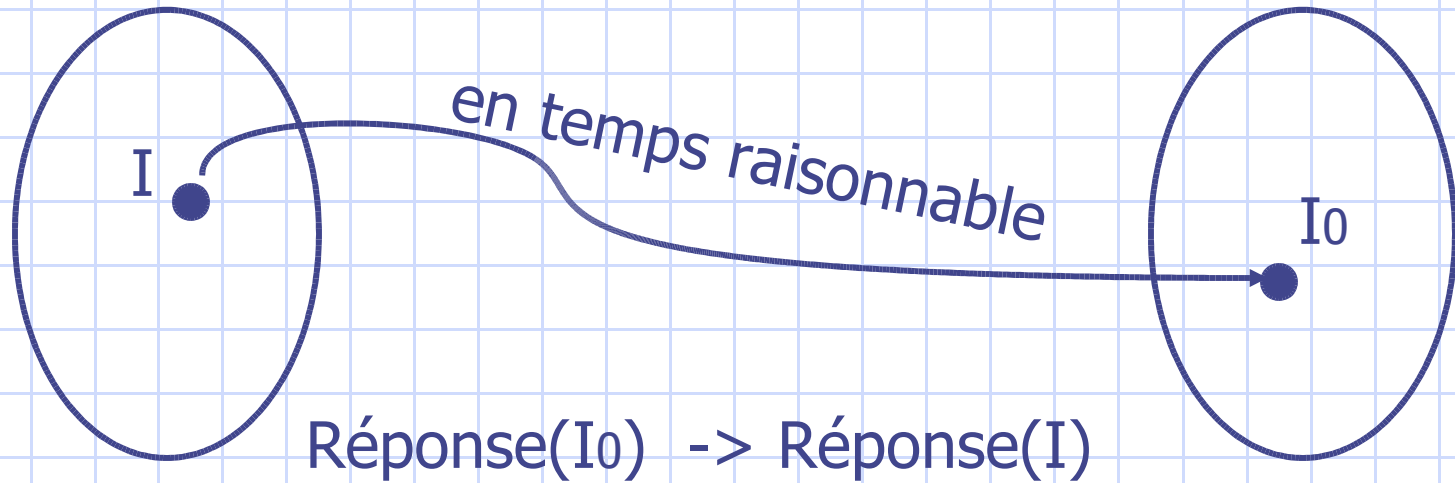
- ◆ Polynomiaux.
- ◆ Réponse exacte en temps raisonnable

### ■ NP

- ◆ Non polynomiaux
- NP-Complet (sous-classe de NP)
  - ◆ Les plus difficiles
  - ◆ Réponse exacte en un temps non raisonnable (en centaines d'années avec la meilleure des machines)



# Preuve de NP-Complétude



Problème difficile

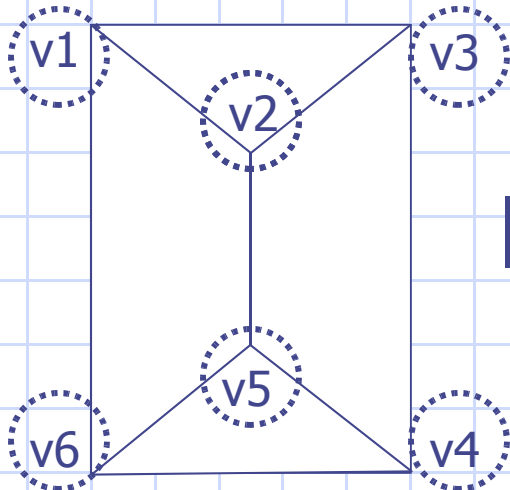
Notre problème  
est difficile

# Complexité de EDIT(Nested, Nested) ?

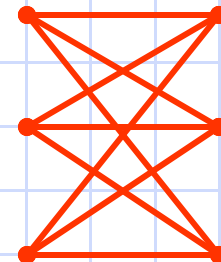
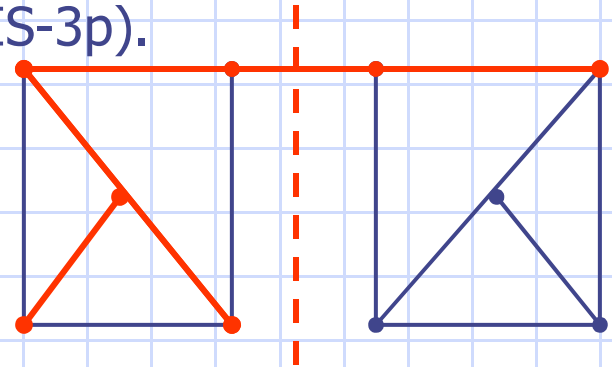
## ◆ Max Independent Set : Ensemble stable de taille maximale

- MIS est NP-Complet sur des graphes cubiques planaires connexes et sans isthme (noté MIS-3p).

