M2 BIM - Lecture 2
Boltzmann equilibrium

Yann Ponty

Bioinformatics Team
École Polytechnique/CNRS/INRIA AMIB – France

http://www.lix.polytechnique.fr/~ponty/index.php?page=bim2015

December 8th, 2015

Outline

1 **Ensemble de Boltzmann**
- Ensemble de Boltzmann
- Nussinov : Minimisation ⇒ Comptage
- Calcul de la fonction de partition
- Échantillonnage statistique

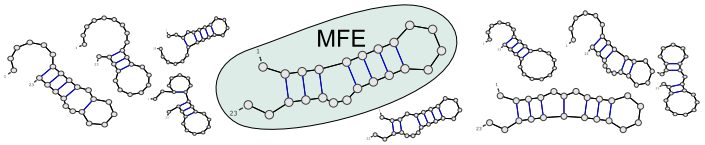
2 **Extensions**
- Validité d'un schéma
- Structures sous-optimales
- Pseudo-noeuds

Ensemble canonique de Boltzmann

L'ARN *respire* ⇒ Il n'existe pas UNE unique conformation native.

Nouveau paradigme

Les conformations d'un ARN **coexistent** dans une **distribution de Boltzmann**.



Conséquence : La probabilité de la MFE peut être négligeable.
⇒ Comprendre les modes d'actions de l'ARN exige de prendre en considération l'ensemble des structures.

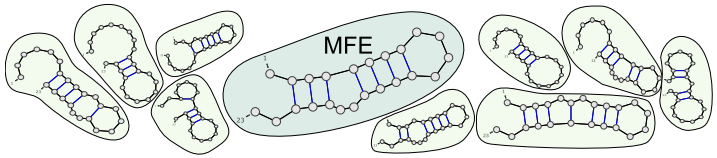
En particulier, des structures proches peuvent se *grouper* et devenir l'hypothèse la plus réaliste dans la recherche d'une conformation fonctionnelle.

Ensemble canonique de Boltzmann

L'ARN *respire* ⇒ Il n'existe pas UNE unique conformation native.

Nouveau paradigme

Les conformations d'un ARN **coexistent** dans une **distribution de Boltzmann**.



Conséquence : La probabilité de la MFE peut être négligeable.
⇒ Comprendre les modes d'actions de l'ARN exige de prendre en considération l'ensemble des structures.

En particulier, des structures proches peuvent se *grouper* et devenir l'hypothèse la plus réaliste dans la recherche d'une conformation fonctionnelle.

Distribution de Boltzmann : Définition

Une distribution de Bolzmann pondère chaque structure S pour un ARN ω par un **facteur de Boltzmann** $\mathcal{B}_{S,\omega} = e^{\frac{-E_{S,\omega}}{RT}}$ où :

- $E_{S,\omega}$ est l'énergie libre de S (kCal.mol^{-1})
- T est la température (K)
- R est la constante des gaz parfaits (1.986.10^{-3} kCal.K^{-1}.mol^{-1})

Distribution renormalisée sur S_ω par la **fonction de partition**

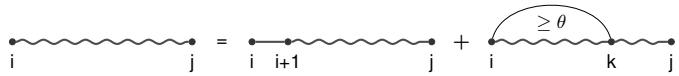
$$\mathcal{Z}_\omega = \sum_{S \in \mathcal{S}_\omega} e^{\frac{-E_{S,\omega}}{RT}}.$$

où $\mathcal{S}_\omega$ est l'ensemble des conformations compatibles avec ω.

La **probabilité de Boltzmann** d'une structure S est alors donnée par

$$P_{S,\omega} = \frac{e^{\frac{-E_{S,\omega}}{RT}}}{\mathcal{Z}_\omega}.$$

Décomposition de Nussinov/Jacobson



Récurrence sur l'énergie minimale d'un repliement :

$$\begin{aligned} N_{i,t} &= 0, \quad \forall t \in [i, i+\theta] \\ N_{i,j} &= \min \begin{cases} N_{i+1,j} & (i \text{ non apparié}) \\ \min_{k=i+\theta+1}^{j} E_{i,k} + N_{i+1,k-1} + N_{k+1,j} & (i \text{ comp. avec } k) \end{cases} \end{aligned}$$

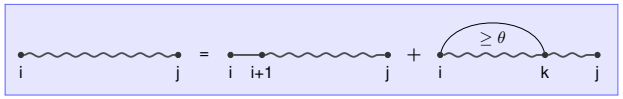
Récurrence de comptage des structures compatibles :

$$\begin{aligned} C_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ C_{i,j} &= \sum \begin{cases} C_{i+1,j} & (i \text{ non apparié}) \\ \sum_{k=i+\theta+1}^{j} 1 \times C_{i+1,k-1} \times C_{k+1,j} & (i \text{ comp. avec } k) \end{cases} \end{aligned}$$

La décomposition est importante, le reste (MFE, comptage...) suit !

