

UNIVERSITE, Faculté : Lille, Faculté des Sciences et Technologies

Domaine scientifique, Spécialité : Métabolites secondaires, bio-informatique, chemo-informatique

Titre de la thèse : Élaboration d'un algorithme de comparaison de structures de biomolécules, basé sur des caractéristiques chimiques et biologiques

Direction de thèse :

Pupin Maude, MCF hors-classe HDR en informatique

Leclère Valérie, Pr en microbiologie

Noé Laurent, MCF en informatique

Laboratoire(s) de Rattachement : CRISAL (UMR 9189)

et UMR-Transfrontalière BIOECOAGRO INRAE N°1158, Institut Charles Violette, Univ Lille, Univ Artois, ULCO, YNCREA

ED : SPI

Co-financements envisagés : région

SUJET DE THESE

1) Le sujet de recherche choisi et son contexte scientifique et économique

La caractérisation structurale des métabolites est une étape importante pour la découverte de nouvelles biomolécules ayant des applications dans divers secteurs comme l'agriculture en recherche de composés pour la protection des cultures ou la santé en demande de nouveaux antibiotiques ou anticancéreux. Des techniques expérimentales permettent d'extraire de nombreux métabolites de milieux naturels. Il faut ensuite élucider la structure et les activités de ces molécules lors de criblages à haut-débit afin de cibler les plus prometteuses. L'utilisation de logiciels dédiés réduit considérablement le recours à des criblages expérimentaux qui sont longs, coûteux et fastidieux à réaliser. De plus, ils nécessitent l'utilisation de matériel jetable, de produits chimiques et matériels biologiques qui sont des facteurs de pollution à cause des déchets, souvent non recyclables, générés.

Les métabolites sont de tailles et de structures variées, certains sont suffisamment petits pour être étudiés à l'aide de logiciels dédiés aux structures chimiques, d'autres sont suffisamment grands pour être étudiés en tant que séquences biologiques (ADN/ARN ou protéines). À notre connaissance, il n'existe pas de logiciel capable de prendre en compte à la fois les structures chimique et biologique des métabolites et donc de caractériser de façon pertinente des métabolites de taille intermédiaire.

Fort de notre expérience pluridisciplinaire, nous avons pour ambition de concevoir un algorithme de comparaison multiple de molécules original et unique car il prendra en compte à la fois les ressemblances biologiques et chimiques.

2) L'état du sujet dans le laboratoire d'accueil

Des membres de CRISAL et de l'Institut Charles Violette (ICV) collaborent depuis 2006 sur un projet de recherche interdisciplinaire. Ce projet est porté par, d'une part, des informaticiens spécialistes de la conception d'algorithmes dédiés à l'analyse de données biologiques (équipe Bonsai) et, d'autre part, des microbiologistes spécialistes de molécules particulières appelées peptides non-ribosomiques (PNR). Ces derniers ont une structure très différente des séquences ADN/ARN ou protéiques : ils peuvent contenir des cycles et/ou des branchements et nous avons répertorié plus de 540 monomères (briques de base) différents.

Ensembles, nous avons constaté l'absence d'outils informatiques dédiés à ces molécules, bien qu'elles aient des applications intéressantes dans le domaine de l'agriculture, de la santé ou encore les bio-détergents. En effet, le célèbre antibiotique pénicilline est un peptide non-ribosomique, tout comme la cyclosporine, le premier médicament utilisé pour limiter les rejets de greffes. L'ICV étudie d'ailleurs le potentiel



de plusieurs PNR pour des applications dans le domaine de l'agriculture, et plus précisément du biocontrôle et a même créé une start-up, LIPOFABRIK, afin de les produire.

Notre collaboration a ainsi abouti à la conception de la ressource logicielle Norine, unique ressource dédiée aux peptides non-ribosomiques, qui répond à un besoin réel de la communauté scientifique. Norine comprend une base de données de PNR, associée à des logiciels d'analyse permettant, entre autre, de rechercher la structure d'un peptide dans la base, de dessiner et visualiser des structures biologiques, d'inférer une structure biologique à partir d'une structure chimique et d'analyser des spectres de masse. Elle a acquis une reconnaissance internationale et des bases de données internationales, dont PDB, référencent nos données. Nos 16 publications cumulent plus de 600 citations et notre serveur-web reçoit environ 2000 requêtes par mois, de scientifiques du monde entier. Cette collaboration fructueuse a été renforcée par le co-encadrement de 2 thèses et des interactions fortes sur 4 autres thèses. Elle a mené à la publication co-signée de 16 articles scientifiques dans des journaux avec comité de lecture et de nombreuses communications sous la forme d'exposés ou de posters dans des congrès nationaux et internationaux que ce soit dans le domaine de la micro-biologie, de la bio-informatique ou de la chemo-informatique. Récemment, Norine a été retenue comme *Service Delivery Plan* de ELIXIR (organisation inter-gouvernementale qui rassemble les ressources dédiées aux sciences du vivant à travers toute l'Europe). Seuls 25 ressources ont été retenues sur les 100 candidatures françaises.

