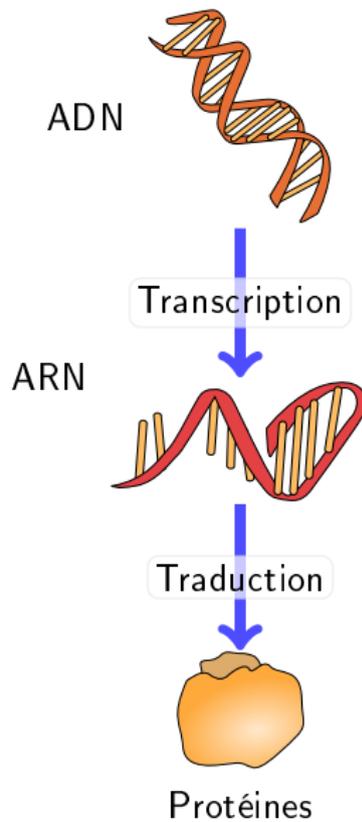


Équipe Bioinfo LIX - INRIA AMIB

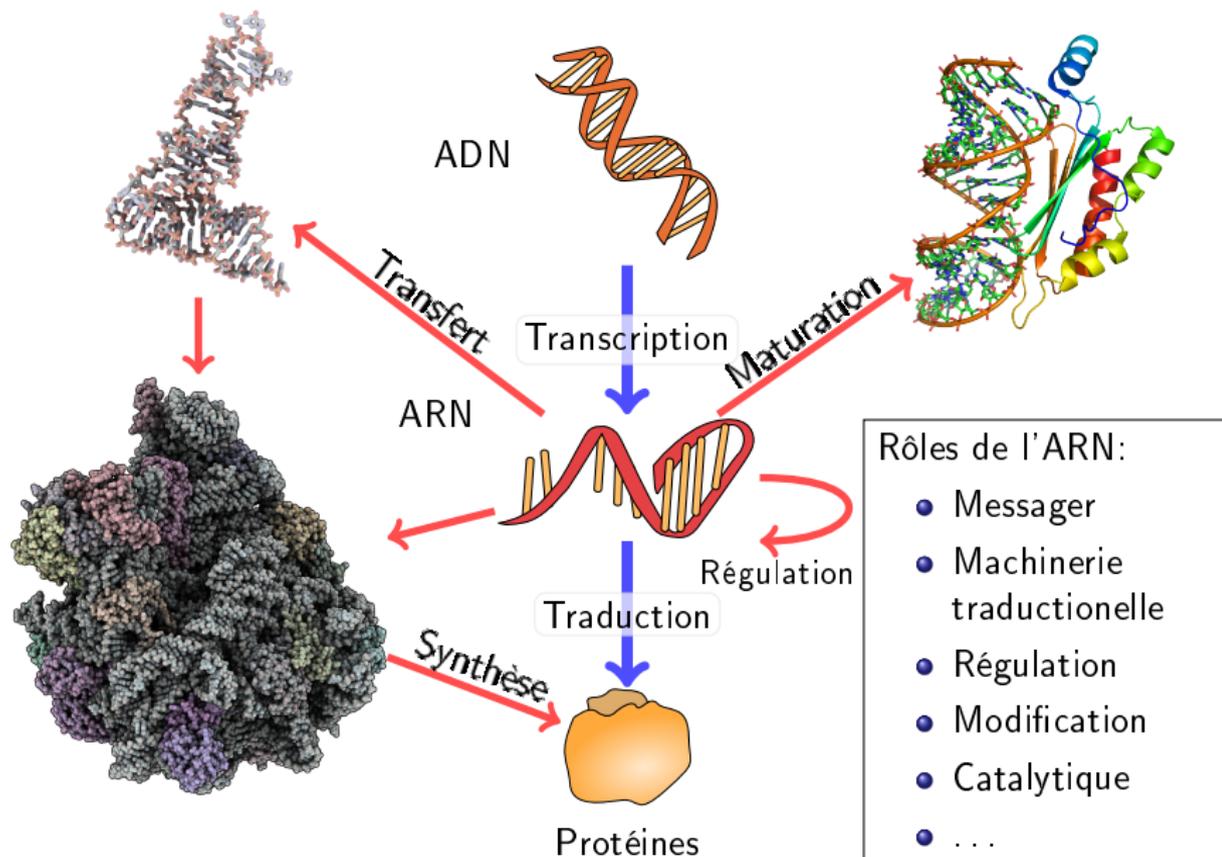
Yann Ponty

Polytechnique/CNRS/INRIA AMIB – France

Dogme fondamental de la biologie moléculaire



Dogme fondamental de la biologie moléculaire

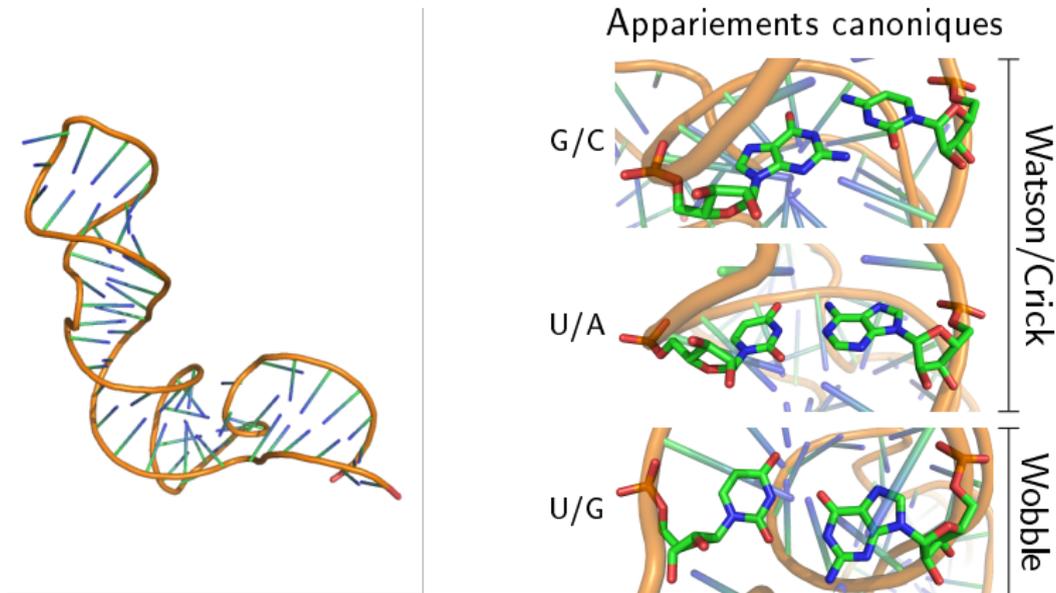


But ultime : Créer automatiquement des connaissances biologiques.
On s'attaque au **quantitatif**, mais aussi parfois au **qualitatif**.

- **Biochimie *in silico*** : Comprendre la biologie *par le bas*
Problèmes : Repliement, Prédiction d'interactions
⇒ Simulation, optimisation, espaces de conformations
- **Ingénierie moléculaire** : Design, nanotechnologies
Problèmes : Optimisation inverse, affinité, autoassemblage
⇒ Recherche locale, évolution dirigée
- **Génomique** : Séquences (ADN, ARN, Acides aminés)
Baisse récente et spectaculaire des coûts de séquençage :
Séquençage haut-débit → 10 000\$ en une semaine !! (Illumina)
Problèmes : Réassemblage, Alignement, Détection de motifs ...
⇒ Algorithmique du texte, combinatoire

Repliement de l'ARN

ARN = Polymère linéaire composé de nucléotides (A,C,G,U)



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

Où est ce que l'informatique apparaît ?

Première approche : Ramener le problème à un problème discret.

