

<https://laurentbloch.org/BlogLB/Enseigner-l-informatique-aux-biologistes>



Enseigner l'informatique aux biologistes

- Les chats ne sont pas de gauche -

Date de mise en ligne : mardi 25 octobre 2022

Copyright © Blog de Laurent Bloch - Tous droits réservés

[Chapitre précédent](#) [\[/Chapitre suivant\]](#)

Tout a commencé avec James D. Watson, Francis Crick, Rosalind Franklin [1] et quelques autres dans les années 1950 : ils nous ont appris que la conformation des êtres vivants se déduisait de l'information génétique codée par l'acide désoxyribonucléique (ADN) de leurs cellules. Les macromolécules (on dit aussi polymères) d'ADN sont de longues chaînes de motifs moléculaires élémentaires, les nucléotides, dont il existe quatre variétés : l'adénine (notée A), la thymine (notée T), la cytosine (notée C) et la guanine (notée G). Le paradigme de la biologie moléculaire postule que l'information génétique est formulée par un texte, le génome, écrit dans un alphabet de quatre lettres, A, T, G, C, et que la connaissance de ce texte permet de connaître les fonctions de l'organisme considéré, sans avoir trop à entrer dans des considérations supplémentaires d'ordre physico-chimique.

Avec l'entrée en vigueur du paradigme de la biologie moléculaire, l'informatique était appelée inéluctablement à prendre un rôle crucial dans la recherche en biologie, puisque la recherche se fondait désormais sur l'analyse d'un texte. C'est aujourd'hui une évidence. Le séquençage du génome, c'est-à-dire l'obtention de son texte au moyen d'appareils nommés séquenceurs, est accessible depuis plusieurs années pour tous les organismes, ainsi pour les 3,4 milliards de nucléotides et les 26 517 gènes du génome humain, et ces données ont bouleversé de fond en comble le métier de biologiste. La connaissance du génome, complétée par la robotisation des manipulations « en phase humide », déplace le chercheur de la paillasse, des boîtes de Pétri et du Pipetman vers l'ordinateur, lui fait retirer sa blouse blanche, et lui donne accès à des méthodes scientifiques entièrement nouvelles pour étudier la génétique, la structure des protéines, le métabolisme, etc. Des investigations qui demandaient des mois de travail répétitif et entaché d'erreurs à la paillasse sont désormais résolues en quelques heures par des méthodes informatiques. La consultation des banques de données qui archivent les résultats exhaustifs du séquençage et des calculs de structure des protéines donne en quelques minutes la réponse à des questions dont la solution directe aurait constitué un thème de recherche à part entière. Ce qui signifie que l'on peut désormais se poser des questions inenvisageables auparavant.

Au début des années 1990 nous sommes à un tournant. Les gens bien informés de la recherche mondiale savent déjà que les choses vont dans cette direction, les mandarins résistent pied à pied, accrochés à leurs paillasses, les institutions de pointe ont déjà lourdement investi en informatique, telles le *National Center for Biotechnology Information* (NCBI) créé en 1988 à Bethesda près de Washington, ou le *Wellcome Trust Sanger Institute* créé en 1992 à Hinxton près de Cambridge en Angleterre. L'Institut Pasteur est en retard, et en prend conscience. C'est en 1991 que François Rougeon, Directeur de la recherche à l'Institut Pasteur, me propose de le rejoindre pour l'aider à mettre sur pied une infrastructure bioinformatique. En fait, avant de faire de la bioinformatique, il faut déjà créer l'informatique, recruter une équipe, installer des serveurs et un réseau, ce qui occupera une bonne part de notre temps les premières années.

Les étudiants en biologie n'apprenaient pas (et en général n'apprenaient toujours pas) l'informatique à l'Université, ce qui en 1993 a suscité chez William Saurin (chercheur dans l'unité de Maurice Hofnung) l'idée d'organiser à leur intention un cours d'informatique destiné à leur procurer les bases de cette science. William m'a soumis cette idée, j'y ai adhéré, elle a aussitôt été soutenue par François Rougeon et Jean-Paul Aubert, alors chef du Département des Enseignements. William, Frédéric Chauveau et moi-même nous sommes mis à la rédaction d'un programme d'enseignements.

