Vérifications du réseau métabolique entier de Tisochrysis lutea

Nicolas Guillaudeux

Institut **N**ational de **R**echerche en **I**nformatique et en **A**utomatique de
RENNES – BRETAGNE ATLANTIQUE
Institut de **R**echerche en **I**nformatique et **S**ystèmes **A**léatoires
Équipe Dyliss
*(**DY**namics, **L**ogics and **I**nference for biological **S**ystems and **S**equences)*

CAMPUS DE BEAULIEU
263 AVENUE GENERAL LECLERC
35042 RENNES

Vérifications du réseau métabolique entier de *Tisochrysis lutea*

Rapport de stage

Master 1 de Bio-informatique et Génomique

Année universitaire 2016/2017

Auteur :

Nicolas GUILLAUDEUX

Encadrants :

Jeanne GOT

Anne SIEGEL

UMR CNRS 6290
Institut de Génétique et Développement de Rennes (IGDR)

Annabelle MONNIER
Equipe Génomique Fonctionnelle Intégrée et Biomarqueurs

UNIVERSITÉ DE RENNES 1

ENGAGEMENT DE NON PLAGIAT

Je, soussigné(e) Nicolas Guillaudeux
étudiant(e) en M1 - Bioinformatique et Génomique
déclare être pleinement informé que le plagiat de documents ou d'une partie de document publiés sur toute forme de support, y comprit l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour la rédaction de ce document.

Date : 12/06/2017

Signature :

Guillaudeux

Document à compléter de manière manuscrite et à insérer obligatoirement en première page du rapport de stage.

Laboratoire de Génomique Médicale
BMT-HC - CHU Pontchaillou

2 rue Henri le Guilloux
35033 Rennes Cedex
FRANCE

Annabelle MONNIER
annabelle.monnier@univ-rennes1.fr
TÉL. 33 (0)2 99 28 92 54

Remerciements

Je tiens à remercier Jeanne Got et Anne Siegel pour m'avoir accordé ce stage, pour leur aides et conseils, leur disponibilité tant sur le stage que sur la rédaction du rapport et l'accompagnement au cours de stage.

Je remercie aussi Camille Trottier, Clémence Frioux et Méziane Aite pour leur précieux conseils sur AuReMe et tout au long du processus de reconstruction du réseau métabolique de *Tisochrysis lutea* et, encore une fois, merci à Méziane pour ses conseils pour le développement et l'optimisation de scripts.

Je remercie également Matthieu Garnier et Grégory Carrier de l'Ifremer pour leur renseignements sur *Tisochrysis lutea*.

Enfin, je remercie l'équipe Dyliss pour leur accueil chaleureux mais aussi mes collègues stagiaires à l'IRISA (Claire, Kévin, Aurélie, Dimitri, Ali, Charlotte, Alexis et Iskander) pour leur bonne humeur et le cadre de travail convivial et leur entraide.

Table des matières

I – Introduction

1. Contexte

1.1. Les réseaux métaboliques

L'étude du métabolisme est une mise en situation des modes de synthèse et de régulation de composés d'un organisme. A partir du génome, il est possible de reconstruire un réseau métabolique dans lequel les gènes produisent une ou plusieurs protéines qui catalysent des réactions pouvant être réversibles ou irréversibles. Celles-ci peuvent contenir des coenzymes et des cofacteurs et sont majoritairement localisées dans le cytosol, les chloroplastes et les mitochondries [13]. Ces réactions s'organisent en voies métaboliques (*pathways*) illustrant ainsi les étapes de transformation d'un métabolite. Un métabolite est une petite molécule organique qui est en mesure d'être un réactant ou un produit. Les métabolites sont des composés essentiels au métabolisme puisqu'ils jouent un rôle dans le déclenchement et la terminaison d'un ensemble de réactions. Les métabolites ont aussi la possibilité d'être des intermédiaires entre des réactions notamment au sein des cascades métaboliques. Dans ce cas précis, les métabolites dits produits seront alors des métabolites réactants.

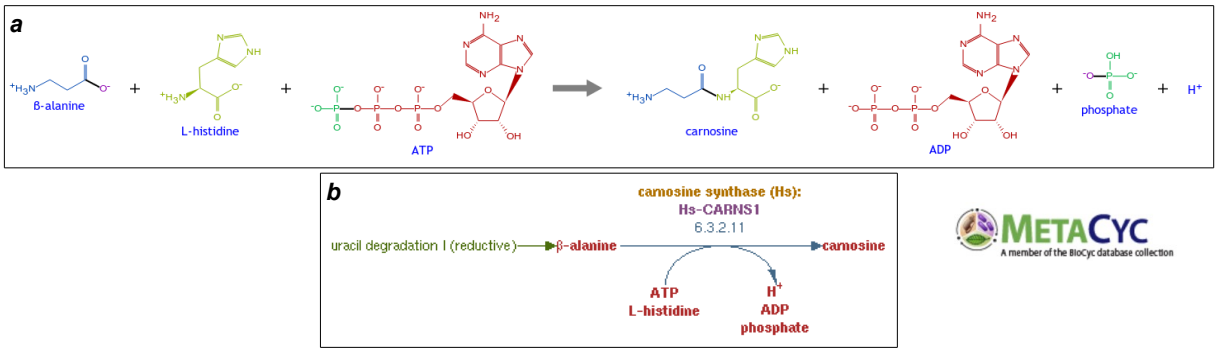

Figure 1. Réaction de la synthèse de la Carnosine, associée à l'E.C. 6.3.2.11 et au *pathway* PWY66-420 (biosynthèse de la carnosine). La réaction (***a***) et le *pathway* (***b***) sont tirés de la base de données MetaCyc version 21.0 à la date de rédaction de ce rapport. Cette réaction est l'unique réaction du *pathway* PWY66-420. La réaction a été mise en évidence chez *Homo sapiens* et est associée au gène CARNS1.

L'interconnexion de ces *pathways* permet ainsi la représentation de l'ensemble de toutes les réactions métaboliques d'un organisme en formant un réseau métabolique. Ces réseaux peuvent donc faire intervenir des enzymes (protéine ayant une activité catalytique et généralement associée à un numéro EC comme présenté dans la figure 1) et des métabolites.

L'objectif de la reconstruction automatique de ces réseaux métaboliques est de pouvoir intégrer de l'information génomique et métabolique issues de séquençage massif, et

d'identifier des gènes candidats à l'échelle du génome [13]. Cette reconstruction permet l'apport de connaissances à la fois sur la biologie de l'organisme, sur sa physiologie mais aussi sur ses interactions biotiques [25]. Dans notre étude, le but est de reconstruire un réseau de façon automatisée sur la microalgue *Tisochrysis lutea* (*cf* paragraphe *1.4.*).

A la différence des espèces modèles parfaitement connues telles que *Saccharomyces cerevisiae*, *Escherichia coli* ou encore *Arabidopsis thaliana*, *T. lutea* est, par opposition, une espèces non modèle du fait que l'on ait moins d'annotations de son génome et moins de connaissance à son sujet. L'équipe Dyliss s'intéresse à ces espèces non modèles.

1.2. Problèmes liés à la modélisation des réseaux métaboliques

Plusieurs problèmes découlent de la modélisation des réseaux métaboliques. Une première difficulté est liée à la réalisation des modèles prédictifs. Ainsi, des limites sont très vites apparues telles que des effets prédits non validés dans la littératures ou l'absence de certains composés comme la biomasse (acides aminés, lipides, nucléotides, *etc*.) importante pour le développement de l'organisme ou encore le manque de certains cofacteurs. L'une des causes de ce manque d'informations est notamment dû à l'absence de réactions dans un organisme annoté, car non identifiée avec l'organisme modèle utilisé. En effet, ceux-ci peuvent être éloignés phylogénétiquement ne permettant pas de combler des génomes incomplets ou des fonctions de gènes encore inconnues qui ne seront alors pas prises en compte dans la reconstruction [19, 22, 24, 25].

