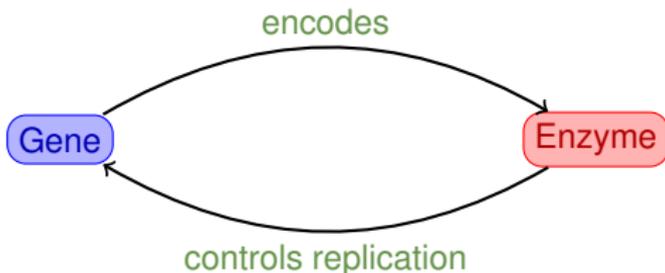


Optimisation énergétique et évolution

Bioinformatique des ARN

Yann Ponty

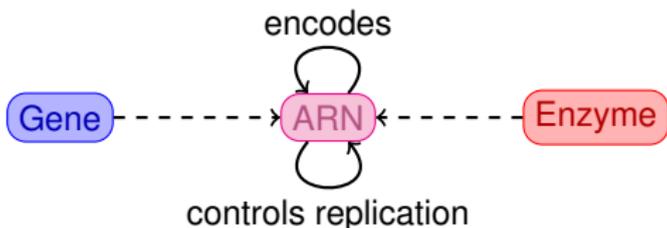
École Polytechnique/CNRS/INRIA AMIB – France



“ A gene big enough to specify an enzyme would be too big to replicate accurately without the aid of an enzyme of the very kind that it is trying to specify. So the system *apparently cannot get started*.

[...] This is the *RNA World*. To see how plausible it is, we need to look at why proteins are good at being enzymes but bad at being replicators ; at why DNA is good at replicating but bad at being an enzyme ; and finally why *RNA might just be good enough at both roles to break out of the Catch-22* ”

R. Dawkins, *The Ancestor's Tale : A Pilgrimage to the Dawn of Evolution*

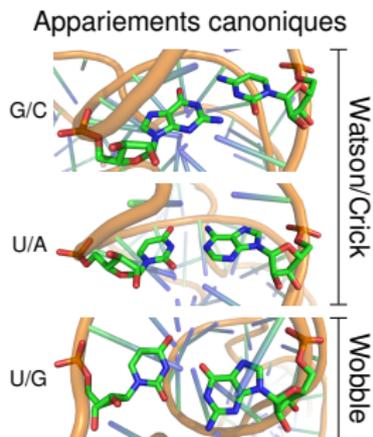
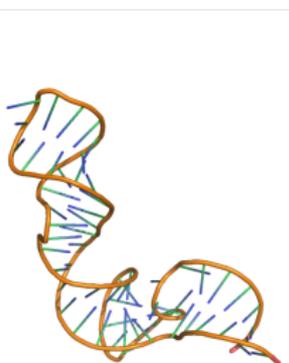


“ A gene big enough to specify an enzyme would be too big to replicate accurately without the aid of an enzyme of the very kind that it is trying to specify. So the system *apparently cannot get started*.

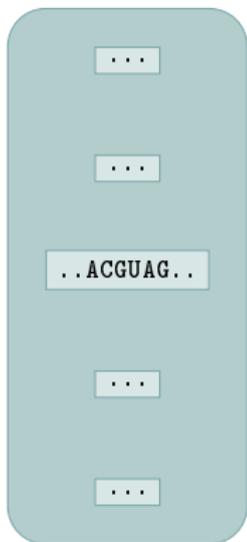
[...] This is the **RNA World**. To see how plausible it is, we need to look at why proteins are good at being enzymes but bad at being replicators ; at why DNA is good at replicating but bad at being an enzyme ; and finally why *RNA might just be good enough at both roles to break out of the Catch-22.*”

R. Dawkins, *The Ancestor's Tale : A Pilgrimage to the Dawn of Evolution*

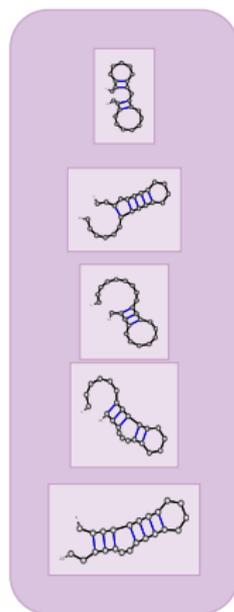
ARN = Biopolymère composé de nucléotides A,C,G et U
A : Adénosine, C : Cytosine, G : Guanine et U : Uracile



Repliement de l'ARN = Processus stochastique continu dirigé par (résultant en) un appariement des nucléotides.

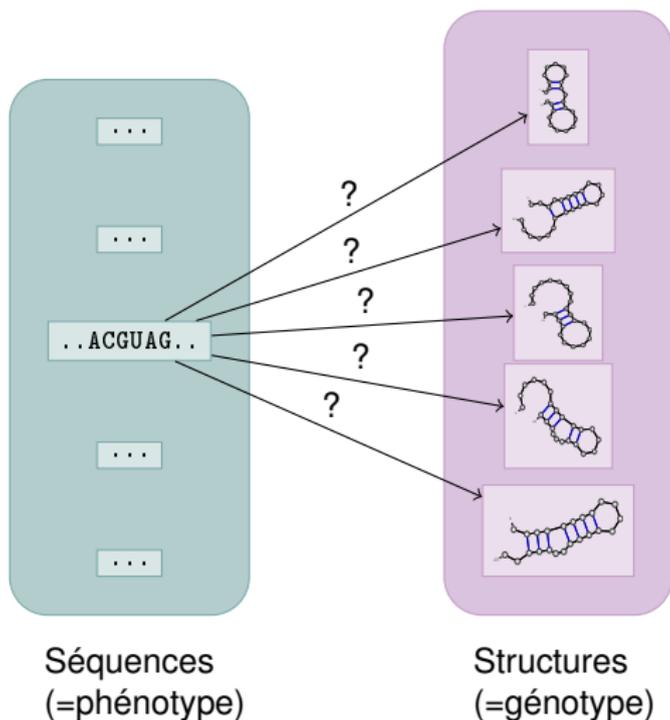


Séquences
(=phénotype)



Structures
(=génotype)

Prédiction du repliement : Quelle(s) structure(s) pour une séquence donnée ?

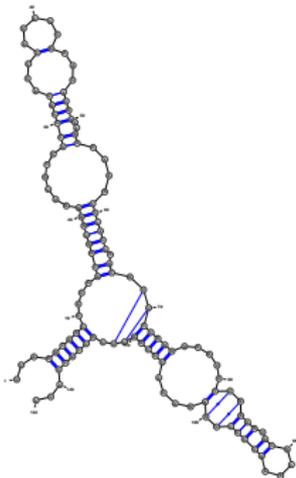


Prédiction du repliement : Quelle(s) structure(s) pour une séquence donnée ?

