



Recherche de motifs : Introduction

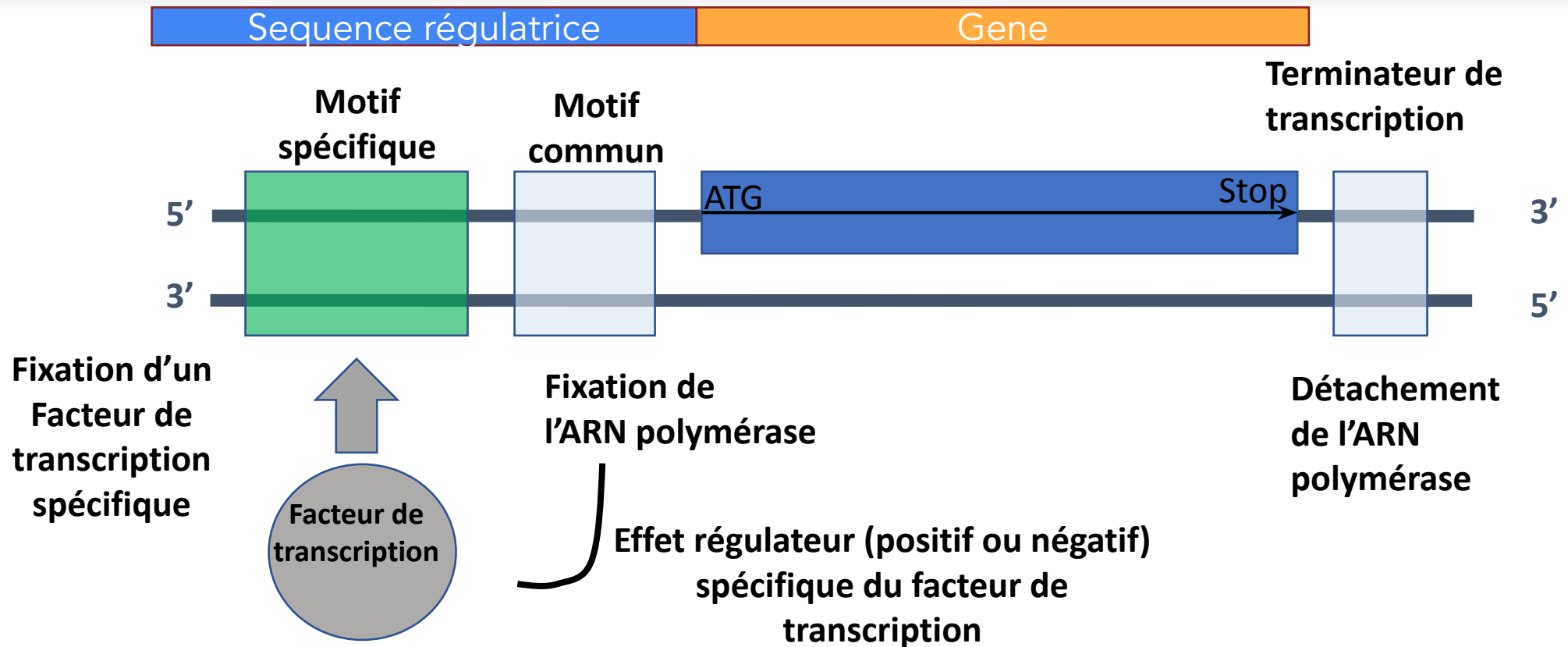
Cours 4



Plan du cours

- Région régulatrice
- Motif spécifique
- Définition formel du problème “Recherche de Motif”
- Comment générer de données artificiel