Fonction de partition

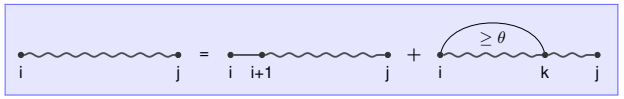
Fonction de partition = **Comptage pondéré** des structures compatibles



$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \begin{cases} \mathcal{Z}_{i+1,j} \\ \sum_{k=i+\theta+1}^{j} 1 \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \end{cases} \end{aligned}$$

Fonction de partition

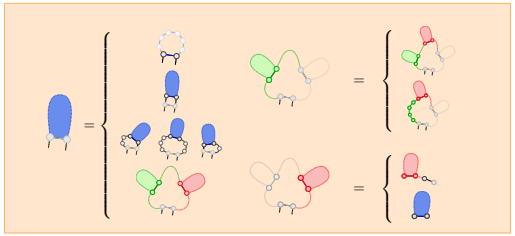
Fonction de partition = **Comptage pondéré** des structures compatibles



$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \begin{cases} \mathcal{Z}_{i+1,j} \\ \sum_{k=i+\theta+1}^{j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \end{cases} \end{aligned}$$

Fonction de partition

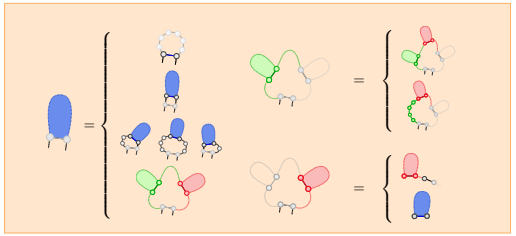
Fonction de partition = **Comptage pondéré** des structures compatibles



$$
\begin{aligned}
\mathcal{M}'_{i,j} &= \text{Min}\begin{cases} E_H(i,j) \\ E_S(i,j) + \mathcal{M}'_{i+1,j-1} \\ \text{Min}(E_{BI}(i,i',j',j) + \mathcal{M}'_{i',j'}) \\ a + c + \text{Min}\left(\mathcal{M}_{i+1,k-1} + \mathcal{M}^1{}_{k,j-1}\right) \end{cases} \\
\mathcal{M}_{i,j} &= \text{Min}\left\{\text{Min}\left(\mathcal{M}_{i,k-1}, b(k-1)\right) + \mathcal{M}^1{}_{k,j}\right\} \\
\mathcal{M}^1{}_{i,j} &= \text{Min}\left\{b + \mathcal{M}^1{}_{i,j-1}, c + \mathcal{M}'_{i,j}\right\}
\end{aligned}
$$

Fonction de partition

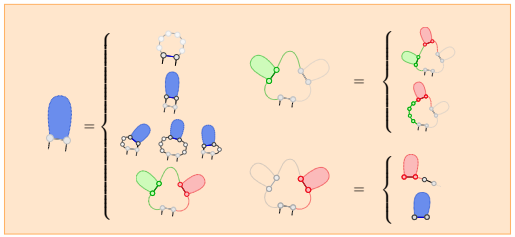
Fonction de partition = **Comptage pondéré** des structures compatibles



$$
\begin{aligned}
\mathcal{M}'_{i,j} &= \text{Min}\begin{cases} e^{\frac{-E_H(i,j)}{RT}} \\ e^{\frac{-E_S(i,j)}{RT}} + \mathcal{M}'_{i+1,j-1} \\ \text{Min}\left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} + \mathcal{M}'_{i',j'}\right) \\ e^{\frac{-(a+c)}{RT}} + \text{Min}\left(\mathcal{M}_{i+1,k-1} + \mathcal{M}^1{}_{k,j-1}\right) \end{cases} \\
\mathcal{M}_{i,j} &= \text{Min}\left\{\text{Min}\left(\mathcal{M}_{i,k-1}, e^{\frac{-b(k-1)}{RT}}\right) + \mathcal{M}^1{}_{k,j}\right\} \\
\mathcal{M}^1{}_{i,j} &= \text{Min}\left\{e^{\frac{-b}{RT}} + \mathcal{M}^1{}_{i,j-1}, e^{\frac{-c}{RT}} + \mathcal{M}'_{i,j}\right\}
\end{aligned}
$$

