Cours M2 BIBS - Séance 2

Programmation dynamique avancée

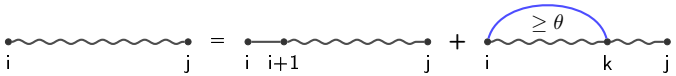
Yann Ponty

Bioinformatics Team
École Polytechnique/CNRS/INRIA AMIB – France

18 janvier 2010

Résumé

Décomposition de Nussinov/Jacobson



Récurrence sur l'énergie minimale d'un repliement :

$$N_{i,t} = 0, \quad \forall t \in [i, i+\theta]$$

$$N_{i,j} = \min \begin{cases} N_{i+1,j} & (i \text{ non apparié}) \\ \min_{k=i+\theta+1}^{j} E_{i,j} + N_{i+1,k-1} + N_{k+1,j} & (i \text{ comp. avec } k) \end{cases}$$

Récurrence de comptage des structures compatibles :

$$C_{i,t} = 1, \quad \forall t \in [i, i+\theta]$$

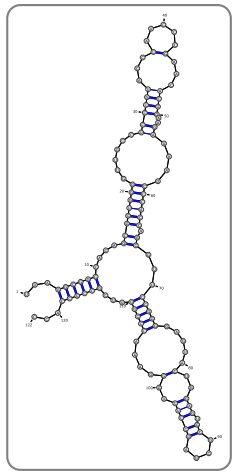
$$C_{i,j} = \sum \begin{cases} C_{i+1,j} & (i \text{ non apparié}) \\ \sum_{k=i+\theta+1}^{j} 1 \times C_{i+1,k-1} \times C_{k+1,j} & (i \text{ comp. avec } k) \end{cases}$$

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

Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2$^{\text{aire}}$:

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



Énergies libres Δ G des boucles dépendent des bases, assymétrie, bases *libres* (dangle) ...

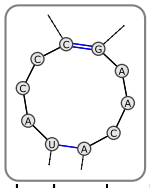
Déterminées expérimentalement
+ Interpolation pour grandes boucles.

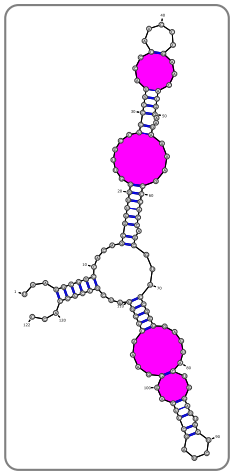
Meilleure résultats grâce à la prise en compte de l'empilement.

Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2^{aire} :

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





Énergies libres Δ G des boucles dépendent des bases, assymétrie, bases *libres* (dangle) . . .

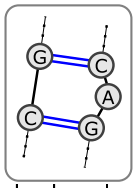
Déterminées expérimentalement
+ Interpolation pour grandes boucles.

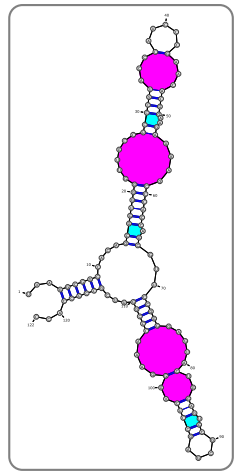
Meilleure résultats grâce à la prise en compte de l'empilement.

Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2^{aire} :

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





Énergies libres Δ G des boucles dépendent des bases, assymétrie, bases *libres* (dangle) . . .

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

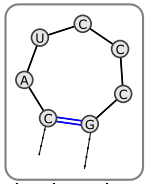
Meilleure résultats grâce à la prise en compte de l'empilement.

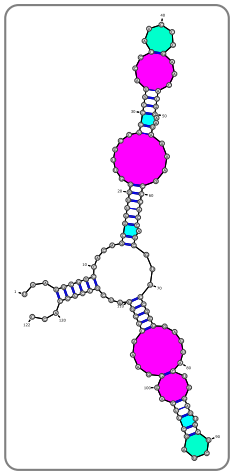
Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2^{aire} :

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





Énergies libres Δ G des boucles dépendent des bases, assymétrie, bases *libres* (dangle) . . .

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

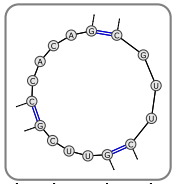
Meilleure résultats grâce à la prise en compte de l'empilement.