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCGAA
CACGGAAGAUAAAGCC
CACCAGCGUUCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC
```

Structure primaire



Structure tertiaire

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

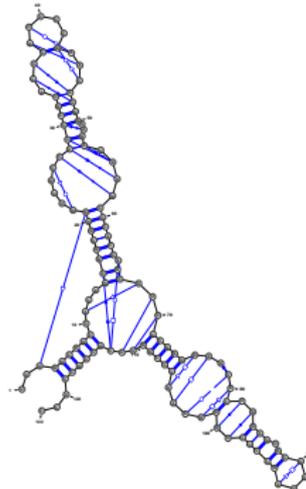
Modèle discret pour repliement

Repliement de l'ARN = Maximisation (pondérée) des appariements.
(\Rightarrow Max. Matching, Max. Independent Set, Context-Free Parsing...)

Première approche : Ramener le problème à un problème discret.

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCGAA
CACGGAAGAUAAAGCC
CACCAGCGUUCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC
```

Structure primaire



Structure secondaire⁺



Structure tertiaire

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

Modèle discret pour repliement

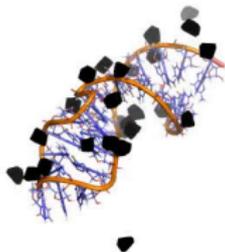
Repliement de l'ARN = Maximisation (pondérée) des appariements.
(\Rightarrow Max. Matching, Max. Independent Set, Context-Free Parsing...)

Deuxième approche : Dynamique moléculaire

Simuler le phénomène du repliement

Avantages : Modèles biochimiques plus complets/complexes

Mais pas de temps ΔT très petit \Rightarrow Simulation très (trop) coûteuse !!



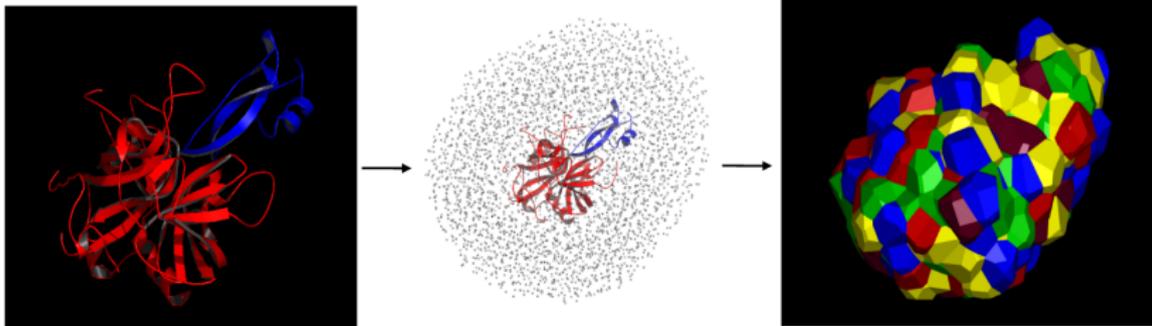
Encore une fois l'informatique peut aider en proposant des simplifications réalistes des espaces de conformations.

Diagrammes de Voronoï

Diagrammes de Voronoï =

Espace géométrique discret de conformations basés sur la distance.

Exemple : Protéines



Diagrammes basés sur la **distance**

⇒ **Plus besoin** de considérer **chaque paire d'atomes** !

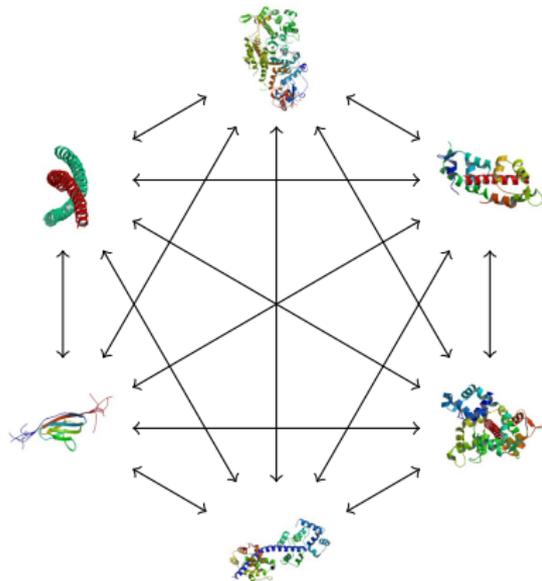
Bonus : Une fois les **modèles d'énergie** réexprimés sous cette forme, on peut utiliser l'arsenal de l'**algorithmique géométrique**.

