

# IFT 3290 hiver 2003

Bioinformatique

Miklós Csűrös

André-Aisenstadt 3253  
csuros@iro.umontreal.ca

<http://www.iro.umontreal.ca/~csuros/IFT3290/>

Préalables :

IFT2010 — Programmation 2

IFT1978 — Probabilités et statistique

Horaire :

cours	mercredi 11 :30–13 :30	Z 205
cours	vendredi 14 :30–15 :30	AA 1360
démo	vendredi 15 :30-17 :30	AA 1360

Démonstrateur ( ? ) : Retiba Terbaoui.

Évaluation : 30% TP, 20% intra, 50% final.

Seuil : au moins 40% au meilleur des deux exams.

# Le but du cours

# La vie

Que'est-ce que la vie ?

(Google "what is life" donne 64500 pages... )

- reproduction
- métabolisme
- évolution

# Diversité et universalité

Diversité de la vie

Universalité au niveau moléculaire : interaction de macromolécules

Protéines et acides nucléiques

# Protéines

Fonctions (exemple du levure)

Métabolisme	650 (10%)
Production et stockage d'énergie	175 (3%)
Transport	250 (4%)
Maintenance d'ADN	175 (3%)
Transcription	400 (6%)
Translation	350 (6%)
Secretion intra-cellulaire	430 (6%)
Structure cellulaire	250 (4%)
???	3520 (57%)

# Example

facteur de croissance EGF : voie de signalisation (www)

recherche dans une base de donnée : SWISS-PROT (www)

(Tapez «EGF receptor» au site Web de SWISS-PROT et suivez le lien vers EGFR\_HUMAN.)

pleine d'information liée : e.g., structure 3D (www)

(Suivez le lien «PDB» dans la ligne «HSSP».)

## Séquence

```
MRPSGTAGAA LLALLAALCP ASRALEEKKV CQGTSNKLTQ  
LGT FEDHFLS LQRMFNNCEV VLG NLEITYV QRNYDLSFLK
```

...

# Séquences protéiques

Protéine : polymères d'acides aminés (polymère est une grosse molécule formée de l'union de molécules plus petites)

Alanine (Ala) Glycine (Gly) Méthionine (Met)

Sérine (Ser) Cystéine (Cys) Histidine (His)

Asparagine (Asn) Thréonine (Thr) Acide aspartique (Asp)

Isoleucine (Ilu) Proline (Pro) Valine (Val)

Acide glutamique (Glu) Lysine (Lys) Glutamine (Gln)

Tryptophane (Trp) Phénylalanine (Phé) Leucine (Leu)

Arginine (Arg) Tyrosine (Tyr)

La séquence détermine la structure

⇒ **Problème de prédiction de structure protéique.**



# Autres problèmes

Comment la structure est établie ?

⇒ Repliement de protéines

La structure détermine la fonction

⇒ Problème de prédiction de fonction protéique.

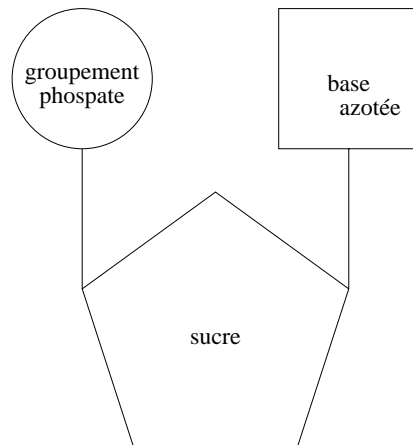
Interaction de protéines

⇒ Analyses de voies métaboliques, chemins de signalisation, réseaux régulatoires

# D'où viennent les protéines ?

Encodage en acides nucléiques : ADN et ARN.

ADN : polymère de nucléotides

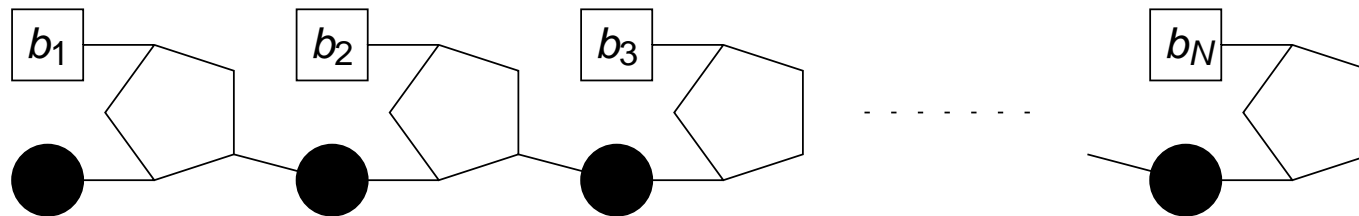


nucléotide :

# ADN 2

4 bases : adénine, thymine, guanine, et cytosine

chaîne de ADN :



$b_i = A, C, G, T.$

# ADN 3

Liaisons entre des bases complémentaires  
A-T et G-C

La molécule d'ADN est formée par deux chaînes (brins).

# Code génétique

20 acides aminés encodés par 4 nucléotides ?

Encodages par triplets : 64 codons

AAA = Phe	AAG = Phe	AAT = Leu	AAC = Leu
AGA = Ser	AGG = Ser	AGT = Ser	AGC = Ser
ATA = Tyr	ATG = Tyr	ATT = FIN	ATC = FIN
	...		

3 triplets d'arrêt, 1 triplet de début (encode aussi Met)

# De ADN à protéine

1. transcription : copie du brin informatif à ARN messenger  
(ARN : utilise Uracile au lieu de Thymine)

2. traduction : ARNm à protéine (par le ribosome) : acides aminés fournis par ARN de transfert

Mécanisme universelle !

⇒ Problème de déterminer la séquence de l'ADN.

# Gènes

Gène : unité d'hérédité, se traduit à une protéine (en général)

Exons et introns

⇒ **Problème de prédiction de gènes.**

Expression d'un gène : quantité de ARNm transcrit ⇒ **Problème d'analyse de l'expression génétique.**

# Comparaisons

Comparaisons de séquences : pour déterminer structure, fonction, ...

⇒ **Problème de comparaison de séquences.**

Grandes bases de données

⇒ **Problème de recherche de séquences.**



# Évolution

# Calcul moléculaire