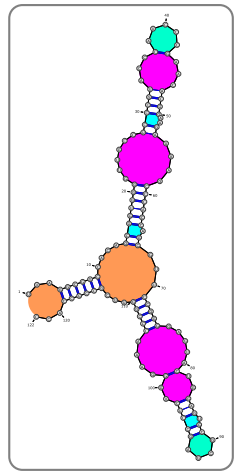
Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2^{aire} :

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





Énergies libres Δ G des boucles dépendent des bases, assymétrie, bases *libres* (dangle) . . .

Déterminées expérimentalement
+ Interpolation pour grandes boucles.

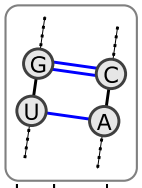
Meilleure résultats grâce à la prise en compte de l'empilement.

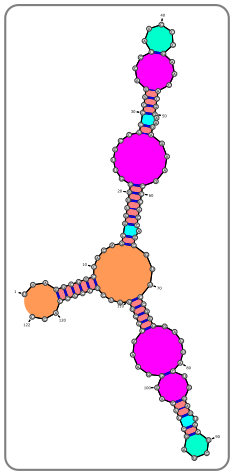
Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Modèle de Turner

Basée sur décomposition non-ambiguë en boucles de la structure 2^{aire} :

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



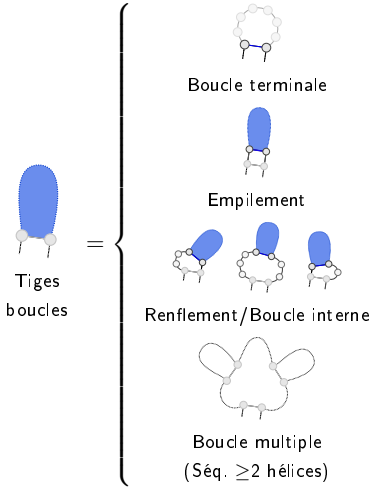


Énergies libres Δ G des boucles dépendent des bases, assymétrie, bases *libres* (dangle) . . .

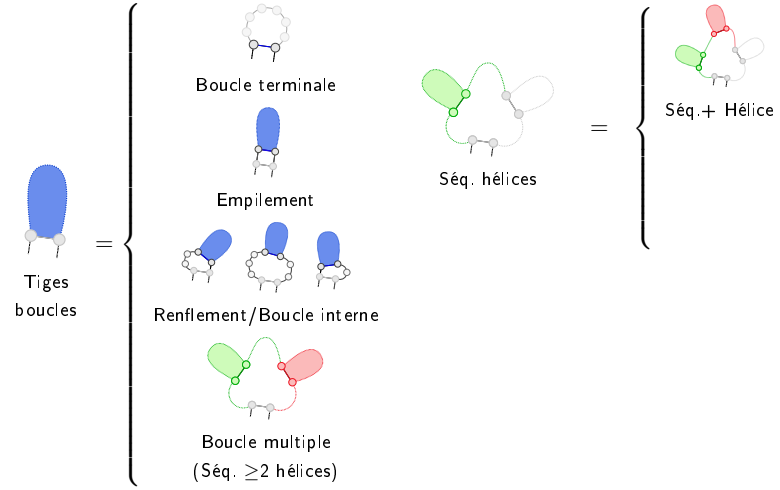
Déterminées expérimentalement
+ Interpolation pour grandes boucles.

Meilleure résultats grâce à la prise en compte de l'empilement.

