1

La vie est une cascade d'interactions !

2

Structure 3D des protéines

Intérêts de la modélisation

agap

Dr Frédéric de Lamotte
UMR AGAP
avec Dr Cécile Fleury

3

Génotype

ADN	Génétique
ARN	Génomique
Protéine	Biochimie
Cellule	Bio. Cell.
Tissu	Physiologie
Plante	Agronomie

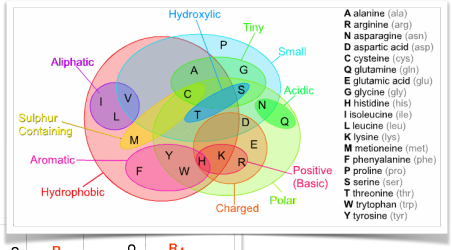
Phénotype

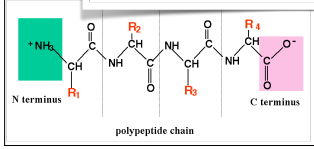
Les protéines :
actrices de la vie cellulaire

- Structure
- Transport
- Signalisation
- Réplication
- Transcription
- Métabolisme
- Défense
- Kératine
- Canal ionique
- Récepteur
- ADN polymérase
- ARN polymérase
- Enzymes
- Toxine

4

Les acides aminés...

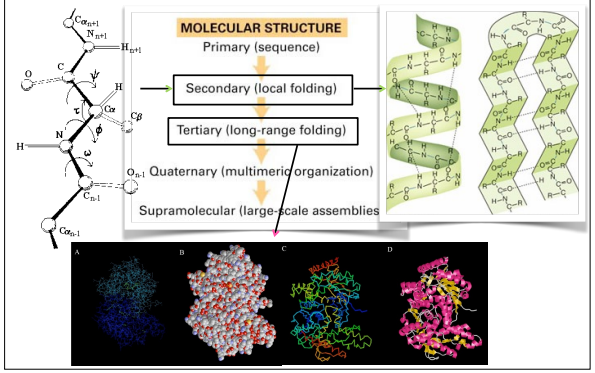




... constituent la séquence protéique (structure primaire)

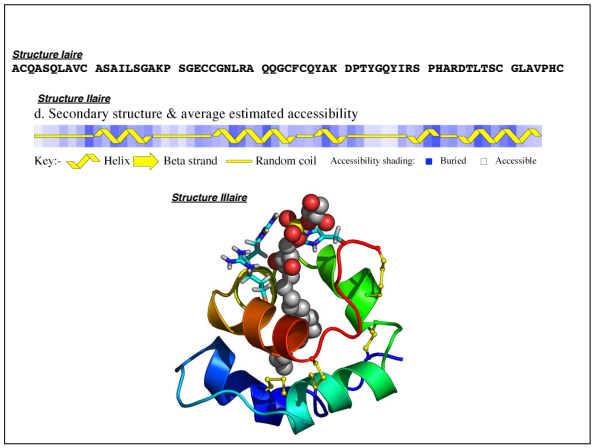
5

La structure des protéines

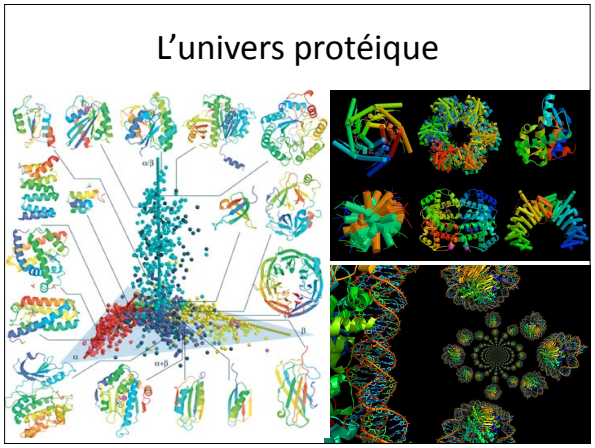


6

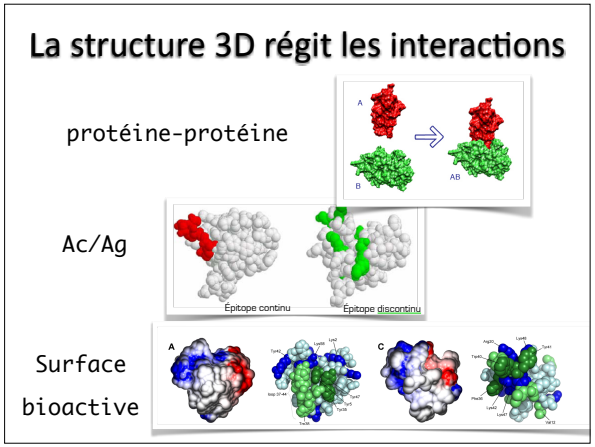
7


Structure laire
ACQASQLAVC ASAILSGAKP SGECCGNLRA QQGCFCQYAK DPTYGQYIRS PHARDTLTSC GLAVPHC
Structure llaire
d. Secondary structure & average estimated accessibility
Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible
Structure llaire

