

## Conseil scientifique de l'institut des sciences biologiques

### RECOMMANDATIONS DU CS INSB POUR UNE STRATEGIE DE DEPLOIEMENT DE LA BIOINFORMATIQUE DANS LES LABORATOIRES DE RECHERCHE

#### Motivation

La biologie recourt désormais à l'informatique pour quasiment tous ses domaines de recherche. En particulier, la possibilité de générer de très grandes quantités de données permet d'aborder de nouvelles questions, ou de généraliser les questions précédentes à des échelles globales. Face à l'explosion quantitative des nouvelles technologies (« Next Generation Sequencing », imagerie), la biologie se retrouve devant un réel goulot en termes d'infrastructure, mais surtout de ressources humaines. Tous reconnaissent la nécessité de l'outil informatique pour stocker et manipuler les données, en extraire l'information pertinente, ainsi que le besoin de méthodes adaptées à la nature et à la masse des données générées. Mais l'évolution de la biologie entraîne de plus en plus l'émergence de questions qui ne sont abordables qu'au moyen d'approches formelles (informatique, mathématiques, physique). Ces approches produisent des modèles explicatifs du vivant qui ont une valeur prédictive, et peuvent soulever des hypothèses testables et réfutables. Souvent, ces méthodes permettent de raisonner sur les données et d'ouvrir le champ à de nouvelles questions qui n'auraient pu se poser précédemment. Plus encore, ces méthodes peuvent contribuer à produire des données biologiques qui compléteront les données expérimentales, et contribueront à construire une vision intégrative des processus d'intérêt. Ceci s'applique par exemple à l'analyse des processus évolutifs (évolution des génomes, du métabolisme, ...) ou des réseaux d'interaction (docking moléculaire, prédictions d'interactions). Dans ce sens, la bioinformatique peut être définie comme un domaine de la recherche en biologie, particulier par ses méthodes plutôt que par ses objets d'analyse.

Les recommandations qui suivent visent à améliorer l'organisation des ressources humaines pour une articulation optimale entre biologie et disciplines formelles qui contribuent désormais de façon cruciale à son développement, en promouvant les aspects suivants.

1. Le développement de la bioinformatique en tant que domaine interdisciplinaire de recherche, en reconnaissant la spécificité de ses questions, approches et défis.
2. Une collaboration effective entre biologistes et bioinformaticiens, mathématiciens, physiciens, afin de créer l'espace où pourra naître la culture multidisciplinaire nécessaire aux développements de la biologie contemporaine.
3. Sensibiliser l'ensemble des sections disciplinaires concernées (non seulement en biologie, mais également physique, mathématiques, informatique) à la nécessité de mettre en œuvre les moyens nécessaires au recrutement de bioinformaticiens, et de se donner les moyens d'évaluer leur dossier scientifique dans sa dimension multidisciplinaire. La CID a été créée à cet effet-là, mais les postes dont elle dispose actuellement ne suffisent pas à rencontrer les besoins des laboratoires en chercheurs bioinformaticiens.
4. Eclairer les politiques de recrutement des ITA (IA, IE, IR) au sein des laboratoires, afin d'optimiser leur intégration dans un processus de développement de la recherche en biologie et en bioinformatique. En particulier, concilier les besoins exprimés par les laboratoires de biologie pour obtenir des ressources humaines spécialisées en analyse de données et la nécessité pour ces bioinformaticiens de maintenir un contact étroit avec leurs domaines disciplinaires (informatique, mathématiques, statistiques, physique) et avec la communauté des bioinformaticiens.

Ces recommandations sont essentiellement orientées vers le domaine de la bioinformatique au sens large (analyse des données génomiques, transcriptomiques, des réseaux, bioinformatique structurale, modélisation mathématique et physique des systèmes biologiques, y compris les modèles dynamiques, inférence phylogénétique, bases de données biologiques, ...). Actuellement, l'analyse d'images biologiques ne relève

## Conseil scientifique de l'institut des sciences biologiques

pas de la définition générale de la bioinformatique, et les communautés d'informaticiens impliquées sont malheureusement encore peu connectées (méthodes, congrès, projets, ...). Il est cependant intéressant de prendre en compte les développements actuels de ce domaine, car il présente avec la bioinformatique un certain nombre de similarités de par la quantité massive des données et la nécessité d'interaction entre biologistes et informaticiens. Des réponses communes pourraient donc être trouvées pour résoudre certaines des difficultés partagées, essentiellement en matière d'infrastructure (stockage, transfert et traitement de données massives) et de mise en place de communautés interdisciplinaires.