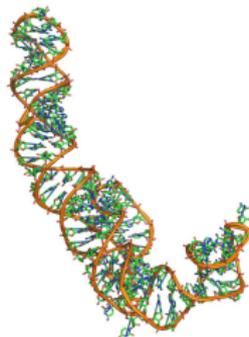
Trois ¹ niveaux de représentation :

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCAUCCGAA
CACGGAAGUAAGCC
CACCAGCGUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCCA
CC
```

Structure primaire



Structure secondaire



Structure tertiaire

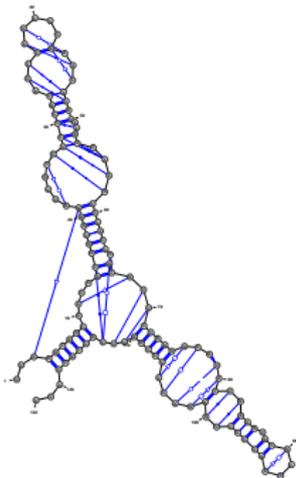
Source : 5s rRNA (PDB 1K73 :B)

1. Enfin, presque ...

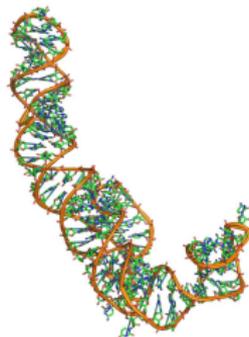
Trois ¹ niveaux de représentation :

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCAUCCGAA
CACGGAAGUAAGCC
CACCAGCGUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCA
CC
```

Structure primaire



Structure secondaire⁺



Structure tertiaire

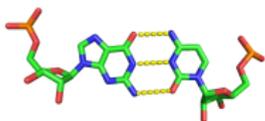
Source : 5s rRNA (PDB 1K73 :B)

1. Enfin, presque ...

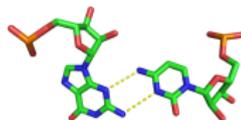
- Appariements non-canoniques

Toute paire de base **autre que** {(A-U), (C-G), (G-U)}

Ou interagissant sur un bord non-standard (WC/WC-Cis) [LW01].



Paire CG canonique (WC/WC-Cis)



Paire CG non canonique (Sucre/WC-Trans)

- Pseudonoeads



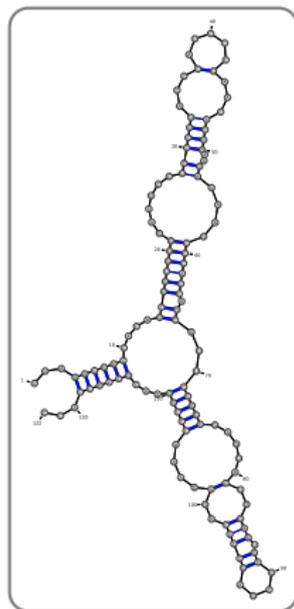
Structure pseudonoead d'un Ribozyme du Groupe I (PDBID : 1Y0Q :A)

Plus expressif, mais repliement général *in silico* avec pseudonoead :

⇒ NP-Compleat [LP00] ... polynomial pour certaines classes [CDR⁺04].

Basée sur décomposition de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples
- Empilements

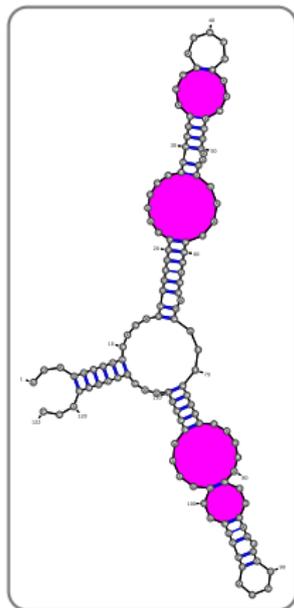
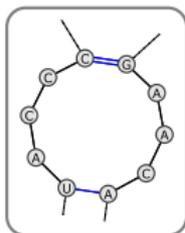


Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) [ZS81] ...

Par une stratégie **diviser pour régner**, on obtient le **repliement d'énergie libre minimale** en temps $\mathcal{O}(n^3)$ ($\sim 1.8^n$ str. sec. compatibles [ZS84])

Basée sur décomposition de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples
- Empilements

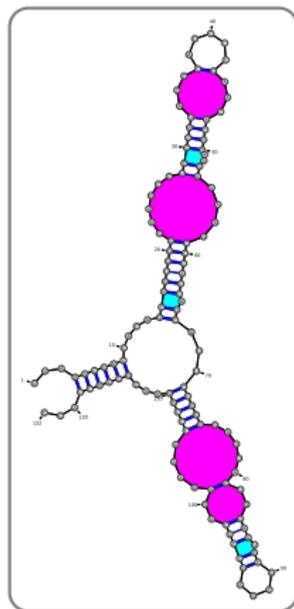
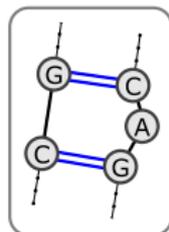


Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) [ZS81] ...

Par une stratégie **diviser pour régner**, on obtient le **repliement d'énergie libre minimale** en temps $\mathcal{O}(n^3)$ ($\sim 1.8^n$ str. sec. compatibles [ZS84])

Basée sur décomposition de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples
- Empilements

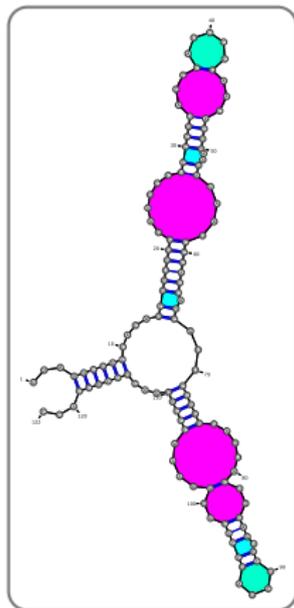
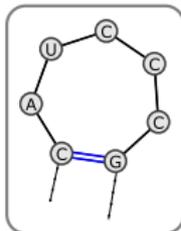


Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) [ZS81] ...

Par une stratégie **diviser pour régner**, on obtient le **repliement d'énergie libre minimale** en temps $\mathcal{O}(n^3)$ ($\sim 1.8^n$ str. sec. compatibles [ZS84])

Basée sur décomposition de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples
- Empilements

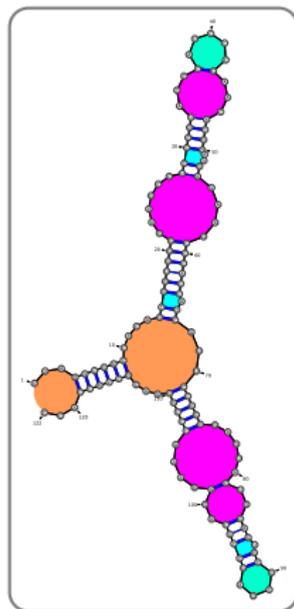
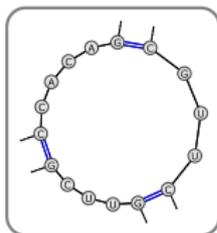


Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) [ZS81] ...