G



$| \text{MIS}(G) | \stackrel{?}{\geq} k$

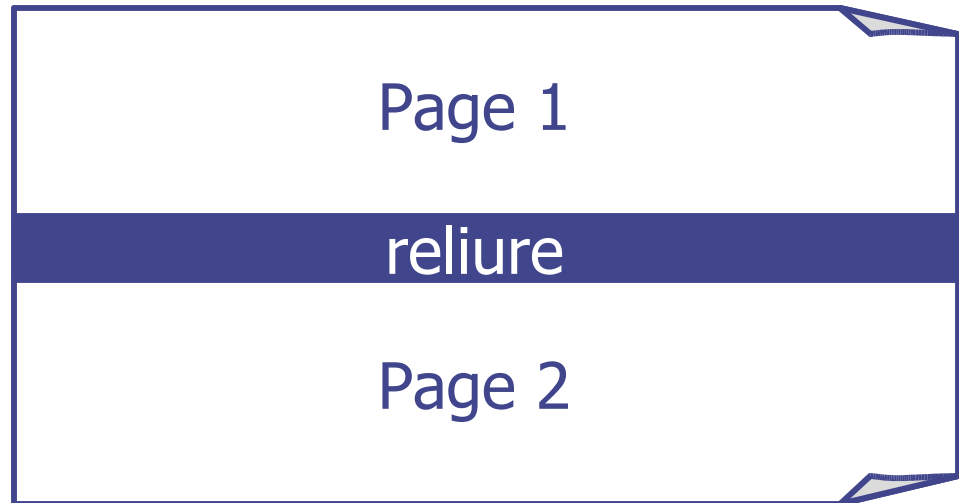
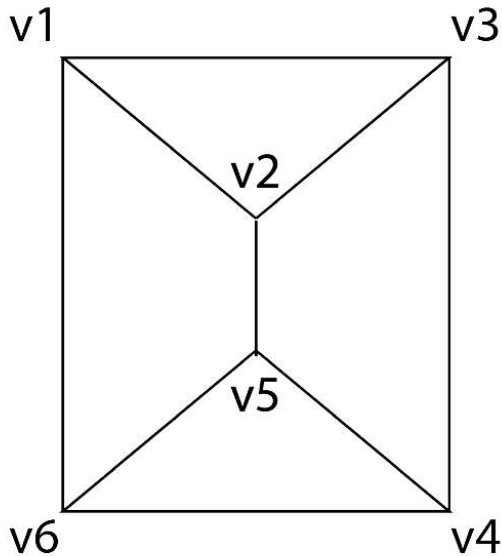


$k_{3,3}$

# Complexité de EDIT(Nested, Nested) ?

◆ MIS-3p -> plongement sur 2 pages

- Chaque sommet n'est pas de degré nul sur une page

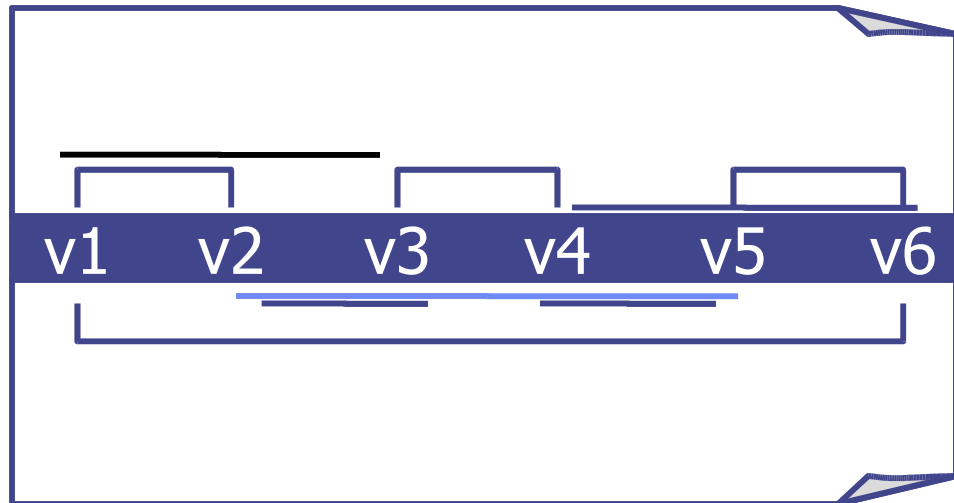
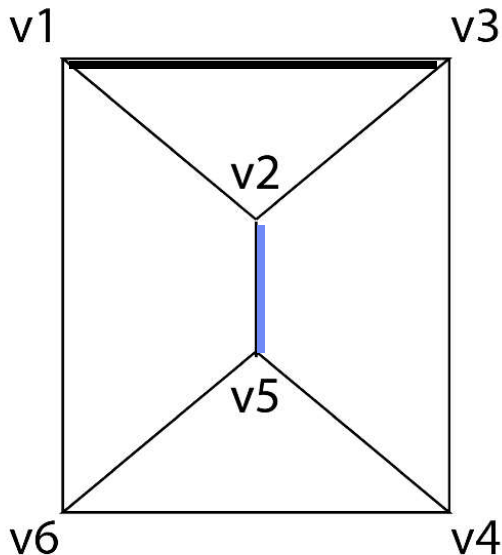


