Université de Rennes

Direction de la Recherche et de l'Innovation

Appel à projets - Année 2024

Contrats doctoraux «Ordinaires»

ECOLE DOCTORALE : Matisse **POLE DE RECHERCHE : Mathématique-Numérique**

Unité de recherche d'accueil

Unité de Recherche : IRISA

Equipe : Dyliss

Encadrement

Directeur de thèse : Anne Siegel **Unité de recherche** : IRISA

Employeur/Établissement d'origine : CNRS

HDR : oui **Nb de thèses en cours** : 3 au 01/2024, dont deux se terminent en 09/24.

Mail : anne.siegel@irisa.fr **Tél.** :

Co-directeur de thèse : Gabriel Markov **Unité de recherche** : Laboratoire de Biologie Intégrative des Modèles Marins (UMR8227), Station Biologique de Roscoff

Employeur/Établissement d'origine : CNRS

HDR : non **Nb de thèses en cours** : 2 (50% 2022-25; 33% 2023-26).

Mail : gabriel.markov@sb-roscoff.fr **Tél.** : 06 40 70 66 01

Sujet

Titre : **Méthodes hybrides pour l'inférence *ab-initio* de voies métaboliques chez des eucaryotes marins**

Mots clés : programmation par contraintes, problèmes de satisfiabilité, raisonnement abductif, apprentissage de fonctions protéiques chemoinformatique, transformations moléculaires

Cotutelle internationale : non

Descriptif (15-30 lignes) :
Contexte socioéconomique et scientifique ; Les hypothèses et questions posées ; Les grandes étapes de la thèse et démarche ; Approches méthodologiques et techniques envisagées

Les méthodes actuellement développées pour construire les cartes du métabolisme associées à des espèces biologiques s'appuient sur les informations contenues dans leurs génomes. Ces méthodes ne permettent pas d'identifier les voies de synthèses de composés qui sont associées à des fonctions trop spécifiques pour être actuellement caractérisées par des approches de génomiques. L'exploration de la « matière noire génomique » (Zhang et Moore, 2015) apparaît comme un enjeu prioritaire des années 2020, en lien avec différents programmes de séquençages massifs de la biodiversité. Par exemple, en biologie marine, le programme Phaeoexplorer a séquencé des dizaines de génomes de macro-algues. Il est précurseur du programme Atlasea qui vise à séquencer des milliers d'eucaryotes marins. Ces programmes visent au final à exploiter la biodiversité pour découvrir de nouvelles voies de synthèses de composés métaboliques spécifique. Le paradoxe ici est que ces voies de synthèses ne pourront pas être identifiées directement à partir des approches de génomique comparative utilisée habituellement (Belcour et al., 2023) puisque ces dernières s'appuient sur des catalogues de fonctions existantes et caractérisées à partir de génomes, et qu'on souhaite précisément explorer les fonctions non encore identifiées.
Une approche complémentaire à celle basée sur des génomes est d'exploiter les connaissances sur la composition chimique et la diversité des molécules identifiées chez différentes espèces proches pour caractériser les différentes voies de biosynthèse qui sont possibles chez une espèce donnée. Dans une première approche dans cette direction, l'approche pathmodel (Belcour et al., 2020) a été développée en collaboration entre l'IRISA et la Station Biologique de Roscoff.

Cette approche consiste à générer manuellement un espace des transformations chimiques possibles à partir de connaissances sur une famille d'espèce, et ensuite de développer un raisonnement abductif modélisé par des problèmes de satisfiabilité pour identifier les chaines de transformation cohérentes avec les mesures expérimentales sur les molécules. Cela a permis de prédire que la voie de synthèse du cholestérol chez les algues rouges pourrait être constituée de blocs de transformations apparaissant chez la tomate et chez les algues brunes (Girard et al., 2021). Ce cas d'étude a ainsi permis de monter que des approches basées sur du raisonnement et des mesures chimiques permettent de capturer la flexibilité adaptative des espèces dans certaines voies de synthèse, propriété qui qui inaccessibles à partir des seules informations génomiques.

Le but de la thèse sera de travailler à l'automatisation et à la généralisation de cette approche visant la prédiction de voies de biosynthèse pour des eucaryotes en tenant compte de la composition chimique (spectrométrie de masse) et de la diversité des molécules identifiées chez différentes espèces proches (connaissances taxonomiques et phylogénétiques). Cette approche basée sur l'inférence de transformations moléculaires est bien connue en santé humaine, en particulier en toxicologie (Conan et al., 2021), et s'appuie sur un grand nombre de mesures réalisés dans différents contextes qui ne sont pas disponibles sur les organismes modèles émergeants. Une autre approche est basée sur l'extraction de connaissances (Duigou et al., 2019) pour fournir des canevas de voies de synthèses mais ne tient pas compte des capacités enzymatiques des espèces. L'objectif de la thèse est d'utiliser en contrepartie les informations sur des organismes proches pour prédire des voies de synthèse. Le cas d'étude concernera l'étude de la variabilité de la synthèse des oxylipines chez les algues brunes, en particulier les laminales et les ectocarpales, pour lesquelles la station biologique de Roscoff dispose à la fois de génomes, de listes de molécules identifiées chez différentes espèces, ainsi que de données métabolomiques (Xing et al., 2023).