Un second problème ne concerne pas uniquement les espèces non modèles mais plutôt le mode de reconstruction qui n'est pas identique d'un modèle à l'autre, et qui apporte une base de connaissance plutôt minimale en terme de prédiction. Dans ce sens, un protocole a été mis en place par Thiele et Palsson (2010) dans le but de permettre, via des étapes précises, une reconstruction de haute qualité en se basant sur quatre étapes : i) la reconstruction d'un réseau métabolique brut ou « *draft* » basée sur l'annotation du génome et les bases de données biochimiques ou métaboliques ; ii) l'amélioration manuelle ; iii) la conversion de la reconstruction en modèle mathématique ; iv) la vérification, l'évaluation et la validation du modèle [28]. Suite à cela, plusieurs approches ont été proposées dans le but d'automatiser ce processus en se basant sur deux étapes : la reconstruction automatisée d'un « *draft* » métabolique et le « remplissage des lacunes » (*gap-filling*) de celui-ci afin d'identifier les réactions manquantes au sein des modèles [25].

1.3. Les bases de données de réseaux

Afin de nous aider à reconstruire des réseaux métaboliques, nous disposons de bases de données métaboliques de références. Plusieurs existent mais dans notre étude, nous nous sommes intéressés à MetaCyc [5] de la librairie BioCyc, KEGG (*Kyoto Encyclopedia of Genes and Genomes*) [14] ou encore BiGG (*Biochemical, Genetic and Genomic*) [16].

1.4. L'espèce d'étude

Cette étude porte sur une microalgue haptophyte de l'ordre des *Isochrysidales*. Les algues haptophytes représentent la source de production la plus importante de biomasse océanique et sont issues d'une endosymbiose secondaire d'une algue rouge à l'intérieur d'un organisme eucaryote non-photosynthétique. Notre espèce a été isolée de Tahiti et désignée sous le nom d'*Isochrysis affinis galbana*. Elle fut rebaptisée plus récemment en *Tisochrysis lutea* (*T. lutea*) en référence à la couleur orange de ses cellules. C'est une espèce largement étudiée en raison de son utilisation dans l'aquaculture en tant que matière première pour les mollusques et les crustacés notamment dans les taux de production d'acides gras polyinsaturés tels que l'acide docosahexaénoïque (DHA). Elle est classifiée dans les *Isochrysidacées* dont l'information sur la reproduction sexuelle et sur le niveau de ploïdie reste encore inconnue [3, 4, 9, 27].

Les microalgues sont des organismes eucaryotes photosynthétiques ayant une part importante dans les processus biologiques. En effet, elles sont capables de fixer le CO_2, de produire de l'oxygène, de recycler des nutriments et ont un taux de croissance rapide. Pour comprendre la biologie de ces organismes notamment, vis à vis de leurs processus métaboliques et de leurs régulations, des études sont réalisées via des approches au niveau du système telles que la reconstruction de réseaux à l'échelle du génome (*Genome Scale Model* ou GSM). Ces réseaux sont reconstruits, complétés et améliorés manuellement [9, 13, 24].

1.5. Les modèles utilisés

Pour permettre la reconstruction de réseaux métaboliques, des organismes modèles sont utilisés. *Arabidopsis thaliana*, une plante terrestre, est l'un des modèle d'étude les plus utilisés en science du fait que ce fut la première plante à fleur à être séquencée et que son génome soit bien annoté. Il peut être utilisé pour représenter à la fois les types cellulaires photosynthétiques et non photosynthétiques [22]. *Chlamydomonas reinhardtii* est une

microalgue de la lignée verte modèle dans l'étude des processus cellulaires spécifiques aux plantes tels que la photosynthèse, la mobilité, le rythme circadien ou encore le contrôle du cycle cellulaire. Son génome est le mieux annoté et le mieux nettoyé ce qui en fait un système idéal dans l'étude du métabolisme des algues [13]. *Ectocarpus siliculosus* est un modèle d'étude biologique chez les algues brunes (Straménopiles) [24]. Enfin, *Synechocystis* sp. PCC 6803 [17] est une cyanobactérie unicellulaire.

2. Objectifs de l'étude

L'objectif de ce stage est de reconstruire le réseau métabolique de *T. lutea* de façon automatisée via les annotations expérimentales et *in silico* et via l'orthologie vis à vis de quatre espèces dont les réseaux métaboliques sont connus : *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Ectocarpus siliculosus*, *Synechocystis sp.* PCC 6803. Le second point d'analyse est de vérifier ce nouveau réseau métabolique par rapport à celui reconstruit manuellement par l'équipe Dyliss en se focalisant sur une voie métabolique spécifique : la voie de biosynthèse de la carnosine.

II – Matériels et Méthodes

1. Outils utilisés

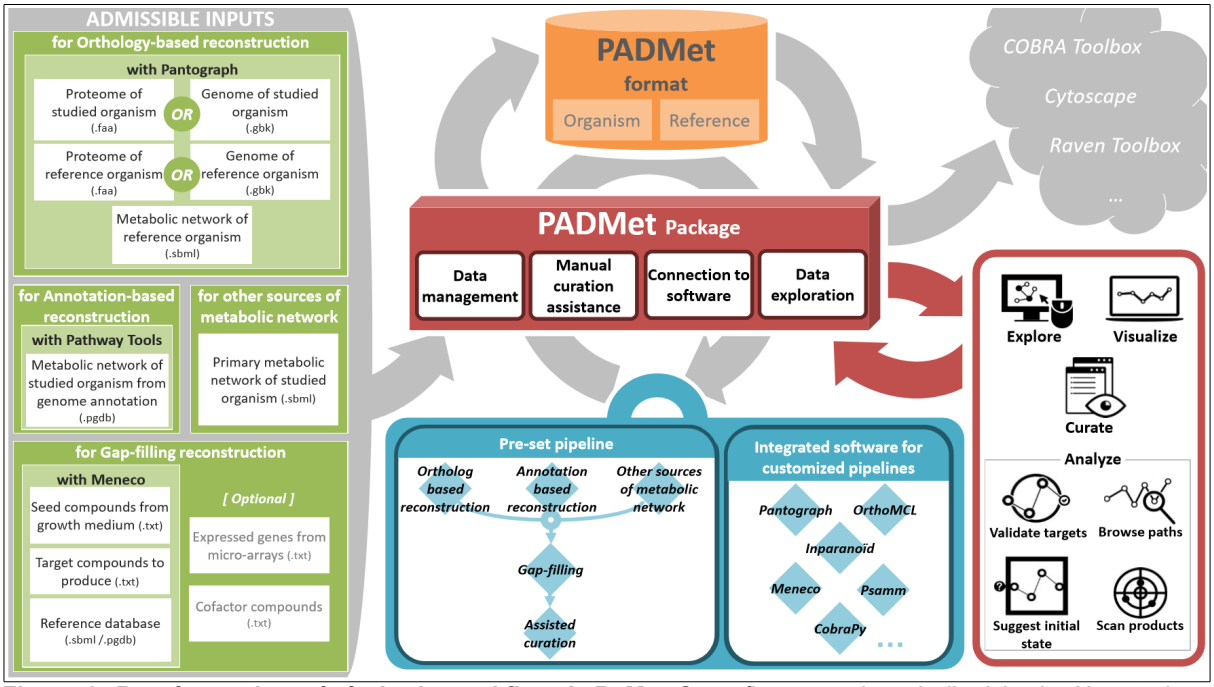

Figure 2. Représentation générale du workflow AuReMe. Cette figure provient de l'article de *Aite et al.* en préparation [1]. Cette figure illustre les entrées possibles, le rôle central du package PADMet, les logiciels inclus, la pipeline de référence, les étapes d'analyse proposées ainsi que le lien avec d'autres outils.