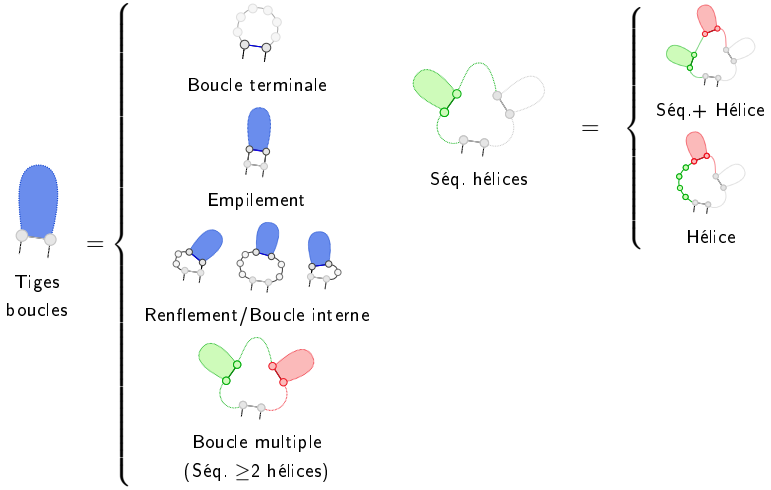
MFE DP equations



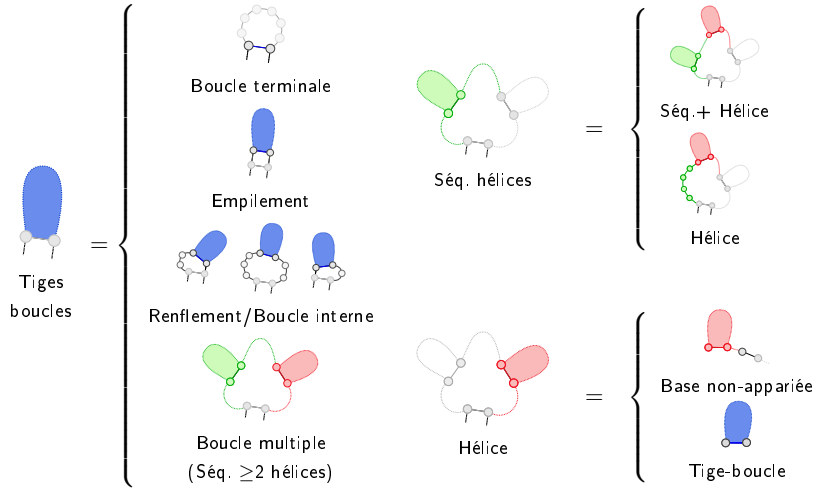
MFE DP equations



MFE DP equations

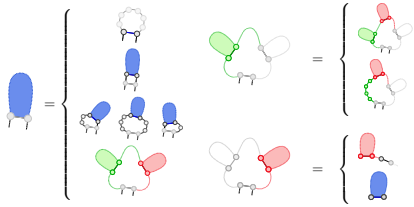


MFE DP equations



MFold Unafold

- $E_H(i,j)$: Energie de boucle terminale *fermée* par une paire (i,j)
- $E_{BI}(i,j)$: Energie de renflement ou boucle interne *fermée* par une paire (i,j)
- $E_S(i,j)$: Energie d'empilement $(i,j)/(i+1,j-1)$
- a,c,b : Pénalité de boucle multiple, hélice et non-appariées dans multiboucle.



Calcul des matrices

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

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

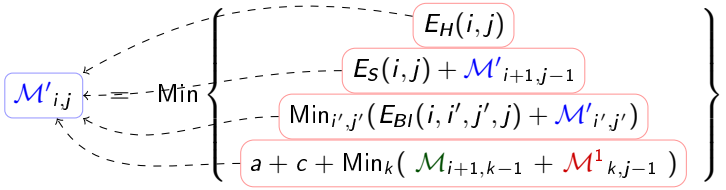
$\mathcal{O}(n)$ contributeurs potentiels au Min :
$\Rightarrow$ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.
Garder les meilleures contributions aux Min $\Rightarrow$ Backtrack en $\mathcal{O}(n)$

Complexités temps/mémoire en $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ pour le précalcul [1]
$\Rightarrow$ `UnaFold` [MZ08] calcule la structure secondaire d'énergie minimale.

1. Avec une astuce pour les bulges/boucles internes ...

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :



$$\begin{aligned}
\mathcal{M}_{i,j} &= \mathrm{Min}_k \left\{ \min\left(\mathcal{M}_{i,k-1}, b(k-1)\right) + {\mathcal{M}^1}_{k,j} \right\} \\
{\mathcal{M}^1}_{i,j} &= \mathrm{Min}_k \left\{ b + {\mathcal{M}^1}_{i,j-1}, c + \mathcal{M}'_{i,j} \right\}
\end{aligned}$$

$\mathcal{O}(n)$ contributeurs potentiels au Min :
$\Rightarrow$ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.
Garder les meilleures contributions aux Min $\Rightarrow$ Backtrack en $\mathcal{O}(n)$

Complexités temps/mémoire en $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ pour le précalcul [1]
$\Rightarrow$ `UnaFold` [MZ08] calcule la structure secondaire d'énergie minimale.