Par une stratégie **diviser pour régner**, on obtient le **repliement d'énergie libre minimale** en temps $\mathcal{O}(n^3)$ ($\sim 1.8^n$ str. sec. compatibles [ZS84])

Basée sur décomposition de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples
- Empilements

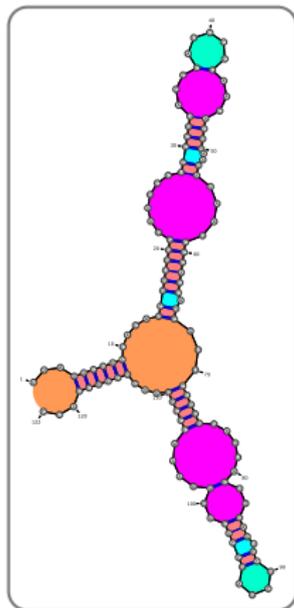
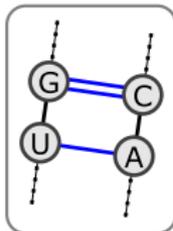


Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) [ZS81] ...

Par une stratégie **diviser pour régner**, on obtient le **repliement d'énergie libre minimale** en temps $\mathcal{O}(n^3)$ ($\sim 1.8^n$ str. sec. compatibles [ZS84])

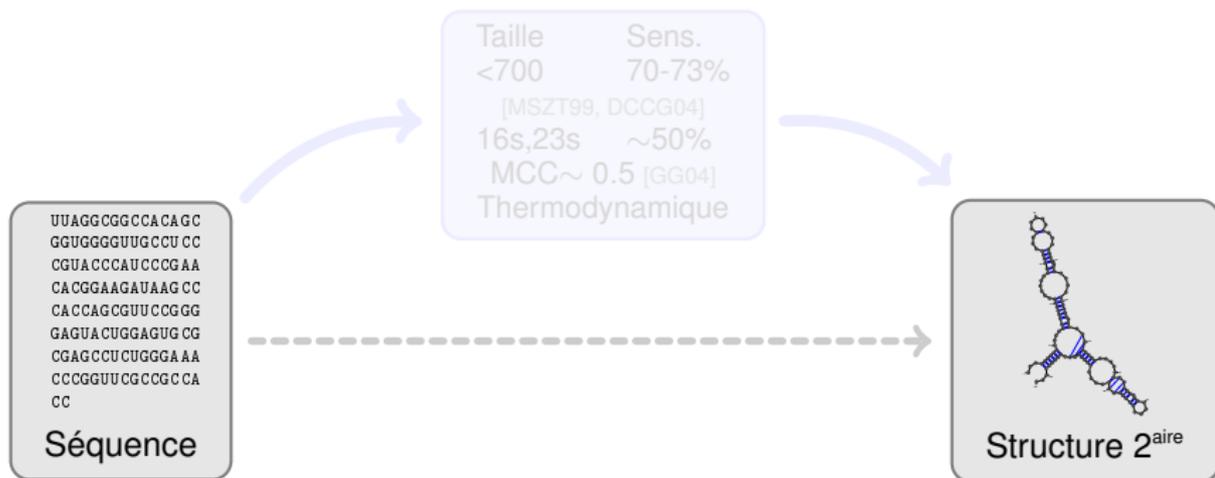
Basée sur décomposition de la structure 2^{aire} :

- Boucles internes
- Renflements
- Boucles terminales
- Boucles multiples
- Empilements

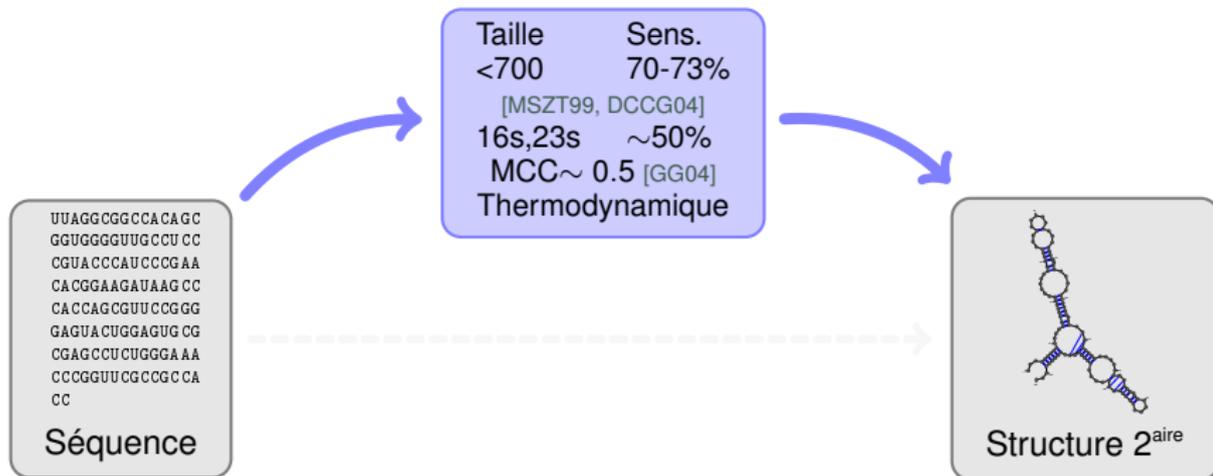


Énergies libres ΔG des boucles dépendent des bases, assymétrie, bases *libres* (dangle) [ZS81] ...

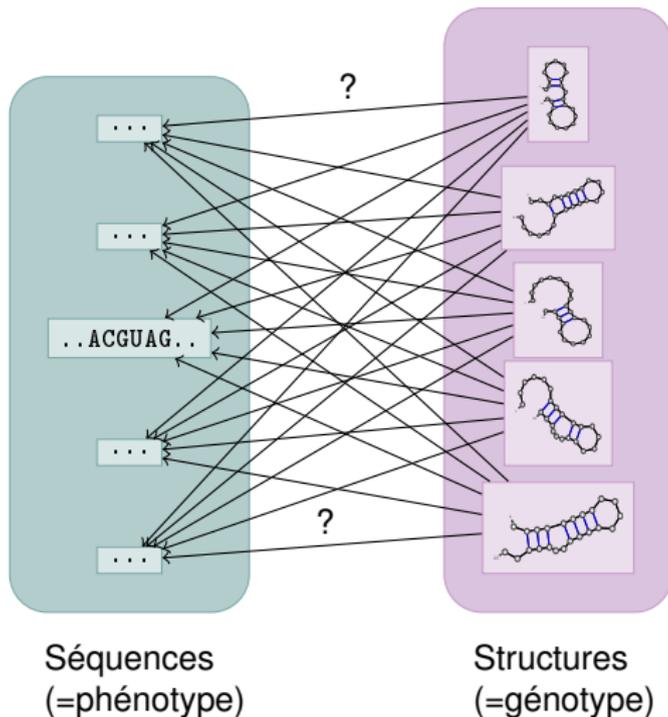
Par une stratégie **diviser pour régner**, on obtient le **repliement d'énergie libre minimale** en temps $\mathcal{O}(n^3)$ ($\sim 1.8^n$ str. sec. compatibles [ZS84])



$$\text{Rappel : } MCC = \frac{t^+t^- - f^+f^-}{\sqrt{(t^+ + f^+)(t^+ + f^-)(t^- + f^+)(t^- + f^-)}}$$



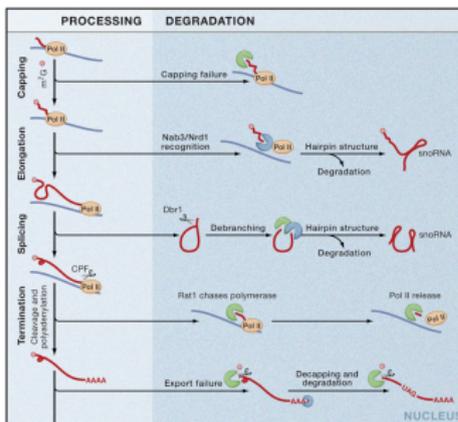
$$\text{Rappel : } MCC = \frac{t^+t^- - f^+f^-}{\sqrt{(t^+ + f^+)(t^+ + f^-)(t^- + f^+)(t^- + f^-)}}$$



Comment utiliser des **propriétés intrinsèques** des ARN pour les détecter dans des génomes ?