Au cours de ce stage, nous avons utilisé des outils situés au niveau du carré bleu à gauche de la figure 2. Au travers de cette partie, un détail de ces outils va vous être présenté.

1.1. Outils de reconstruction des réseaux métaboliques

Afin de reconstruire de façon automatique un « *draft* » de réseau métabolique via les annotations génomiques, le logiciel Pathway-Tools a été utilisé. Celui-ci est implémenté de l'outil PathoLogic qui permet de créer une nouvelle base de données de voies métaboliques/génomes (*Pathway/Genome DataBase* ou PGDB) propre à notre organisme à partir du génome annoté de ce dernier. Cette PGDB contient toute l'information génomique de cet organisme telle que les gènes, les protéines, les réactions biochimiques et les voies métaboliques prédites pour un organisme donné. PathoLogic permet la reconstruction de GSM en se basant sur une PGDB de référence à savoir MetaCyc [15]. Son utilisation dépend de la taxonomie. Celle-ci correspond à celle contenue sur le site NCBI Taxonomy (*https://www.ncbi.nlm.nih.gov/taxonomy*) soit pour notre étude *Cellular organisms > Eukaryota > Haptophyceae > Isochrysidales > Isochrysidaceae > Tisochrysis > Tisochrysis lutea*. La version du logiciel utilisée au moment de ce stage est la 20.5.

Quant à la reconstruction du « *draft* » basé sur l'orthologie, l'outil Pantograph a été utilisé au sein du workflow AuReMe qui contient aussi OrthoMCL [20] et InParanoïd [26]. Ces deux derniers logiciels identifient les paires de protéines orthologues au travers de BLAST [2] fournissant ainsi des sorties qui sont ensuite combinées en utilisant Pantograph [21] pour créer un réseau métabolique. Pantograph se sert du SBML d'un réseau modèle (*template*), d'une table contenant les orthologies démontrées entre les gènes du *template* et les gènes de l'organisme étudié (cette table est fournie par les sorties d'OrthMCL et d'InParanoïd), d'une table contenant les caractéristiques phénotypiques d'une condition d'étude donnée et d'une table contenant les améliorations réalisées manuellement.

Les réseaux ainsi reconstruits sont contenus dans des fichiers au format SBML (*System Biology Markup Language*) [12]. Ces fichiers présentent les réactions, les compartiments et les métabolites de l'organisme.

1.2. Outils permettant de « combler les lacunes » (Gap-filling)

Le *gap-filling* est une étape nécessaire aux reconstructions de réseaux métaboliques. En effet, de nombreux *pathways* sont incomplets à l'issue des logiciels Pathway-Tools ou

Pantograph, et cette étape permet d'ajouter des réactions absentes du réseau qui pourraient être retrouvées si l'annotation était de meilleure qualité. L'un des problèmes inhérent au *gap-filling* est que les identifiants de réactions et des métabolites entre les modèles utilisés ne proviennent pas toujours des mêmes bases de données. Pour permettre une unification de ces identifiants, l'équipe Dyliss a développé l'outil Samifier. Cet outil est une plateforme d'aide à la décision possédant une interface graphique. Il permet ainsi de réaliser une étape d'unification de différents identifiants sur une base de données commune, ici, MetaCyc pour ensuite permettre l'étape de *gap-filling* [29].

Cette étape est effectuée avec l'outil Meneco (*Metabolic Network Completion*) qui réalise un *gap-filling* topologique, c'est à dire la complétion du graphe [25]. Il est employé après la fusion des réseaux reconstruits par annotation et par orthologie au travers du *workflow* AuReMe. Il suggère des réactions à ajouter, à partir d'une base de données de référence et de composés limitatifs du milieux de croissance, appelés *seeds*. Ces réactions sont suggérées dans le but de savoir si il est possible de satisfaire un critère de productibilité afin d'aboutir à un réseau métabolique topologiquement fonctionnel. Ce critère correspond à un ensemble de métabolites de la biomasse cibles, nommées *targets*, démontrées produites expérimentalement au travers de cofacteurs déjà présent dans la cellule (par exemple : ATP, ADP, NAD(P), Cytochrome c, *etc*.).

1.3. Curation manuelle

Après l'étape de *gap-filling*, il arrive que la biomasse ne soit toujours pas produite. Un travail de raffinement sera donc nécessaire afin de produire de la biomasse. La curation manuelle consiste à apporter des modifications au modèle dans le but d'inclure une connaissance approfondie des biologistes et de la littérature permettant ainsi l'enrichissement de la qualité d'une reconstruction d'un modèle.

1.4. Outil d'analyse des réseaux métaboliques

Concernant l'analyse de ces reconstructions, ici, on a utilisé l'analyse de l'équilibre des flux (*Flux-balance analysis* ou FBA) qui calcule le flux de métabolites à travers le réseau métabolique pour permettre de prédire le taux de croissance d'un organisme. Cette analyse est une méthode pour confirmer que chacun des composants de la biomasse peut-être synthétisé par le réseau métabolique dans une condition environnementale donnée ce qui en fait un moyen permettant la validation des modèles. Ce procédé se base sur des contraintes

quantitatives à savoir les coefficients stœchiométriques associés au flux de métabolites à travers le réseau [23]. En plus de la FBA, l'analyse de la variabilité de flux (*flux variability analysis* ou FVA) est aussi employée pour tester les flux en regardant quelles sont les réactions essentielles, les réactions alternatives et les réactions bloquées du flux. Ainsi la FVA est un moyen de tester la robustesse du modèle [10]. Ces deux méthodes d'optimisation des flux sont employés au travers de la boîte à outil COBRApy [17] également incluse dans AuReMe.

1.5. Espace de travail

AuReMe (*Automatic Reconstruction of Metabolic model*) [1] est un workflow comprenant une boîte à outils nommé PADMet (*PortAble Database for Metabolism*) [6], contenant à la fois des logiciels développés par l'équipe Dyliss et des outils conçus par d'autres équipes de recherche pour la reconstruction de réseaux. Ces outils permettent de reconstruire, de compléter, de manipuler, d'analyser et de visualiser des réseaux métaboliques. Ce workflow nous offre la possibilité de stocker les métadonnées favorisant ainsi leur exploration et leur distribution. Le but est de garantir une traçabilité et une reproductibilité des reconstructions. Cet environnement conduit de façon pratique le suivi de quatre étapes que sont la modélisation basée sur l'annotation, la modélisation basée sur l'orthologie, le *gap-filling* et la curation manuelle s'appuyant sur le package PADMet. AuReMe permet aussi la génération d'un wiki local pour la visualisation des données du modèle reconstruit tout en le reliant à une base de données de réaction telle que MetaCyc [28]. AuReMe est encapsulé dans une image Docker [7] avec des versions de bases de données (MetaCyc 20.0 et BIGG 2.3) facilitant sa distribution auprès de la communauté scientifique. L'utilisation de conteneurs Docker permet ainsi une utilisation isolée et autonome du workflow.

2. Données utilisées

2.1. Données pour T. lutea

Les données utilisées pour *T. lutea* [9] correspondent à son génome pour lequel nous disposons d'annotations expérimentales et *in silico*, et à son protéome expérimental et *in silico*. Les annotations et le génome sont utilisés pour la reconstruction du réseau par annotation avec le logiciel Pathway-Tools. Quant aux protéomes, ils servent aux reconstructions par orthologie, grâce au logiciel Pantograph. Les protéomes ont été au

préalable fusionnés du fait que l'utilisation du protéome expérimental n'est pas suffisante pour produire un réseau fonctionnel. Les modèles *in silico* se basent sur des hypothèses notamment en terme de découverte de nouveaux rôles métaboliques des gènes individuels [22]. Les données comportent également un premier réseau métabolique, nommé réseau cœur reconstruit expérimentalement par Caroline Baroukh (*INRA*) correspondant au réseau primaire de *T. lutea*. Il contient approximativement 300 réactions. Ces données proviennent d'une collaboration avec le laboratoire *Physiologie et Biotechnologie des Algues* de l'*IFREMER*.

