

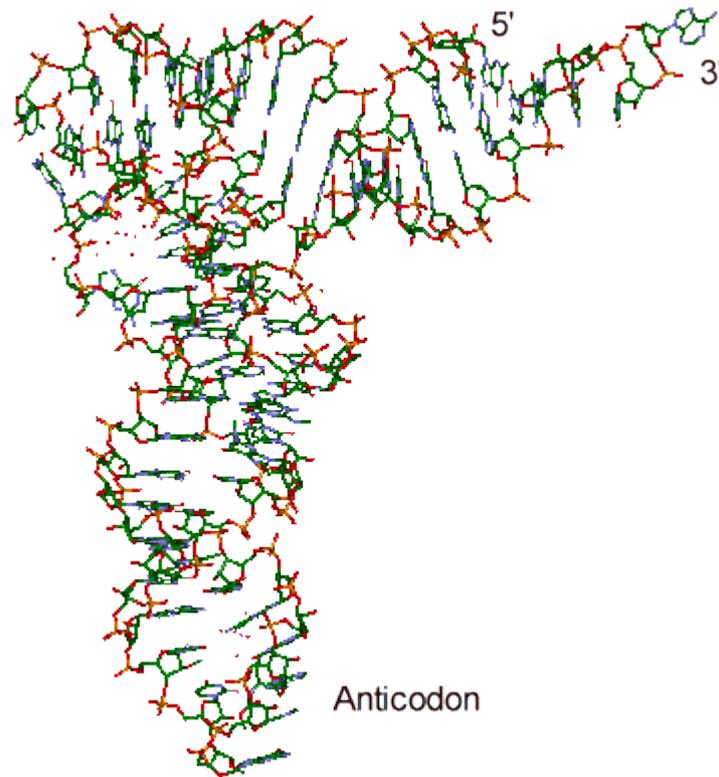
# Énumération de structures ARN avec pseudonoeuds.

**Cédric Saule**, Mireille Régnier, Jean-Marc Steyaert  
Alain Denise.  
(LRI, Université Paris-sud, CNRS, INRIA)

ALEA 2010

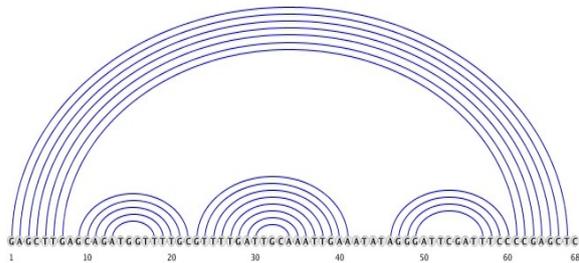
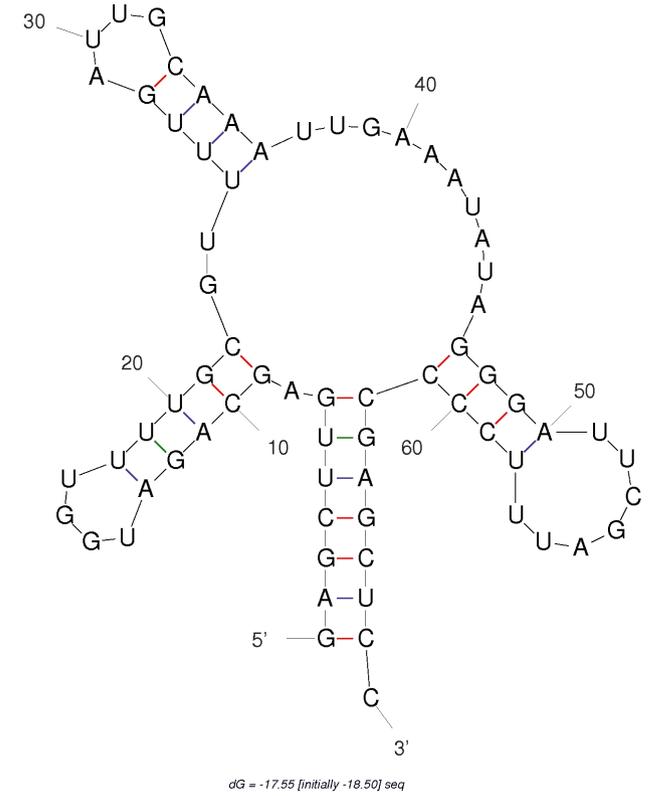
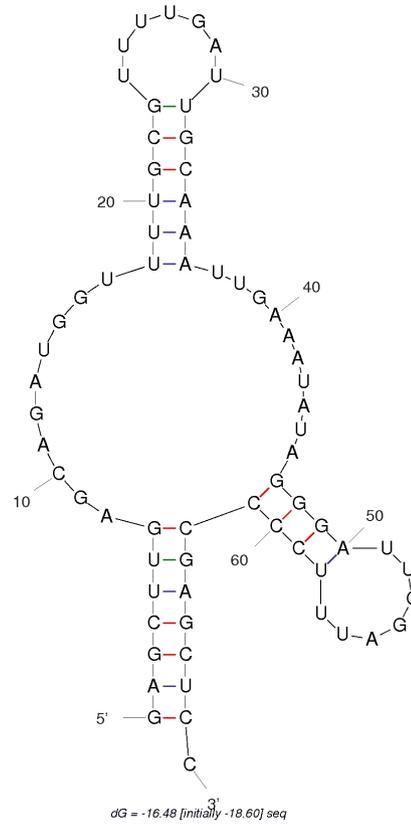
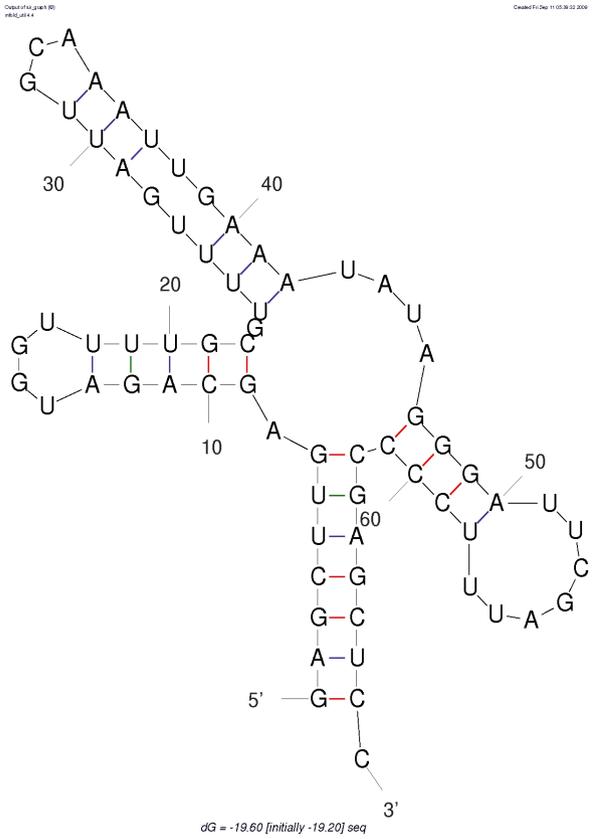
# Prédiction de structures ab initio