Le contenu initial du cours a été largement déterminé par les trois personnes qui en avaient eu l'idée, non sans conseils extérieurs, notamment de Harald Wertz, professeur à l'Université de Paris 8, et d'Éric Gressier, maître de conférences au CNAM. Ils seront rejoints plus tard par Manuel Serrano, de l'INRIA, et Christian Queinnec, de l'Université Pierre et Marie Curie.

L'idée est la suivante : la biologie moléculaire informatique connaît un essor impétueux et ses progrès théoriques s'incarnent dans des logiciels. Contribuer à cet essor (et comment imaginer que l'Institut Pasteur pût s'y soustraire ?) suppose de créer des logiciels. La création de logiciels demande de savoir programmer. Le cœur de la compétence informatique est la programmation. Si l'on ne veut pas se contenter de former des techniciens qui utiliseraient des logiciels sans être capable d'en comprendre le fonctionnement, de le critiquer et d'en réaliser (ou d'en proposer) des modifications qui correspondent à des innovations dans le domaine de la science, il faut que notre cours soit une formation informatique sérieuse et systématique, c'est-à-dire fondée sur l'apprentissage de la programmation.

Le cours est lancé en janvier 1994. Il dure trois mois et demi, à plein temps, trois heures de cours le matin, trois heures de travaux pratiques l'après-midi. Et comme la salle de travaux pratiques est contiguë aux locaux du Service d'informatique scientifique, où travaillent une bonne partie des enseignants, les travaux personnels des étudiants peuvent se prolonger dans la soirée. Les promotions sont de quinze étudiants, sélectionnés sur dossier et par audition pour ceux qui peuvent se déplacer. Dès la première session nous avons en effet un effectif international, bien que les enseignements soient en français.

Enseigner l'informatique ne saurait se réduire à l'algorithme et à la programmation : le programme comporte également des cours d'introduction aux réseaux, aux bases de données, à la modélisation moléculaire, au système d'exploitation, à la logique et à l'histoire de l'informatique. Comme je l'ai déjà signalé, ce programme est largement inspiré de celui du cycle A d'informatique du Cnam, dont Frédéric et moi-même sommes familiers. Pour l'algorithme et la programmation, le premier langage est Scheme (un dialecte de Lisp), le second Ada, puisqu'il est clair pour nous qu'il faut enseigner au moins deux langages ; en fait les étudiants en apprennent un troisième, le *shell* Unix, et même un quatrième si l'on ajoute *make* et *Makefile*. En tout cas il nous semble indiscutable que l'on ne saurait enseigner un algorithme sans qu'il soit programmé ; je sais qu'il y a des gens qui font autrement, je ne dirai pas ce que j'en pense.

Nos premiers étudiants viennent souvent des marges de la biologie, parfois après avoir découvert que les manipulations à la paillasse, passionnantes les premiers mois, deviennent vite répétitives et fastidieuses. Mais très vite la crise de l'emploi en biologie renouvelle notre public. En effet, les effectifs des cursus de biologie sont très excessifs en regard des possibilités d'emploi, et nous voyons arriver des étudiants munis d'une thèse, voire d'un stage postdoctoral dans une université étrangère prestigieuse, et qui n'ont pas de travail. À tous, notre formation a ouvert les portes d'un emploi, et souvent d'une très belle carrière.

La promotion 1997 est restée dans toutes les mémoires : elle est très internationale, avec une Argentine, une Mexicaine, une Vénézuelienne, une Thaïlandaise, une Allemande, un Canadien et un Vietnamien. Les latino-américaines mènent une ambiance torride, il y a une soirée « tequila frappée » dont les participants ont gardé des souvenirs confus mais enthousiastes.

Pendant les premières années le diplôme délivré est uniquement pasteurien, mais à partir de 1998 il devient un diplôme universitaire (DU), co-habité par les universités Pierre et Marie Curie et Denis Diderot.