Fonction de partition

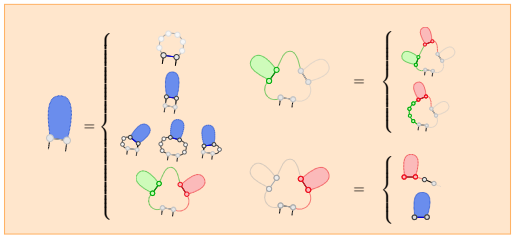
Fonction de partition = **Comptage pondéré** des structures compatibles



$$
\begin{aligned}
\mathcal{M}'_{i,j} &= \text{Min}\begin{cases} e^{\frac{-E_H(i,j)}{RT}} \\ e^{\frac{-E_S(i,j)}{RT}} \mathcal{M}'_{i+1,j-1} \\ \text{Min}\left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{M}'_{i',j'}\right) \\ e^{\frac{-(a+c)}{RT}} \text{Min}\left(\mathcal{M}_{i+1,k-1} \mathcal{M}^1{}_{k,j-1}\right) \end{cases} \\
\mathcal{M}_{i,j} &= \text{Min}\left\{\text{Min}\left(\mathcal{M}_{i,k-1}, e^{\frac{-b(k-1)}{RT}}\right) \mathcal{M}^1{}_{k,j}\right\} \\
\mathcal{M}^1{}_{i,j} &= \text{Min}\left\{e^{\frac{-b}{RT}} \mathcal{M}^1{}_{i,j-1}, e^{\frac{-c}{RT}} \mathcal{M}'_{i,j}\right\}
\end{aligned}
$$

Fonction de partition

Fonction de partition = **Comptage pondéré** des structures compatibles



$$
\begin{aligned}
\mathcal{Z}'(i,j) &= \sum\begin{cases} e^{\frac{-E_H(i,j)}{RT}} \\ e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) \\ + \sum\left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j')\right) \\ + e^{\frac{-(a+c)}{RT}} \sum\left(\mathcal{Z}(i+1,k-1)\mathcal{Z}^1(k,j-1)\right) \end{cases} \\
\mathcal{Z}(i,j) &= \sum\left(\mathcal{Z}(i,k-1) + e^{\frac{-b(k-1)}{RT}}\right) \mathcal{Z}^1(k,j) \\
\mathcal{Z}^1(i,j) &= e^{\frac{-b}{RT}} \mathcal{Z}^1(i,j-1) + e^{\frac{-c}{RT}} \mathcal{Z}'(i,j)
\end{aligned}
$$

Fonction de partition

Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Fonction de partition

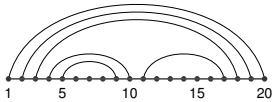
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = \mathcal{B}_{1,20}$

Fonction de partition

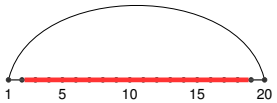
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{1}{RT}} . \mathcal{B}_{2,19}$

Fonction de partition

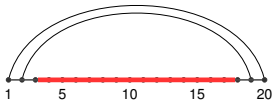
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{-1}{RT}} . e^{\frac{1}{RT}} . \mathcal{B}_{3,18}$

Fonction de partition

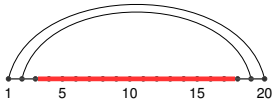
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{2}{RT}} . \mathcal{B}_{3,18}$

Fonction de partition

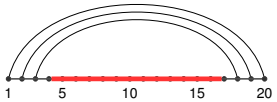
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{3}{RT}} . \mathcal{B}_{4,17}$

Fonction de partition

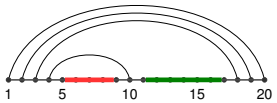
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{4}{RT}} . \mathcal{B}_{5,9} . \mathcal{B}_{11,17}$

Fonction de partition

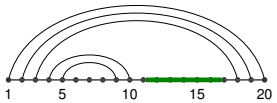
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{5}{RT}} . \mathcal{B}_{11,17}$

Fonction de partition

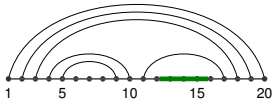
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{6}{RT}} . \mathcal{B}_{12,16}$

Fonction de partition

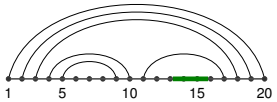
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{6}{RT}} . \mathcal{B}_{13,16}$

Fonction de partition

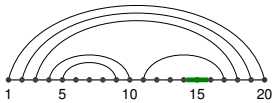
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{6}{RT}} . \mathcal{B}_{14,16}$

Fonction de partition

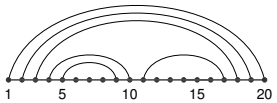
Fonction de partition = **Comptage pondéré** des structures compatibles

$$\begin{aligned} \mathcal{Z}_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ \mathcal{Z}_{i,j} &= \sum \left\{ \sum_{k=i+\theta+1}^{\mathcal{Z}_{i+1,j}\ j} e^{\frac{-E_{bp}(i,k)}{RT}} \times \mathcal{Z}_{i+1,k-1} \times \mathcal{Z}_{k+1,j} \right. \end{aligned}$$