1. Avec une astuce pour les bulges/boucles internes ...

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

$$\mathcal{M}_{i,j} = \text{Min}_k\left\{\min\left(\mathcal{M}_{i,k-1}, b(k-1)\right) + {\mathcal{M}^1}_{k,j}\right\}$$

$$\mathcal{M}^1_{i,j} = \text{Min}_k\left\{b + {\mathcal{M}^1}_{i,j-1}, c + \mathcal{M}'_{i,j}\right\}$$

$\mathcal{O}(n)$ contributeurs potentiels au Min :
⇒ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.
Garder les meilleures contributions aux Min ⇒ Backtrack en $\mathcal{O}(n)$

Complexités temps/mémoire en $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ pour le précalcul [1]
⇒ `UnaFold` [MZ08] calcule la structure secondaire d'énergie minimale.

1. Avec une astuce pour les bulges/boucles internes ...

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

$$\mathcal{M}_{i,j} = \text{Min}_k\left\{\min\left(\mathcal{M}_{i,k-1}, b(k-1)\right) + {\mathcal{M}^1}_{k,j}\right\}$$

$$\mathcal{M}^1_{i,j} = \text{Min}_k\left\{b + {\mathcal{M}^1}_{i,j-1}, c + \mathcal{M}'_{i,j}\right\}$$

$\mathcal{O}(n)$ contributeurs potentiels au Min :
⇒ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.
Garder les meilleures contributions aux Min ⇒ Backtrack en $\mathcal{O}(n)$

Complexités temps/mémoire en $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ pour le précalcul [1]
⇒ `UnaFold` [MZ08] calcule la structure secondaire d'énergie minimale.

1. Avec une astuce pour les bulges/boucles internes ...

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Remontée (Backtracking)

Reconstruction de la structure d'énergie minimale :

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

$$\mathcal{M}_{i,j} = \text{Min}_k\left\{\min\left(\mathcal{M}_{i,k-1}, b(k-1)\right) + {\mathcal{M}^1}_{k,j}\right\}$$

$$\mathcal{M}^1_{i,j} = \text{Min}_k\left\{b + {\mathcal{M}^1}_{i,j-1}, c + \mathcal{M}'_{i,j}\right\}$$

$\mathcal{O}(n)$ contributeurs potentiels au Min :
⇒ Complexité au pire en $\mathcal{O}(n^2)$ pour un backtrack naif.
Garder les meilleures contributions aux Min ⇒ Backtrack en $\mathcal{O}(n)$

Complexités temps/mémoire en $\mathcal{O}(n^3)/\mathcal{O}(n^2)$ pour le précalcul [1]
⇒ `UnaFold` [MZ08] calcule la structure secondaire d'énergie minimale.

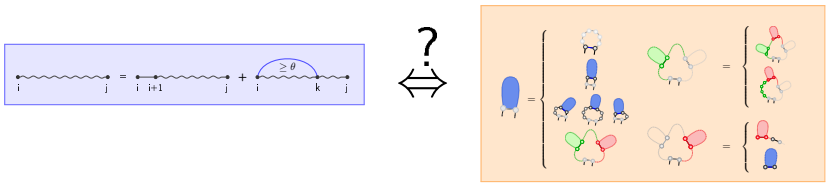
1. Avec une astuce pour les bulges/boucles internes ...

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Validité d'un schéma

Une preuve de correction possible :

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



Forte certitude mais pas encore preuve (Séries génératrices).

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Validité d'un schéma

Une preuve de correction possible :

Calcul correct localement

\+ Toutes les conformations sont parcourues

⇒ Algorithme correct (Induction)

$$C_{i,t} = 1, \quad \forall t \in [i, i+\theta]$$
$$C_{i,j} = \sum \left\{ \sum_{k=i+\theta+1}^{C_{i+1,j}}{}_{j} 1 \times C_{i+1,k-1} \times C_{k+1,j} \right.$$

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

$\Leftrightarrow$?