Dégradation des ARN :

- Chez *E. coli*, structures en tiges-boucles dans les régions 3'-UTR (vs exonucléases) et 5'-UTR (RNase-E) agissent sur la demi-vie.
- **Eukaryotes** : petits ARN nucléolaires *sauvés* des exonucléases par la présence de structures secondaires



Houseley/Tollervey, Cell 2009 [HT09]

En général, des motifs de structures secondaires sont associés à une dégradation retardée.

⇒ Signal énergétique pour détecter des ARN dans les génomes ?

ARN non-codant \Rightarrow Structurés (Exception petits ARN).

ARN messagers \Rightarrow Stabilisés par une structure secondaire

\Rightarrow *Utiliser l'énergie pour détecter de nouveaux ARN ?*

Problème : L'énergie libre des ARN est-elle **significativement** faible ?

- **Oui** : Le et al. 1990. NAR [LMCM90], Seffens & Digby. 1999. NAR [SD99], Clote et al. 2005 RNA [CFKK05].
- **Non** : Workman & Krogh. 1999. NAR [wk99], Rivas & Eddy. 2000. Bioinformatics [RE00]

Significativité dépend crucialement du modèle de séquences aléatoires ...

(Ici Markov ordre 0,1, permutations préservant les dinucléotides/codons)

Mais on peut toujours combiner le modèle de séquence et l'énergie !!

Exemple : GC% suffisant chez les hyperthermophiles AT-riches [RJE01]

Morale : Ne laissez jamais les stats aux mathématiciens !

Et, en temps que matheux, n'acceptez jamais sans discuter le modèle d'un biologiste.

ARN non-codant \Rightarrow Structurés (Exception petits ARN).

ARN messagers \Rightarrow Stabilisés par une structure secondaire

\Rightarrow *Utiliser l'énergie pour détecter de nouveaux ARN ?*

Problème : L'énergie libre des ARN est-elle **significativement** faible ?

- **Oui** : Le et al. 1990. NAR [LMCM90], Seffens & Digby. 1999. NAR [SD99], Clote et al. 2005 RNA [CFKK05].
- **Non** : Workman & Krogh. 1999. NAR [WK99], Rivas & Eddy. 2000. Bioinformatics [RE00]

Significativité dépend cruciallement du **modèle de séquences aléatoires** ...
(Ici Markov ordre 0,1, permutations préservant les dinucléotides/codons)

Mais on peut toujours combiner le modèle de séquence et l'énergie !!

Exemple : GC% suffisant chez les hyperthermophiles AT-riches [RKJE01]

Morale : Ne laissez jamais les stats aux mathématiciens !

Et, en temps que matheux, n'acceptez jamais sans discuter le modèle d'un biologiste.

ARN non-codant \Rightarrow Structurés (Exception petits ARN).

ARN messagers \Rightarrow Stabilisés par une structure secondaire

\Rightarrow *Utiliser l'énergie pour détecter de nouveaux ARN ?*

Problème : L'énergie libre des ARN est-elle **significativement** faible ?

- **Oui** : Le et al. 1990. NAR [LMCM90], Seffens & Digby. 1999. NAR [SD99], Clote et al. 2005 RNA [CFKK05].
- **Non** : Workman & Krogh. 1999. NAR [WK99], Rivas & Eddy. 2000. Bioinformatics [RE00]

Significativité dépend cruciallement du **modèle de séquences aléatoires** ...

(Ici Markov ordre 0,1, permutations préservant les dinucléotides/codons)

Mais on peut toujours combiner le modèle de séquence et l'énergie !!

Exemple : GC% suffisant chez les hyperthermophiles AT-riches [RKJE01]

Morale : Ne laissez jamais les stats aux mathématiciens !

Et, en temps que matheux, n'acceptez jamais sans discuter le modèle d'un biologiste.

Definition

ANR structuraux Les ARN structuraux sont les ARN dont la fonction est indissociable de la formation d'une structure précise.

Exemples : tRNA, rRNA, RNase-P, or SRP RNA...

Chez les ARN structuraux, des mutations compensatoires sont sélectionnées positivement au cours de l'évolution, préservant la structure.

Exemple : Tombusviridae, région 3' de l'ARNm TBSV

```

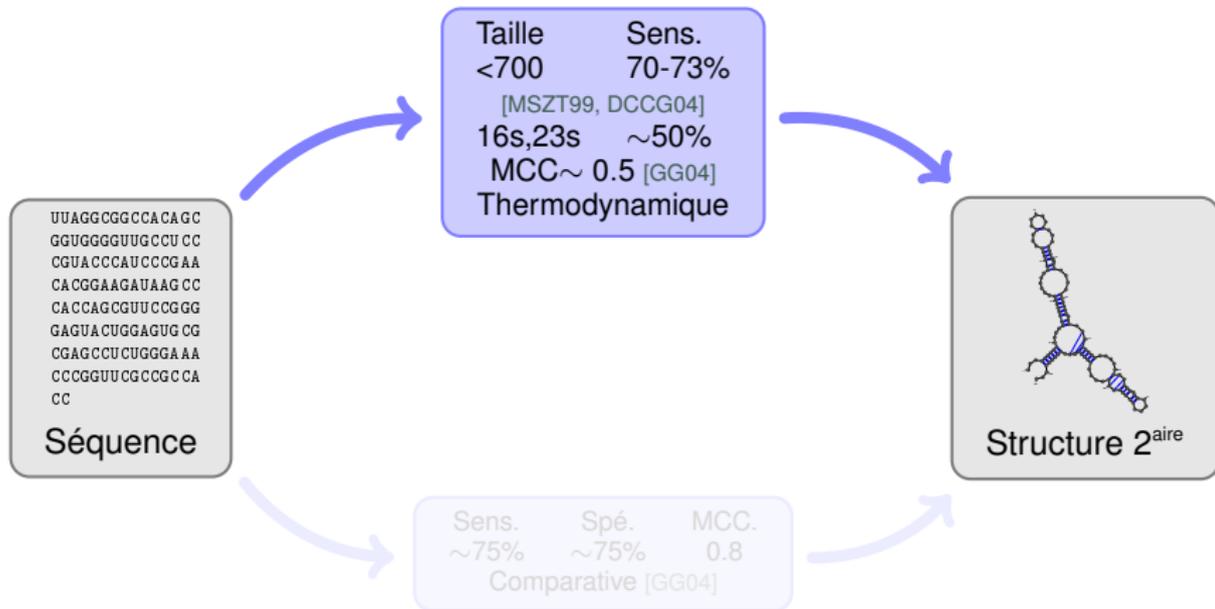
TBSVc 5' ..CGGGCUCCGCACUAGGUUUGGUCGCCUAGGGAUGGAGAUAG..
TBSVp  ..CGGGCUCCGCACUAGGUUUGGUCGCCUAGGGGAUGGAGAUAG..
TBSVs  ..CGGGCUCCGCACU -GGUUGGUC -CC -AGGGGAUGGAGAUAG..
AMCV  ..CGGGCUCCGCAC -CGGUUUGGUCGCC -GGGGAUGGAGAUAG..
CBLV  ..GGUGGU -CGACCU -CAG -CUGGUACCU -GGACUG -G -AGGACG..
CIRV  ..CGGGCUCCGCACUAGGUUUGGUC -CCUAGGGGAUGGAGAUAG..
CNV   ..CGGGC -C -AUCU UAGGUUUGGUC -CCUAGGGGACGG -GAUAG..
CymRSV ..CGGGCUCCGCACU -GGUUGGUC -CC -AGGGGAUGGAGAUAG..
LNSV  ..CGGGCUC -ACG -U -GGUUGGUCGCC -A -GGGACGGAGAUAG..
PeLV  ..CGGGC -CCG -C -UAGGAUUGGUACCUA -GGAUUGG -GAUAG..
PNSV  ..CAGGC -CCGCGC -AGGUUUGGUC -CCU -GGGGAUGG -GACACG..
    
```