8


L'univers protéique

9

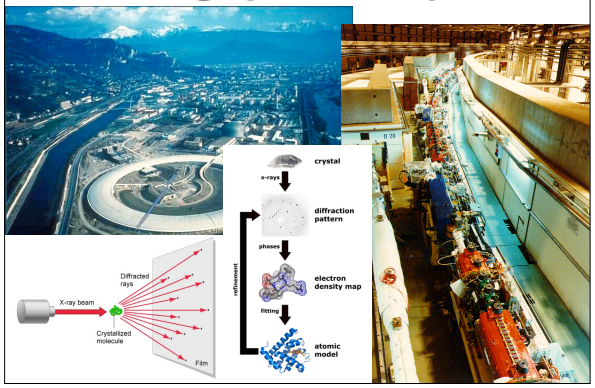

La structure 3D régit les interactions
protéine-protéine
Ac/Ag
Surface bioactive
Épitope continu
Épitope discontinu

Détermination de la structure d'une protéine

- Méthodes expérimentales
- Alternative bioinformatique

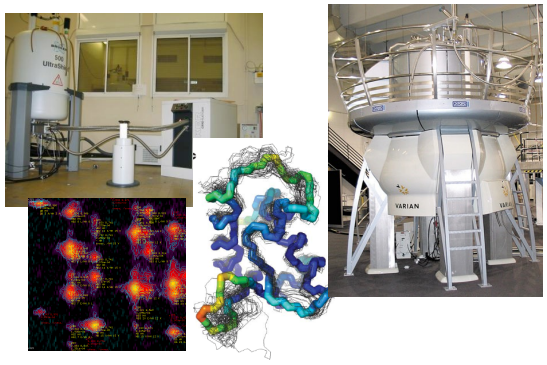
10

Cristallographie aux rayons X



11

Résonance Magnétique Nucléaire



12

Techniques expérimentales - limites -

13

- Cristallographie
  - Quelques **mg** d'une protéine **très pure**
  - Cristallisation, diffraction, phasage
  - 1 synchrotron = 1G€
  - 1 structure nette = 20 000 à 40 000€
- RMN
  - Quelques **100 µg** d'une protéine **pure**
  - 1 spectromètre = 500 000 à 2 000 000€
  - 1 structure dynamique = 100 000€

Le paradoxe de Levinthal

14

If a chain of a 100 amino acids is considered and it assumed each amino acid can exist in one of 3 conformations, extended, helical or loop, then there are **3^{100} possible ways to arrange this chain**.

This is roughly **10^{48} conformations**. Bond rotation can be estimated to occur at a rate of roughly 10^{14} s-1. This means that search for the right conformation through random searching alone would take the order of 10^{34} s or **10^{26} years**, several orders of magnitudes greater than the age of the universe!

(C. Levinthal, Chim. Phys., 65 (1968), 44-45)

Les protéines ne se replient pas au hasard
Il existe une Loi du repliement

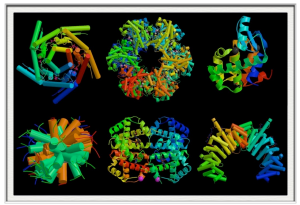
L'alternative bioinformatique

15

- Avantages
  - Pas besoin de matériel biologique
  - Coût réduit (temps, ressources)
  - Compatibles analyses large échelle
- Inconvénients
  - Précision dépend des *templates*
  - Validation nécessaire

Les méthodes prédictives

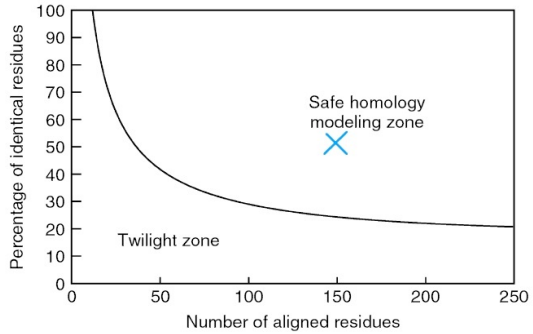
- Modélisation moléculaire comparative :
- Séquence + structure connue
- homologie
- *threading*



- Prédiction *ab initio* :
  - Séquence seule (+ statistiques)

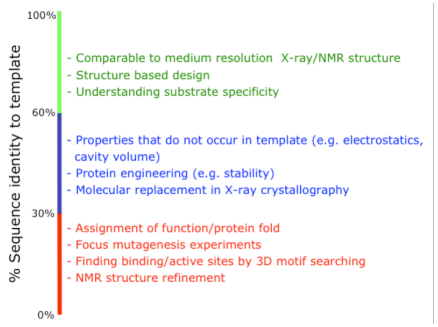
16

Modélisation « *safe* »



17

Applications



18

19

Modélisation comparative :
méthodologie

- Recherche d'homologues avec structure 3D connue
- Recherche de structures à la topologie similaire
- Sélection de *template(s)*
- Alignement cible/*template*
- Calcul du modèle théorique
- Evaluation et visualisation

20

Les bases de données du TD

- PDB (*Protein Data Bank*) : données expérimentales sur les structures 3D connues => coordonnées atomiques
- Pfam (*Protein Family*) : familles de protéines ayant des domaines structuraux en commun
- SCOP : structural classification of proteins
- Uniprot (uniprotkb) : protein sequence and functional information
- ProteinModelPortal : recherche de modèle de protéines connus
- PSIPRED : protein structure prediction server (structure 2aire)