# Complexité de EDIT(Nested, Nested) ?

## ◆ MIS-3p -> plongement sur 2 pages

- Chaque sommet n'est pas de degré nul sur une page
- ∃ algorithme polynomial pour cette classe de graphe

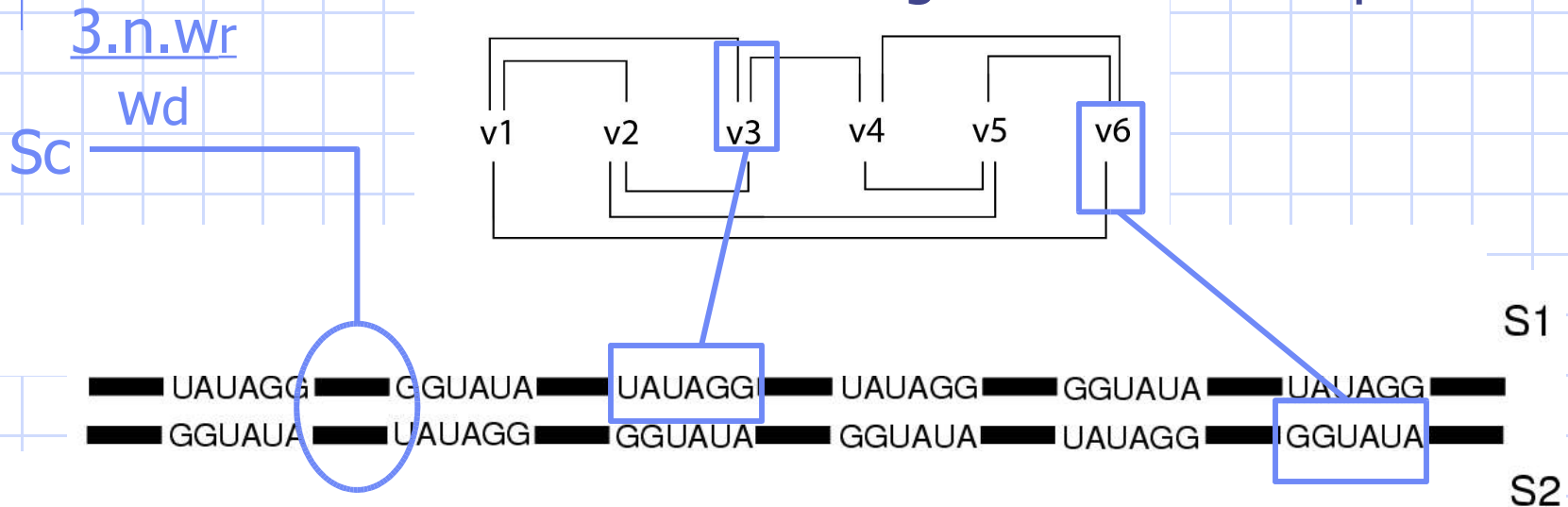
[Lin,Chen,Jiang,Wen – ICALP01]



# Construction

- ◆ Sommet degré 2 : UAUAGG
- ◆ Sommet degré 1 : GGUAUA

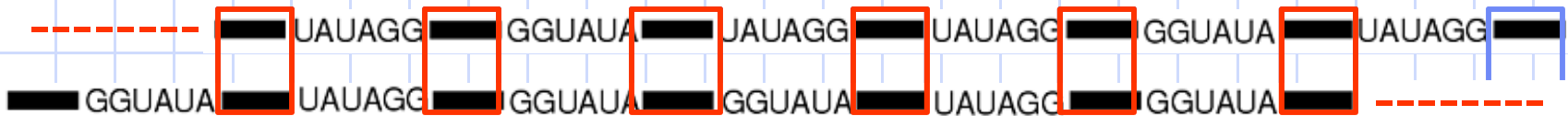
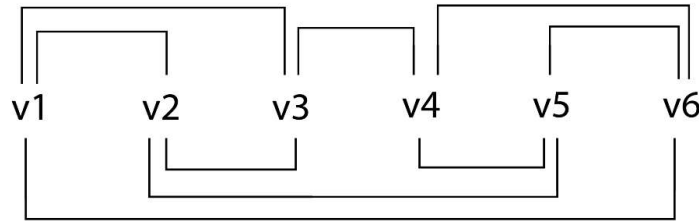
Alignement canonique