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

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

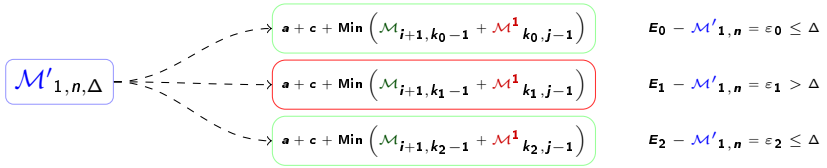
Forte certitude mais pas encore preuve (Séries génératrices).

Repliement sous-optimal

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

⇒ 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)

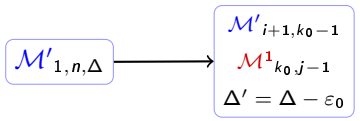


Repliement sous-optimal

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

⇒ 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)

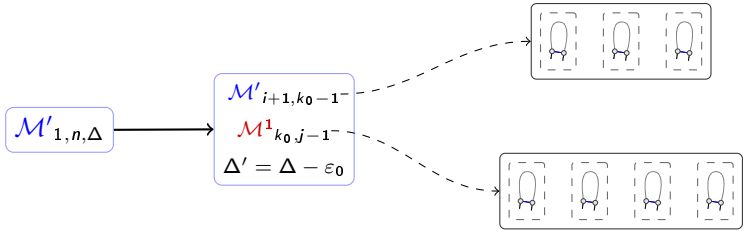


Repliement sous-optimal

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

⇒ 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)

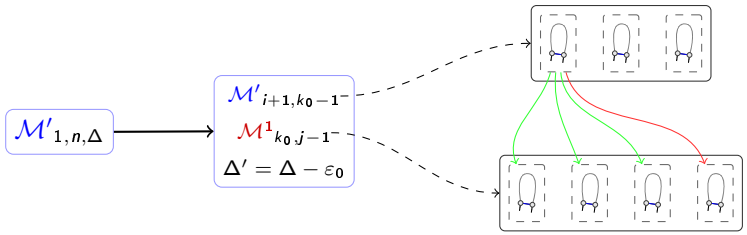


Repliement sous-optimal

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

⇒ 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)

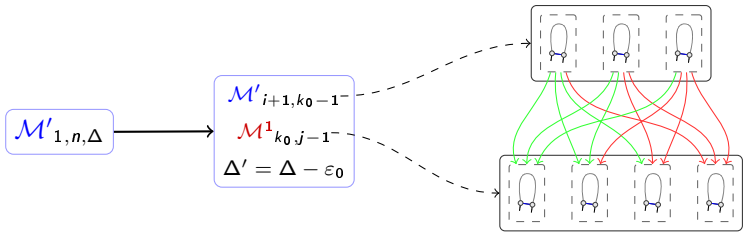


Repliement sous-optimal

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

⇒ 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.)
⇒ La structure native (fonctionnelle) pourrait être ignorée.

⇒ 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)

⇒ Complexité en temps (Tri) : $\mathcal{O}(n^3 + nk\log(k))$
(k croît exponentiellement sur Δ, mais bon...)

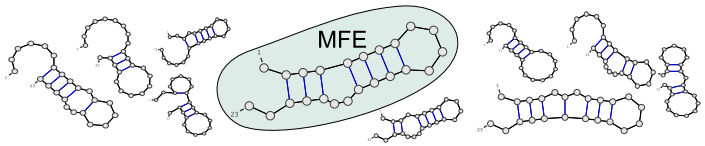
Résumé

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.



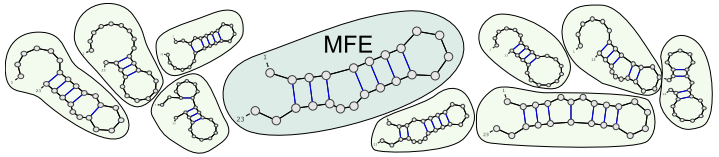
Conséquence : La probabilité de la MFE peut être négligeable.
⇒ La structure native (fonctionnelle) est alors à chercher dans les sous-optimales, où des structures très proches peuvent se *grouper*.

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.
⇒ La structure native (fonctionnelle) est alors à chercher dans les sous-optimales, où des structures très proches peuvent se *grouper*.