2.2. Données pour les modèles

Afin de permettre une reconstruction du réseau par orthologie, quatre organismes modèles ont été choisis en tant que référence : *Chlamydomonas reinhardtii* pour sa proximité taxonomique avec *T. lutea* ; *Ectocarpus siliculosus* car c'est une algue brune dont le réseau a été reconstruit par l'équipe Dyliss ; *Arabidopsis thaliana* car c'est l'organisme modèle chez les plantes ; *Synechocystis sp* car il a des propriétés de voies de synthèse des acides gras similaires à *T. lutea* [18].

Un descriptif de ces modèles est présenté dans le tableau 1. Les données utilisées correspondent à leur protéome et leur réseau métabolique qui servent d'entrée pour la reconstruction par orthologie avec Pantograph.

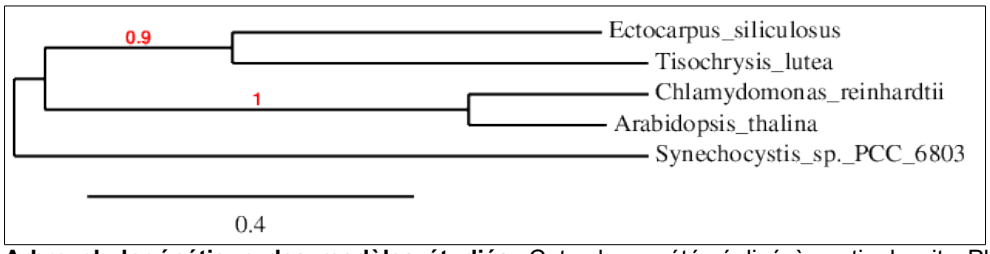

Figure 3. Arbre phylogénétique des modèles étudiés. Cet arbre a été réalisé à partir du site Phylogeny.fr (*http://www.phylogeny.fr/*) en se basant sur la séquence protéique *Tisochrysis_lutea_Proteine_23688* et la séquence ayant la *E-value* la plus faible et le score le plus élevé des protéomes des modèles lors du Blastp réalisé pour la réaction HISTAMINOTRANS-RXN.

Tableau 1. Informations sur les modèles d'étude utilisés. Le tableau présente les bases de données et les références associées à chaque modèle. Le nombre de réactions et de métabolites contenus dans les modèles sont figurés. La notation « *Mapping* » indique si un mapping sur la version MetaCyc 20.0 a été nécessaire.

Modèle	Base de données	Référence *(organisme)*	Réactions	Métabolites	Mapping
AraGEM	KEGG	*De Oliveira Dal'Molin et al. (2010) [22] (Arabidopsis thaliana)*	1601	1769	Oui
*i*Cre1355 version 5.5	BiGG	*Imam et al. (2015) [13] (Chlamydomonas reinhardtii)*	2394	1845	Oui
EctoGEM	MetaCyc 18.5	*Aite et al. (in preparation) [1] (Ectocarpus siliculosus)*	1906	2104	Non

Synechocystis sp PCC 6803	KEGG	*Knoop et al. (2013) [17] (Synechocystis sp)*	603	586	Oui

3. Pipeline de la réalisation du projet d'étude

3.1. Reconstruction automatique du réseau métabolique de Tisochrysis lutea

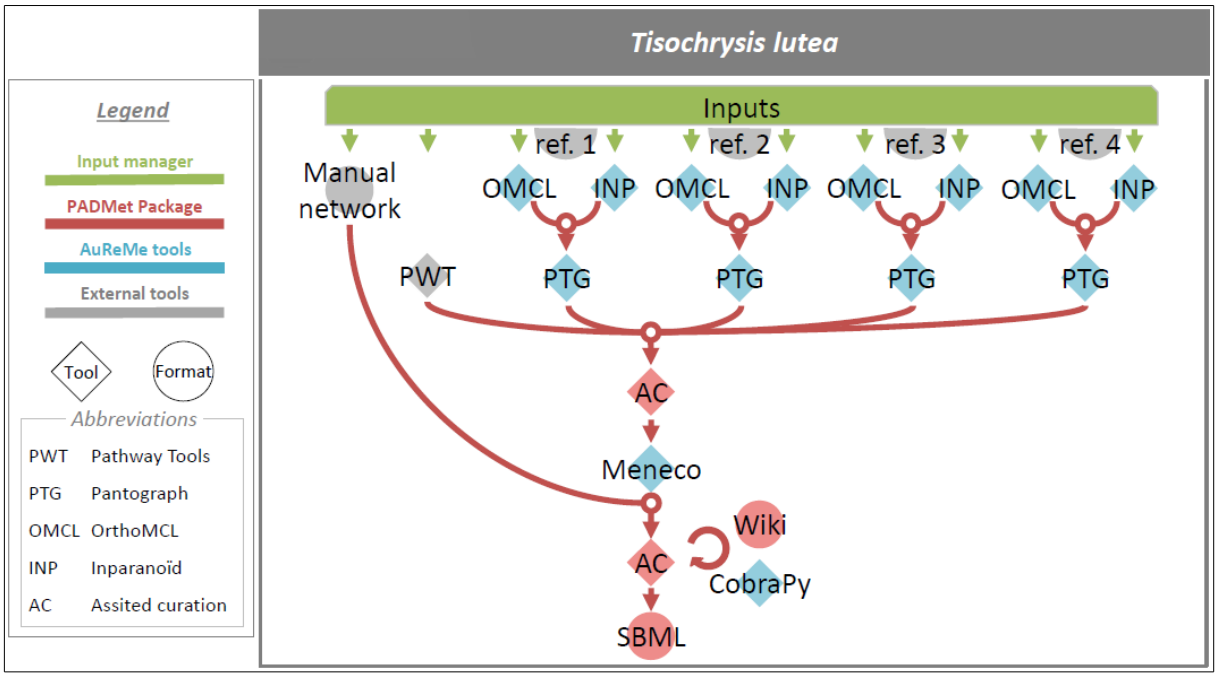

Figure 4. Pipeline de reconstruction de GSM employé dans le cadre de cette étude. Cette figure provient de l'article de *Aite et al.* en préparation [1]. Deux sources d'entrée sont utilisées, d'une part le génome et les annotations de l'espèce pour Pathway-Tools et d'autre part les protéomes et réseaux des quatre modèles correspondant aux références (ref.) 1 à 4 et le protéome de *T. lutea* pour Pantograph.

La reconstruction des réseaux métaboliques basée sur l'annotation a été effectuée avec Pathway-Tools pour le génome et les annotations expérimentale et *in silico*. Les résultats ont été récupérés au format PGDB servant d'entrée à AuReMe. Ces deux entrées PGDB ont été reconstruites et fusionnées à travers AuReMe pour obtenir un fichier SBML. D'un autre côté, le protéome expérimental et le protéome *in silico* on été fusionnés sous un même fichier. Celui-ci a été placé comme entrée tout comme les protéomes et les réseaux métaboliques des modèles utilisés pour la reconstruction par orthologie avec Pantograph après vérification de la validité de chacun. Une fois les réseaux reconstruits avec Pantograph, un mapping sur la base de données MetaCyc 20.0 a été effectué pour les réseaux ayant des identifiants de réactions de bases de données autres que MetaCyc. A la suite de cela, une fusion de l'ensemble des reconstructions a été effectuée, soit une fusion de six réseaux : le réseau par annotation, les quatre réseaux par orthologie et le réseau primaire. Suite à cette fusion, une étape de *gap-filling* a été réalisée avec Meneco. Puis, une étape de curation manuelle a été effectuée dans le

but d'ajouter une réaction consommant la biomasse en l'exportant dans le milieu de culture de manière à éviter l'accumulation de métabolites. Le fichier final au format SBML correspond au réseau métabolique de *T. lutea*. En parallèle, un wiki a été généré dans le but de mettre à jour le wiki déjà existant vis à vis de la reconstruction manuelle du réseau de *T. lutea* (*http://tisogem.irisa.fr/wiki/index.php/Main_Page*).

