

Les notes de cours et les feuilles de TD sont disponibles (avec un peu de retard par rapport au déroulement du cours) à l'adresse web suivante :

http://www.i3s.unice.fr/~bernot/Enseignement/GB3_Python1/

1 Domaines de la bio-informatique

L'objectif de ce cours de *Bonnes pratiques de la programmation* est d'apprendre ces bonnes pratiques au travers de la partie dite « impérative » du langage Python. Ce qui motive cet enseignement dans une formation de génie biologique est qu'à l'heure actuelle aucun cadre scientifique, quelle que soit sa discipline de base, ne peut prétendre être un *ingénieur* s'il ne connaît pas les principes de la programmation. À l'issue de ce cours, vous saurez programmer en Python. Python est le langage de programmation le plus utilisé par les biologistes. Il faut de plus être conscient que les principes de programmation sont les mêmes pour la plupart des langages de programmation. Ainsi, à la sortie de ce cours, vous serez en fait capables de programmer avec presque n'importe quel langage de programmation, pourvu que vous ayez sous la main un petit résumé des mots-clés de ce langage.

Ce cours de programmation n'est pas un cours de bio-informatique. La bio-informatique est une activité où la biologie et l'expérimentation sont à la base de la *prédiction* de propriétés biologiques. Si la programmation entre en jeu en bio-informatique, c'est uniquement de manière minoritaire parce que les données et des connaissances manipulées sont d'un volume tel que leur gestion requiert un ordinateur. Ce qui fait un bon ingénieur en bio-informatique n'est pas la programmation ; c'est un *mode de raisonnement* avec beaucoup de recul et adapté au vivant pour produire de bonnes prédictions. Cela suppose une bonne compréhension des protocoles expérimentaux aussi bien *in vivo* que *in vitro* et *in silico*.

Pour éviter la confusion fréquente entre la programmation et la bio-informatique, il semble utile d'inventorier comment la bio-informatique intervient dans tous les domaines de la biologie.

1.1 Domaines de la biologie (panels ERC)

(ERC = European Research Council)

- *Biologie moléculaire et cellulaire* (biologie structurale, signalisation, métabolisme, cycles divers, apoptose...) : la bio-informatique permet d'assembler et annoter les séquences biologiques, elle prédit la structure de repliement des bio-molécules à partir de connaissances sur les séquences, sur l'énergie de repliements élémentaires, *etc.* Elle établit des modèles d'interactions et d'échanges de signaux entre entités biologiques, propose des comportements dynamiques possibles, inventorie les comportements stables, *etc.*
- *Génomique fonctionnelle* (omics, génétique, épigénétique, réseaux...) : la bio-informatique classe les mesures acquises, évalue leur qualité, en déduit des profils d'expression, quantifie la pertinence de certaines interactions putatives, prédit la ou les fonction(s) d'un gène ou d'un réseau génétique, optimise les stratégies expérimentales en fonction des objectifs, *etc.*
- *Organes et physiologie* (systèmes, métabolisme et pathologies, infection et immunité, vieillissement...) : la bio-informatique fournit des cadres méthodologiques pour étudier les systèmes biologiques complexes, établit des lois comportementales prédictives, offre des simulations crédibles, limite les expérimentations animales, assiste l'analyse d'images, *etc.*
- *Neurosciences* (neuro..., imagerie, cerveau,...) : la bio-informatique met en place des modèles mathématiques explicatifs, prédit certains comportements, gère des simulations *in silico*, réfute des hypothèses sur la base des comportements émergents, analyse les mesures, reconstruit la structure spatiale, *etc.*
- *Populations* (évolution, phylogénie, écologie, écotoxicologie, environnement, santé...) : la bio-informatique organise les données, en extrait des connaissances, classe, établit des graphes de phylogénie, simule les individus pour prédire des phénomènes collectifs émergents, établit des normes de sécurité, prédit les effets d'une perturbation, permet de tester des modifications planifiées de l'environnement (e.g. autoroute traversant une forêt), *etc.*
- *Médecine santé* (pharmacologie, diagnostic, informatique médicale) : la bio-informatique met en place des systèmes experts, établit des modèles prédictifs de comportement tant au niveau cellulaire que tissulaire ou organique, permet le criblage de molécules, organise et stocke les informations médicales, les exploite pour extraire des connaissances, *etc.*

1.2 Domaines pertinents de l'informatique dans le cadre de la biologie

La bio-informatique exploite presque tous les domaines de l'informatique.