Prédiction d'interactions

Constat : **Fonction biologiques** naissent de l'**interaction des molécules**.

Données disponibles :

- **Génomiques :**
Pression sélective pesant sur le réseau
⇒ Conservation, mutations comp.
- **Biochimiques :**
Interaction fondée au niveau moléculaire
⇒ Simulation de l'assemblage (Docking)

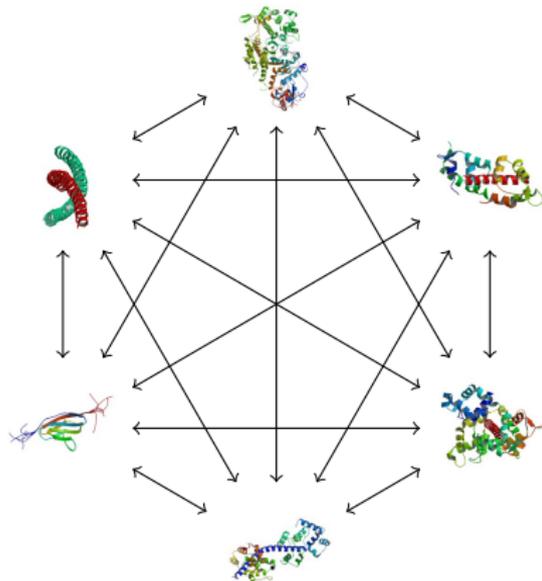


Prédiction d'interactions

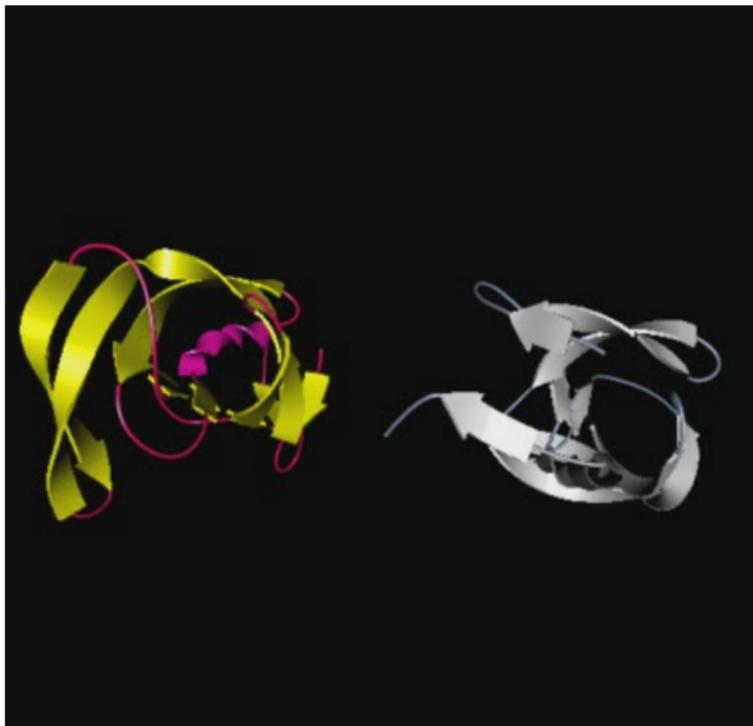
Constat : **Fonction biologiques** naissent de l'**interaction des molécules**.

Données disponibles :

- **Génomiques** :
Pression sélective pesant sur le réseau
⇒ Conservation, mutations comp.
- **Biochimiques** :
Interaction fondée au niveau moléculaire
⇒ Simulation de l'assemblage (Docking)