# Construction

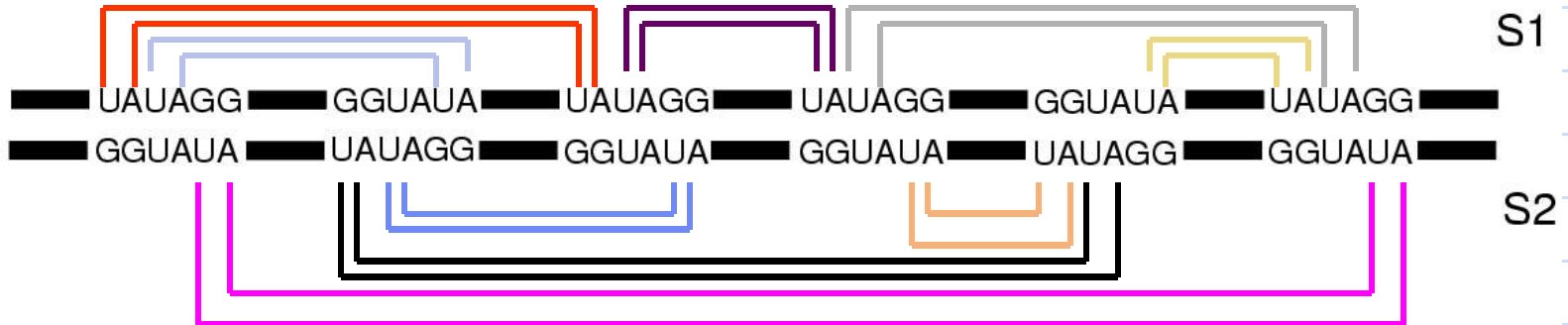
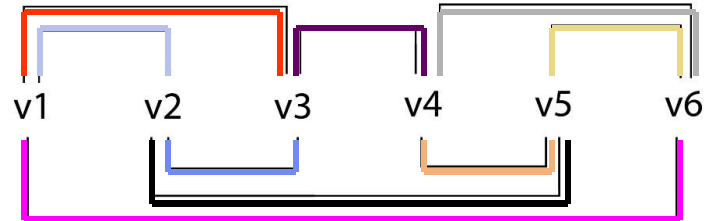
Alignement non canonique = Alignement non optimal  
Alignement avec des substitutions = Alignement non optimal