D'un point de vue méthodologique, l'objectif sera d'utiliser des approches basées sur des technologies du web sémantique pour extraire des sources de données publiques (ChEBI, Rhea, Retrorules) les informations sur les transformations chimiques existant dans le vivant et définir une banque de transformations moléculaires possibles dans le vivant. Il faudra ensuite classifier ces réactions en grande famille, comme cela est fait en santé humaine ou en biologie végétale (Conan et al., 2021, Elser et al., 2023), et associer à chaque réaction un score de faisabilité chez un organisme considéré, en tenant compte des informations connues chez des espèces proches. On pourra ainsi formuler sous la forme de problème d'optimisation la recherche d'une voie de synthèse en étendant l'approche développée dans pathmodel (Belcour et al., 2020). Ces travaux ouvriront la possibilité de tenir compte d'autres informations pour affiner les espaces de recherche, en particulier les structure des enzymes impliquées dans les chaines de transformation, en lien avec les méthodes de prédictions de fonctions enzymatiques développées dans l'équipe Dyliss (Buton et al, 2023).

Bibliographie

Belcour A, Girard J, Aite M, Delage L, Trottier C, Marteau C, Leroux C, Dittami SM, Sauleau P, Corre E, Nicolas J, Boyen C, Leblanc C, Collén J, **Siegel A**, & **Markov GV**. (2020). Inferring Biochemical Reactions and Metabolite Structures to Understand Metabolic Pathway Drift. *iScience*, 23(2), 100849. https://doi.org/10.1016/j.isci.2020.100849

Belcour A, Got J, Aite M, Delage L, Collén J, Frioux C, Leblanc C, Dittami SM, Blanquart S, **Markov GV** and **Siegel A**. (2023) *Genome Research*, https://doi.org/10.1101/gr.277056.122

Buton N, Coste F, Le Cunff Y (2023), Predicting enzymatic function of protein sequences with attention, Bioinformatics 10.1093/bioinformatics/btad620

Conan M, Théret N, Langouet S, & **Siegel A**. (2021). Constructing xenobiotic maps of metabolism to predict enzymes catalyzing metabolites capable of binding to DNA. *BMC bioinformatics*, 22(1), 450. https://doi.org/10.1186/s12859-021-04363-6

Duigou T, du Lac M, Carbonell P, & Faulon JL. (2019). RetroRules: a database of reaction rules for engineering biology. *Nucleic acids research*, 47(D1), D1229–D1235. https://doi.org/10.1093/nar/gky940

Elser D, Pflieger D, Villette C, Moegle B, Miesch L, & Gaquerel E. (2023). Evolutionary metabolomics of specialized metabolism diversification in the genus *Nicotiana* highlights N-acylnornicotine innovations. *Science advances*, 9(34), eade8984. https://doi.org/10.1126/sciadv.ade898

Girard J, Lanneau G, Delage L, Leroux C, Belcour A, Got J, Collén J, Boyen C, **Siegel A**, Dittami SM, Leblanc C, & **Markov GV**. (2021). Semi-Quantitative Targeted Gas Chromatography-Mass Spectrometry Profiling Supports a Late Side-Chain Reductase

Université de Rennes

Direction de la Recherche et de l'Innovation

Appel à projets - Année 2024

Contrats doctoraux «Ordinaires»

Cycloartenol-to-Cholesterol Biosynthesis Pathway in Brown Algae. *Frontiers in plant science*, 12, 648426. https://doi.org/10.3389/fpls.2021.648426

Xing Q, Cabioch L, Desrut A, Le Corguillé G, Rousvoal S, Dartevelle L, Rolland E, Guitton Y, Potin P, **Markov GV**, Faugeron S, Leblanc C. (2023). Aldehyde perception induces specific molecular responses in *Laminaria digitata* and affects algal consumption by a specialist grazer. *The Plant journal* https://doi.org/10.1111/tpj.16450

Zhang JJ, & Moore BS. (2015). Digging for biosynthetic dark matter. *eLife*, 4, e06453. https://doi.org/10.7554/eLife.06453

➢ **Ce document est à renvoyer par la direction de l'unité de Recherche aux animateurs du Pole Recherche et à l'école Doctorale pour le 9 février 2024 au plus tard**