### Les pratiques de la bioinformatique

Du fait du caractère intrinsèquement interdisciplinaire du domaine, les bioinformaticiens ont depuis toujours été recrutés à partir de diverses disciplines scientifiques : biochimistes, biologistes, informaticiens, mathématiciens, statisticiens, physiciens, ... Le développement du domaine présente une similarité évidente avec celui de la biologie moléculaire, durant les années 1960-1980.

Le nom « bioinformatique » donne parfois l'impression que ce domaine consiste à juxtaposer l'informatique et la biologie. Certes, ces deux disciplines sont essentielles, mais insuffisantes à former des bioinformaticiens. La discipline comporte également les composantes suivantes : (i) statistiques (y compris méthodes spécialisées pour l'analyse des données à haut débit), (ii) mathématiques (analyse des systèmes dynamiques, analyse des réseaux), (iii) physique biologique (biophysique des macromolécules), (iv) informatique (algorithmique, intelligence artificielle)

Au-delà des complémentarités disciplinaires, il existe plusieurs façons de « faire » de la bioinformatique. On peut en première approche distinguer les catégories suivantes

1. **Bioinformaticien-développeur.** Personne dont l'activité principale consiste à concevoir des approches originales pour répondre à des questions biologiques. Ces approches serviront ensuite de fondement conceptuel pour implémenter des outils logiciels accessibles à la communauté. Ces activités sont généralement déployées dans des équipes spécialisées de bioinformatique, que ce soit au sein de laboratoires de biologie, d'informatique ou de mathématiques. On trouve aussi des bioinformaticiens-développeurs dans des structures intégratives combinant des équipes de biologistes expérimentaux, de théoriciens et de bioinformaticiens.
2. **Bioinformaticien-analyste.** Personne dont l'activité principale consiste à utiliser de façon éclairée des méthodes existantes afin d'analyser des données biologiques. Par « utilisation éclairée », nous entendons la capacité à évaluer, comparer, choisir les paramètres appropriés à chaque étude, identifier les cas où les méthodes ne sont pas encore satisfaisantes, et proposer des pistes pour les développements futurs. On trouve des bioinformaticiens-analystes dans différentes structures : sur les plateformes de séquençage et de bioinformatique, dans des laboratoires de biologie ou de bioinformatique. Il faut noter que dans certains laboratoires orientés vers des activités de génomique ou métagénomique, l'analyse des données représente désormais un temps plus important que la genèse des données.
3. **Bioinformaticien-curateur.** Développement et maintien des bases de données (création des structures et des interfaces, annotation).

La plupart des bioinformaticiens couvrent au moins partiellement plusieurs de ces activités, ou passent de l'une à l'autre en fonction des projets. En effet, le développement de nouvelles approches trouve sa pertinence dans la confrontation quotidienne avec données et les questions biologiques. Par ailleurs l'analyse de données ne se satisfait généralement pas de la simple application d'outils existants : chaque projet de recherche génère des besoins spécifiques, et appelle à l'adaptation voire au développement d'approches dédiées. Cependant, le travail de bioinformaticiens-analystes

## Conseil scientifique de l'institut des sciences biologiques

dépend en grande partie de l'existence d'outils éprouvés, efficaces, fiables et stables, qui dépendent du travail de bioinformaticiens-développeurs.

En effet, il est important de distinguer d'une part les logiciels *ad hoc* développés par des bioinformaticiens pour répondre aux besoins ponctuels de leur laboratoire, et qui ne sont généralement utilisables que par leur développeur, et d'autre part les outils bioinformatiques conçus pour une communauté d'utilisateurs : équipés d'interfaces intuitives, documentés, testés, maintenus et susceptibles d'évoluer à long terme. De tels outils demandent un investissement en ressources humaines à plus long terme. En effet, la précarité des développeurs (souvent doctorants ou post-doctorants) nuit à la pérennité et à l'évolution des outils développés. Il est donc essentiel d'adapter les mécanismes du financement de la recherche pour soutenir des projets de développement d'outils bioinformatiques.