## Alignement canonique



# Construction

- ◆ Copie la configuration des arcs du plongement en doublant chacun des arcs

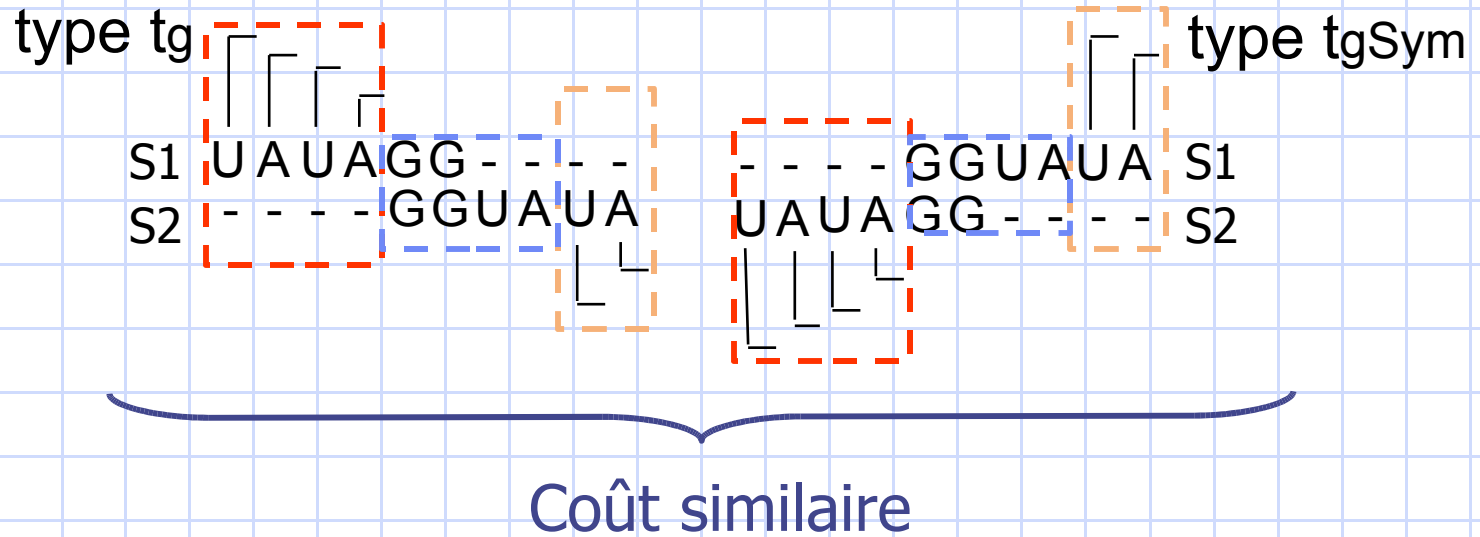


# Les 18 types d'alignements locaux

<p>type tg</p> <pre>                     UAUAGG- - - -       - - - -GGUAUA                         </pre>	<p>type t1</p> <pre>                     UAUAG-G - - - -       - - - -GG-UAUA                         </pre>	<p>type t2</p> <pre>                     UAUAGG- - - - -       - - - - -GGUAUA                         </pre>
<p>type tua</p> <pre>                     - - UAUAGG       GGUAUA - -                         </pre>	<p>type t3</p> <pre>                     - - - UAU - - GG       GGU - - - AUA - -                         </pre>	<p>type t4</p> <pre>                     - - - - UAUAGG       GGUAU - - - A - -                         </pre>
<p>type t5</p> <pre>                     - - - UAU - AGG       GGU - A - UA - -                         </pre>	<p>type t6</p> <pre>                     - - UAU - - - GG       GGU - - - AUA - -                         </pre>	<p>type t7</p> <pre>                     - - - - UAUAGG-       GGUAU - - - - -A                         </pre>
<p>type t8</p> <pre>                     - - UAU - - GG       GGUA - - - UA - -                         </pre>	<p>type t9</p> <pre>                     - - - - UAUAGG       GGUAUA - - - -                         </pre>	<p>type t10</p> <pre>                     - - U - - AUAGG       GGUAUA - - - -                         </pre>
<p>type t11</p> <pre>                     - - - UAUAGG-       GGU - AU - - - A                         </pre>	<p>type t12</p> <pre>                     - - UA - UAGG       GGUAU - A - -                         </pre>	<p>type t13</p> <pre>                     - - - UAUAGG       GGU - AUA - -                         </pre>
<p>type t14</p> <pre>                     - - U - AUAGG-       GGUA - U - - - A                         </pre>	<p>type t15</p> <pre>                     - - UAUAGG-       GGUAU - - - A                         </pre>	<p>type t16</p> <pre>                     - - U - AUAGG       GGUA - UA - -                         </pre>