GAGCTTGAGCAGATGGTTTTGCGTTTTGATTGCAAATTGAAATATAGGGATTTCGATT  
CCCCGAGCTCC

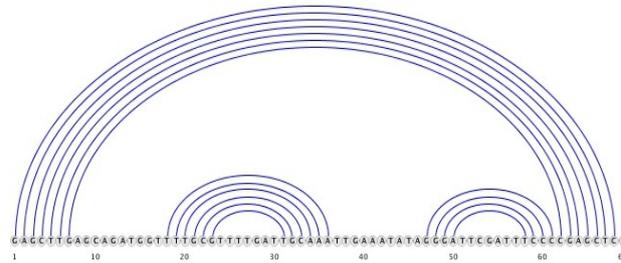


- Recherche la structure thermodynamiquement la plus stable
- Minimisation de l'énergie de repliement
- Exemple : Modèle d'énergie de Turner

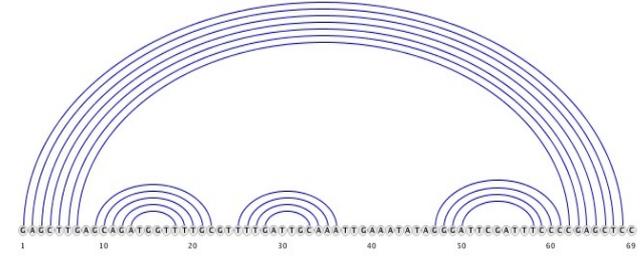
# Structures secondaires sans pseudonoeuds



User defined 3

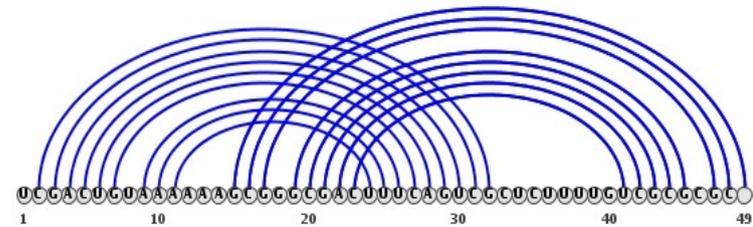
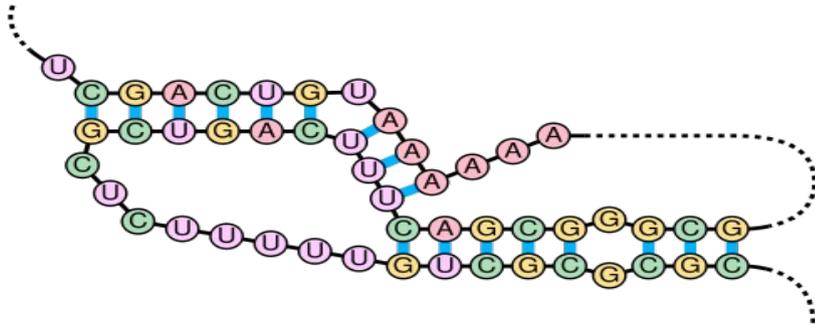


User defined 4



User defined 5

- Structure secondaires avec pseudonoeud



User defined 3

- Prédire les structures secondaires avec pseudonoeud est NP-difficile en toute généralité.
- On contraint les pseudonoeuds.

# Histoire d'algorithmes

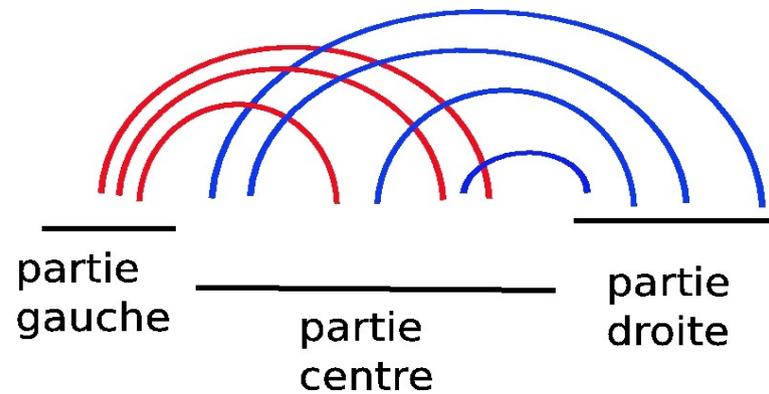
Année	Nom	Complexité	Exact	Pseudonoéuds
1978	Nussinov	$O(n^3)$	Oui	Non
1981	Zuker et Stiegler	$O(n^3)$	Oui	Non
1999	Rivas et Eddy	$O(n^6)$	Oui	Oui
1999	Uemura	$O(n^5)$	Oui	Oui
2000	Lyngsø et Pedersen	$O(n^5)$	Oui	Oui
2000	Akutsu	$O(n^5)$	Oui	Oui
2003	Dirks et Pierces	$O(n^5)$	Oui	Oui
2004	Reeder et Giegerich	$O(n^4)$	Oui	Oui
2009	Cao et Chen	$O(n^6)$	Oui	Oui
2003	Ruan, Stormo et Zhang	$O(n^4)$	Non	Oui
2005	Xayaphoummine et al	$O(n^3)$	Non	Oui
2007	Meyer et Miklos	??	Non	Oui
2008	Chen et al	??	Non	Oui
2008	Parisien et Major	??	Non	Oui

- En 2004 Condon A. et al publient une classification.

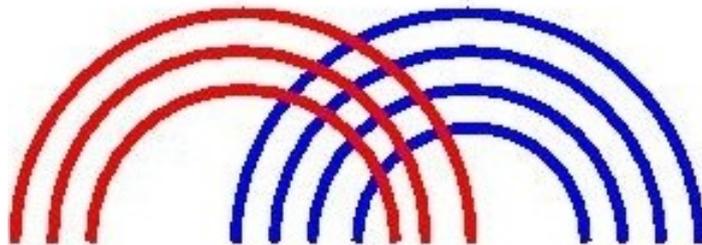
# Présentation des Classes

# Modèle de pseudonoeuds

- Pseudonoeud Simple



- Pseudonoeud de type H

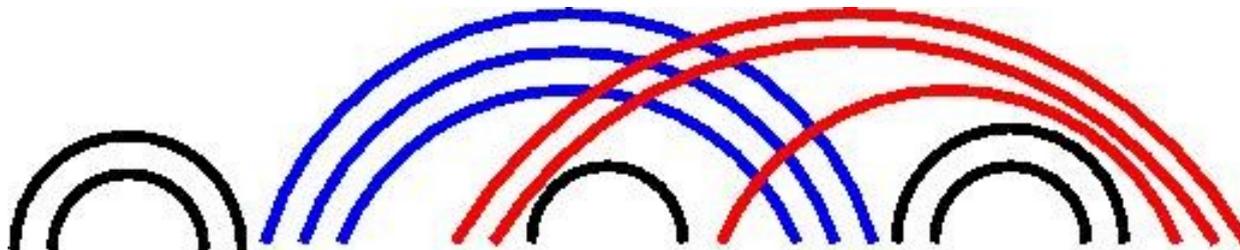


# Lyngso-Pedersen (L&P)

- Structures ne contenant qu'un seul pseudonoeud de type H.

