

Éditorial: La bio-informatique restera-t-elle une exception ou deviendra-t-elle la normalité dans la communauté scientifique ?

Robert Brasseur

Centre de Biophysique moléculaire numérique. Faculté universitaire des Sciences agronomiques de Gembloux. 2, passage des Déportés. B-5030 Gembloux (Belgique). E-mail : brasseur.r@fsagx.ac.be

La bio-informatique est une discipline née de la rencontre de la biologie et de l'informatique.

L'explosion actuelle des connaissances biologiques permet aux scientifiques non seulement de faire d'immenses progrès dans la compréhension des phénomènes qui nous entourent, mais également de devenir acteurs de leurs transformations ; les OGM, le clonage sont deux exemples de ces transformations. Cependant, il est clair que le chercheur doit, tant au plan intellectuel qu'au plan éthique contrôler ce qu'il induit : pour cela, il faut qu'il reste au courant des nouveautés, c'est-à-dire qu'il assimile et digère une multitude de connaissances. Seul l'outil informatique peut l'aider dans cette tâche en lui permettant d'accéder rapidement aux données, de les analyser et d'en extraire les parties pertinentes. Examinons les deux grands domaines de la bio-informatique, la gestion des bases de données, de plus en plus nombreuses et volumineuses, et la modélisation c'est-à-dire la construction et l'analyse tridimensionnelle de molécules de plus en plus complexes.

Les premières bases de données de séquences de protéines et d'acides nucléiques datent des années 60. En 1965, Dayhoff et ses collaborateurs ont rassemblé toutes les séquences d'acides aminés de l'époque. Cette banque représentait moins de 1,44 MB, soit la capacité d'une simple disquette. La même banque fait actuellement de l'ordre de 250 MB, soit une progression d'un facteur 173 en un peu plus de 30 ans. En 1982, la première banque nucléotidique a été créée à l'initiative du laboratoire européen de biologie, l'EMBL. La connaissance complète du génome humain prévue pour les années 2005 devrait enrichir considérablement ces deux banques, protéique et nucléique. Par ailleurs, une multitude de banques spécialisées existe : banque de séquences de sites d'interactions protéine-protéine, banque de séquences de motifs divers, banque de données sur le virus HIV... Leur nombre et leurs dimensions dépassent maintenant de très loin les possibilités humaines d'assimilation. Seul l'usage de l'ordinateur et de

logiciels spécialisés devrait permettre de les analyser, de les trier pour en extraire les informations pertinentes susceptibles de répondre aux questions que l'homme se pose. Concevoir et développer des outils d'analyse puissants, versatiles et conviviaux est devenu un défi. Actuellement, un grand nombre de logiciels apparaissent, malheureusement ces logiciels ont été développés pour répondre ponctuellement à une question. Il faudrait concevoir des outils plus universels, plus versatiles, ouvrant l'exploration de nouvelles voies à un public de chercheurs de plus en plus large. Ce qu'il faudra concevoir pour la bio-informatique, c'est l'équivalent de ce que sont les bons logiciels de traitement de texte et de dessin à la bureautique.

La modélisation moléculaire ne peut exister quant à elle que grâce au développement de l'informatique. En effet, sans l'évolution de la puissance des ordinateurs, la prédiction de la structure des protéines aurait été impossible. Si la modélisation moléculaire a longtemps été l'apanage de chercheurs spécialisés, ces dernières années, le développement d'outils de plus en plus simples d'emploi a permis à un grand nombre de découvrir le monde tridimensionnel des protéines, des acides nucléiques, etc. La plupart des études de modélisation moléculaire comportent trois étapes. Dans la première, un protocole est sélectionné pour décrire les interactions intra- et intermoléculaires du système. Les deux protocoles les plus couramment utilisés en modélisation sont la mécanique quantique et la mécanique moléculaire. Ces méthodes ont déjà aidé à la découverte et à la conception de nouvelles molécules et à l'analyse de relations structure-activité. À l'aide d'un champ de forces spécifique, on peut calculer l'énergie de n'importe quel arrangement atomique et moléculaire et déterminer comment cette énergie varie quand la position des atomes et des molécules change. La deuxième étape de l'étude de modélisation moléculaire est une minimisation de l'énergie, que ce soit par dynamique moléculaire, simulation de "Monte Carlo" ou par recherche conformationnelle. À la fin de cette étape on obtient une

structure modèle probable. Enfin, dans la troisième étape, on calcule les propriétés de cette structure et on propose des vérifications expérimentales pour tester ses propriétés et le bien-fondé de l'analyse théorique.

On évalue l'évolution de la bio-informatique de ces dernières 30 années par la dissémination des ordinateurs dans le monde de la recherche scientifique. Ces ordinateurs deviennent aussi communs que bien d'autres techniques scientifiques. Après une période de maturation, les méthodes bio-informatiques deviennent accessibles à un très grand nombre de chercheurs. Elles possèdent, grâce à Internet, un avantage considérable : leur dissémination extrêmement rapide, ce qui n'est pas le cas des techniques expérimentales. Une méthode, des données structurelles, une banque d'informations peuvent s'échanger entre tout point du globe, le village terre, pratiquement instantanément. Il n'est pas nécessaire de faire appel à un transporteur

coûteux pour envoyer de son propre laboratoire un élément à San Francisco, et *vice versa*.

Tout progrès significatif en bio-informatique repose sur le rassemblement et le tri des données brutes. Ces données sont celles obtenues laborieusement par les expérimentateurs. En effet, si l'on exclut les secteurs industriel et commercial qui gardent en point de mire le profit à court terme, la communauté bio-informatique doit être ouverte aux échanges en général, plus que la plupart des secteurs expérimentaux de la biologie. On observe lors de congrès internationaux que la communauté scientifique voit s'ouvrir un tel potentiel de découvertes fondamentales ou appliquées que les chercheurs sont là, avides d'entendre et anxieux de savoir si leur intelligence associera les bonnes données au bon moment : le maître-mot étant, à la veille du 21^e siècle, communication et souplesse intellectuelle.