La réussite de ce projet interdisciplinaire reposera sur les expertises complémentaires des membres des deux laboratoires. L'équipe Bonsai de CRISAL a acquis un savoir-faire dans l'algorithmique des graphes dédiés aux molécules. Nous avons déjà conçu des logiciels de comparaison de PNR représentés sous leur forme biologique et d'inférence de la structure biologique à partir de la structure chimique (point de départ nécessaire à la nouvelle comparaison). Les membres de l'ICV ont acquis un savoir-faire dans la production, l'analyse structurale et la caractérisation des PNR. Ils pourront ainsi participer à la définition des critères de ressemblance entre molécules et pourront évaluer la pertinence des résultats obtenus par le logiciel.

3) Les objectifs visés, les résultats escomptés

De nombreux secteurs, dont l'agriculture et la santé, sont demandeurs d'outils pour élucider les structures chimiques des métabolites lors de criblages haut-débit et permettre ainsi la découverte de nouvelles molécules. La Pistoia Alliance, organisation mondiale sans but lucratif, rassemble des entreprises pharmaceutiques (Pfizer, Roche...), des bases de données internationales (ChEMBL pour l'Europe, PubChem pour les USA) et d'autres acteurs majeurs de la chimie du vivant qui ont pour ambition de réduire les obstacles à l'innovation. Parmi leurs projets, HELM est une notation associée à des logiciels opensource. Elle permet de représenter les biomolécules qui ne sont pas analysables via les logiciels dédiés aux petites molécules ou aux séquences. Nous avons établi des interactions fortes avec la Pistoia Alliance en participant au groupe de travail qui a défini les règles pour déterminer si une molécule est un monomère, la brique de base de la notation HELM. L'objectif est de combler le fossé qui existe entre l'analyse chimique et l'analyse biologique car la notation HELM associe les deux approches. L'enjeu est une meilleure compréhension de ces molécules afin d'en découvrir de nouvelles ayant des débouchés industriels.

Concernant la comparaison structurale de molécules, il existe seulement une preuve de concept pour la recherche d'une molécule représentée avec la notation HELM dans une base de données. Nous avons l'ambition de concevoir un logiciel de comparaison via la notation HELM qui inclut non seulement la recherche de structures similaires dans une base de molécules, et aussi l'alignement multiple de molécules. Ce logiciel inédit répond à un besoin réel puisque la comparaison de molécules est une étape indispensable à leur analyse, notamment en vue de faire du criblage haut-débit pour identifier de nouveaux composés pour des applications dans le domaine de l'agriculture et de la santé, entre autre. La comparaison est nécessaire à la détermination de la diversité des molécules étudiées, l'identification des



molécules de la même famille, la caractérisation de fragments impliqués dans une fonction, *etc.* Actuellement, les logiciels de comparaison de structures chimiques et de séquences coexistent. L'objectif de la thèse est d'innover en concevant un logiciel hybride qui allie les points forts des approches dédiées aux petites molécules et celles dédiées aux séquences. Ce logiciel sera mis à disposition de la communauté scientifique sur la ressource Norine, et pourra aussi être déployé sur les serveurs d'autres bases de données internationales de molécules telles que ChEMBL, gérée par le *European Bioinformatics Institute*. Ainsi, nous pourrons l'utiliser dans nos projets de découverte de nouvelles biomolécules capables de protéger les cultures. De plus, les scientifiques du monde entier pourront l'utiliser dans leurs projets appliqués à d'autres débouchés liés aux métabolites qu'ils étudient.

Date de recrutement envisagée : septembre ou octobre

Contact (adresse e-mail) : maude.pupin@univ-lille.fr, valerie.leclere@univ-lille.fr