Validité de la fonction de partition :

- Exhaustivité/non ambiguïté du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant
  $\Big(e^{-a/RT} \cdot \mathcal{Z}^1 \cdot \mathcal{Z}' = e^{-a/RT} \cdot \sum_x e^{-E_x/RT} \cdot \sum_y e^{-E_y/RT}$
  $= \sum_{x,y} e^{-a/RT} \cdot e^{-E_x/RT} \cdot e^{-E_y/RT} = \sum_{x,y} e^{-(a+E_x+E_y)/RT}\Big)$

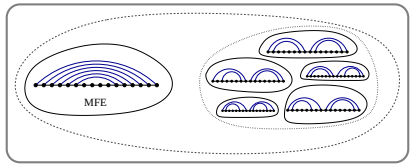
Exemple :

$S =$ 

$\Rightarrow \mathcal{B}_{1,20} = e^{\frac{6}{RT}} = e^{\frac{-E_S}{RT}}$

Échantillonnage statistique de structures d'ARN

La MFE (Probabilité maximale) peut être **largement dominée** par un ensemble $\mathcal{B}$ de sous-optimaux **structurellement similaires**.
$\Rightarrow$ Conformation fonctionnelle trouvée plus probablement dans $\mathcal{B}$.



Expérience : [DCL05]

- Échantillonner des structures selon une probabilité de Boltzmann
- Effectuer un clustering
- Construire structure consensus dans le plus lourd cluster

$\Rightarrow$ Amélioration relative pour spécificité (+17.6%) et sensibilité (+21.74%, sauf Introns du groupe II)

Problème

Comment engendrer des structures dans la distribution de Boltzmann ?

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$

Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_w$ est engendrée uniquement (Unambiguïté de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$

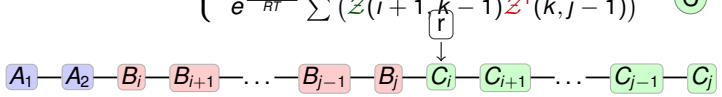
Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_w$ est engendrée uniquement (Unambiguïté de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$



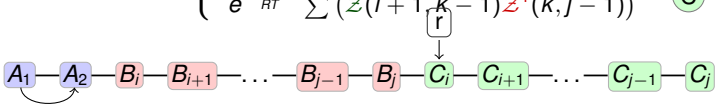
Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_w$ est engendrée uniquement (Unambiguïté de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$



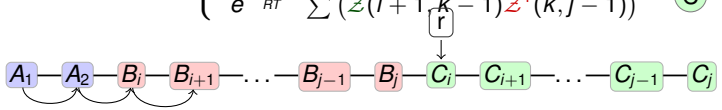
Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_\omega$ est engendrée uniquement (Unambiguïté de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$



Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_\omega$ est engendrée uniquement (Unambiguïté de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc

Yann Ponty M2 BIM - Lecture 2 - Beyond MFE folding

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$

Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_\omega$ est engendrée uniquement (Unambiguïté de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc

Yann Ponty M2 BIM - Lecture 2 - Beyond MFE folding

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$

Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_\omega$ est engendrée uniquement (Unambiguïté de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc

Yann Ponty M2 BIM - Lecture 2 - Beyond MFE folding

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$

Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_\omega$ est engendrée uniquement (Unambiguité de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc $p_S = \frac{\mathcal{B}(E_1)}{\mathcal{B}(\mathcal{S}_w)} \cdot \frac{\mathcal{B}(E_2)}{\mathcal{B}(E_1)} \cdot \frac{\mathcal{B}(E_3)}{\mathcal{B}(E_2)} \cdots \frac{\mathcal{B}(\{S\})}{\mathcal{B}(E_m)}$

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$

Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_\omega$ est engendrée uniquement (Unambiguité de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc $p_S = \frac{1}{\mathcal{B}(\mathcal{S}_w)} \cdot \frac{1}{1} \cdot \frac{1}{1} \cdots \frac{\mathcal{B}(\{S\})}{1}$

Yann Ponty M2 BIM - Lecture 2 - Beyond MFE folding

Remontée stochastique

Algorithme (Reformulation `SFold` [DL03])
Précalcul : Calculer les matrices $(\mathcal{Z}, \mathcal{Z}', \mathcal{Z}^1)$ des fonctions de partition.
Remontée stochastique :

