

Option *BIMB* :  
**Bio-Informatique et Modélisation**  
pour la **Biologie**



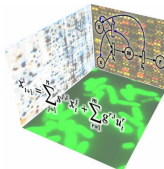
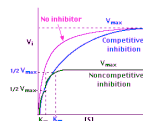
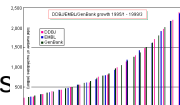
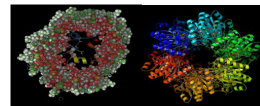
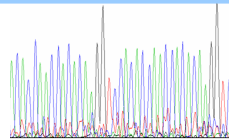
Resp. Jean-Paul Comet

Jean-Paul.Comet@univ-cotedazur.fr



## Quatre grands domaines

- Analyse de la séquences :
  - Séquence nucléique – séquence protéique
  - Structure 2D, 3D
- Modélisation moléculaire
- Données
  - Séquences, transcriptomique, protéomique,
  - Pyroséquençage...
  - Apprentissage : régulations, paramètres cinétiques
- Modélisation des réseaux : métaboliques, génétiques, signalisation
  - Simulations



## Vous avez dit « Bio-informatique » ?

- « Domaine lié à l'application des méthodes et outils de l'informatique à la biologie et la médecine »
- **BUT:** résoudre un problème posé par la *biologie*
  - séquençage, l'analyse du génome, des séquences
  - modélisation moléculaire,
  - reconstruction d'arbres phylogénétiques (phylogénie).
  - analyse de l'expression de gènes
  - analyse d'images,
  - ...
  - modélisation de l'évolution d'une pop. dans un environn donné,

champ d'activité **multi-disciplinaire** :  
**biologie, informatique, mathématique, physique/chimie**

⇒ C'est la « **biologie in silico** »,

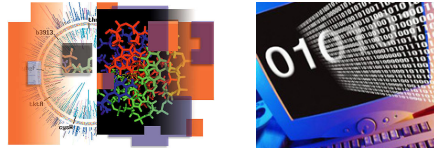
## Domaines pertinents de l'informatique pour la bio-informatique

- **Algorithmique**  
analyse de séquences, structure 2D des ARN, Docking...
- **Systèmes d'information**  
stockage et exploitation des données omics, traçabilité, ...  
statistiques, analyse de données...
- **Intelligence artificielle**  
fouille de données, diagnostic, classification, ...
- **Modélisation et simulation**  
modélisation moléculaire, biologie des systèmes, cellules virtuelle...

## Objectifs de la formation

- Former des *biologistes* qui dominent les deux langages : biologie / informatique.
  - interactions avec les métiers de l'informatique
  - animer une équipe sur un projet pluridisciplinaire.

- Liste Bioinfo / SFBI



## Pourquoi cette formation ?

L'informatique est partout, même en biologie - 2

Débouchés classés par domaines (European Research council):

- **Population** (évolution, phylogénie, écologie, environnement, santé..): organisation des données, extraction de connaissances, classification, reconstruction, phénomènes émergents, normes de sécurité, biodiversité...
- **Médecine – santé** (pharmacologie, diagnostic, info. Méd.): modèles prédictifs, systèmes experts, aide au diagnostic, criblage de molécules, design de nouveaux médicaments, stockage de données, extraction de connaissance, systèmes d'informations médicales, confidentialité...
- **Biotechnologiques** : biologie structurale, biologie à grande échelle, bio-statistique et bio-informatique, biologie des systèmes et biologie synthétique, etc.

## Pourquoi cette formation ?

L'informatique est partout, même en biologie - 1

Débouchés classés par domaines (European Research council) :

- **Biologie mol. et cell.** : assemblage & annotation de séq, repliement, modèles d'interactions, ...
- **Génomique fonctionnelle** : qualité des données, stockage, prédiction de la fonction, optimisation des expériences...
- **Organes et Physiologie** (métabolisme et pathologie, infection et immunité, vieillissement) : cadre méthodologique d'études, prédictions, simulation, limite de l'expérimentation animale...
- **Neuroscience** : modèles mathématiques explicatifs, prédictions, simulations, analyse des mesures, reconstruction spaciae...

## Secteurs d'activité

**Biotechnologies (R,V,B)**: optimisation des bioréacteurs, puces à ADN, pyroséquenceurs, stockage et analyse de données...