Docking : Définition par l'exemple

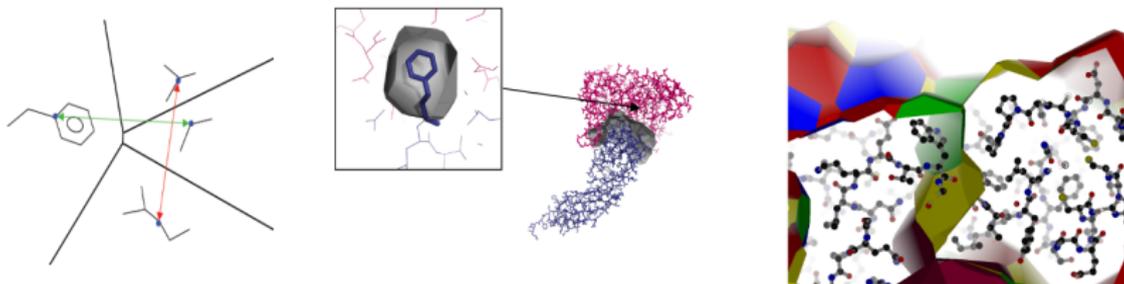


O. Kreylos *et al*, IEEE Visualization 2003

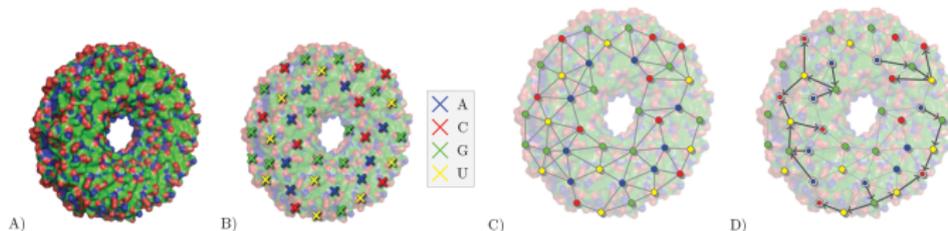
Modèles discrets pour les interactions

Ici aussi, utiliser des conformations discrètes aide beaucoup !!!

Exemple 1: Voronoï pour interactions protéines/protéines



Exemple 2: Modèles *graphes* pour interactions protéines/ARN



- Biochimie *in silico* : Comprendre la biologie *par le bas*
- Génomique : Séquences (ADN, ARN, Acides aminés)
Baisse récente et spectaculaire des coûts de séquençage :
Séquençage haut-débit → 10 000\$ en une semaine !! (Illumina)
Problèmes : Réassemblage, Alignement, Détection de motifs ...
⇒ Algorithmique du texte, combinatoire
- Ingénierie moléculaire : Design, nanotechnologies
Problème : Optimisation inverse, affinité, autoassemblage
⇒ Recherche locale, évolution dirigée

Multiplication des données/Modèles de plus en plus précis (et complexes)
⇒ L'informatique peut (et doit) aider la biologie à faire face.

Venez nous aider à l'aider !

- Biochimie *in silico* : Comprendre la biologie *par le bas*
- Génomique : Séquences (ADN, ARN, Acides aminés)
Baisse récente et spectaculaire des coûts de séquençage :
Séquençage haut-débit → 10 000\$ en une semaine !! (Illumina)
Problèmes : Réassemblage, Alignement, Détection de motifs ...
⇒ Algorithmique du texte, combinatoire
- Ingénierie moléculaire : Design, nanotechnologies
Problème : Optimisation inverse, affinité, autoassemblage
⇒ Recherche locale, évolution dirigée

Multiplication des données/Modèles de plus en plus précis (et complexes)
⇒ L'informatique peut (et doit) aider la biologie à faire face.

Venez nous aider à l'aider !

- Biochimie *in silico* : Comprendre la biologie *par le bas*
- Génomique : Séquences (ADN, ARN, Acides aminés)
Baisse récente et spectaculaire des coûts de séquençage :
Séquençage haut-débit → 10 000\$ en une semaine !! (Illumina)
Problèmes : Réassemblage, Alignement, Détection de motifs ...
⇒ Algorithmique du texte, combinatoire
- Ingénierie moléculaire : Design, nanotechnologies
Problème : Optimisation inverse, affinité, autoassemblage
⇒ Recherche locale, évolution dirigée

Multiplication des données/Modèles plus en plus précis (et complexes)
⇒ L'informatique peut (et doit) aider la biologie à faire face.

Venez nous aider à l'aider !