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$

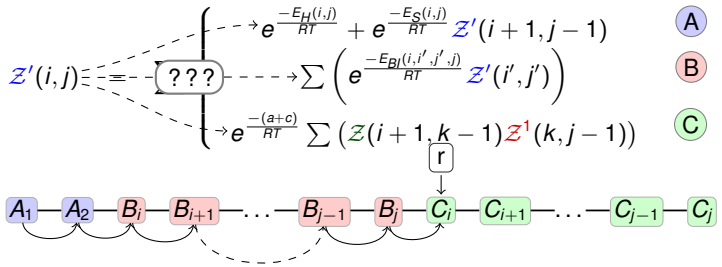
Correction : Chaque terme de la décomposition engendre $\mathcal{T} \in \{\mathcal{A}_1, \ldots, \mathcal{C}_j\}$, et est choisi selon son facteur de Boltzmann cumulé $\mathcal{B}(\mathcal{T})/\mathcal{Z} = \sum_{S \in \mathcal{T}} e^{-E/RT}/\mathcal{Z}$ (Par récurrence).
Chaque structure $S \in \mathcal{S}_\omega$ est engendrée uniquement (Unambiguité de Turner) par une séquence de choix d'ensembles $\mathcal{S}_w \supset E_1 \supset E_2 \supset \ldots \supset \{S\}$.
La probabilité d'engendrer S est donc $p_S = \frac{\mathcal{B}(\{S\})}{\mathcal{B}(\mathcal{S}_w)} = \frac{e^{-E_S/RT}}{\mathcal{Z}} = P_{S,\omega}$

Yann Ponty M2 BIM - Lecture 2 - Beyond MFE folding

Complexité

Algorithme (Reformulation `SFold` [DL03])

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

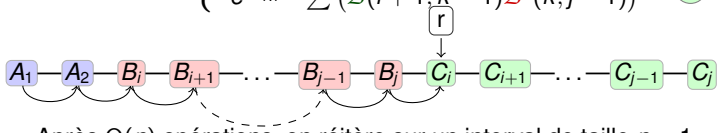


Complexité

Algorithme (Reformulation `SFold` [DL03])

1. Générer un nombre aléatoire r dans $[0, \mathcal{Z}'(i,j))$
2. Retirer à r les contributions à $\mathcal{Z}'(i,j)$, jusqu'à ce que $r < 0$
3. Réitérer sur les sous-structures

$$\mathcal{Z}'(i,j) = \sum \begin{cases} e^{\frac{-E_H(i,j)}{RT}} + e^{\frac{-E_S(i,j)}{RT}} \mathcal{Z}'(i+1,j-1) & \text{A} \\ \sum \left(e^{\frac{-E_{BI}(i,i',j',j)}{RT}} \mathcal{Z}'(i',j') \right) & \text{B} \\ e^{\frac{-(a+c)}{RT}} \sum \left(\mathcal{Z}(i+1,k-1) \mathcal{Z}^1(k,j-1) \right) & \text{C} \end{cases}$$



Après $\Theta(n)$ opérations, on réitère sur un interval de taille $n-1$
$\Rightarrow$ Complexité du cas au pire en $\mathcal{O}(n^2 k)$ pour k échantillons

Remarque : Instance pondérée d'un problème de génération aléatoire par la méthode *récursive* [Pon08].
Complexité en moyenne en $\Theta(n\sqrt{n})$ dans l'hypothèse tout appariement.
Adaptation d'un parcours Boustrophedon $\Rightarrow \mathcal{O}(n \log nk)$ au pire.

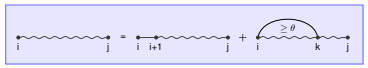
Outline

1. Ensemble de Boltzmann
   - Ensemble de Boltzmann
   - Nussinov : Minimisation $\Rightarrow$ Comptage
   - Calcul de la fonction de partition
   - Échantillonnage statistique
2. Extensions
   - Validité d'un schéma
   - Structures sous-optimales
   - Pseudo-noeuds

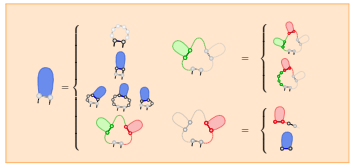
Validité d'un schéma

Une preuve de correction possible :

Calcul correct localement
+ Toutes les conformations sont parcourues
$\Rightarrow$ Algorithme correct (Induction)



$\overset{?}{\Leftrightarrow}$



Forte certitude **mais** pas encore preuve (Séries génératrices).