**Pharmacologie** : drug design, cosmétique, toxicologie, sécurité...

**Agro-alimentaire** : OGM, sécurité, toxicologie...

**Environnement, écologie** : dévelopt durable, toxicologie, sécurité...

**Biologie de synthèse** : biocarburants, diagnostics, ...

**Recherche en biologie**, bio-informatique : publique / privée

**ESN / SSII** (Société de Service en Ingénierie Informatique)

**Sociétés de Consulting** : qualité des données / lims...

## Pourquoi la bio-info est-elle une discipline amenée à se développer ?

- **Interface** de 2 disciplines : biologie et informatique.
- chaque laboratoire peut générer énormément de données.
- La taille des BD génomiques **augmente** exponentiellement *stockage, gestion, analyse, visualisation, interprétation...*
- Besoin de classer / annoter / confronter analyse de séquences, structure des protéines, phylogénie, génomique fonctionnelle, transcriptomique...
- **Extraction** de connaissances + **prédictions**

**Leem** : « Pour répondre aux nouveaux enjeux en bio-informatique, les entreprises du médicament requièrent de nouveaux profils, les bio-informaticiens à l'interface entre le monde de la biologie et celui de l'informatique. »

**Le savant doit ordonner; on fait la science avec des faits comme une maison avec des pierres; mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison.**

Henri Poincaré, 1908

## Stages GB5 : sociétés

### Modélisation moléculaire (3DS)

Modélisation du passage de la barrière cutanée par un modèle QSPR (Galderma)  
Modélisation moléculaire de nano-particule Drug design basé sur les fragments (Sanofi)  
Prédiction de propriétés pharmacocinétique (Sanofi)

### Toxicologie prédictive (Sobios), toxicogenomics (3DS)

Prédiction de toxicité (Bayer crop Science) / Modélisation mécanistique (Bayer crop Science)  
approche QSAR dans le cadre d'une étude de la génotoxicité de petites molécules (Galderma)  
Conception et réalisation d'un modèle de l'évolution de l'infection par le virus de l'hépatite B (Cosmo)

**Data**: Mise en place des outils de data management et statistiques sur des tests d'évaluation (Episkin)  
Développement d'outils pour l'interprétation de résultats (Genomic Vision)  
Dév. base de connaissances - puce à ADN (BioMérieux)  
Fédération de bases de données biol. (Dassault systèmes)  
Gestion des expériences à l'échelle d'un institut (Bioaster)  
Caractérisation génétique de souches virales (Transgene)  
Agrégation et modélisation de méta-connaissances (3DS)

### Extraction de connaissances :

Mise en évidence de données d'inhibition de fragments (Sanofi)  
Data mining pour la recherche de nouveaux peptides thérapeutiques (Nuritas)

**Autres** : Évaluation des stocks globaux de carbone par scénarios sylvicoles (FCBA Institut technologique)  
Alignement de protéines contraints par la structure (Medit)  
Pharmacological validation of an automated phenotyping system (Roche)  
Identification de nouvelles cibles thérapeutiques pour le cancer (Servier)  
Propriété intellectuelle (Galderma) Planification fonctionnelle (Galderma)  
Optimisation de flux de données (L'oreal)  
Standardisation des résultats de laboratoires (Clinical Research Organization)  
Reconstruction de modèles métaboliques (Bioaster)  
Optimisation d'un pipeline d'analyse (BioMérieux)

## Stages GB5 : centres de recherche

### Modélisation :

Température de fusion des Acides Nucl. (EMBL-EBI)  
Signatures transcriptomiques pour le diagnostic des maladies rares musculaires (IMMG)  
Modélisation du cycle cellulaire (lab CNRS)  
Modélisation qualitative de la régulation métabolique d'une cellule cancéreuse pancréatique (lab CNRS)  
Analyse et intégration de données multi-omiques pour la classification et la comparaison de tumeurs et de lignées cellulaires cancéreuses

### Autres :

Développement d'un outils de visualisation en métabolomique (EMBL-EBI)  
Représentation graphiques de modèles pour la biol. des systèmes (EMBL-EBI)  
Modélisation de la régulation transcriptionnelle par du DNA supercoiling chez les bactéries (CNRS)

### Analyses :

Pipeline de reconstruction de réseau métabolique pour des organismes non-classiques (INRIA)  
Étude de génomique fonctionnelle sur des génomes de rongeurs (EMBL-EBI)  
Intégration de pipelines d'analyse (INSERM)  
Analyse de données transcriptomique (lab CNRS)  
Recherche de modulateurs de l'activité des canaux sodiques (lab CNRS)  
Première annotation du transcriptome de l'esturgeon européen (lab CNRS)  
Analyse bioinfo RNA-Seq (lab CNRS)  
Utilisation de données RNA-seq (INRA)  
Méthode d'analyse pour l'optimisation fonctionnelle de réseaux métaboliques (lab CNRS)

## Et après ?

### Postes d'ingénieurs :

EMBL – EBI, Glaxo-Smith-Kline, Bioaster (partenariat public/privé), Evotec, CNRS, Soladis, 3DS, ...