# Alignements locaux symétriques

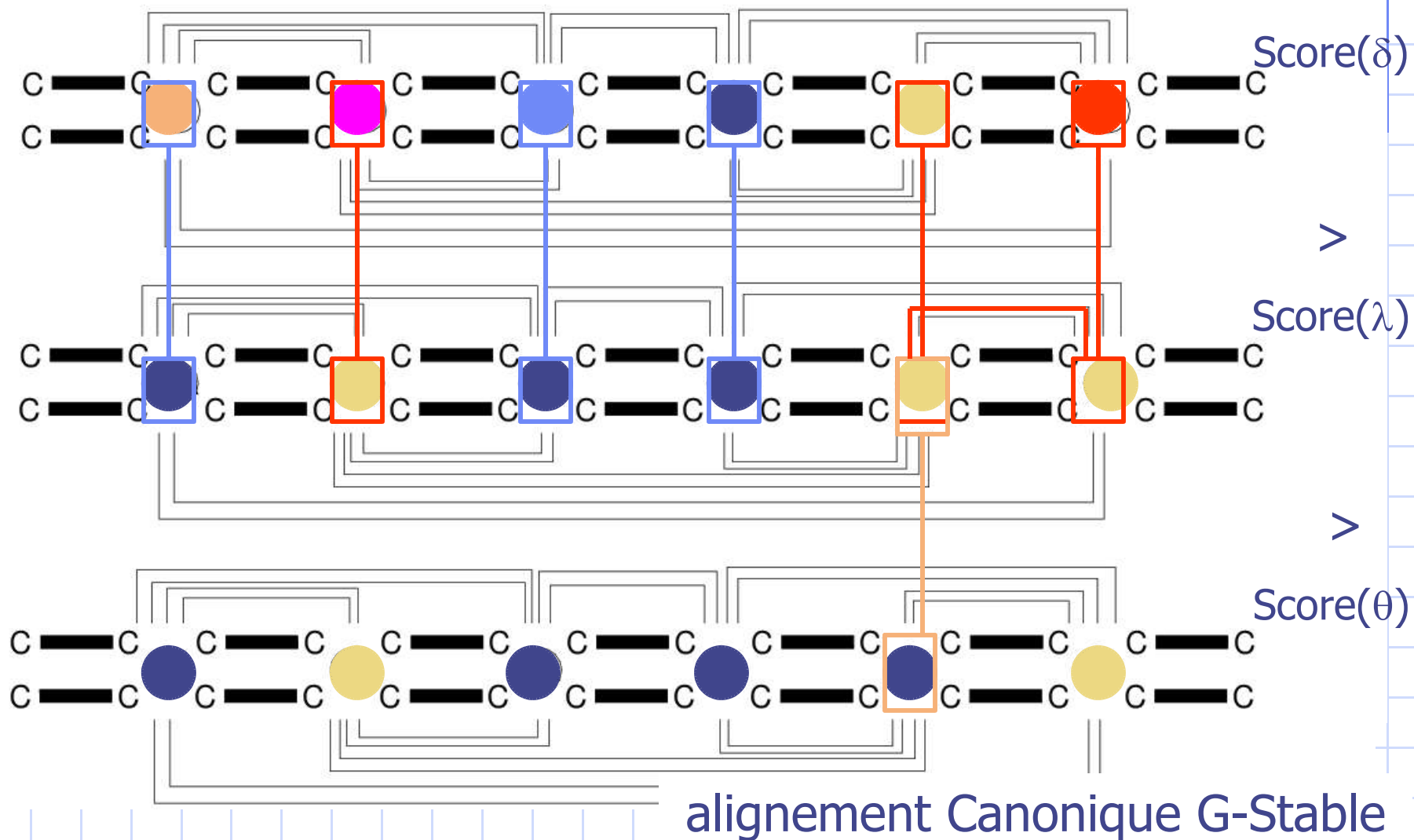


# Démarche

---

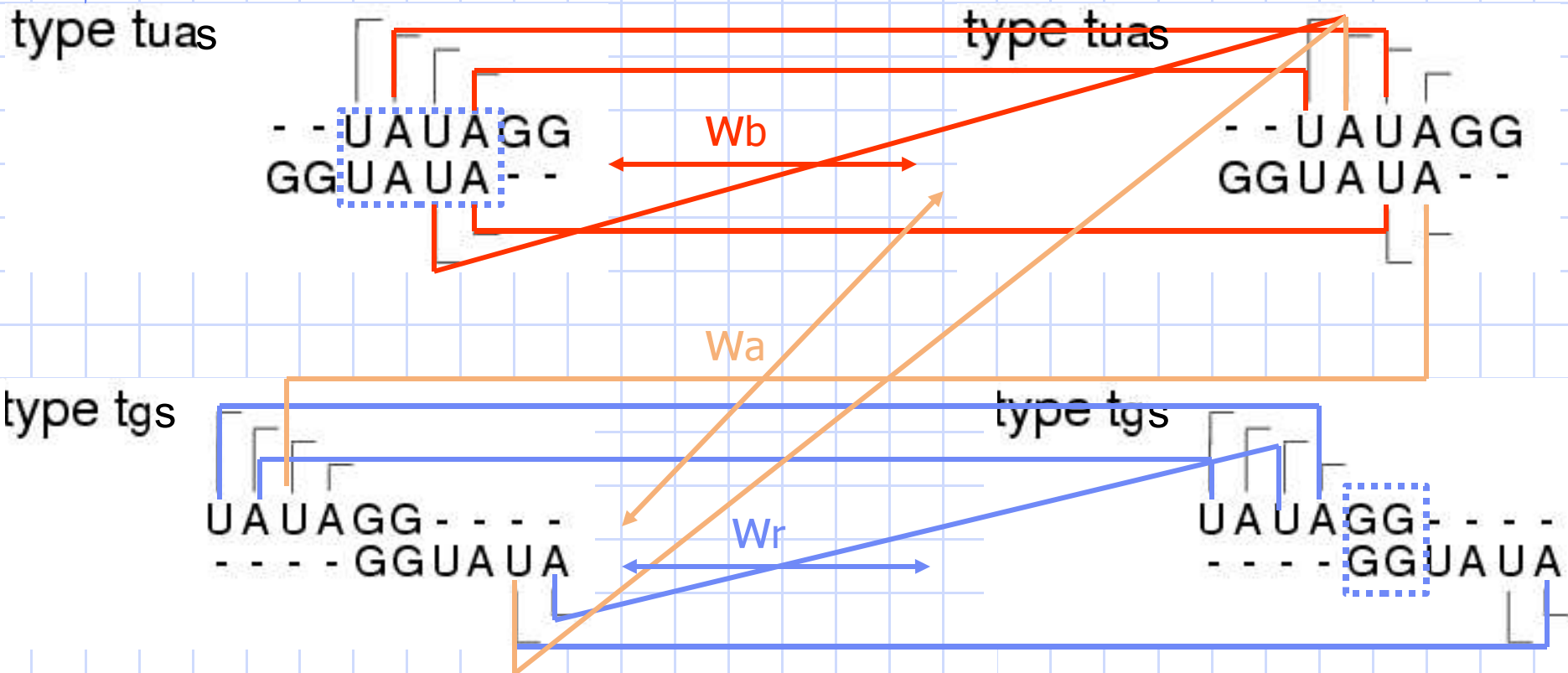
- ◆ Un des 18 types représente les sommets de l'ensemble stable
  - N'avoir que des alignements locaux de deux types
    - ◆ L'autre type représentant les sommets n'appartenant pas à l'ensemble stable

# Démarche



# Choix des deux types

- ◆  $t_{uas}$  et  $t_{gs}$  (le type des sommets de l'ensemble stable)
  - Idée : pénaliser l'arc-removing