Fabian & White, JBC 2004 [FW04]

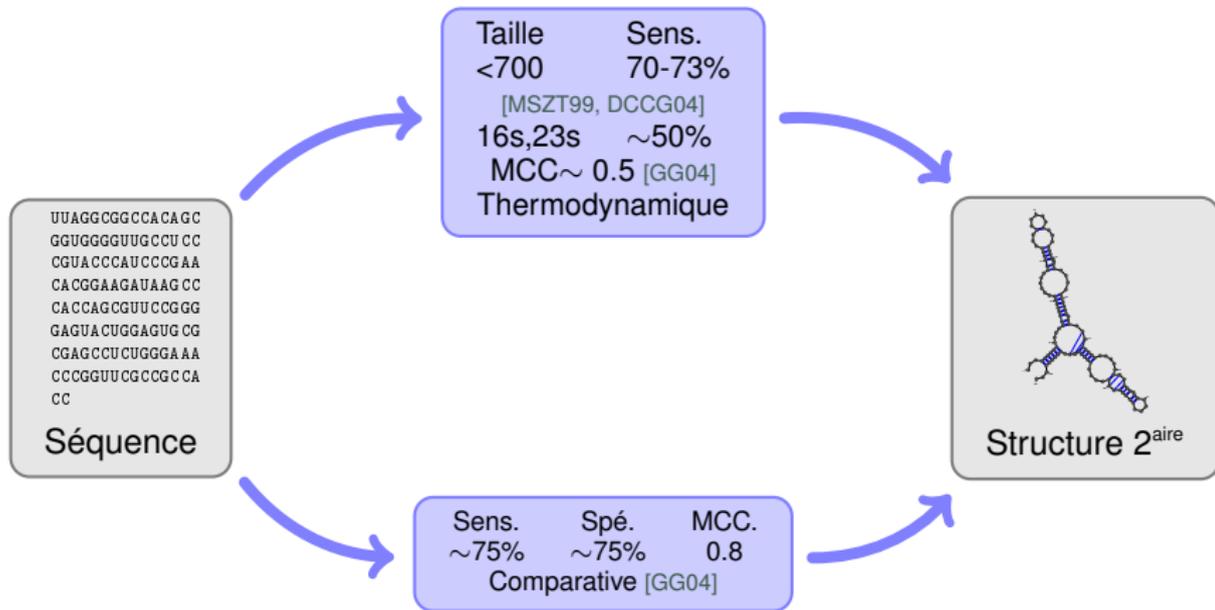
Observation exploitable *in silico* pour la détection d'ARN non-codant.

⇒ Gold standards : QRNA [RE01], RNAz [WHS05], EvoFold [PBS+06] ...

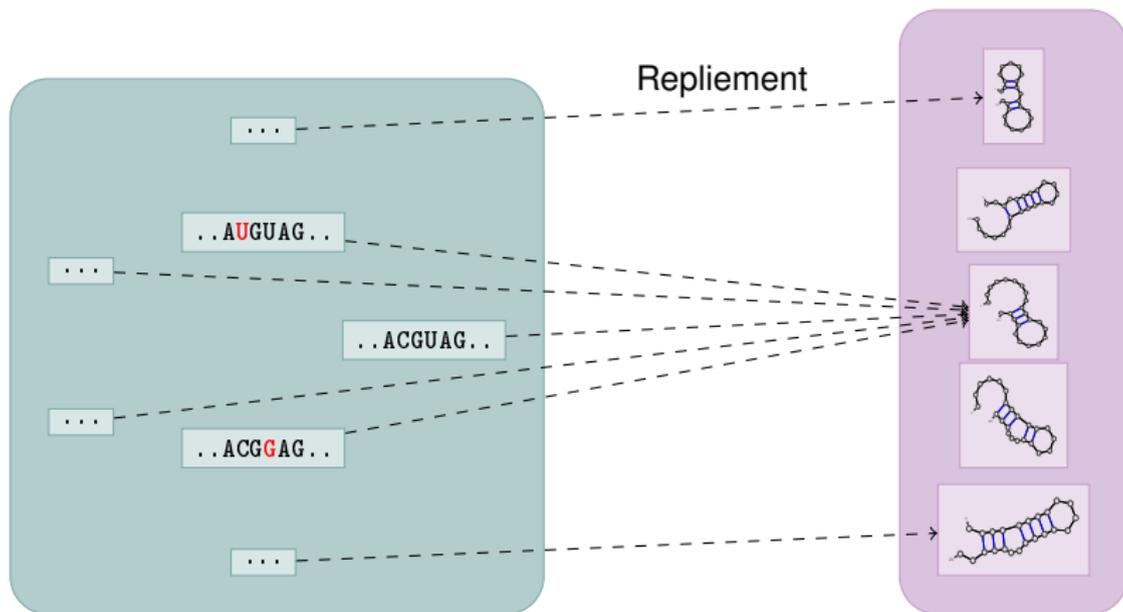
... mais aussi algorithmes pour la prédiction du repliement !