3.2. Vérification du nouveau réseau métabolique

Dans l'objectif de vérifier ce nouveau réseau métabolique, une recherche des *pathways* associés aux métabolites réactant de la carnosine s'est effectuée par exploration du wiki généré à la reconstruction du réseau métabolique de *T. lutea* réalisée manuellement et finalisée en septembre 2016. Ce réseau a été obtenu par la combinaison manuelle des méthodes utilisées au cours de ce stage avec les versions identiques des modèles à l'exceptions de la version d'*Ectocarpus siliculosus* basée sur une version MetaCyc 17.0 [24]. Ensuite, des Blastp [2] ont été réalisés pour aligner une séquence protéique orthologue de *T. lutea* concernée par une réaction de ces *pathways* contre les protéomes des quatre organismes modèles dans le but de vérifier et de compléter les résultats visualisés dans le wiki. Pour ces Blastp, les paramètres utilisés sont ceux proposés par défaut par le logiciel du NCBI.

III – Résultats

1. Reconstruction automatique

A partir du procédé de reconstruction automatisé au travers du workflow AuReMe, les résultats suivants ont étés obtenus. Ils sont présentés dans le tableau 2 au fur et à mesure du processus de reconstruction. Parmi l'ensemble de ces résultats, la fusion du réseau expérimental et du réseau *in silico* issue de Pathway-Tools (logiciel extérieur à AuReMe), a permis d'obtenir 1788 réactions, 2168 métabolites pour 1966 gènes. 81,54 % des réactions sont associées à des gènes. L'étape de *gap-filling* a permis l'ajout de 18 réactions. Le réseau final, après la fusion, le *gap-filling* et la curation manuelle, a donné un réseau métabolique de 2490 réactions, 2790 métabolites et 3380 gènes dont 74,10 % des réactions sont associées à des gènes. Ce réseau obtient un taux de croissance de 74.8554653249 par l'analyse de la FBA. Quant à l'analyse de la FVA, elle donne les résultats suivant : 149 réactions essentielles, 352 réactions alternatives et 1989 réactions bloquées.

Tableau 2. Résultats de la reconstruction automatisée du réseau de *T. lutea*. Le nombre de métabolites, de gènes et de réactions (notamment les réactions conservés après Pantograph et après unification des résultats entre les différents modèles, *i.e* mapping), le pourcentage de réactions associées aux gènes, le nombre de cibles produites topologiquement et via les flux ainsi que la FBA y sont présentés.

Réseaux	Réactions		Métabolites	Gènes après/avant mapping	%Gènes-Réactions	Flux activant des cibles	Cibles produites par topologie	FBA
	Avant mapping	Après mapping						
Merge annotations expérimentales – *in silico*	1788		2168	1966	81,54 %	0/39	0/39	0
Orthologie avec *AraGEM*	905/1601	474/905	729/1769	594/845	100 %	0/39	0/39	0
Orthologie avec *iCre1355*	1406/2394	378/1406	651/1845	446/952	100 %	0/39	0/39	0
Orthologie avec *EctoGEM*	1199/1906		1578/2104	1650	100 %	0/39	0/39	0
Orthologie avec *Synechocystis sp.* PCC 6803	328/603	212/328	313/586	209/496	100 %	0/39	0/39	0
Données ajoutées du réseau primaire	261		281	0	0 %	0/39	0/39	0
Réseau *T. lutea* après merge	2471		2781	3380	74,67 %	10/39	9/39	0
Réseau *T. lutea* final après *gap-filling* et curation manuelle	2490		2790	3380	74,10 %	19/39	20/39	74,86

2. Analyse d'un pathway particulier : la voie de biosynthèse de la carnosine

Le *pathway* associé à la carnosine concerne le *pathway* PWY66-420. Celui-ci est composé d'une unique réaction, la réaction CARNOSINE-SYNTHASE-RXN, illustrée à la figure 1, qui est à l'origine de la biosynthèse de la carnosine. La carnosine est un métabolite antioxydant essentiel qui empêche la neurodégénescence du cerveau chez les mammifères. Elle est retrouvée en grande concentration dans les muscles des animaux mais aussi chez les algues marines [11]. Ce métabolite est un un dipeptide constitué de β-alanine et de L-histidine ce qui a montré un intérêt à examiner leur voie de biosynthèse. Dans ce but, une figure basée sur l'exploration du wiki a été effectuée au cours de ce stage avant sa mise à jour avec le nouveau réseau. Celle-ci est faite de manière à la présenter dans l'article en préparation [1] dont je figure parmi les co-auteurs. Pour chaque réaction associée à une séquence protéique orthologue chez *T. lutea*, des Blastp ont été réalisés. Les résultats des Blastp des séquences protéiques de *T. lutea* orthologues concernées par une réaction ont été retenus si la *E-value* était inférieure à 10^{-10} et si le pourcentage d'identité était supérieur à 25 %. Ils ont été

représentés par un carré de couleur, associé au modèle correspondant au protéome utilisé pour le Blastp, sur la figure.