# Les Conditions posées par le remplacement

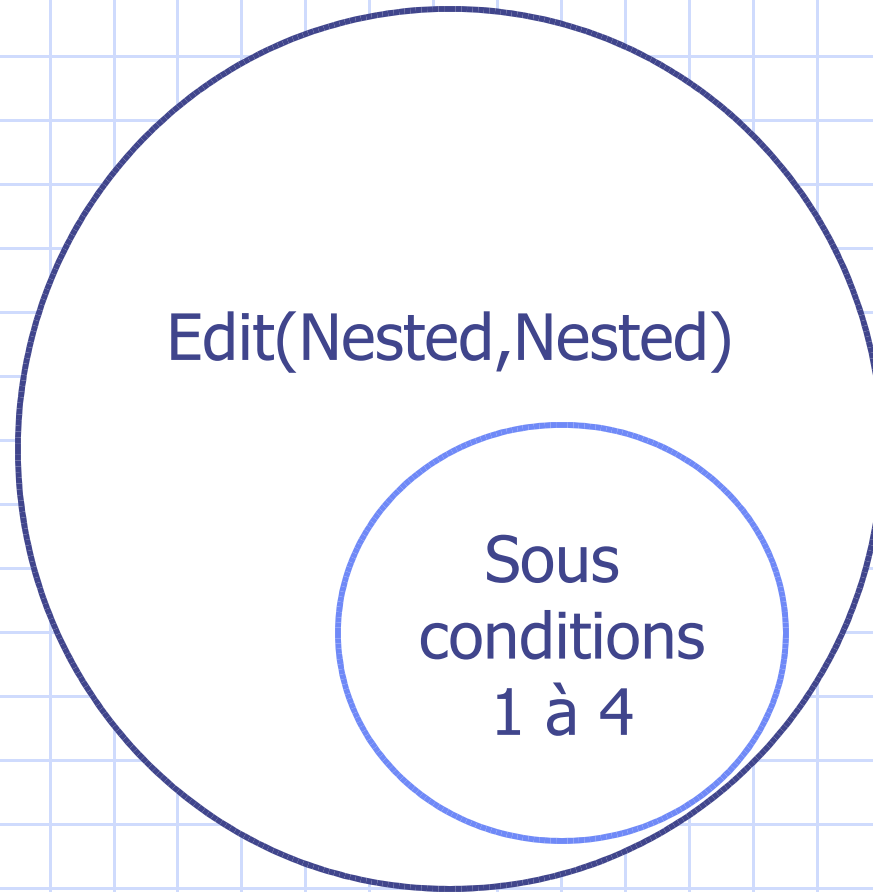
$$w_a > w_b > w_d > 0 \quad (1)$$

$$w_r > w_a + w_d \quad (2)$$

$$w_b + \frac{w_d}{3} > w_a \quad (3)$$

$$w_m > 2w_r \quad (4)$$

# Les Conditions posées par le remplacement



Cas général est difficile



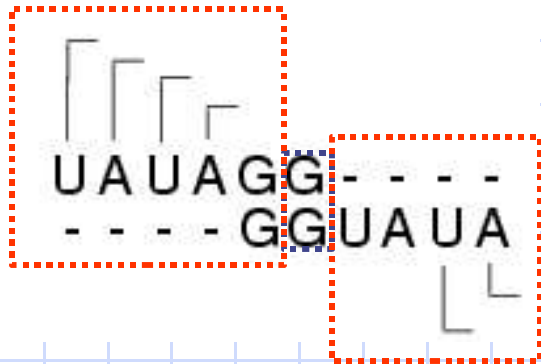
Cas spécifique est difficile

# Exemple

Un alignement contenant un sommet de type t1 n'est pas optimal

$\theta$

type tg



$\varepsilon$

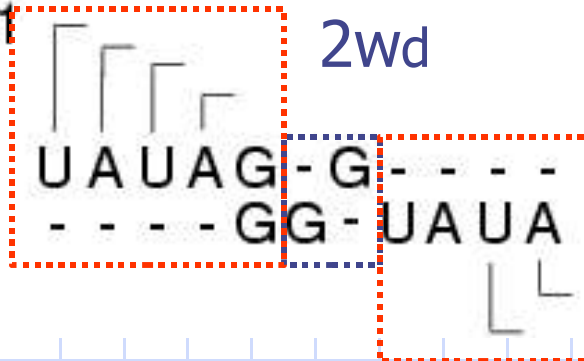
$$\text{Score}(\theta) < \text{Score}(\delta) ?$$

$$0 + \varepsilon < 2wd + \varepsilon ?$$

$$0 < wd$$

$\delta$

type t1

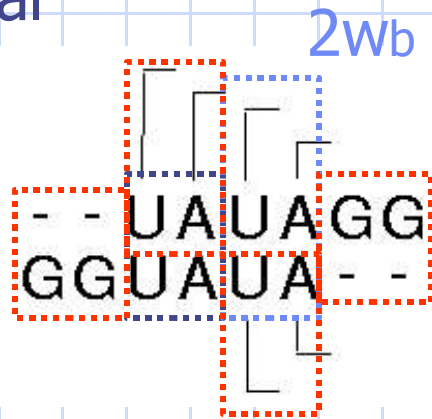


$2wd$

# Exemple

Un alignement contenant un sommet de type t9 n'est pas optimal

$\theta$   
type tua

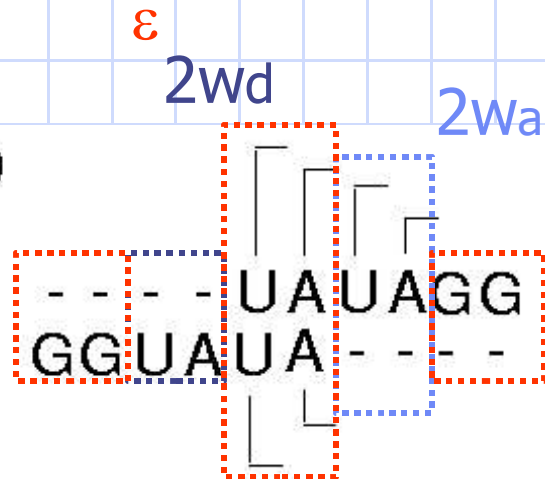


Score( $\theta$ ) < Score( $\delta$ ) ?