Rappel :
$$MCC = \frac{t^+t^- - f^+f^-}{\sqrt{(t^++f^+)(t^++f^-)(t^-+f^+)(t^-+f^-)}}$$

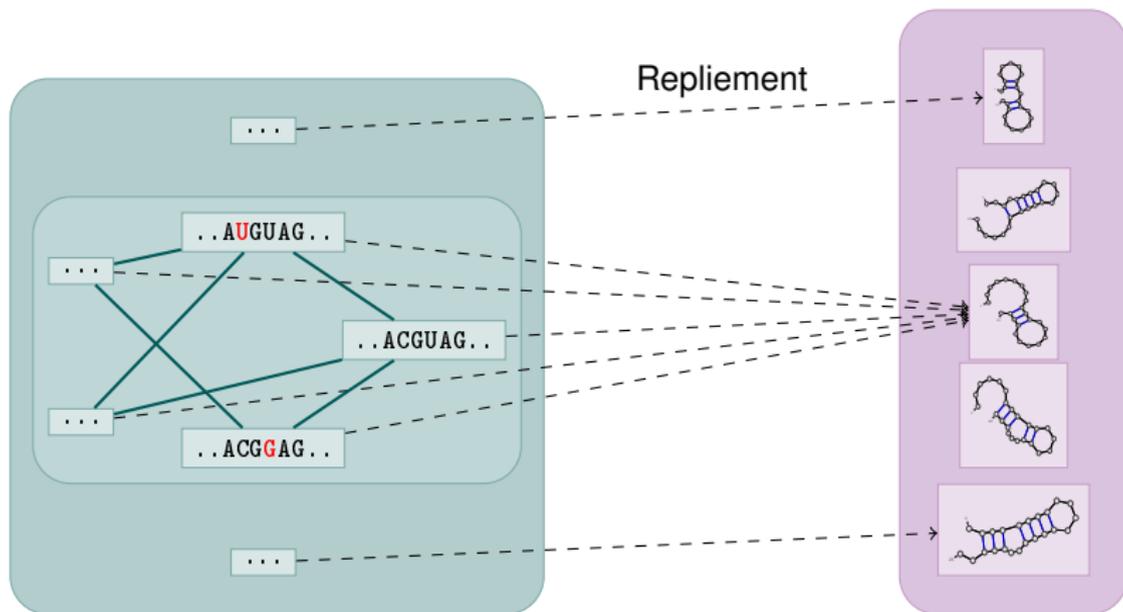


Rappel :
$$MCC = \frac{t^+t^- - f^+f^-}{\sqrt{(t^++f^+)(t^++f^-)(t^-+f^+)(t^-+f^-)}}$$



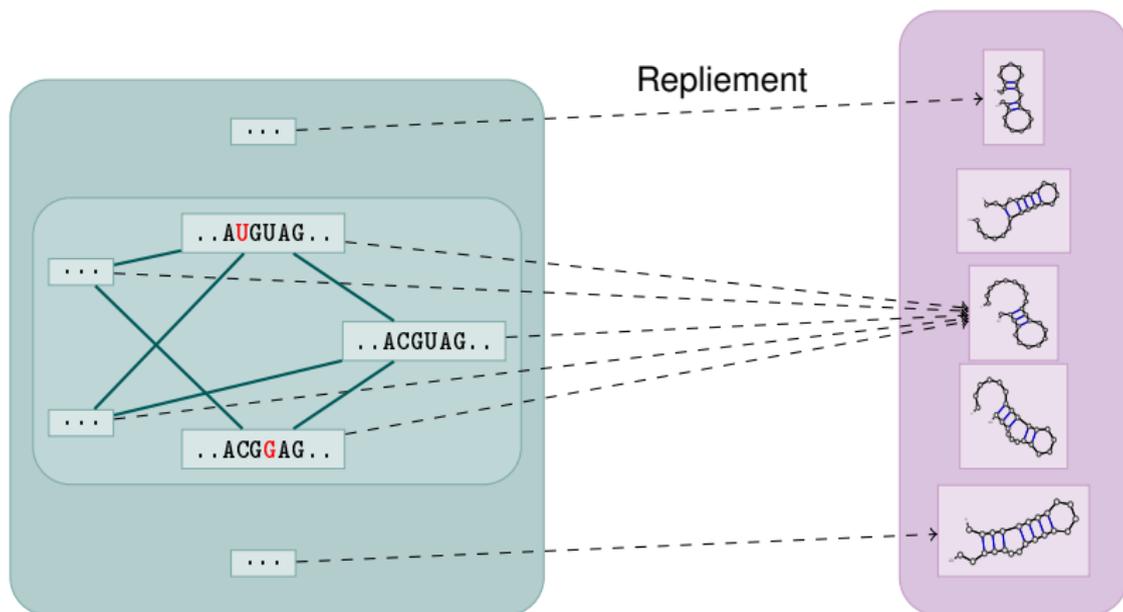
En fixant la **structure**, on peut aborder l'**évolution des ARN** :

- Neutral Network : réseau des mutations préservant la structure
- On remplace une vision optimisatrice de l'évolution par une vision neutrale, plus propice au changement de phénotype.



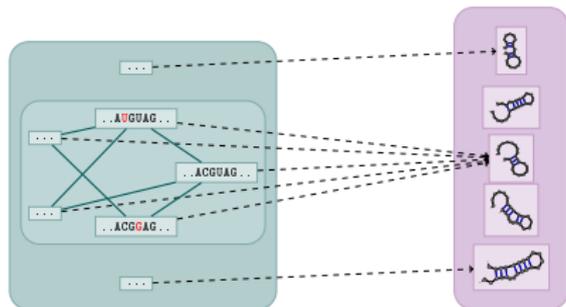
En fixant la **structure**, on peut aborder l'**évolution des ARN** :

- **Neutral Network** : réseau des mutations préservant la structure
- On remplace une vision **optimisatrice** de l'évolution par une vision **neutralité**, plus propice au changement de phénotype.



En fixant la **structure**, on peut aborder l'**évolution des ARN** :

- **Neutral Network** : réseau des mutations préservant la structure
- On remplace une vision **optimisatrice** de l'évolution par une vision **neutralité**, plus propice au changement de phénotype.



Structure du réseau [SFSH94, RSS97] :

- 4^n séquences pour $2.6^n / n\sqrt{n}$ structures
- Répartition des tailles des réseaux en *loi de Zipf*

$$\phi(r) = A(1 + r/B)^{-\gamma} \text{ avec } A, B > 0 \text{ et } \gamma > 1$$

- Certaines structures sont sans séquence (cf repliement inverse)
- Réseaux non-nécessairement connectés
- Diamètre faible

⇒ Diversité et réactivité aux changements de contexte.

Diamètre faible \Rightarrow Possibilité d'être *au bord du réseau*, où une simple mutation modifie le phénotype.

Exemple : La β -Thalassemia,

- SNP C33G identifié dans le 5' UTR de la β -globin.
- Pas de modification des niveaux de transcription.
- Pas de changement notable de la structure d'énergie minimale prédite.

Comment expliquer le **changement de phénotype** ?

Considérons toutes structures compatibles avec l'ARN dans un **ensemble de Boltzmann**.

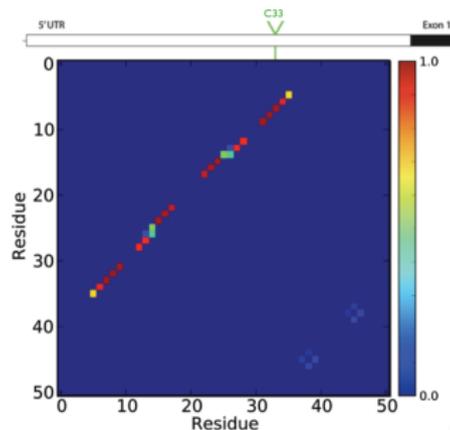
Probabilité d'une structure = $e^{-\frac{E}{RT}} / \mathcal{Z}$.