- Molecular Modeler, Research Informatics
- Ingénieur bio-informaticien CNRS
- Data Manager - Ingénieur Biostatisticienne, Soladis chez L 'Oréal
- Computational biologist (Nuritas LTD)
- Intitek (SSII, en mission chez BioMérieux), INRA / CHU
- Ingénieur d'études en Bioinformatique, Faculté de médecine de Nancy
- Ingénieur Bio-informaticienne (support) chez Interactive BioSoftware
- Consultant en Solutions - Spécialisé Life Science chez TEMIS
- Data Scientist, IA Scientist, ...

### Recherche :

- Laboratoires de recherche (Barcelone, Montpellier, )
- CEA Saclay (modélisation moléculaire)
- Thèses (EMBL, Irlande, Allemagne, Univ. Nice, Montpellier, Cifre, INRIA , Lyon)

### Consulting :

- Néo 2 consulting (Iims), Excelya consulting / IPSEN R&D (validation de données)

### Poursuite d'études :

- CEIPI puis Patent attorney, Roquette Frères
- Masters de management, de bio-statistiques

# Contenu de la formation

**Fondements** de la bio-informatique classique

**Les aspects les plus récents** de la modélisation assistée par ordinateur sont abordés :

- modélisation moléculaire, modélisation des systèmes biol., extraction de connaissances...

**4 domaines de compétence :**

- Algorithmique et programmation
- Systèmes d'information
- Intelligence artificielle (statistique & explicable):  
extraction de connaissances, diagnostic, pharmacovigilance...
- Modélisation et simulation :  
biologie des systèmes, réseaux biol., cellules virtuelle

# Les enseignements spécifiques (1)

**Semestre 8 :**

|   |     |
|---|-----|
| Algorithmique pour la biologie  | 27h |
| Administration systèmes et réseaux  | 18h |
| Techniques de simulation de syst. Biologiques (plusieurs paradigmes de la modélisation en biologie) | 25h |
| Modélisation des réseaux biol. complexes  | 27h |
| Programmation Objet et Java   | 27h |
| Toxicologie cellulaire et moléculaire   | 27h |

# Les enseignements spécifiques (2)

**Semestre 9**

|  |              |
|--|--------------|
| Modélisation moléculaire & drug design         | 40h          |
| Génie logiciel et UML                          | 30h + projet |
| Bases de données avancées et interfaces        | 30h + projet |
| Introduction à l'IA pour la biologie           | 24h          |
| IA symbolique pour les réseaux biol. Complexes | 27h          |
| Biologie intégrative                           | Projet       |
| Omiques  | 12h          |

# Stages GB4 : Sujets

**Biologie moléculaire :**

*Assemblage génomique, annotation, analyse Analyse bioinfo de la structure peptidique, découverte de peptides membranaires Biotechnologies moléculaires*

**Santé :**

*Drug resistance in ovarian cancer and lymphoma Obesity, Diabete, Breast Cancer and Ageing Role of SORL1 in Alzheimer's disease Neuroscience, Aging, and Stem Cell Research Facteurs de transcriptions et vieillissement*

**Développements d'outils :**

*Interface intranet d'une BD d'expériences Développement d'outils web pour la bioinfo Création d'une BD génomique (cancer)*

**Analyse de données :**

*Méta analyse des données du transcriptome Classification des protéines hub et non hub Prédiction des sites de liaison des microARN, Interactions et désordre moléculaire Prédiction de localisation des protéines Analyse de données pré-clinique*

**Modélisation :**

*Reconstruction de réseaux de régulation Systems biology, dynamic modelling, cancer Modélisation d'un neurone Analyse de la dynamique de chromosomes Modélisation personnalisée de la réponse au dommage ADN (p53)*

**Compétences acquises :**

*python, perl, Java Modélisation / simulation, Représentation graphique*

*Tests statistiques, Bases de données, ...*