Validité d'un schéma

Une preuve de correction possible :

Calcul correct localement
+ Toutes les conformations sont parcourues
$\Rightarrow$ Algorithme correct (Induction)

$$\begin{aligned} C_{i,t} &= 1, \quad \forall t \in [i, i+\theta] \\ C_{i,j} &= \sum \begin{cases} C_{i+1,j} \\ \sum_{k=i+\theta+1}^{j} 1 \times C_{i+1,k-1} \times C_{k+1,j} \end{cases} \end{aligned}$$

Homopolymère (Toute paire autorisée) + $\theta = 1$
$\Rightarrow C_{1,n} = \mathbf{1}, 1, 1, 2, 4, 8, 17, 32, 82, 185, 423, \ldots$

$\overset{?}{\Leftrightarrow}$

$$\begin{aligned} \mathcal{C}'_{i,j} &= \sum \begin{cases} 1 \\ \mathcal{C}'_{i+1,j-1} \\ \sum_{i',j'} \mathcal{C}'_{i',j'} \\ \sum_k C_{i+1,k-1} \times \mathcal{C}^1{}_{k,j-1} \end{cases} \\ C_{i,j} &= \sum_k \left((C_{i,k-1} + 1) \times \mathcal{C}^1{}_{k,j} \right) \\ \mathcal{C}^1{}_{i,j} &= \mathcal{C}^1{}_{i,j-1} + \mathcal{C}'_{i,j} \end{aligned}$$

Homopolymère + $\theta = 1$
$\Rightarrow \mathcal{C}'_{1,n} = \mathbf{0}, 1, 1, 2, 4, 8, 17, 32, 82, 185, 423, \ldots$

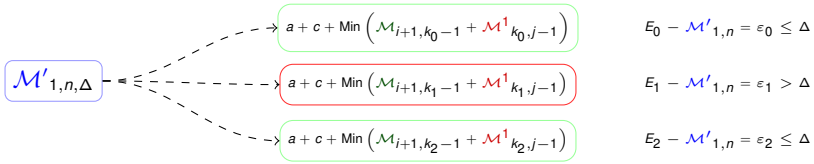
Forte certitude **mais** pas encore preuve (Séries génératrices).

Repliement sous-optimal

Prob. : Simplifications de l'énergie (Pseudo-noeuds, non-can.)
$\Rightarrow$ La structure **native** (fonctionnelle) pourrait être **ignorée**.

$\Rightarrow$ Engendrer des repliements sous-optimaux (`RNASubopt` [WFHS99]),
i.e. construire toutes les structures à Δ KCal.mol^{-1} de la MFE :

- Calculer la matrice des énergies minimales
- Effectuer un Backtrack sur toutes les contributions à $\leq \Delta$ de la MFE
- Mettre à jour Δ t.q. les futurs backtracks donnent ≥ 1 struct.
- Engendrer (Rec.) les sous-ensembles et combiner (brutal ou Tri)

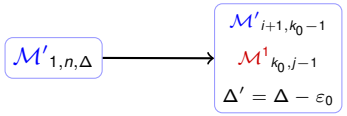


Repliement sous-optimal

Prob. : Simplifications de l'énergie (Pseudo-noeuds, non-can.)
$\Rightarrow$ La structure **native** (fonctionnelle) pourrait être **ignorée**.

$\Rightarrow$ Engendrer des repliements sous-optimaux (`RNASubopt` [WFHS99]),
i.e. construire toutes les structures à Δ KCal.mol^{-1} de la MFE :

- Calculer la matrice des énergies minimales
- Effectuer un Backtrack sur toutes les contributions à $\leq \Delta$ de la MFE
- Mettre à jour Δ t.q. les futurs backtracks donnent ≥ 1 struct.
- Engendrer (Rec.) les sous-ensembles et combiner (brutal ou Tri)

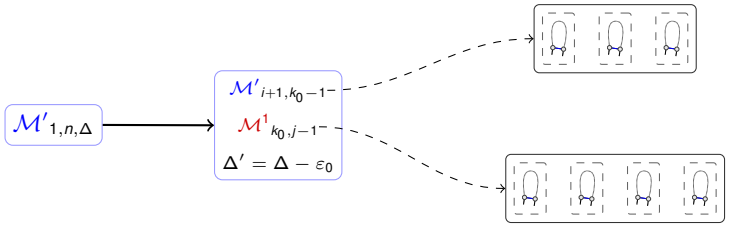


Repliement sous-optimal

Prob. : Simplifications de l'énergie (Pseudo-noeuds, non-can.)
$\Rightarrow$ La structure **native** (fonctionnelle) pourrait être **ignorée**.

$\Rightarrow$ Engendrer des repliements sous-optimaux (`RNASubopt` [WFHS99]),
i.e. construire toutes les structures à Δ KCal.mol^{-1} de la MFE :

- Calculer la matrice des énergies minimales
- Effectuer un Backtrack sur toutes les contributions à $\leq \Delta$ de la MFE
- Mettre à jour Δ t.q. les futurs backtracks donnent ≥ 1 struct.
- Engendrer (Rec.) les sous-ensembles et combiner (brutal ou Tri)

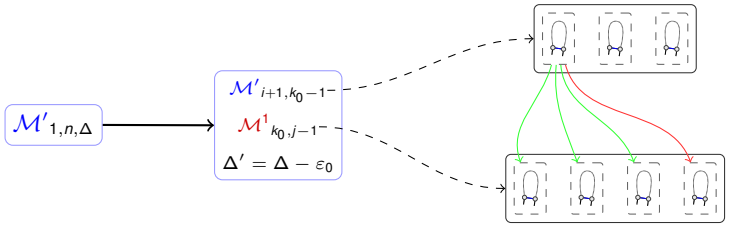


Repliement sous-optimal

Prob. : Simplifications de l'énergie (Pseudo-noeuds, non-can.)
$\Rightarrow$ La structure **native** (fonctionnelle) pourrait être **ignorée**.