\Rightarrow Probabilité de paire de base.

Perturbation **significative** de l'ensemble de Boltzmann provoquée par la mutation.

(Pas d'explication systémique . . .)

Halvorsen *et al*, Plos genetics 2010 [HMBL10].



Diamètre faible \Rightarrow Possibilité d'être *au bord du réseau*, où une simple mutation modifie le phénotype.

Exemple : La β -Thalassemia,

- SNP C33G identifié dans le 5' UTR de la β -globin.
- Pas de modification des niveaux de transcription.
- Pas de changement notable de la structure d'énergie minimale prédite.

Comment expliquer le **changement de phénotype** ?

Considérons toutes structures compatibles avec l'ARN dans un **ensemble de Boltzmann**.

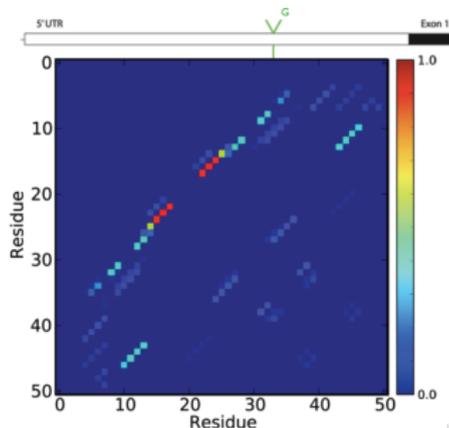
Probabilité d'une structure = $e^{-\frac{E}{RT}} / \mathcal{Z}$.

\Rightarrow Probabilité de paire de base.

Perturbation **significative** de l'ensemble de Boltzmann provoquée par la mutation.

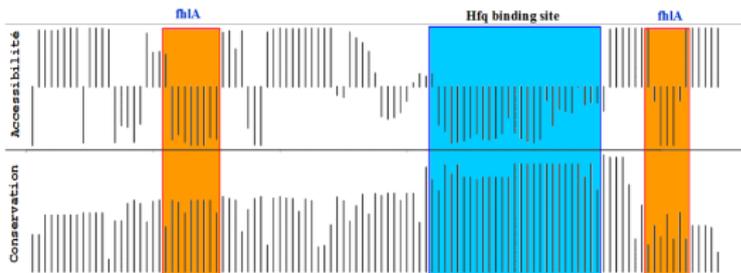
(Pas d'explication systémique . . .)

Halvorsen *et al*, Plos genetics 2010 [HMBL10].



- Pas de bons modèles pour la prédiction des interactions ARN/Protéines
- Mais pressions de sélection multiples (exploitables pour la prédiction)

Exemple : ARN *oxyS*, produit en réponse à un stress d'oxydation chez *E. coli*, provoque la répression de la transcription en interagissant avec ARNm *fhlA* aidé par la protéine Hfq.



Gautheret & Ott, à paraître

Bons signaux de conservation/non-structuration en **séquence uniquement** !
Approches similaires côté protéines [ELSM⁺08].

Bioinformatique des ARN sait déjà faire :

- Prédiction de structure dominante
- Degré de structuration
- Prédiction de cibles (ARN interférence)
- Intégration des données expérimentales
- Exploration exhaustive du mapping structure/séquence
- Découverte d'homologues

... à un coût algorithmique *minimal*!

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire :

- Prédiction de structure dominante
- Degré de structuration
- Prédiction de cibles (ARN interférence)
- Intégration des données expérimentales
- Exploration exhaustive du mapping structure/séquence
- Découverte d'homologues

... à un coût algorithmique *minimal*!

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire :

- Prédiction de structure dominante
- Degré de structuration
- Prédiction de cibles (ARN interférence)
- Intégration des données expérimentales
- Exploration exhaustive du mapping structure/séquence
- Découverte d'homologues

... à un coût algorithmique *minimal*!

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire :

- Prédiction de structure dominante
- Degré de structuration
- Prédiction de cibles (ARN interférence)
- Intégration des données expérimentales
- Exploration exhaustive du mapping structure/séquence
- Découverte d'homologues

... à un coût algorithmique *minimal*!

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire :

- Prédiction de structure dominante
- Degré de structuration
- Prédiction de cibles (ARN interférence)
- Intégration des données expérimentales
- Exploration exhaustive du mapping structure/séquence
- Découverte d'homologues

... à un coût algorithmique *minimal*!

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire :

- Prédiction de structure dominante
- Degré de structuration
- Prédiction de cibles (ARN interférence)
- Intégration des données expérimentales
- Exploration exhaustive du mapping structure/séquence
- Découverte d'homologues

... à un coût algorithmique *minimal*!

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire ... à un coût algorithmique *minimal*!

Axes de développement actuels :

- Découvrir de nouveaux ARN [RE01, WHS05, PBS⁺06]
- Intégrer des données externes (évolution/sondage enzymatique [DLP⁺05])
- Enrichir des espaces de conformations [RE99, RG04, RRCH05]
⇒ Vers un repliement 3D des ARNs [PM08]
- Prédire des interactions [AKN⁺05, HQRS10, SBS10]
- Cinétique des interactions (Dur ! [MTSC09]) ⇒ Vitesses de réaction

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire ... à un coût algorithmique *minimal*!

Axes de développement actuels :

- Découvrir de nouveaux ARN [RE01, WHS05, PBS⁺06]
- Intégrer des données externes (évolution/sondage enzymatique [DLP⁺05])
- Enrichir des espaces de conformations [RE99, RG04, RRCH05]
⇒ Vers un repliement 3D des ARNs [PM08]
- Prédire des interactions [AKN⁺05, HQRS10, SBS10]
- Cinétique des interactions (Dur ! [MTSC09]) ⇒ Vitesses de réaction

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire ... à un coût algorithmique *minimal*!

Axes de développement actuels :

- Découvrir de nouveaux ARN [RE01, WHS05, PBS⁺06]
- Intégrer des données externes (évolution/sondage enzymatique [DLP⁺05])
- Enrichir des espaces de conformations [RE99, RG04, RRCH05]
⇒ Vers un repliement 3D des ARNs [PM08]
- Prédire des interactions [AKN⁺05, HQRS10, SBS10]
- Cinétique des interactions (Dur ! [MTSC09]) ⇒ Vitesses de réaction

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire ... à un coût algorithmique *minimal*!

Axes de développement actuels :

- Découvrir de nouveaux ARN [RE01, WHS05, PBS⁺06]
- Intégrer des données externes (évolution/sondage enzymatique [DLP⁺05])
- Enrichir des espaces de conformations [RE99, RG04, RRCH05]
⇒ Vers un repliement 3D des ARNs [PM08]
- Prédire des interactions [AKN⁺05, HQRS10, SBS10]
- Cinétique des interactions (Dur ! [MTSC09]) ⇒ Vitesses de réaction

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire ... à un coût algorithmique *minimal*!