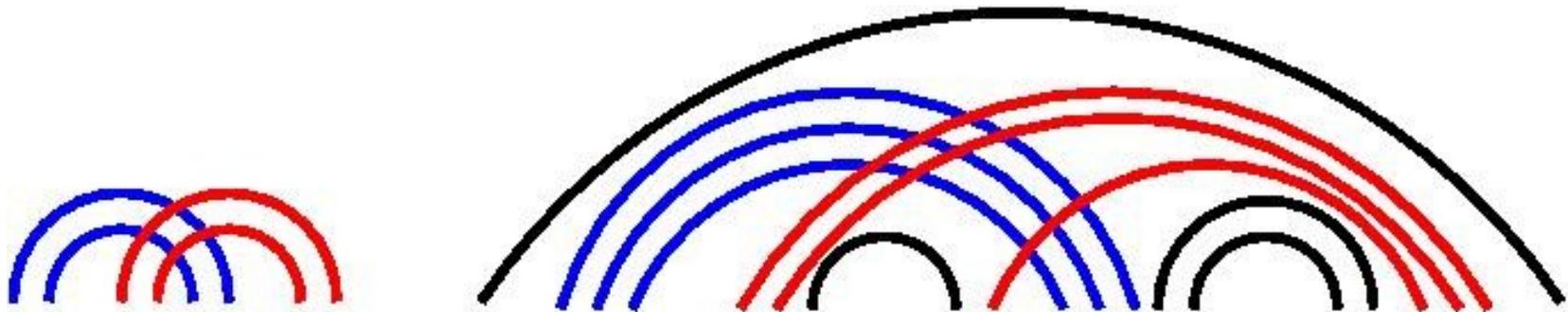


- Le pseudonoeud n'est pas empilé sous une autre structure.



# Dirks-Pierce (D&P), Cao-Chen (C&C), Reeder-Giegerich (R&G)

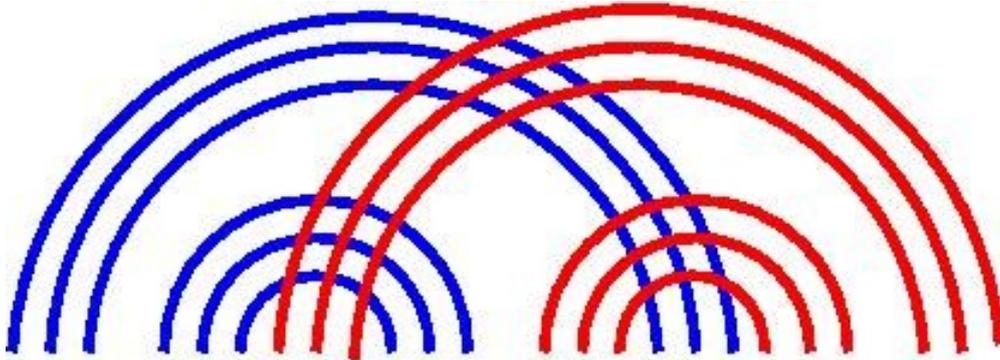
- Nombre et position quelconque de pseudonoeuds de type H.



- Cao et Chen, Reeder et Giegerich : pseudonoeuds similaires à Dirks et Pierce mais plus contraints.

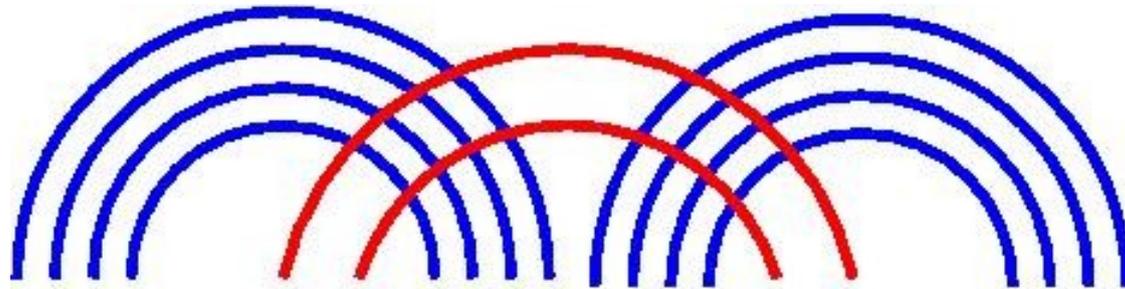
# Akutsu-Uemura (A&U)

- Pseudonoëud simples et récursifs



# Rivas-Eddy (R&E)

- Pseudonoeuds « chevauchants » et certains pseudonoeuds non planaires.



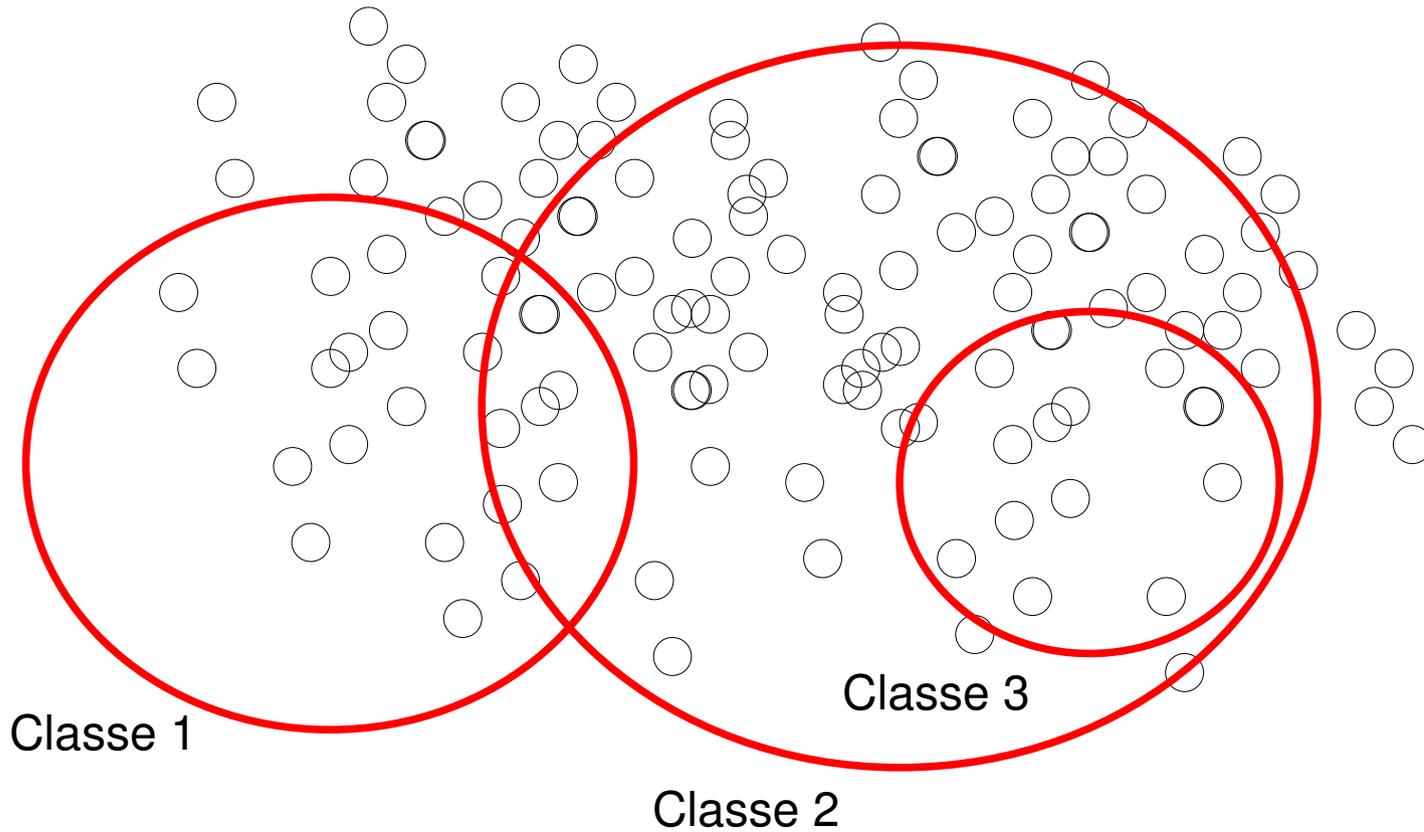
- Une structure optimale peut se trouver dans une classe et pas dans les autres.
- Condon *et al* se posent la question de la relation entre ces classes  $\rightarrow$  PKF  $\subset$  L&P  $\subset$  D&P  $\subset$  A&U  $\subset$  R&E.
- Nous nous posons plus précisément deux questions
  - Comment placer les classes C&C et R&G
  - de la relation entre la complexité et le nombre de structures prédites.