Distribution de Boltzmann : Définitions

Une distribution de Bolzmann pondère chaque structure S pour un ARN ω par un facteur de Boltzmann $e^{\frac{-E_{S,\omega}}{RT}}$ 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

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

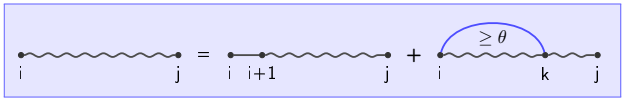
où $\mathcal{S}_\omega$ est l'ensemble des conformations compatibles avec ω.

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

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

Fonction de partition

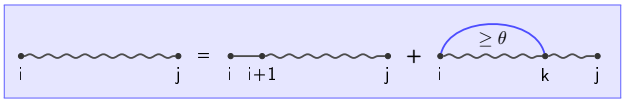
Fonction de partition = Comptage pondéré des structures compatibles



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

Fonction de partition

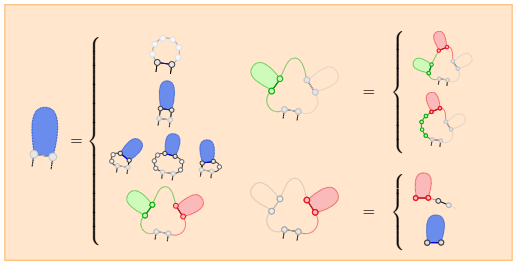
Fonction de partition = Comptage pondéré des structures compatibles



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

Fonction de partition

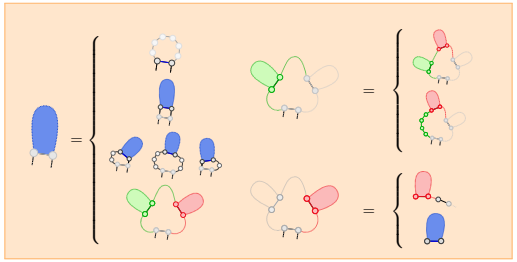
Fonction de partition = Comptage pondéré des structures compatibles



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

Fonction de partition

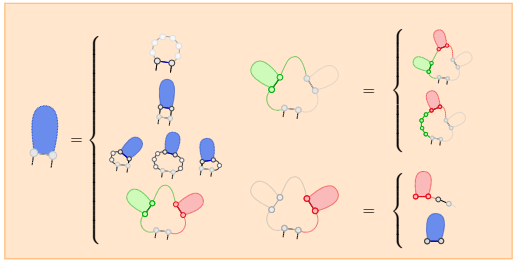
Fonction de partition = Comptage pondéré des structures compatibles



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

Fonction de partition

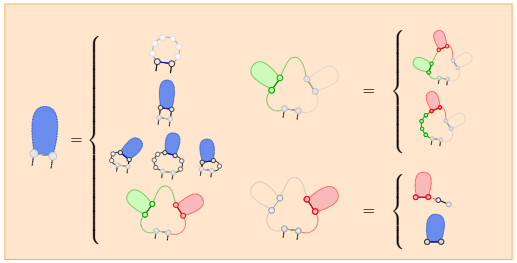
Fonction de partition = Comptage pondéré des structures compatibles



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

Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles



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

Fonction de partition

Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Fonction de partition

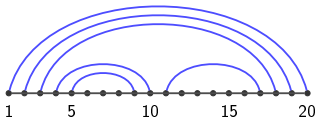
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

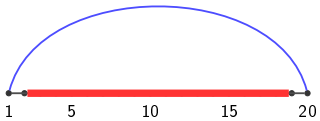
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

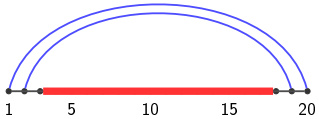
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

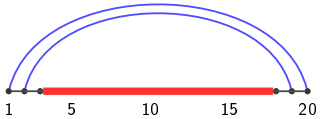
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

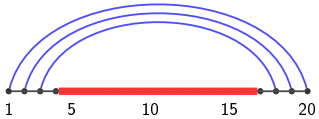
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

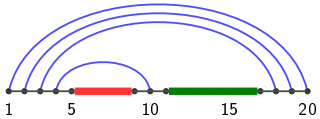
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