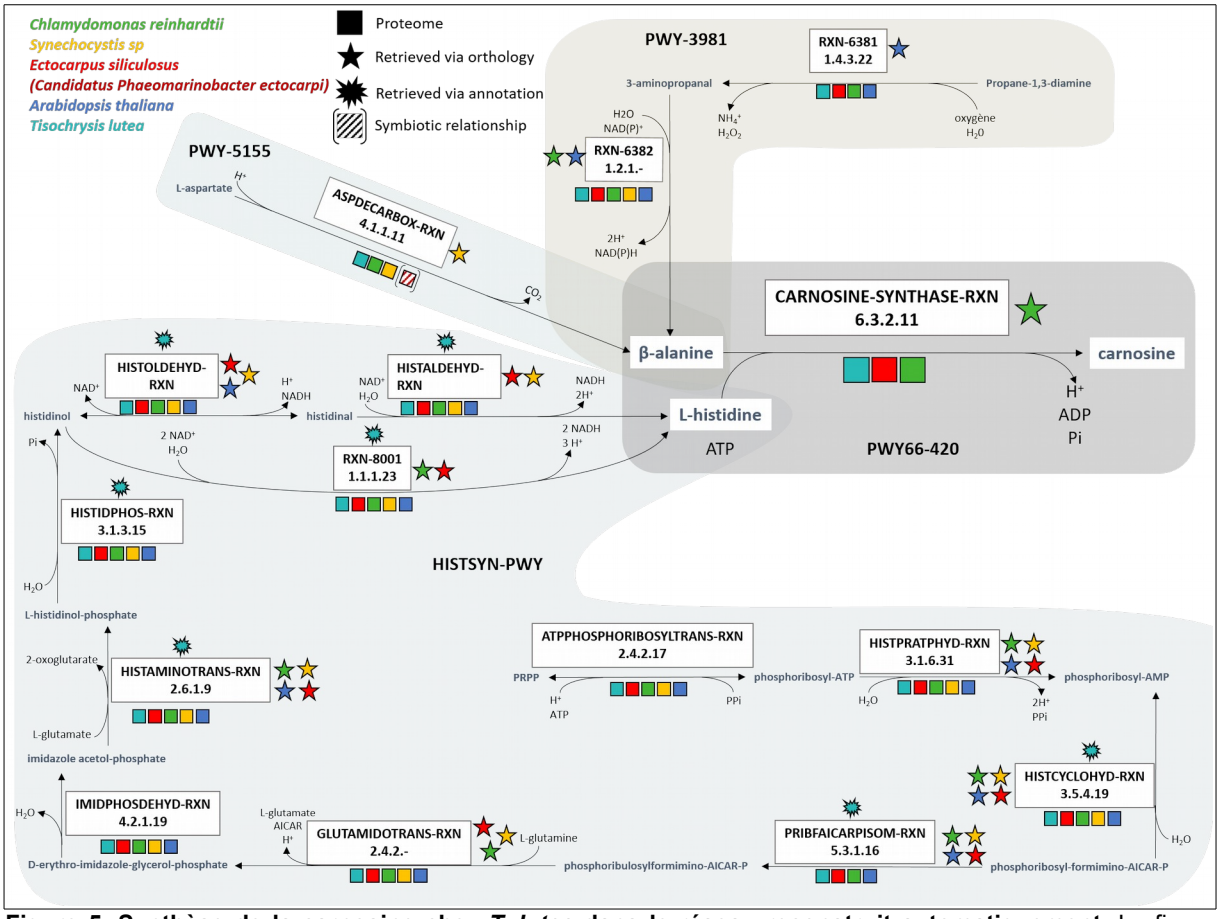

Figure 5. Synthèse de la carnosine chez *T. lutea* dans le réseau reconstruit automatiquement. La figure illustre la reconstruction du *pathway* de la synthèse de la carnosine chez *T. lutea* à partir de trois sources de données utilisées : les annotations du génome expérimental et *in silico*, l'orthologie via quatre modèles métaboliques de références (*A. thaliana*, *C. reinhardtii*, *E. siliculosus* et *Synechocystis sp.*) et les protéomes complets de ces quatre organismes utilisés pour l'alignement des séquences. La présence de relation symbiotique est aussi figurée. Cette figure fait partie de l'article *Aite et al (in preparation)* [1].

L'exploration du wiki généré a permis de montrer que deux *pathways* complets permettaient la synthèse de β-alanine : PWY-3981 et PWY-5155. Le *pathway* PWY-5155 met en avant une aspartate décarboxylase (ASPDECARBOX-RXN) retrouvée chez *Chlamydomonas reinhardtii* (par Blastp) et *Synechocystis sp.* PCC 6803 (par Pantograph). Le *pathway* PWY-3981 met en exergue deux enzymes : une diamine oxydase (RXN-6381) et une aminopropionaldéhyde déshydrogénase (RXN-6382). Quant à la L-histidine, un *pathway* complet est retrouvé : HISTSYN-PWY. Celui-ci regroupe 10 réactions dont 8 sont validées par les logiciels Pathway-Tools ou Pantograph avec les 4 organismes modèles utilisés. Les deux réactions manquantes (IMIDPHOSDEHYD-RXN et ATPPHOSPHORIBOSYLTRANS-RXN) ont été ajoutés manuellement après vérification par Blastp de la séquence protéique

contre les protéomes des espèces. Ces deux réactions étaient présentes dans le réseau primaire de *T. lutea*, et ont ajouté par Meneco pour produire la biomasse.

IV – Discussion

1. Vérification de la reconstruction automatique

Au travers de ce stage, un réseaux a été produit pour *T. lutea* par reconstruction au travers du workflow AuReMe comme expliqué dans la partie *Matériels et Méthodes*. Concernant ce réseau reconstruit, les résultats obtenus par l'analyse des flux nous permettent de déduire que le réseau de *T. lutea* croît sur son milieu de culture. Une vérification des résultats entre le réseau reconstruit manuellement et celui reconstruit de façon automatique au cours de ce stage a été entreprise en se basant sur l'exemple de la carnosine. Cette vérification, réalisée à partir du fichier au format padmet généré lors de la reconstruction automatique du réseau métabolique, a permis de reprendre et de corriger la figure. Cette figure est présentée ici comme figure 5. A la différence de la version précédente, la réaction RXN-6381 du *pathway* PWY-6381 n'est pas associée par orthologie avec *Ectocarpus siliculosus* du fait que la version du réseau utilisée ne soit pas la même. Une autre différence concerne la réaction HISTAMINOTRANS-RXN du *pathway* HISTSYN-PWY qui était associée à l'orthologie de trois des modèles en plus d'un résultat de Pathway-Tools. Avec le nouveau réseau, cette réaction est aussi associée au quatrième modèle : *Chlamydomonas reinhardtii*.

La création de cette figure permet d'illustrer le côté pratique du wiki et du format padmet. En effet, ils permettent de naviguer au sein du réseau métabolique tout en conservant les sources d'origines de chacune des données renfermées facilitant ainsi la traçabilité et la reproductibilité du processus de reconstruction.

2. Intérêt de la combinaison de données et de méthodes hétérogènes

Avec cette figure, on peut montrer que les reconstructions basées sur une méthodologie unique ne sont pas suffisantes dans le sens où l'on passerait à côté de réactions qui sont pourtant présentes. En effet, avec des Blastp réalisés entre la séquence protéique orthologue retrouvée dans le réseau reconstruit et les protéomes des modèles d'étude, certaines réactions enzymatiques manquantes, a priori, dans des modèles sont identifiées. Un exemple peut être la réaction HISTALDEHYD-RXN du *pathway* HISTSYN-PWY. En effet, cette réaction est identifiée chez *Ectocarpus siliculosus* et *Chlamydomonas reinhardtii*.

Pourtant, l'utilisation de Blatp avec les deux autres modèles montre sa présence. Ceci illustre une des limites des reconstructions automatisées avec les méthodes actuelles. La conséquence provient du fait que certaines réactions ne sont pas associées à des gènes même au sein des réseaux métaboliques pourtant considérés comme modèles. Ceci montre donc que la complémentation de données hétérogènes est un apport à la découverte de nouvelles données tant sur le modèle d'étude que sur les modèles de référence. L'exemple de la CARNOSINE-SYNTHASE-RXN (*pathway* PWY66-420) est probant. En effet, cette réaction semble absente chez tous nos modèles à l'exception de *Chlamydomonas reinhardtii* (par Pantograph). Si la reconstruction ne s'était effectuée que par orthologie avec *Arabidopsis thaliana* ou sur *Synechocystis sp.*, cette réaction aurait été totalement absente.

3. Développement de nouvelles connaissances

Ces reconstructions sont aussi un moyen d'apport de connaissance sur les organismes modèles comme par exemple *Ectocarpus siliculosus* qui ne produit pas de β-alanine au travers de la L-aspartate. Pourtant, ce composé est retrouvé au sein de son organisme. *Ectocarpus siliculosus* devrait être capable de synthétiser de la carnosine avec la seconde voie de synthèse de la β-alanine. Une recherche avancée montre que la β-alanine peut-être synthétisée au travers de la L-aspartate grâce à une alpha-protéobactérie de l'ordre des *Rhizobiales*. Cette dernière est en effet un symbiote obligatoire de la paroi d'*Ectocarpus siliculosus* nommée *Candidatus Phaeomarinobacter ectocarpi* ou Ec32. La seconde voie de synthèse de la β-alanine n'a pas été mise en évidence chez cette algue brune [8, 25]. La réaction ASPDECARBOX-RXN a été perdue chez *Ectocarpus siliculosus* mais cette relation symbiotique avec cette bactérie permet le maintient de cette réaction.

L'autre point intéressant qui ressort de notre étude est d'établir le fait que l'association et la nécessité d'employer plusieurs méthodes de reconstruction est non seulement un bon moyen de réduire le besoin de *gap-filling*, mais aussi de permettre la reconstruction d'un réseau plus complet. L'utilisation de plusieurs modèles métaboliques, nous permet de vérifier la présence de certaines réactions que nous aurions omis autrement. En effet, la reconstruction du réseau métabolique de *T. lutea* à partir de plusieurs modèles d'étude est un moyen d'identifier des voies de biosynthèse spécifique. La complétion de méthodes hétérogènes est donc un moyen de permettre de compléter des *pathways*.