# Quelles(s) classe(s) couvrent le plus de structures potentielles ?

Séquence



Structures compatibles

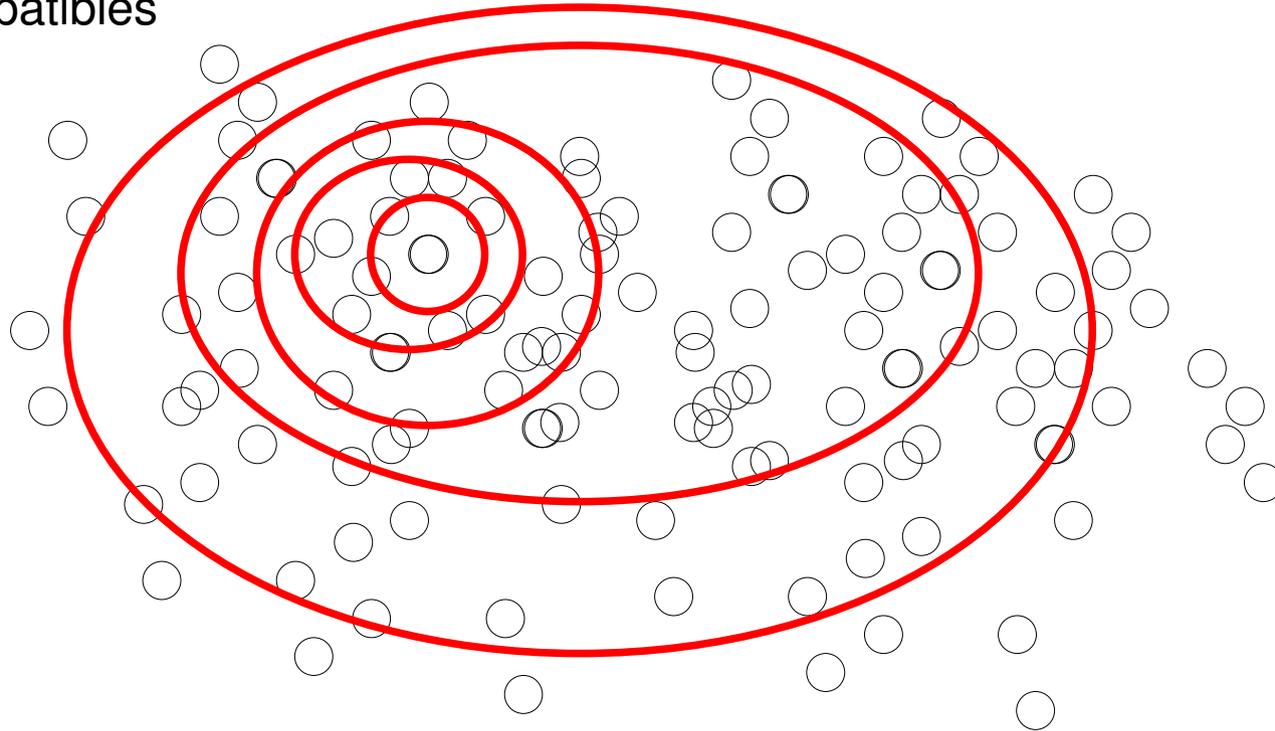


# Condon et al. 2004 : relations d'inclusion entre les classes

Séquence



Structures compatibles



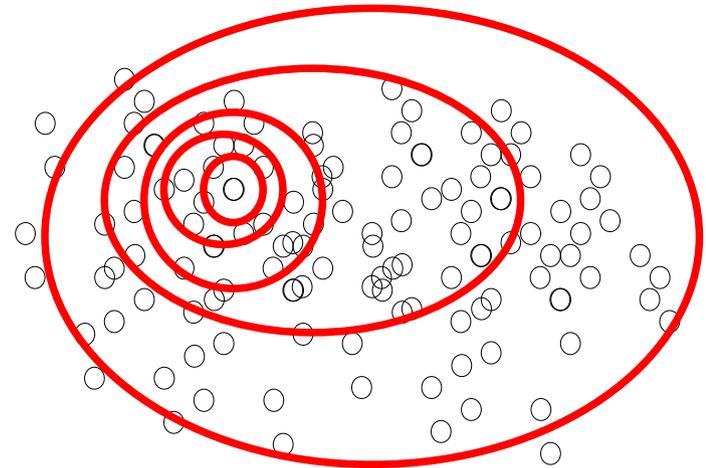
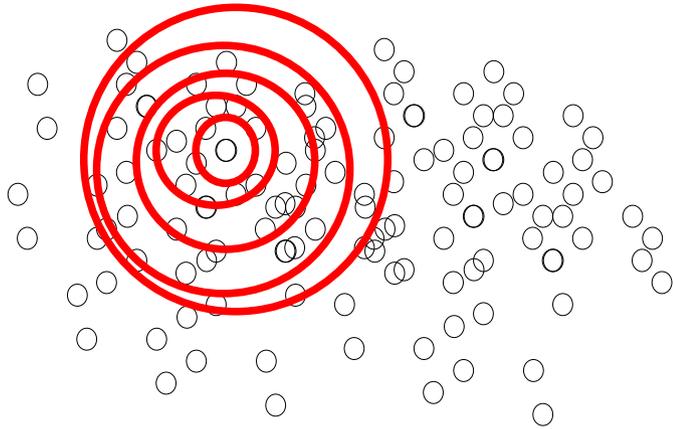
$$\text{PKF} \subset \text{L\&P} \subset \dots \subset \text{P \&E}$$