- *Algorithmique* (comment gérer une question reposant sur beaucoup de données en entrée pour la résoudre en un temps acceptable) : par exemple pour le traitement et la comparaison des séquences génomiques, la classification des données, le calcul de repliement de biomolécules, l'étude systématique de grand nombre de modèles comportementaux, *etc*
- *Systèmes d'information* (gestion des bases de données et des interfaces pour les exploiter) : gestion de toutes les données omics, représentation des grands réseaux biologiques, mémorisation des protocoles expérimentaux, suivis divers, patients, gestion de bibliographie, tracabilité, *etc*.
- *Intelligence artificielle* (méthodes heuristiques pour résoudre la plupart du temps des questions qui seraient hors de portée d'une machine si elles étaient traitées exhaustivement et de manière exacte) : extraction de connaissances à partir de données biologiques, classification fonctionnelle, diagnostic à partir de données à grande échelle, pharmacovigilance, *etc*.
- *Modélisation et simulation* (créer des modèles mathématiques ou virtuels d'objets réels, raisonner automatiquement dessus ou les simuler) : biologie des systèmes, biologie intégrative, réseaux biologiques, embryologie 3D, reconstruction d'organes, cellules virtuelles, prédictions en chronobiologie, chronopharmacologie, *etc*.

1.3 Domaines de l'ingénierie biologique (panels ERC)

Les domaines d'activité des ingénieurs bio-informaticiens couvrent tous les domaines du génie biologique car tous ont une composante bio-informatique. Pour reprendre la classification ERC, on peut citer :

- *Biotechnologies* : tous les domaines (biotechnologie rouge, verte, blanche ou bleue) doivent faire appel à la bio-informatique pour organiser les données, prédire l'effet des modifications planifiées du génome, optimiser les bioréacteurs, la bio-informatique fait l'objet de nombreuses « niches » des startups, gère de bout en bout les puces à ADN et les pyroséquenceurs, *etc*.
- *Pharmacologie* : drug design, cosmétique, toxicologie, sécurité, *etc*.
- *Environnement, écologie* : développement durable, toxicologie, sécurité, *etc*.
- *Agro-alimentaire* : OGM, sécurité, toxicologie, *etc*.
- *Biologie de synthèse* : biocarburants, diagnostiques, antibiotiques, *etc*.
- *Recherche en biologie* : publique ou privée.

... Et il faut aussi mentionner de nombreuses sollicitations en *informatique classique*, en raison de la forte pénurie d'informaticiens diplômés, en France et dans le monde, qui rend les DRH du secteur informatique avides de jeunes scientifiques ayant touché de près ou de loin à l'informatique.

2 Métiers de la bio-informatique

Tout ce que ni un informaticien non biologiste, ni un biologiste non bio-informaticien, ne saura jamais faire :

Maître d'ouvrage de projets multi-disciplinaires, chef de projet : de nombreux projets de biologie moderne font appel à toutes les autres disciplines scientifiques (physique, chimie, robotique, informatique, mathématique) voire aux sciences humaines (droit, éthique). Pour mener à bien de tels projets, les entreprises ont besoin de maîtres d'ouvrage rompus à la communication entre biologie et autres sciences. La bio-informatique est le lieu où sont formés ces ingénieurs, habitués à la cette communication pluri-disciplinaire.

Responsable de service bio-informatique : une part notable de ce métier est de participer à des réunions de direction où les besoins informatiques ne sont abordés que sous l'angle des utilisateurs biologistes ou commerciaux, de faire comprendre ce que l'informatique peut apporter dans ces problématiques, de traduire les besoins de l'entreprise pharmacologique, de dégager ce qu'il est possible ou non de faire, d'assurer la qualité des produits logiciels. Le responsable de service bio-informatique est capable de diriger des informaticiens « pur jus » et d'assurer l'utilité pour la biologie des projets développés.