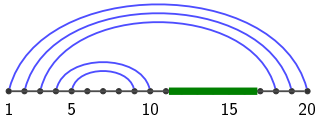
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

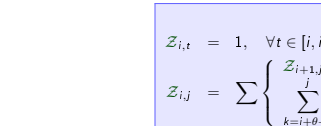
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

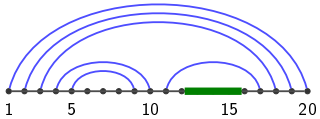
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

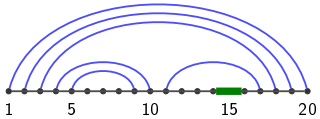
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

$S =$ 

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

Fonction de partition

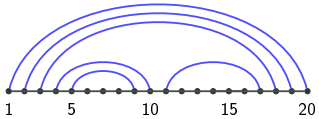
Fonction de partition = Comptage pondéré des structures compatibles

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

Validité de la fonction de partition :

- Exhaustivité/non ambiguité du schéma
- Correction du facteur de Boltzmann
  Facteur d'un backtrack = Produit des facteurs de ses parties
  Contributions énergétiques passent à l'exposant

Exemple :

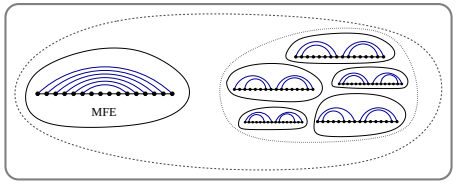
$S =$ 

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

Échantillonnage statistique de structures d'ARN

Motivation de l'échantillonnage statistique

La MFE (Probabilité de Boltzmann max) $\mathcal{M}$ peut être isolée et moins probable qu'un ensemble $\mathcal{B}$ de sous-optimaux similaires structurellement. Alors, la structure native est plus probablement dans $\mathcal{B}$ que $\mathcal{M}$ [DCL05].



Approche :

- 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)

Remontée stochastique

Algorithme 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

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

Remontée stochastique

Algorithme 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

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

Remontée stochastique

Algorithme 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

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

r ↓

$A_1 | A_2 | B_i | B_{i+1} | \ldots | B_{j-1} | B_j | C_i | C_{i+1} | \ldots | C_{j-1} | C_j$

Remontée stochastique

Algorithme 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

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

r ↓

$A_1 | A_2 | B_i | B_{i+1} | \ldots | B_{j-1} | B_j | C_i | C_{i+1} | \ldots | C_{j-1} | C_j$

Remontée stochastique

Algorithme 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

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

r ↓

$A_1 | A_2 | B_i | B_{i+1} | \ldots | B_{j-1} | B_j | C_i | C_{i+1} | \ldots | C_{j-1} | C_j$

Remontée stochastique

Algorithme 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

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

r ↓

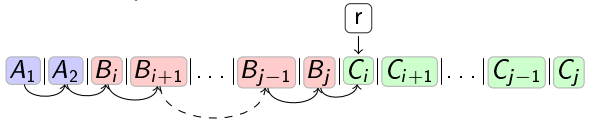
$A_1 | A_2 | B_i | B_{i+1} | \ldots | B_{j-1} | B_j | C_i | C_{i+1} | \ldots | C_{j-1} | C_j$

Remontée stochastique

Algorithme 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

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

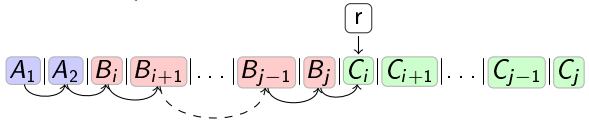


Remontée stochastique

Algorithme 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*. $\Rightarrow \mathcal{O}(n \log nk)$ au pire [Pon08].

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Conclusion

Biologie/biochimie

- Structure aide à la compréhension de la fonction
- Minimiser une fonction d'énergie n'est pas suffisant
- Boltzmann = équilibre $\Rightarrow$ Cinétique = convergence

$\Rightarrow$ Éviter une vision trop statique de la relation séquence/structure.

Bioinformatique

- Programmation dynamique (DP) est à la frontière du polynomial
- DP résout exactement des problèmes complexes
- Mais implémentation parfois problématiques
- Séparer espace des conformations/énergie fait gagner du temps

Chercher des ensembles de conformations discrets et simples[2].