$$0 + 2w_b + \epsilon < 2w_d + 2w_a + \epsilon ?$$

$$w_b < w_a$$

$\delta$   
type t9

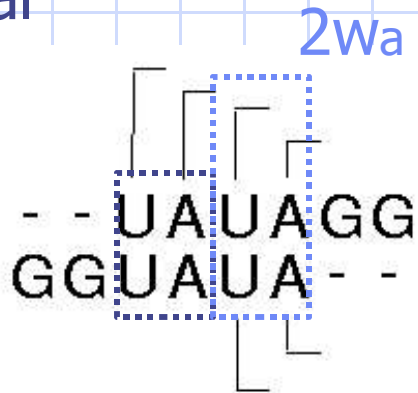




# Exemple

Un alignement contenant un sommet de type t9 n'est pas optimal

$\theta$   
type tua

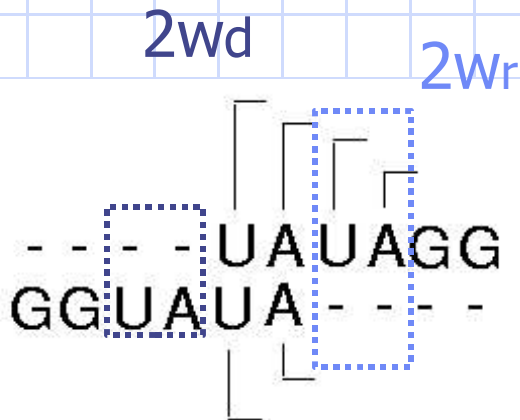


$$\text{Score}(\theta) < \text{Score}(\delta) ?$$

$$0 + 2W_a + \varepsilon < 2W_d + 2W_r + \varepsilon ?$$

$$W_a + W_d < W_r$$

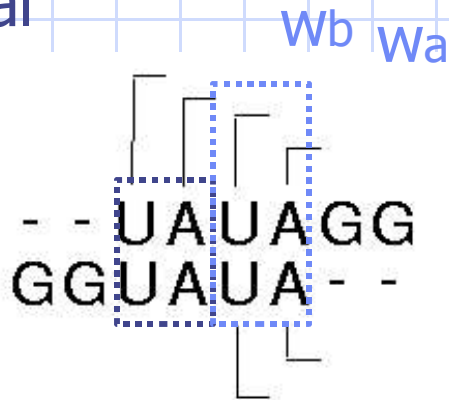
$\delta$   
type t9



# Exemple

Un alignement contenant un sommet de type t9 n'est pas optimal

$\theta$   
type tua

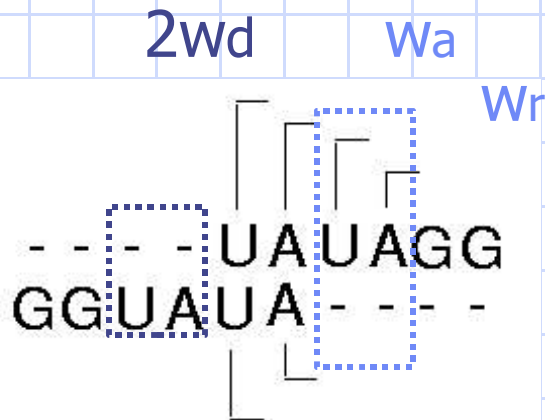


$$\text{Score}(\theta) < \text{Score}(\delta) ?$$

$$0 + W_a + W_b + \varepsilon < 2W_d + W_r + W_a + \varepsilon ?$$

$$W_b < W_r$$

$\delta$   
type t9



# Preuve de l'équivalence des problèmes

On veut maximiser  $k$  dans les deux problèmes

$G$  possède un ensemble  
indépendant  $V'$   
tel que  
 $|V'| \geq k$

$$\text{Edit}(S1, S2) \leq s$$

où  $s =$

$$3n \left( w_b + \frac{4w_d}{3} \right) - k(6w_b + 2w_d - 6w_a)$$

coût de l'alignement  
Canonique G-Stable  
où  $k$  est le nombre  
de sommets  
de type  $t_{gs}$

# Conclusion et perspectives

- ◆ (1) On peut trouver en temps raisonnable une instance de Edit(Nested,Nested) à partir d'un graphe cubique connexe planaire sans isthme
- ◆ (2) Les instances construites sont biologiquement correctes
  - Alphabet {U,C,G,A}
  - les arcs U-A représentent les liens hydrogènes
- ◆ (3) Démarche théorique
  - Plonger un graphe intéressant vers une instance de notre problème
  - Contraindre les coûts pour favoriser deux types de sommets
  - Empêcher les liens entre deux sommets de l'ensemble stable

# Conclusion et perspectives

- ◆ On apporte une réponse au problème ouvert dans [Lin, Ma, Zhang – RECOMB01] en faisant intervenir des notions non triviales issues de domaines diverses.
- ◆ Déterminer la complexité d'autres mesures de similarité sur l'ARN.

# Questions?

## NP-complétude de EDIT(Nested, Nested)

G.Blin, G. Fertin, I. Rusu, C. Sinoquet

Institut de Recherche en  
Informatique de Nantes