# Quelles sont les relations de cardinalité entre ces classes ?

Séquence



Structures compatibles



Deux situations très différentes

# Méthodes

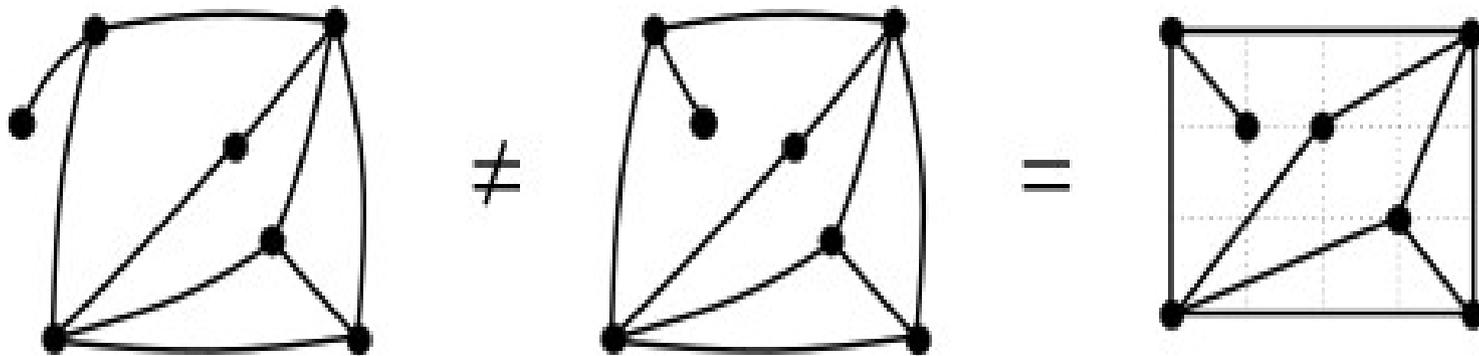
- Etablir des bijections avec des objets combinatoires connus.
- Encodage par des mots décrits par une grammaire algébrique → série génératrice.

# L&P

- Première bijection : Les structures de la classe L&P
- Théorème : Les structures de la classe L&P de ayant  $n$  arcs sont en bijection avec les **cartes planaires enracinées sans isthme** à 1 ou 2 sommets ayant  $n$  arêtes

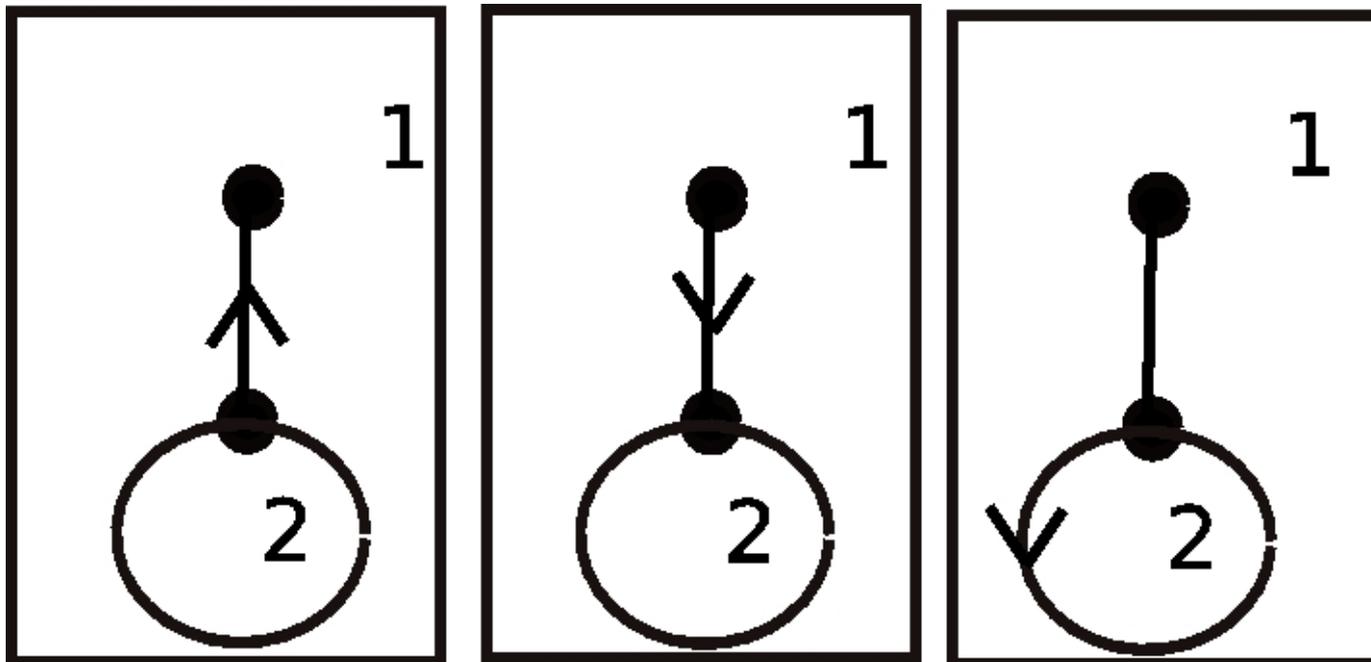
# Cartes planaires.

- Plongement d'un graphe planaire connexe dans le plan
- Les cycles forment des faces internes tandis que le reste du plan forme la face externe



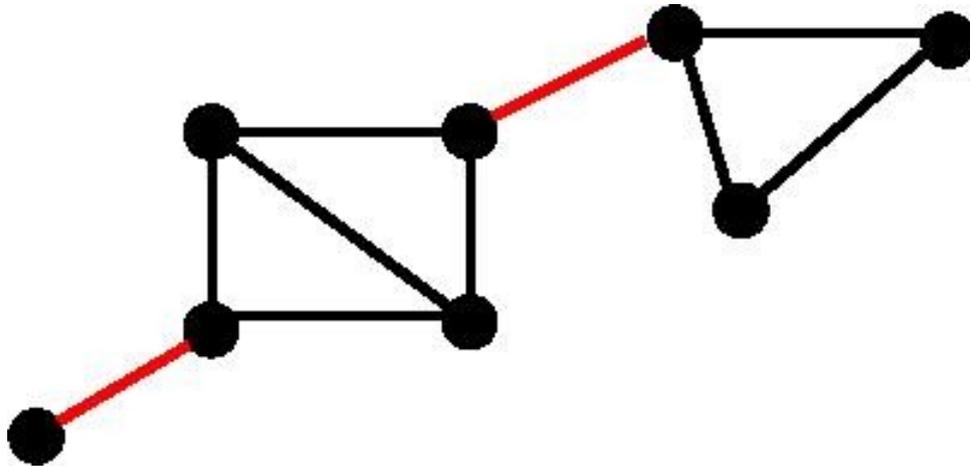
# Cartes planaires.