Architecte et administrateur de systèmes d'information : partant du savoir-faire et des données spécifiques d'une entreprise en pharmacologie, seul un ingénieur pluri-disciplinaire peut dégager les informations pertinentes, les structurer, les regrouper pour faciliter l'extraction de connaissances et finalement définir l'architecture adéquate du système d'information scientifique de l'entreprise.

Responsable de conception et réalisation web, multi-média : seul un bio-informaticien peut séparer judicieusement l'intranet de l'extranet de telle sorte que les informations passées à l'extérieur prouve la compétence de l'entreprise mais ne permettent pas à une autre entreprise de s'approprier cette compétence. Il s'agit de concevoir un site web attractif tout en assurant la confidentialité.

Chef de projet de calcul intensif : (fermes de calcul, simulations, extraction de connaissances, ...); demandez à un ingénieur non biologiste d'extraire des connaissances ou des prédictions à partir de données biologiques, et après un usage intensif de puissances de calcul énormes, les résultats que vous obtiendrez seront, au mieux, des évidences pour un biologiste, émaillées de bon nombre d'inepties. Seul un bio-informaticien peut identifier les approximations pertinentes d'un point de vue biologique qui rendront les calculs incommensurablement plus efficaces et les résultats valorisables.

Responsable d'un service de biostatistiques : les statistiques reposent d'abord sur le choix des indicateurs mesurés, la compréhension critique des protocoles expérimentaux est donc un préalable à toute démarche biostatistique. De plus, il est bien connu que sans un œil critique lors de l'interprétation des résultats, on peut faire dire ce que l'on veut aux statistiques. La connaissance à la fois du contexte biologique et des limites théoriques sont donc indispensables au professionnel.

Architecte et administrateur systèmes et réseaux : dans des entreprises ou laboratoires de petite taille, le volume du parc informatique ne justifie pas, parfois, d'employer un informaticien à plein temps pour gérer les systèmes et réseaux. Cette tâche incombe alors aux bio-informaticiens de l'entreprise. Il s'agit d'une part d'assurer la sécurité des données, leur confidentialité, de protéger le réseau d'entreprise contre les attaques malveillantes, et d'autre part de jouer un rôle de conseil. . . Cette activité est beaucoup moins complexe qu'il n'y paraît (un cours de 15h permet d'en acquérir les principes de base).

Responsable de la robotisation : (protocole, bases de données, images...); des activités comme le screening font un large usage de robots. Leur pilotage inclut la gestion de la bibliothèque des produits utilisés par le robot, les protocoles définissant les actions successives du robot, l'analyse des résultats en série, *etc.* Cette robotisation nécessite une parfaite coordination avec les expérimentateurs et là encore seuls les bio-informaticiens de formation peuvent mettre en place ces protocoles expérimentaux d'une manière cohérente avec les objectifs biologiques.

Consultant pluridisciplinaire : après quelques années d'expérience dans l'un (ou plusieurs) des métiers qui précèdent, les ingénieurs qui souhaitent acquérir une plus grande indépendance de travail peuvent devenir consultants. La pénurie de ressources humaines dans les métiers proches de l'informatique laisse nécessairement une place à cette activité de consultance. Par ailleurs certaines PMI ou PME peuvent aussi faire appel à des consultants pour les aider à définir et/ou externaliser leurs besoins en bio-informatique.

Métiers de la recherche : La plupart des laboratoires en biologie, qu'ils soient publics ou privés, ont identifié la bio-informatique comme un point clef de leur développement. D'une part, les gros laboratoires se munissent tous d'un service de bio-informatique généralement dirigé par un ingénieur de recherche, et d'autre part, le besoin d'une équipe de recherche de bio-informatique ouvre d'assez nombreux postes de chercheurs à des jeunes diplômés possédant un doctorat de bio-informatique.