Région régulatrice pour l'activation génique



Séquence régulatrice



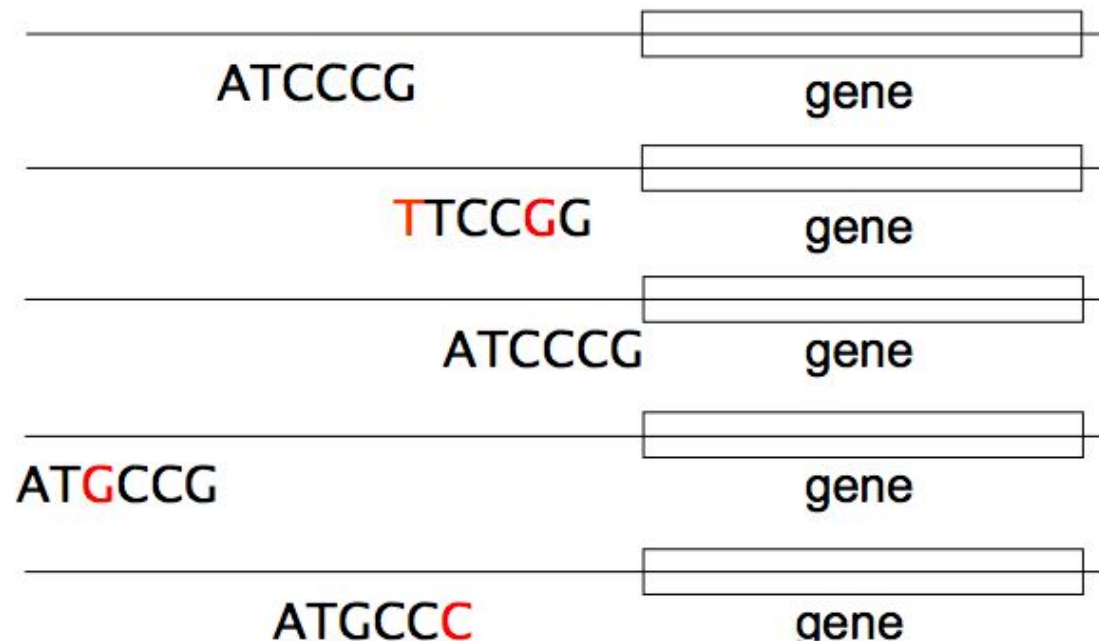
- Chaque gène contient une région régulatrice s'étendant typiquement de 100 à 1000 pb en amont du site d'initiation de la transcription.

Séquences régulatrices



- Les sites de liaison de facteurs de transcription (les motifs) sont spécifiques à un facteur de transcription donné situés dans la région régulatrice.

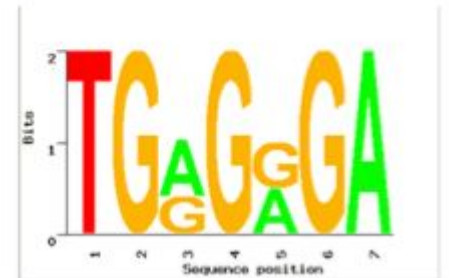
Le Motif spécifique peut être situé n'importe où dans la région régulatrice



Les motifs peuvent muter sur des bases non importantes.

- Exemple:
 - Les cinq motifs de cinq gènes différents ont des mutations en position 3 et 5.
 - Ces représentations appelées logos illustrent les régions conservées et variables d'un motif.

TG**GG**GA
TG**AG**GA
TG**GG**GA
TG**AG**GA
TG**AG**GA



Identifier les motifs dans les régions régulatrices

- La découverte de motifs similaires dans les régions régulatrices de plusieurs gènes suggère une relation de régulation entre ces gènes.
- **Problèmes**
 - Nous ne connaissons pas la séquence du motif à l'avance.
 - Nous ne savons pas où se situe le motif par rapport au début des gènes.
 - Un motif peut différer légèrement d'un gène à l'autre.

Identifier les motifs dans les régions régulatrices

- La découverte de motifs similaires dans les régions régulatrices de plusieurs gènes suggère une relation de régulation entre ces gènes.
- **Problèmes**
 - Nous ne connaissons pas la séquence du motif à l'avance.
 - Nous ne savons pas où se situe le motif par rapport au début des gènes.
 - Un motif peut différer légèrement d'un gène à l'autre.
- Notre question : **Comment identifier les vrais motif lié à de factor de transcription?**

Recherche de Motif : The Motif Finding Problem

- Nous allons voir une série d'**algorithmes** pour chercher des motifs.
- On cherche toujours **un motif de taille fixe**, identique ou contenant des variations.
- Pour pouvoir bien tester les algorithmes **nous allons générer de données artificiel** (ou de contrôle)