- On marque un noeud et une arête à l'interface avec la face externe qui donnent l'orientation de la carte. Si l'arc est une boucle, l'enracinement se fait dans le sens anti horaire



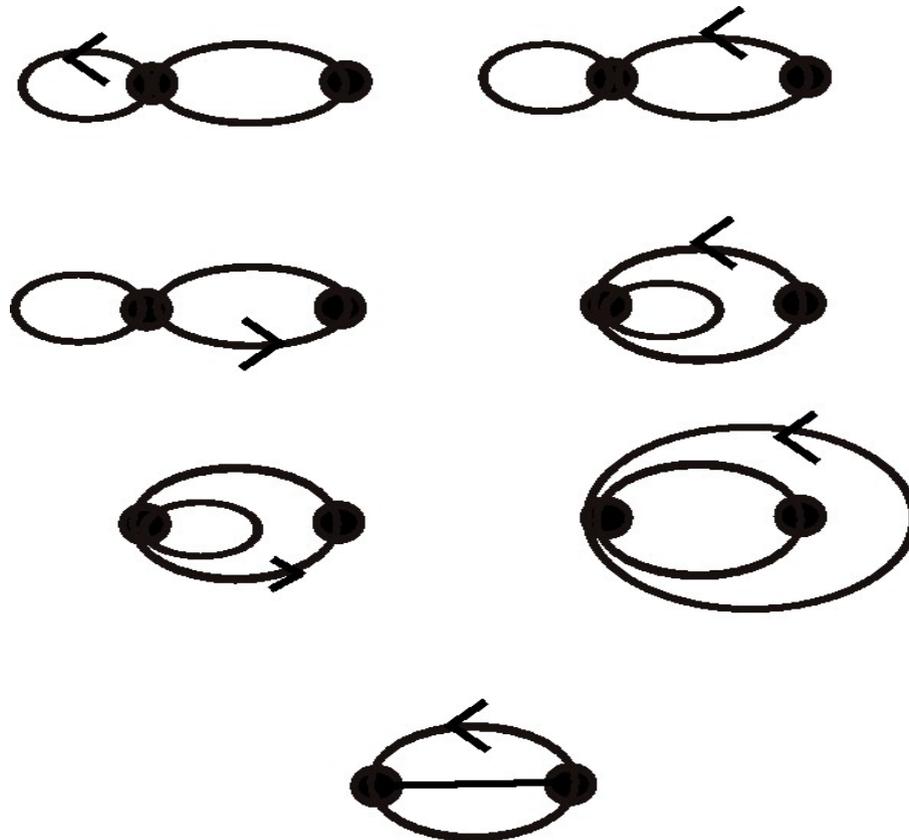
# Cartes planaires.

- Un isthme est un arc qui déconnecte le graphe s'il est brisé.



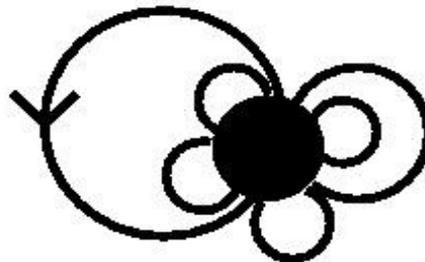
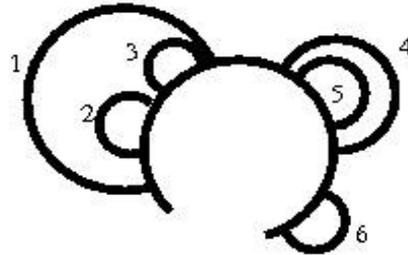
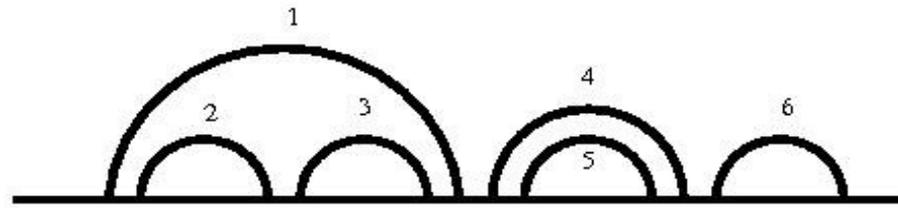
# Cartes planaires

- Les 7 cartes planaires enracinées sans isthmes à deux sommets ayant 3 arêtes.



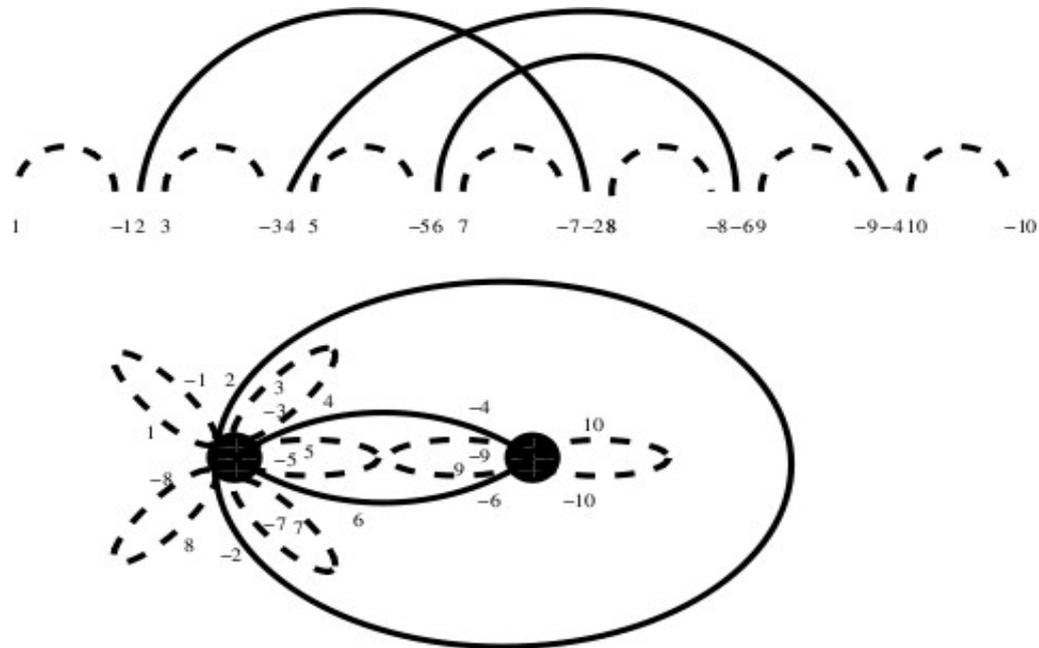
# Bijection entre structure sans pseudonoeud et carte

- Si pas de pseudonoeud, résultat connu. La bijection se fait vers les cartes à 1 sommet.