Une formation supérieure en bioinformatique doit bien sûr permettre aux étudiants d'acquérir la maîtrise des méthodes et des outils d'analyse et de modélisation à la pointe des avancées actuelles. Elle ne peut pas se limiter à ce premier objectif : le chercheur, l'ingénieur d'une industrie de pointe doivent posséder les compétences qui leur permettront de mettre en œuvre les méthodes et les outils du moment, mais il faut également qu'ils soient en mesure de se frayer un chemin dans la bioinformatique en devenir, et aussi, pour ceux qui le souhaitent, de contribuer à l'élaboration de ce devenir, de créer leurs propres méthodes et les logiciels qui les implémentent.

Pour atteindre ces deux objectifs, le cursus envisagé ne peut faire l'économie d'un volet consacré à un véritable enseignement initial des bases de l'informatique. Le Cours Pasteur a permis de conférer une véritable double

compétence à de futurs chercheurs ou ingénieurs, ce qui leur a ouvert des horizons professionnels impossibles à imaginer autrement. Ce cours a formé plusieurs promotions de bioinformaticiens actifs dans le domaine aujourd'hui, et a permis à nombre de biologistes de se reconvertis comme chercheurs ou ingénieurs dans le domaine de l'informatique, biologique ou autre, alors qu'il est de plus en plus difficile à un jeune diplômé en biologie de trouver un emploi s'il ne possède pas de compétences informatiques suffisantes.

En même temps se développe au sein de l'équipe une véritable activité de recherche au sens académique du terme, notamment après la venue de Marie-France Sagot, qui venait de soutenir une thèse remarquable, ainsi qu'avec les travaux de Catherine Letondal.

Dire que le Service d'informatique scientifique est pendant cette décennie un des principaux facteurs d'évolution de l'Institut Pasteur n'est pas exagéré. Le passage à l'informatique, qui se heurte à de fortes réticences des chercheurs « installés » dont cela menace les positions, est une question de survie scientifique pour l'institution, menacée de ridicule pour son retard en 1991, et dont six ans plus tard le site Web, avec son offre de logiciels scientifiques et de bases de données en libre accès, est une référence au niveau mondial, auquel ne se compare que celui de Washington University à Saint-Louis (Missouri).

De ces expériences je retire que si les ingénieurs laissés à eux-mêmes sont menacés par la tentation du conservatisme technique et de la routine, dans un domaine tel que l'informatique, où les réalisations concrètes comptent, des chercheurs laissés à eux-mêmes risquent de tourner en rond faute de moyens suffisants pour mettre leurs idées à l'œuvre. La réunion de chercheurs et d'ingénieurs est beaucoup plus productive, les ingénieurs sont stimulés par les idées des chercheurs, qui trouvent dans le même couloir le soutien nécessaire aux réalisations techniques qui prolongent leurs travaux. Pour que cela marche, il faut bien sûr que cela ne se passe pas, comme il est de règle dans les institutions de recherche françaises, selon un rapport de sujétion des frères convers aux membres du clergé.

Une grande partie des activités des équipes mixtes évoquées ci-dessus étaient de la recherche : pas toujours au sens académique du terme, certes, plutôt de la R&D, avec à la clé non des articles, mais des choses novatrices qui fonctionnent, ce qui ne me semble pas déshonorant. Je dois constater que les auteurs de ces travaux n'ont pas été très bien traités par le mandarinat universitaire.

L'institution a préféré, quelques années plus tard, que les torchons ne soient pas mélangés aux serviettes, elle a créé une vaste structure académique vouée à l'écriture d'articles, relégué les ingénieurs et les techniciens aux tâches ancillaires, et oublié que pour être crédible en informatique, fût-elle bioinformatique, il faut aussi être capable d'écrire des logiciels. Les brillants résultats obtenus lors de la crise de la Covid illustrent le bien-fondé de cette orientation.

[\[/Chapitre suivant/\]](#)

[1] L'histoire a retenu les noms de Crick et Watson : Rosalind Elsie Franklin a joué dans cette découverte un rôle aussi important, mais elle est morte avant l'attribution du prix Nobel.