### Recommandations

1. Pour répondre aux nouveaux défis de la biologie (*Next Generation Sequencing*, biologie des réseaux, métagénomique, etc), il est essentiel d'investir simultanément dans le déploiement de services en bioinformatique, les analyses au sein des laboratoires de biologie, et le développement d'outils originaux.
2. Différents contextes de travail demandent l'intégration des bioinformaticiens : laboratoires de recherche en biologie, laboratoires à l'interface entre sciences formelles et biologie. Les bioinformaticiens post-doctorants intégrés dans les laboratoires de biologie sont souvent confrontés au danger d'être considérés comme collaborateurs techniques sur des projets scientifiques de leurs collègues, sans que soit défini un projet de recherche propre. Cette situation génère typiquement des CV de second auteur, qui, dans le contexte trop restrictif de la sélection bibliométrique, constituent un frein à la pérennisation des postes. Plusieurs stratégies pourraient être combinées pour éviter l'isolation intellectuelle qui pourrait survenir dans la longue durée, pour les bioinformaticiens accueillis dans des laboratoires de recherche en biologie : mise en place de réseaux trans-laboratoires qui maintiendront une culture disciplinaire, double rattachement à des équipes de biologie et de bioinformatique.
3. Il serait opportun de promouvoir la création de nouvelles unités caractérisées par la combinaison d'une masse critique de théoriciens (informaticiens, mathématiciens et physiciens) intéressés à des problèmes de biologie et intéressés à développer des nouvelles méthodes pour la biologie, couplées à des équipes de biologistes axés sur des problématiques particulières (par exemple l'UMR7238). Il serait bon que les différents instituts de biologie en création en France réfléchissent au développement de contextes de travail qui pourront regrouper des théoriciens interagissant avec les expérimentateurs au sein de leur structure (par exemple l'IBPS).
4. Un problème récurrent en bioinformatique est lié à la volatilité des ressources humaines. Typiquement, un doctorant ou un post-doctorant développe un logiciel ou une base de données, puis quitte le laboratoire pour poursuivre ailleurs sa carrière, et la ressource n'est plus maintenue. Pour éviter ce gaspillage, il serait utile de renforcer le support technique (CDI d'IE et IR) pour les groupes qui manifestent une activité importante de développement de ressources bioinformatiques originales et utiles à la communauté.
5. Il existe au sein des plates-formes un potentiel élevé de développement d'approches bioinformatiques, mais leur visibilité est parfois réduite du fait d'un manque d'intégration et de mise en commun des ressources développées, qui occasionne une certaine redondance des efforts. Il serait utile de promouvoir l'accessibilité, la visibilité et l'interopérabilité des outils développés par les différentes équipes.
6. Afin de renforcer les structures françaises performantes en bioinformatique (équipes de recherche et plates-formes), il serait stratégique de favoriser leur intégration au sein de réseaux fédératifs plutôt que de concentrer les ressources sur une seule

## Conseil scientifique de l'institut des sciences biologiques

structure centralisée. Ceci s'applique au niveau national mais également au niveau international. A cet égard, le Conseil s'interroge quant à la pertinence d'un investissement de la France dans le projet européen Elixir, qui revient à organiser l'ensemble des services de bioinformatique européens en centralisant sur un seul site non seulement des moyens matériels, mais également les ressources humaines et les structures décisionnelles.

7. Le recrutement de bioinformaticiens pose actuellement une difficulté du fait de la rareté des personnes disposant de la double compétence en sciences de la vie et sciences formelles. Afin de rencontrer les besoins présentes et futurs, il est crucial de renforcer les formations. Ceci concerne non seulement le développement de parcours interdisciplinaires (bioinformatique, sciences génomiques), mais également l'insertion, dans les formations des biologistes, d'unités d'enseignement adaptées à l'évolution des technologies à haut débit. Il serait bénéfique d'ouvrir dès les licences des filières de formation interdisciplinaires, à cheval entre sciences de la vie et sciences formelles. Il serait également souhaitable de favoriser l'ouverture des formations de second et troisième cycle aux étudiants d'origine étrangère en prévoyant des cursus en anglais.

Texte préparé par Jacques van Helden, Alessandra Carbone et Francis-André Wollman

**Recommandation adoptée le 31 mars 2014, (16 votants, 15 oui, 1 abstention)**

**Destinataire :**

Mme Catherine JESSUS, Directrice de l'INSB