V – Conclusion et Perspectives

Les méthodes de reconstruction de réseaux métaboliques proposées actuellement ne sont pas suffisantes indépendamment pour générer un réseau métabolique complet. En effet, l'utilisation compilée de plusieurs méthodes hétérogènes comme celles basées sur l'annotation et celles basées sur l'orthologie est un moyen de reconstruire de meilleurs modèles. Le workflow AuReMe est une solution apportée par l'équipe Dyliss pour permettre de mettre en œuvre ce procédé. Son utilisation à travers la reconstruction du réseau métabolique de *Tisochrysis lutea* illustre l'intérêt de la compilation de ces méthodes notamment sur l'exemple de la biosynthèse de la carnosine. Le nouveau réseau réalisé automatiquement permet l'obtention de résultats sensiblement similaires au réseau reconstruit manuellement.

L'exemple de la voie de synthèse de la carnosine pose les bases de nouvelles trajectoires d'étude sur le fait qu'une relation symbiotique permet la croissance d'un organisme tel qu'*Ectocarpus siliculosus*. Une étude pourrait être entreprise pour rechercher ce genre de relation avec des symbiotes de *T. lutea*. De plus, ce *pathway* n'est pas unique. Une comparaison sur une plus grande échelle devrait être effectuée entre les réseaux métaboliques reconstruits manuellement et automatiquement pour *T. lutea*. De même, il pourrait être intéressant de comparer ce réseau entier avec le réseau primaire.

Un autre direction d'étude pourrait être de modifier la réaction de biomasse sur le réseau métabolique récemment produit avec de nouvelles données que l'Ifremer devrait apporter.

Enfin, il pourrait être enrichissant de comprendre les voies de synthèse des acides gras polyinsaturés et de comprendre pourquoi la souche mutée produit plus de lipide que la souche sauvage de *Tisochrysis lutea* [9].

Glossaire

ADP	*Adénosine DiPhosphate*
ATP	*Adénosine TriPhosphate*
AuReMe	*Automatic Reconstruction of Metabolic model*
BiGG	*Biochemical, Genetic and Genomic*
BLAST	*Basic Local Alignment Search Tool*
CO2	*Carbon dioxide*
COBRApy	*Constraints-Based Reconstruction and Analysis for Python*
DHA	*DocosaHexaenoic Acid*
EC	*Enzyme Commission number*
FBA	*Flux-Balance Analysis*
FVA	*Flux Variability Analysis*
GSM	*Genome Scale Model*
KEGG	*Kyoto Encyclopedia of Genes and Genomes*
Meneco	*Metabolic Network Completion*
NAD(P)	*Nicotinamide Adénine Dinucléotide (Phosphate)*
NCBI	*National Center for Biotechnology Information*
PADMet	*PortAble Database for Metabolism*
PGDB	*Pathway/Genome DataBase*
T. lutea	*Tisochrysis lutea*

Bibliographie

[1] Aite, M. et al. In preparation. Traceability, reproducibility and wiki-exploration of genome-scale metabolic model reconstructions. (In preparation).

[2] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215, 3 (Oct. 1990), 403–410.

[3] Bendif, E.M., Probert, I., Schroeder, D.C. and Vargas, C. de 2013. On the description of Tisochrysis lutea gen. nov. sp. nov. and Isochrysis nuda sp. nov. in the Isochrysidales, and the transfer of Dicrateria to the Prymnesiales (Haptophyta). *Journal of Applied Phycology*. 25, 6 (Dec. 2013), 1763–1776.

[4] Carrier, G., Garnier, M., Le Cunff, L., Bougaran, G., Probert, I., De Vargas, C., Corre, E., Cadoret, J.-P. and Saint-Jean, B. 2014. Comparative transcriptome of wild type and selected strains of the microalgae Tisochrysis lutea provides insights into the genetic basis, lipid metabolism and the life cycle. *PloS One*. 9, 1 (2014), e86889.

[5] Caspi, R. et al. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*. 44, D1 (Jan. 2016), D471–D480.

[6] Chevallier, M., Aite, M., Got, J., Collet, G., Loira, N., Cortes, M.-P., Frioux, C., Laniau, J., Trottier, C., Maas, A. and others 2016. Handling the heterogeneity of genomic and metabolic networks data within flexible workflows with the PADMet toolbox. *Jobim 2016: 17ème Journées Ouvertes en Biologie, Informatique et Mathématiques* (2016).

[7] Di Tommaso, P., Palumbo, E., Chatzou, M., Prieto, P., Heuer, M.L. and Notredame, C. 2015. The impact of Docker containers on the performance of genomic pipelines. *PeerJ*. 3, (2015), e1273.

[8] Dittami, S.M., Barbeyron, T., Boyen, C., Cambefort, J., Collet, G., Delage, L., Gobet, A., Groisillier, A., Leblanc, C., Michel, G., Scornet, D., Siegel, A., Tapia, J.E. and Tonon, T. 2014. Genome and metabolic network of "Candidatus Phaeomarinobacter ectocarpi" Ec32, a new candidate genus of Alphaproteobacteria frequently associated with brown algae. *Frontiers in Genetics*. 5, (Jul. 2014).

[9] Garnier, M., Bougaran, G., Pavlovic, M., Berard, J.-B., Carrier, G., Charrier, A., Le Grand, F., Lukomska, E., Rouxel, C., Schreiber, N., Cadoret, J.-P., Rogniaux, H. and Saint-Jean, B. 2016. Use of a lipid rich strain reveals mechanisms of nitrogen limitation and carbon partitioning in the haptophyte Tisochrysis lutea. *Algal Research*. 20, (Dec. 2016), 229–248.

[10] Gudmundsson, S. and Thiele, I. 2010. Computationally efficient flux variability analysis. *BMC Bioinformatics*. 11, (Sep. 2010), 489.

[11] Holdt, S.L. and Kraan, S. 2011. Bioactive compounds in seaweed: functional food applications and legislation. *Journal of Applied Phycology*. 23, 3 (Jun. 2011), 543–597.

[12] Hucka, M. et al. 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics (Oxford, England)*. 19, 4 (Mar. 2003), 524–531.

[13] Imam, S., Schäuble, S., Valenzuela, J., López García de Lomana, A., Carter, W., Price, N.D. and Baliga, N.S. 2015. A refined genome-scale reconstruction of Chlamydomonas

metabolism provides a platform for systems-level analyses. *The Plant Journal: For Cell and Molecular Biology*. 84, 6 (Dec. 2015), 1239–1256.

[14] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*. 45, D1 (Jan. 2017), D353–D361.

[15] Karp, P.D., Paley, S. and Romero, P. 2002. The Pathway Tools software. *Bioinformatics*. 18, suppl_1 (Jul. 2002), S225–S232.

[16] King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O. and Lewis, N.E. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*. 44, D1 (Jan. 2016), D515–D522.

[17] Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W. and Steuer, R. 2013. Flux balance analysis of cyanobacterial metabolism: the metabolic network of Synechocystis sp. PCC 6803. *PLoS computational biology*. 9, 6 (2013), e1003081.

[18] Kotajima, T., Shiraiwa, Y. and Suzuki, I. 2014. Functional screening of a novel Δ15 fatty acid desaturase from the coccolithophorid Emiliania huxleyi. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. 1841, 10 (Oct. 2014), 1451–1458.

[19] Latendresse, M., Krummenacker, M., Trupp, M. and Karp, P.D. 2012. Construction and completion of flux balance models from pathway databases. *Bioinformatics*. 28, 3 (Feb. 2012), 388–396.

[20] Li, L., Stoeckert, C.J. and Roos, D.S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*. 13, 9 (Sep. 2003), 2178–2189.

[21] Loira, N., Zhukova, A. and Sherman, D.J. 2015. Pantograph: A template-based method for genome-scale metabolic model reconstruction. *Journal of Bioinformatics and Computational Biology*. 13, 02 (Apr. 2015), 1550006.

[22] de Oliveira Dal'Molin, C.G., Quek, L.-E., Palfreyman, R.W., Brumbley, S.M. and Nielsen, L.K. 2010. AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis. *Plant Physiology*. 152, 2 (Feb. 2010), 579–589.