$\Rightarrow$ Engendrer des repliements sous-optimaux (`RNASubopt` [WFHS99]),
i.e. construire toutes les structures à Δ KCal.mol^{-1} de la MFE :

- Calculer la matrice des énergies minimales
- Effectuer un Backtrack sur toutes les contributions à $\leq \Delta$ de la MFE
- Mettre à jour Δ t.q. les futurs backtracks donnent ≥ 1 struct.
- Engendrer (Rec.) les sous-ensembles et combiner (brutal ou Tri)

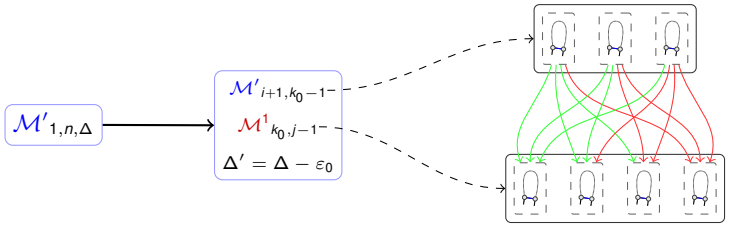


Repliement sous-optimal

Prob. : Simplifications de l'énergie (Pseudo-noeuds, non-can.)
$\Rightarrow$ La structure **native** (fonctionnelle) pourrait être **ignorée**.

$\Rightarrow$ Engendrer des repliements sous-optimaux (`RNASubopt` [WFHS99]),
i.e. construire toutes les structures à Δ KCal.mol^{-1} de la MFE :

- Calculer la matrice des énergies minimales
- Effectuer un Backtrack sur toutes les contributions à $\leq \Delta$ de la MFE
- Mettre à jour Δ t.q. les futurs backtracks donnent ≥ 1 struct.
- Engendrer (Rec.) les sous-ensembles et combiner (brutal ou Tri)



Repliement sous-optimal

Prob. : Simplifications de l'énergie (Pseudo-noeuds, non-can.)
$\Rightarrow$ La structure **native** (fonctionnelle) pourrait être **ignorée**.

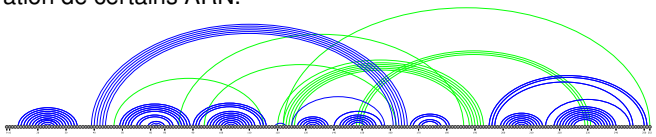
$\Rightarrow$ Engendrer des repliements sous-optimaux (`RNASubopt` [WFHS99]),
i.e. construire toutes les structures à Δ KCal.mol^{-1} de la MFE :

- Calculer la matrice des énergies minimales
- Effectuer un Backtrack sur toutes les contributions à $\leq \Delta$ de la MFE
- Mettre à jour Δ t.q. les futurs backtracks donnent ≥ 1 struct.
- Engendrer (Rec.) les sous-ensembles et combiner (brutal ou Tri)

$\Rightarrow$ Complexité en temps (Tri) : $\mathcal{O}(n^3 + nk\log(k))$
(k croît exponentiellement sur Δ !)

Repliement avec pseudo-noeuds

Les pseudo-noeuds (et vrais noeuds) sont des constituants essentiels à la structuration de certains ARN.



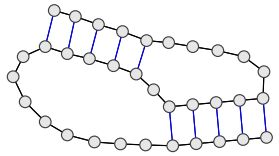
Ribozyme du groupe I

Leur absence historique au sein algorithmes de repliement est liée à la difficulté algorithmique des problèmes associés.
(Présence de croisements interdit une hypothèse d'indépendance des repliements).

Type	Complexité	Référence
Structures secondaires	$\mathcal{O}(n^3)$	[MSZT99]
L&P	$\mathcal{O}(n^5)$	[LP00]
D&P	$\mathcal{O}(n^5)$	[DP03]
A&U	$\mathcal{O}(n^5)$	[Aku00]
R&E	$\mathcal{O}(n^6)$	[RE99]
Généraux	NP-complet	[LP00]

Algorithme d'Akutsu/Uemura

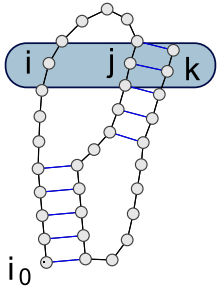
But : Capturer des catégorie de pseudo-noeuds *simples*, mais très représentés.