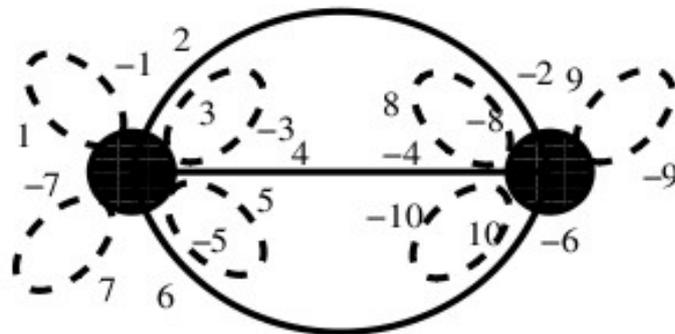
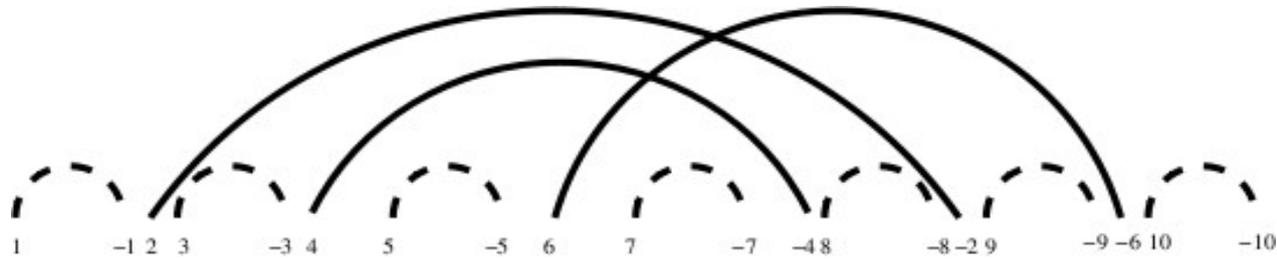
# Bijection structure/carte

- S'il y a un pseudonoeud :
  - Cas 1 : La partie droite du pseudonoeud contient plus d'un arc.



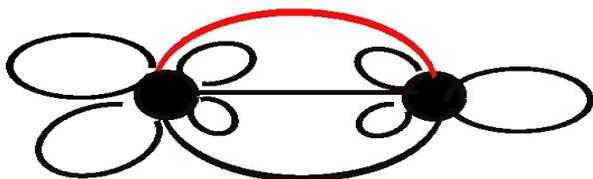
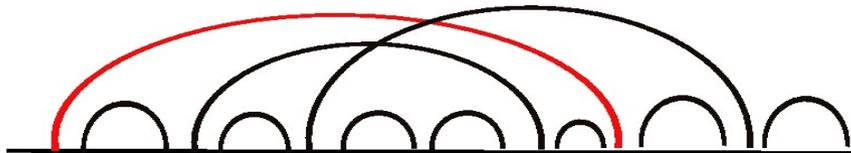
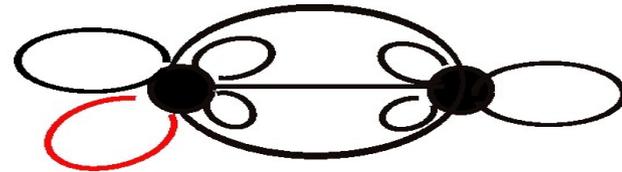
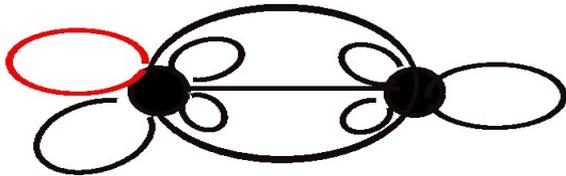
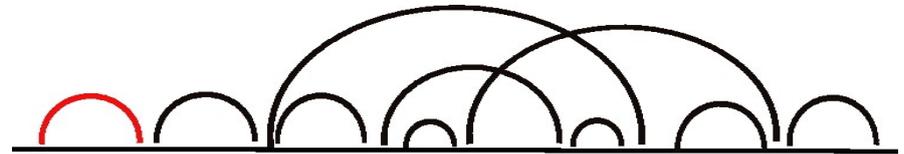
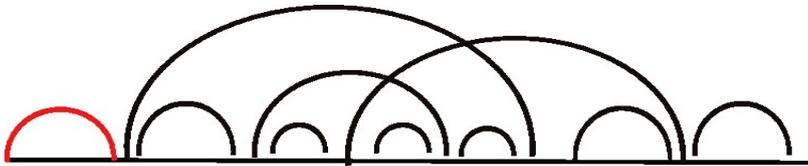
# Bijection structure/carte

- Cas 2 : Il n'y a qu'un arc dans la partie droite du pseudonoeud.



# Bijection structure/carte

- Enracinement de la carte.



...

# Nombre de structures L&P

Exact	Asymptotic
$2^{2n+1} - \binom{2n+3}{n+1} + \binom{2n+1}{n} + \frac{1}{n+1} \binom{2n}{n}$	$2.4^n$

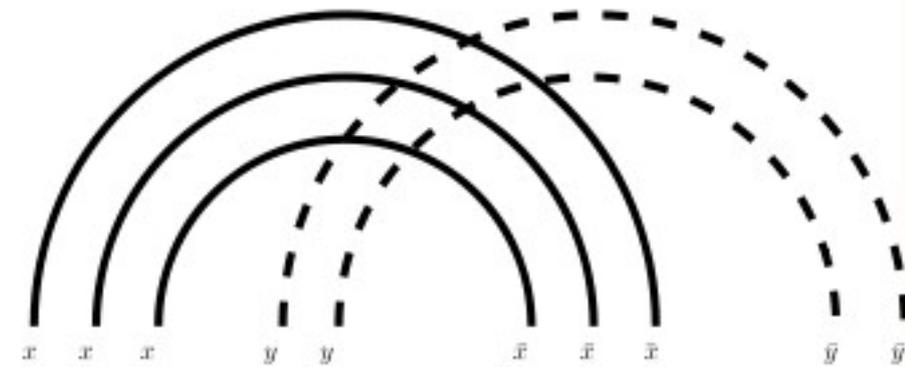
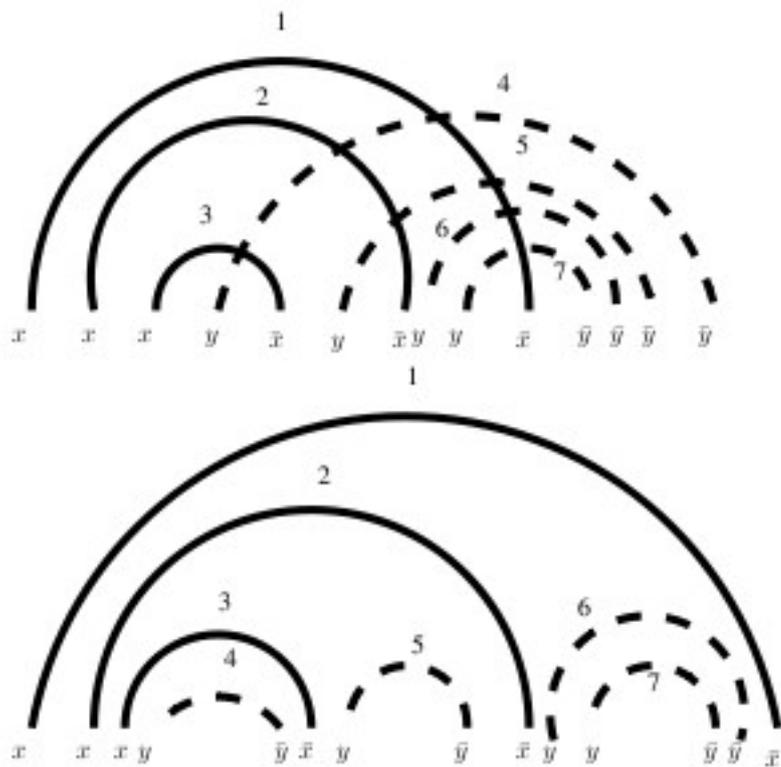


Nombre de structures avec  
pseudonoed

Nombre de structures sans  
pseudonoed

- Comment compter les autres classes ?
  - Encoder les structures par des mots décrits par des grammaires algébriques.
  - Différencier les arcs.
  - Décroiser les structures.
  - Pour cela on va réaliser une deuxième bijection.