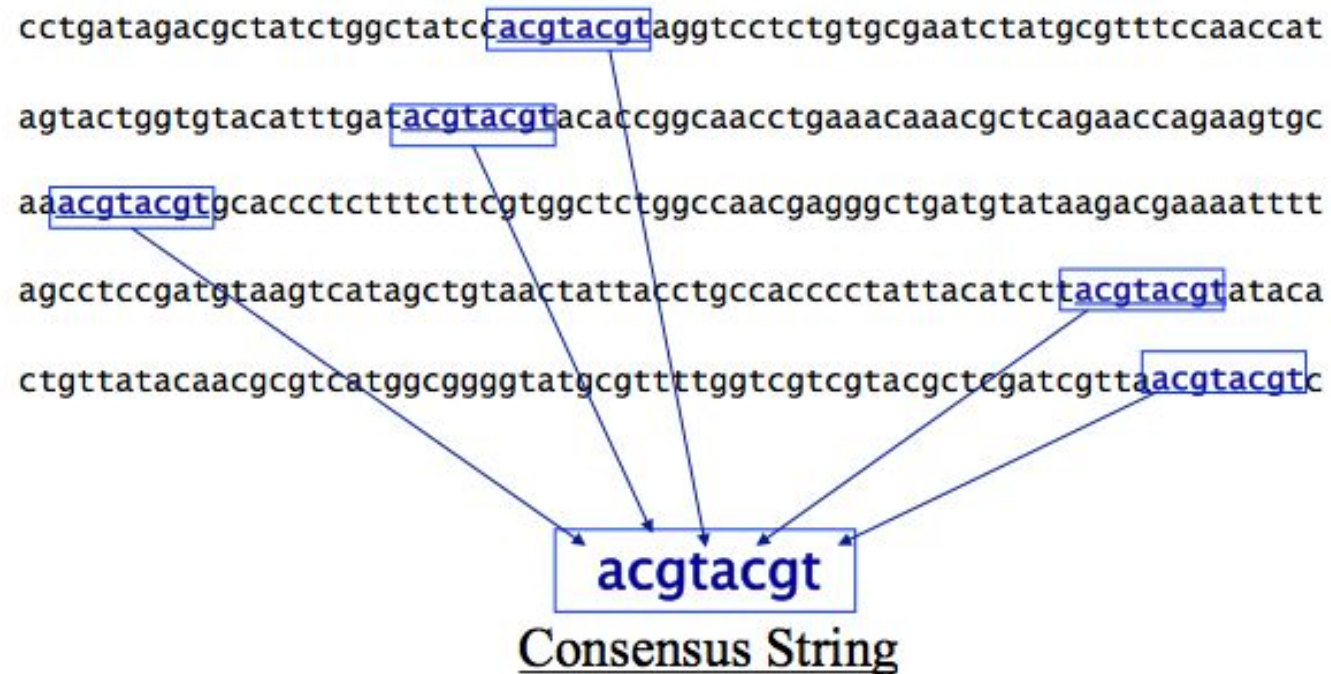
Motifs implantés

- Générer un échantillon aléatoire de séquences d'ADN

```
cctgatagacgctatctggctatccacgtacgtaggctcctctgtgcgaatctatgcgtttccaaccat  
agtactggtgtacatttgatacgtacgtacaccggcaacctgaaacaaacgctcagaaccagaagtgc  
aaacgtacgtgcaccctctttcttcgtggctctggccaacgagggctgatgtataagacgaaaatttt  
agcctccgatgtaagtcatacgtgtaactattacctgccaccctattacatcttacgtacgtataca  
ctgttatacaacgcgctcatggcgggggatgcgttttggtcgtcgtacgctcgatcgttaacgtacgtc
```

Motifs implantés

- Implanter un motif aléatoirement dans chaque séquences



A retenir

- Nous voulons chercher de motif lié à de Facteur de transcription
- Ces motifs se retrouvent dans les régions régulatrices, ils peuvent différer légèrement d'un gène à l'autre.
- Pour tester nos algorithmes nous allons générer des séquences d'ADN aléatoires et implanté un motif.