Idée : Quand on *retourne* ce type de pseudonoeuds, il suffit de précalculer les meilleures configurations *en dessous* d'un triplet (i, j, k) pour obtenir son énergie minimale.

Algorithme d'Akutsu/Uemura

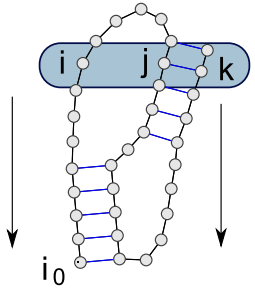
But : Capturer des catégorie de pseudo-noeuds *simples*, mais très représentés.



Idée : Quand on *retourne* ce type de pseudonoeuds, il suffit de précalculer les meilleures configurations *en dessous* d'un triplet (i, j, k) pour obtenir son énergie minimale.

Algorithme d'Akutsu/Uemura

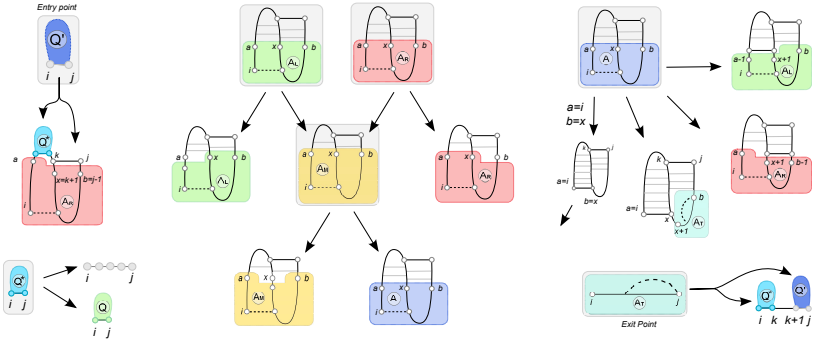
But : Capturer des catégorie de pseudo-noeuds *simples*, mais très représentés.



Idée : Quand on *retourne* ce type de pseudonoeuds, il suffit de précalculer les meilleures configurations *en dessous* d'un triplet (i, j, k) pour obtenir son énergie minimale.

Yann Ponty M2 BIM - Lecture 2 - Beyond MFE folding

Akutsu/Uemura : Programmation dynamique



Application/Problème	Weight fun.	Time/Space	Ref.
Minimisation d'énergie	π_{bp}	$\mathcal{O}(n^4)/\mathcal{O}(n^4)$	[Aku00]
Fonction de partition	$e^{\frac{-\pi_{bp}}{RT}}$	$\mathcal{O}(n^4)/\mathcal{O}(n^4)$	$\Theta(n^6)$[CC09]
Probabilité de paires de bases	$e^{\frac{-\pi_{bp}}{RT}}$	$\mathcal{O}(n^4)/\mathcal{O}(n^4)$	–
Échantillonnage (k-struct.)	$e^{\frac{-\pi_{bp}}{RT}}$	$\mathcal{O}(n^4 + kn\log n)/\mathcal{O}(n^4)$	–

Exercice : Ecrire l'équation de programmation dynamique associée pour le repliement, le comptage et la fonction de partition.

References I

Tatsuya Akutsu.
Dynamic programming algorithms for rna secondary structure prediction with pseudoknots.
Discrete Appl. Math., 104(1-3) :45–62, 2000.

S. Cao and S-J Chen.
Predicting structured and stabilities for h-type pseudoknots with interhelix loop.
RNA, 15 :696–706, 2009.

Y. Ding, C. Y. Chan, and C. E. Lawrence.
RNA secondary structure prediction by centroids in a boltzmann weighted ensemble.
RNA, 11 :1157–1166, 2005.

Y. Ding and E. Lawrence.
A statistical sampling algorithm for RNA secondary structure prediction.
Nucleic Acids Research, 31(24) :7280–7301, 2003.

Robert M Dirks and Niles A Pierce.
A partition function algorithm for nucleic acid secondary structure including pseudoknots.
J Comput Chem, 24(13) :1664–1677, Oct 2003.

R. B. Lyngsøand C. N. S. Pedersen.
RNA pseudoknot prediction in energy-based models.
Journal of Computational Biology, 7(3-4) :409–427, 2000.

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.

Y. Ponty.
Efficient sampling of RNA secondary structures from the boltzmann ensemble of low-energy : The boustrophedon method.
Journal of Mathematical Biology, 56(1-2) :107–127, Jan 2008.

References II

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

S. Wuchty, W. Fontana, I.L. Hofacker, and P. Schuster.
Complete suboptimal folding of RNA and the stability of secondary structures.
Biopolymers, 49 :145–164, 1999.