2. Même si ça fait râler les biologistes/biochimistes ...

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Conclusion

Biologie/biochimie

- Structure aide à la compréhension de la fonction
- Minimiser une fonction d'énergie n'est pas suffisant
- Boltzmann = équilibre $\Rightarrow$ Cinétique = convergence

$\Rightarrow$ Éviter une vision trop statique de la relation séquence/structure.

Bioinformatique

- Programmation dynamique (DP) est à la frontière du polynomial
- DP résout exactement des problèmes complexes
- Mais implémentation parfois problématiques
- Séparer espace des conformations/énergie fait gagner du temps

Chercher des ensembles de conformations discrets et simples[2].

2. Même si ça fait râler les biologistes/biochimistes ...

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Conclusion

Biologie/biochimie

- Structure aide à la compréhension de la fonction
- Minimiser une fonction d'énergie n'est pas suffisant
- Boltzmann = équilibre ⇒ Cinétique = convergence

⇒ Éviter une vision trop statique de la relation séquence/structure.

Bioinformatique

- Programmation dynamique (DP) est à la frontière du polynomial
- DP résout exactement des problèmes complexes
- Mais implémentation parfois problématiques
- Séparer espace des conformations/énergie fait gagner du temps

Chercher des ensembles de conformations discrets et simples [2].

2. Même si ça fait râler les biologistes/biochimistes ...

Conclusion

Biologie/biochimie

- Structure aide à la compréhension de la fonction
- Minimiser une fonction d'énergie n'est pas suffisant
- Boltzmann = équilibre ⇒ Cinétique = convergence

⇒ Éviter une vision trop statique de la relation séquence/structure.

Bioinformatique

- Programmation dynamique (DP) est à la frontière du polynomial
- DP résout exactement des problèmes complexes
- Mais implémentation parfois problématiques
- Séparer espace des conformations/énergie fait gagner du temps

Chercher des ensembles de conformations discrets et simples [2].

2. Même si ça fait râler les biologistes/biochimistes ...

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Conclusion

Biologie/biochimie

- Structure aide à la compréhension de la fonction
- Minimiser une fonction d'énergie n'est pas suffisant
- Boltzmann = équilibre ⇒ Cinétique = convergence

⇒ Éviter une vision trop statique de la relation séquence/structure.

Bioinformatique

- Programmation dynamique (DP) est à la frontière du polynomial
- DP résout exactement des problèmes complexes
- Mais implémentation parfois problématiques
- Séparer espace des conformations/énergie fait gagner du temps

Chercher des ensembles de conformations discrets et simples [2].

2. Même si ça fait râler les biologistes/biochimistes ...

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Conclusion

Biologie/biochimie

- Structure aide à la compréhension de la fonction
- Minimiser une fonction d'énergie n'est pas suffisant
- Boltzmann = équilibre ⇒ Cinétique = convergence

⇒ Éviter une vision trop statique de la relation séquence/structure.

Bioinformatique

- Programmation dynamique (DP) est à la frontière du polynomial
- DP résout exactement des problèmes complexes
- Mais implémentation parfois problématiques
- Séparer espace des conformations/énergie fait gagner du temps

Chercher des ensembles de conformations discrets et simples [2].

2. Même si ça fait râler les biologistes/biochimistes ...

Yann Ponty Cours M2 BIBS - Séance 1 - Repliement de l'ARN

Conclusion

Biologie/biochimie

- Structure aide à la compréhension de la fonction
- Minimiser une fonction d'énergie n'est pas suffisant
- Boltzmann = équilibre ⇒ Cinétique = convergence

⇒ Éviter une vision trop statique de la relation séquence/structure.

Bioinformatique

- Programmation dynamique (DP) est à la frontière du polynomial
- DP résout exactement des problèmes complexes
- Mais implémentation parfois problématiques
- Séparer espace des conformations/énergie fait gagner du temps

Chercher des ensembles de conformations discrets et simples[2].

2. Même si ça fait râler les biologistes/biochimistes . . .

References I


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.


N. R. Markham and M. Zuker.
Bioinformatics, chapter UNAFold, pages 3–31.
Springer, 2008.


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.


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.