Axes de développement actuels :

- Découvrir de nouveaux ARN [RE01, WHS05, PBS⁺06]
- Intégrer des données externes (évolution/sondage enzymatique [DLP⁺05])
- Enrichir des espaces de conformations [RE99, RG04, RRCH05]
⇒ Vers un repliement 3D des ARNs [PM08]
- Prédire des interactions [AKN⁺05, HQRS10, SBS10]
- Cinétique des interactions (Dur ! [MTSC09]) ⇒ Vitesses de réaction

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !

Bioinformatique des ARN sait déjà faire ... à un coût algorithmique *minimal*!

Axes de développement actuels :

- Découvrir de nouveaux ARN [RE01, WHS05, PBS⁺06]
- Intégrer des données externes (évolution/sondage enzymatique [DLP⁺05])
- Enrichir des espaces de conformations [RE99, RG04, RRCH05]
⇒ Vers un repliement 3D des ARNs [PM08]
- Prédire des interactions [AKN⁺05, HQRS10, SBS10]
- Cinétique des interactions (Dur ! [MTSC09]) ⇒ Vitesses de réaction

Challenge de la bioinformatique des ARN

Rattraper la biologie des systèmes à partir des génomes, tout en préservant une qualité d'explication mécanique.

Merci !



C. Alkan, E. Karakoç, J. H. Nadeau, S. C. Sahinalp, and K. Zhang.
RNA-RNA Interaction Prediction and Antisense RNA Target Search.
In Proceedings of RECOMB'05, 2005.



A. Condon, B. Davy, B. Rastegari, S. Zhao, and F. Tarrant.
Classifying RNA pseudoknotted structures.
Theoretical Computer Science, 320(1) :35–50, 2004.



Peter Clote, Fabrizio Ferré, Evangelos Kranakis, and Danny Krizanc.
Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency.
RNA, 11(5) :578–591, May 2005.



K. Doshi, J. J. Cannone, C. Cobaugh, and R. R. Gutell.
Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for rna secondary structure prediction.
BMC Bioinformatics, 5(1) :105, 2004.



R. Das, A. Laederach, S. M. Pearlman, D. Daniel, and R. B. Altman.
SAFA : Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments.
RNA, 11(3) :344–354, 2005.



S. Engelen, L. Trojan, S. Sacquin-Mora, R. Lavery, and A. Carbone.
Joint evolutionary trees : detection and analysis of protein interfaces.
Accepted to PLoS Computational Biology, 2008.



Marc R Fabian and K. Andrew White.
5'-3' RNA-RNA interaction facilitates cap- and poly(A) tail-independent translation of tomato bushy stunt virus mrna : a potential common mechanism for tombusviridae.
J Biol Chem, 279(28) :28862–28872, Jul 2004.



P. Gardner and R. Giegerich.
A comprehensive comparison of comparative RNA structure prediction approaches.
BMC Bioinformatics, 5(1) :140, 2004.



Matthew Halvorsen, Joshua S Martin, Sam Broadaway, and Alain Laederach.

Disease-associated mutations that alter the RNA structural ensemble.

PLoS Genet, 6(8) :e1001074, Aug 2010.



Fenix W D Huang, Jing Qin, Christian M Reidys, and Peter F Stadler.

Target prediction and a statistical sampling algorithm for RNA-RNA interaction.

Bioinformatics, 26(2) :175–181, Jan 2010.



Jonathan Houseley and David Tollervey.

The many pathways of RNA degradation.

Cell, 136(4) :763–776, Feb 2009.



S. Y. Le, M. H. Malim, B. R. Cullen, and J. V. Maizel.

A highly conserved RNA folding region coincident with the Rev response element of primate immunodeficiency viruses.

Nucleic Acids Res, 18(6) :1613–1623, Mar 1990.



R. B. Lyngsø and C. N. S. Pedersen.

RNA pseudoknot prediction in energy-based models.

Journal of Computational Biology, 7(3-4) :409–427, 2000.



N. Leontis and E. Westhof.

Geometric nomenclature and classification of RNA base pairs.

RNA, 7 :499–512, 2001.



D.H. Mathews, J. Sabina, M. Zuker, and D.H. Turner.

Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure.

J Mol Biol, 288 :911–940, 1999.



Ján Maňuch, Chris Thachuk, Ladislav Stacho, and Anne Condon.

Np-completeness of the direct energy barrier problem without pseudoknots.

pages 106–115, 2009.



Jakob Skou Pedersen, Gill Bejerano, Adam Siepel, Kate Rosenbloom, Kerstin Lindblad-Toh, Eric S Lander, Jim Kent, Webb Miller, and David Haussler.
Identification and classification of conserved RNA secondary structures in the human genome.
PLoS Comput Biol, 2(4) :e33, Apr 2006.



M. Parisien and F. Major.
The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.
Nature, 452(7183) :51–55, 2008.



E. Rivas and S.R. Eddy.
A dynamic programming algorithm for RNA structure prediction including pseudoknots.
J Mol Biol, 285 :2053–2068, 1999.



E. Rivas and S. R. Eddy.
Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs.
Bioinformatics, 16(7) :583–605, Jul 2000.



E. Rivas and S. R. Eddy.
Noncoding RNA gene detection using comparative sequence analysis.
BMC Bioinformatics, 2 :8, 2001.



J. Reeder and R. Giegerich.
Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics.
BMC Bioinformatics, 5 :104, 2004.



E. Rivas, R. J. Klein, T. A. Jones, and S. R. Eddy.
Computational identification of noncoding RNAs in E. coli by comparative genomics.
Curr Biol, 11(17) :1369–1373, Sep 2001.



J. Ren, B. Rastegari, A. Condon, and H. H. Hoos.
HotKnots : Heuristic prediction of RNA secondary structures including pseudoknots.
RNA, 11(10) :1494–1504, 2005.



C. Reidys, P. F. Stadler, and P. Schuster.

Generic properties of combinatorial maps : neutral networks of rna secondary structures.

Bull Math Biol, 59(2) :339–397, Mar 1997.



Raheleh Salari, Rolf Backofen, and S. Cenk Sahinalp.

Fast prediction of RNA-RNA interaction.

Algorithms Mol Biol, 5 :5, 2010.



W. Seffens and D. Digby.

mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences.

Nucleic Acids Res, 27(7) :1578–1584, Apr 1999.



P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker.

From sequences to shapes and back : a case study in rna secondary structures.

Proc Biol Sci, 255(1344) :279–284, Mar 1994.



Stefan Washietl, Ivo L Hofacker, and Peter F Stadler.

Fast and reliable prediction of noncoding RNAs.

Proc Natl Acad Sci U S A, 102(7) :2454–2459, Feb 2005.



C. Workman and A. Krogh.

No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution.

Nucleic Acids Res, 27(24) :4816–4822, Dec 1999.



M. Zuker and P. Stiegler.

Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.

Nucleic Acids Res, 9 :133–148, 1981.



M. Zuker and D. Sankoff.

RNA secondary structures and their prediction.

Bull Math Bio, 46 :591–621, 1984.