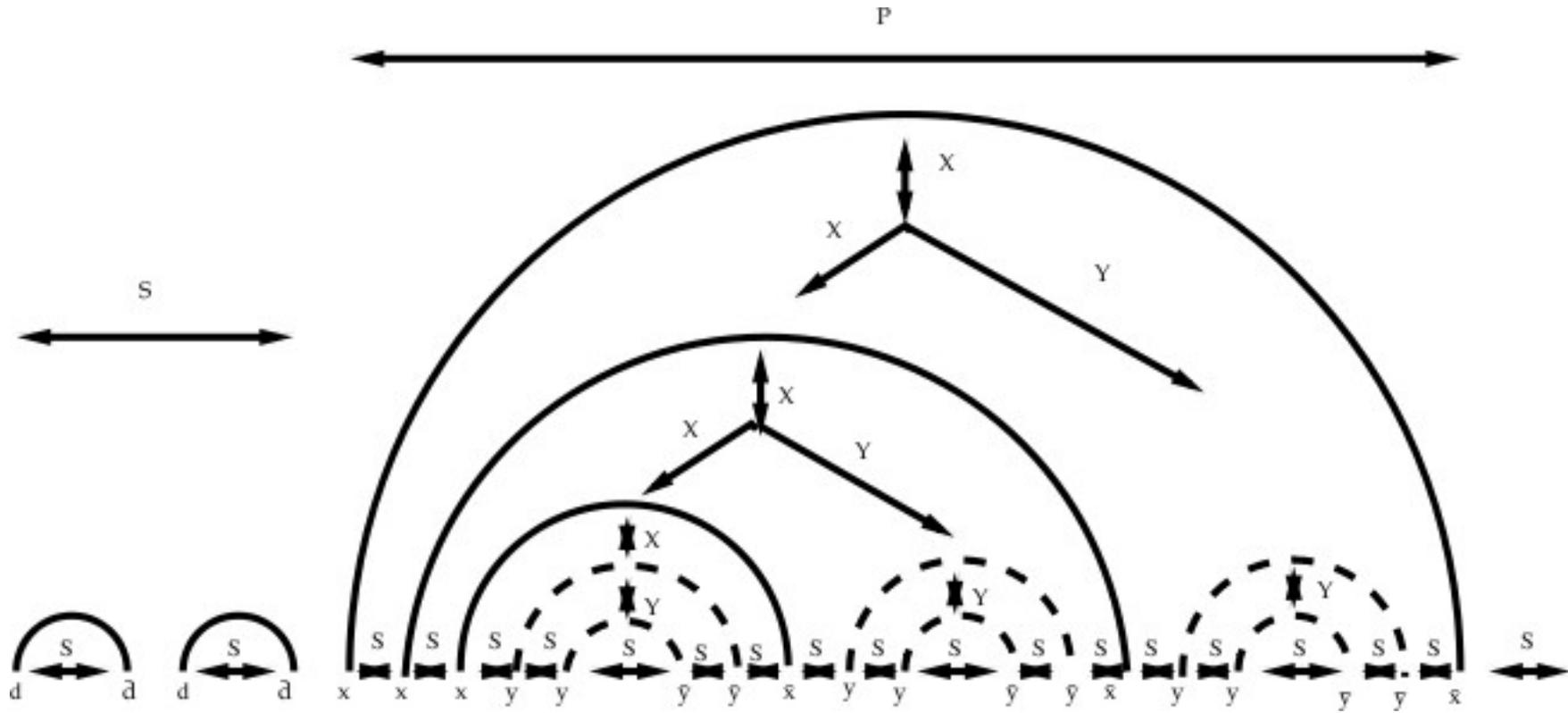
# Bijection des structures croisées vers empilées



(a) Simple pseudoknot

(b) H-type pseudoknot

# Construction de la grammaire



# Grammaire algébrique pour les structures A&U.

$$S \rightarrow dS\bar{d}S|P$$

$$S(z) = zS^2(z) + P(z)$$

$$P \rightarrow xSX\bar{x}S|\epsilon$$

$$P(z) = zS^2(z)X(z) + 1$$

$$X \rightarrow xSX\bar{x}SY|yYS\bar{y}S$$

$$X(z) = zS^2(z)Y(z)(X(z) + 1)$$

$$Y \rightarrow ySY\bar{y}S|\epsilon$$

$$Y(z) = zS^2(z)Y(z) + 1.$$

$$F(z, S) = z^2S^4 - 2zS^3 + zS^2 + S - 1 = 0,$$

$$[z^n]S(z) = \sqrt{\frac{2\rho_1\partial F/\partial z|_{z=\rho_1, S=\sigma_1}}{\partial^2 F/\partial S^2|_{z=\rho_1, S=\sigma_1}}} \frac{1}{2\sqrt{\pi}} \rho_1^{-n} n^{-3/2} (1 + O(1/n)).$$

$$\sigma_1 = 1.403556586\dots$$

$$\rho_1 = 0.1324975681\dots$$

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

- Passage de PKF à L&P : Faible gain en nombre de structures mais saut de deux ordres de complexité. Saut conceptuel (sans / avec pseudonoeuds)
- Passage de C&C à R&G : Gain exponentiel en nombre de structures prédites et gain de deux ordres de complexité. Cependant C&C ne calcule également la fonction de partition selon un modèle élaboré.

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

- Passage de D&P à A&U : Gain exponentiel en nombre de structures sans augmentation de complexité. A&U n'apporte pas de nouvelles structures bio tandis que D&P calcule la fonction de partition.
- Les restrictions des grammaires et le nombre de structures nous apprennent que :  $C\&C \subset R\&G \subset D\&P$

# Remerciements

- Mes co-auteurs :
  - Alain Denise (LRI / IGM Université Paris-Sud, CNRS, INRIA)
  - Mireille Régnier (LIX Ecole Polytechnique and CNRS, INRIA)
  - Jean-Marc Steyaert (LIX, Ecole Polytechnique and CNRS, INRIA)
- Equipe bio-informatique du LRI :
  - Dirigée par Christine Froidevaux (LRI Université Paris-Sud, CNRS, INRIA)