[23] Orth, J.D., Thiele, I. and Palsson, B.Ø. 2010. What is flux balance analysis? *Nature Biotechnology*. 28, 3 (Mar. 2010), 245–248.

[24] Prigent, S., Collet, G., Dittami, S.M., Delage, L., Ethis de Corny, F., Dameron, O., Eveillard, D., Thiele, S., Cambefort, J., Boyen, C., Siegel, A. and Tonon, T. 2014. The genome-scale metabolic network of Ectocarpus siliculosus (EctoGEM): a resource to study brown algal physiology and beyond. *The Plant Journal*. 80, 2 (Oct. 2014), 367–381.

[25] Prigent, S., Frioux, C., Dittami, S.M., Thiele, S., Larhlimi, A., Collet, G., Gutknecht, F., Got, J., Eveillard, D., Bourdon, J., Plewniak, F., Tonon, T. and Siegel, A. 2017. Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. *PLoS Computational Biology*. 13, 1 (Jan. 2017).

[26] Remm, M., Storm, C.E.V. and Sonnhammer, E.L.L. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons1. *Journal of Molecular Biology*. 314, 5 (Dec. 2001), 1041–1052.

[27] Shi, Q., Araie, H., Bakku, R.K., Fukao, Y., Rakwal, R., Suzuki, I. and Shiraiwa, Y. 2015. Proteomic analysis of lipid body from the alkenone-producing marine haptophyte alga Tisochrysis lutea. *Proteomics*. 15, 23–24 (Dec. 2015), 4145–4158.

[28] Thiele, I. and Palsson, B.Ø. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*. 5, 1 (2010), 93–121.

[29] Vignet, P. 2016. Développement d'une plateforme d'aide à la décision pour l'uniformisation des identifiants dans un réseau métabolique.

Annexes 1 – Présentation de la structure d'accueil

L'équipe DYLISS (*Dynamics, Logics and Inference for biological Systems and Sequences*) est située dans les locaux de l'INRIA (*Institut National de Recherche en Informatique et en Automatique*) de Rennes – Bretagne Atlantique sur le campus scientifique de Beaulieu, l'un des huit centre INRIA. L'équipe a pour but de caractériser les acteurs génétiques en utilisant des systèmes formels pour analyser les séquences et la biologie des systèmes en se basant sur des méthodes combinant de l'intégration de données à partir de programmation logique par contrainte, les dynamiques symboliques et les systèmes formels. L'équipe s'intéresse aux espèces non modèles mais aussi à l'intégration sur plusieurs échelles.

Le centre INRIA de Rennes – Bretagne Atlantique, créé en 1980, est constitué de 3 sites (Rennes, Nantes et Lannion) regroupant 730 personnes réparties sur 34 équipes de recherche. Il bénéficie de partenariats avec des universités (Rennes 1 et Rennes 2, Nantes), des écoles (Supélec, Ecoles des Mines de Nantes, ENS, …), des organismes (INSA, CNRS, Inserm, INRA, Ifremer, Institut Curie, …).

L'INRIA est un établissement public à caractère scientifique et technologique rattaché au ministère de la recherche et au ministère de l'industrie. Il a été créé en 1967 et son siège social est situé à Rocquencourt . Il est à la tête de 172 équipes de recherche. Chaque année, l'institut est à l'origine de 4600 publications scientifiques et 300 thèses soutenues. L'INRIA est composé de huit centres basés à Paris, Rennes, Sophia Antipolis, Grenoble, Nancy, Bordeaux, Lille et Saclay. Le budget total de l'institut s'élève à hauteur de 230 millions d'euros.

Annexes 2 – Bilan personnel du stage

Au travers ce stage, j'ai pu réaliser le besoin de mettre en lien l'informatique et les données biologiques. Ce projet était une véritable opportunité de mettre en avant la pratique à partir des connaissances jusque là très théoriques.

Même si ce stage a comporté une certaine part de biologie, il m'a permis de rester dans mon domaine de discipline initial et d'y apporter un côté analytique avec l'informatique. J'ai pu améliorer mes connaissances informatiques grâce au développement de scripts en python, via les conseils de mes collègues stagiaires et de l'équipe.

J'ai pris conscience de la part importante et chronophage que représente les articles bibliographiques dans le domaine de la recherche et j'ai pu améliorer ma compréhension de l'anglais via ce projet qui m'a vraiment intéressé. De plus, j'ai pu me rendre compte de l'importance de développer des outils et des formats uniformes pour une discipline précise comme celle de la reconstruction de réseau métabolique. Malgré le développement du format SBML, les données contenus et les bases de données différentes n'y sont pas correctement représentées. L'utilisation du format padmet m'a permis de voir cet intérêt et la reproductibilité que la communauté scientifique devrait acquérir.

Ce projet m'a appris à développer un travail et un esprit d'équipe mais aussi mes connaissances sur les réseaux métaboliques et ma vision sur l'importance de l'étude d'organismes encore peu étudiés.

En dernier point, j'ai pu apprendre la manière de rédiger un rapport de stage grâce à la patience et aux nombreuses relectures de mon encadrante. J'ai pu comprendre comment bien mettre en forme et comment mettre en avant les informations principales de mon stage.

Résumé : Vérifications du réseau métabolique entier de *Tisochrysis lutea.* La reconstruction de réseau métabolique est une nouvelle étape essentielle pour faire le lien entre les données issues du séquençage massif et les informations biochimiques et métaboliques renfermées dans des bases de données. En ce sens, ce stage a eu pour but de reconstruire de manière automatique le réseau métabolique de la microalgue *Tisochrysis lutea* (*T. lutea*) au travers du workflow AuReMe développé par l'équipe Dyliss. Ce workflow combine plusieurs méthodes telles que la reconstruction basée sur l'annotation génomique et la reconstruction basée sur l'orthologie. Il permet aussi bien le remplissage des lacunes (*gap-filling*) et la curation manuelle que l'analyse du flux de ces réseaux. Le processus de reconstruction automatique du réseau métabolique de *T. lutea* a été réalisé en partant de deux annotations de génomes de deux protéomes, l'un expérimental et l'autre *in silico*, et de quatre réseaux métaboliques connus (*Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Ectocarpus siliculosus* et *Synechocystis sp. PCC 6803*). Le réseau métabolique obtenu croît sur son milieu de culture. Suite à cela, une vérification de ce réseau a été effectuée sur l'analyse d'une voie de biosynthèse spécifique : la voie de biosynthèse de la carnosine. La figure résultat fait partie d'un article en préparation dont je figure parmi des co-auteurs.

Mots clés : Reconstruction, analyse topologique, analyse de flux, AuReMe, modèles à l'échelle du génome

Abstract: Validations of the whole metabolic network of *Tisochrysis lutea.* Metabolic network reconstruction is a new critical step to connect massive sequencing and biochemical data with metabolic information contained in databases. In this context, the aim of this internship was to reconstruct the metabolic network of the *Tisochrysis lutea* (*T. lutea*) microalgae through a workflow developed by the Dyliss team. This workflow combines several methods such as genomic annotation based reconstruction and orthology based reconstruction. It allows the *gap-filling* and manual curation as well as the analysis of the flux of these networks. The metabolic network reconstruction of *T. lutea* was carried out on one hand, from the combination of two distinct annotations, one obtained experimentally and the other *in silico*, on the other hand from the orthology informations of four organisms (*Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Ectocarpus siliculosus* et *Synechocystis* sp. PCC 6803). The metabolic network obtained grows on its growth medium. Following this, a verification of this network was carried out on the analysis of a specific biosynthetic pathway: the carnosine biosynthesis pathway. The resulting figure is part of an article in preparation which co-authors I figure among.

Keywords : Metabolic network reconstruction, topological analysis, flux balance analysis, AuReMe, Genome-scale model